Essays in Modern Time Series Econometrics with Applications in Macroeconomics and Finance

Dissertation

zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaften (Dr. rer. pol.)

vorgelegt von
Julie Schnaitmann

Sektion Politik – Recht – Wirtschaft
Fachbereich Wirtschaftswissenschaften
Konstanz, 2021

Tag der mündlichen Prüfung: 16. Juli 2021

1. Referent: Prof. Dr. Ralf Brüggemann
2. Referent: Prof. Dr. Winfried Pohlmeier
3. Referent: Prof. Dr. Joachim Grammig

Danksagung

An dieser Stelle möchte ich mich bedanken für die Unterstützung und Begleitung bei der Erstellung meiner Dissertation.

Zuerst gilt mein Dank meinem Doktorvater Herrn Prof. Dr. Ralf Brüggemann für die Betreuung meiner Promotion. Danke für die wertvollen Anmerkungen zu meinen Projekten und das Schaffen einer sehr angenehmen Arbeitsatmosphäre am Lehrstuhl. Außerdem möchte ich mich für die Finanzierung meiner Promotion und die Möglichkeit, auf viele Konferenzen zu fahren, bedanken. Desweiteren danke ich Herrn Prof. Dr. Winfried Pohlmeier für die Übernahme der Zweitbetreuung. Auch möchte ich mich für das hilfreiche Feedback in Seminaren und die Gespräche über meine Forschung und über Publikationsthemen bedanken. Danke an Jun.-Prof. Dr. Stephan Maurer für die Übernahme des ersten Vorsitz in der mündlichen Prüfung. Zudem bedanke ich mich bei Herrn Prof. Dr. Joachim Grammig für das Heranführen an die Ökonometrie, die vielen Gespräche und Hilfestellungen, die Koautorentätigkeit und die Übernahme der Drittbetreuung. Außerdem möchte ich mich bei Frau Prof. Dr. Roxana Halbleib bedanken, die mir mit Rat und Tat zur Seite steht und mir wertvolle Kontakte vermittelt hat und bei Herrn Prof. Dr. Alastair Hall, der mich in Manchester empfangen hat und ein wichtiger Mentor für mich geworden ist.

Herzlich bedanken möchte ich mich auch bei meinen Koautoren: Danke Maurizio Daniele für die Zusammenarbeit, das zusammen Schreiben und Coden und alle sonstigen Fragen, die ich über die Jahre mit dir besprechen konnte. Danke Timo Dimitriadis für die gemeinsam durchgeführten Projekte, die vielen Telefonate und Besprechungen - es war und ist mir ein Vergnügen. Und zuletzt, danke Dalia Elshiaty und Joachim Grammig für die Zusammenarbeit, die mir immer wieder gute Gründe gab an meine alte Wirkungsstätte zurückzukehren, die vielen Diskussionen, das Feilschen um jeden Satz und die vielen Gespräche, die über das Projekt hinausgingen.

Ein großer Dank gilt allen Mitgliedern des Lehrstuhls: Ralf Brüggemann, Dominik Bertsche, Robin Braun, Tilmann Härtl und allen aktuellen und ehemaligen studentischen

Hilfskräften und Tutoren. Die Zusammenarbeit hat mir immer viel Spaß bereitet. Ein besonderer Dank gilt auch Verena Kretz, die uns immer den Rücken freihält und zu allen universitären Fragen Antwort weiß. Außerdem danke ich allen Beteiligten an unseren Konstanzer Ökonometrieseminaren für Diskussionen, Feedback und Unterstützung, insbesonderen den Lehrstuhlmitgliedern von Herrn Pohlmeier, Roxana Halbleib und Lyudmila Grigoryeva. Danke auch an die Ökonometriegruppe aus Tübingen, die mich regelmäßig beherbergt und mir über die Jahre viel Input gegeben hat. Danke Consti, Dalia, Jantje, Jo, Johannes und Thomas.

Für die Finanzierung zahlreicher Konferenzen und Summer Schools danke ich der Graduate School of Decision Science und Herrn Brüggemann. Danke Christian Mücher, Maurizio Daniele und Timo Dimitriadis für die zahlreichen gemeinsamen Konferenzbesuche. Es war mir eine Freunde mit euch dort das eine oder andere Kaltgetränk zu verköstigen.

Meinen Freunden an der Uni, der Donnerstagsgang und den vielen Mittagessen-, Kaffeepausen- und Feierabend(bier)begleitern bin ich sehr dankbar für die gemeinsame Zeit und das gegenseitige Aufmuntern. Danke Henrike Biehl, Timo Dimitriadis, Thomas Hattenbach, Patrick Hauf, Jan Hausfeld, Tim Hermann, Mark Kapteina, Marlene Koch, Mario Krauser, Manuel Leiniger, Marco Menner, Frederic Menninger, Timm Prein, Florian Röser, Annika Schürle, Sina Streicher und meinen Lehrstuhlkollegen. Ein besonderer Dank gilt Simon Stehle, meinem wichtigsten Ansprechpartner an der Uni.

Außerdem danke ich meinen Konstanzer, Tübinger und Stuttgarter Freunden für die gemeinsamen Erlebnisse während der letzten vier Jahre, den Abenden, Wochenenden, Bergausflügen, Urlauben und den vielen Gesprächen. Danke an Isabelle Kalkusch, Simon Stehle, Nele Klinkenberg, Sabrina Ruess, Mark Kapteina, Marlene Koch, Jonas Allerbeck, Fabian Paschke, Falco Ruppert, Tilmann Härtl, Tobias Lemke, Tobias Schmidt, Leonie Mack, Anne Butscher, Steffen Dengler, Tim Raschka, Natascha Bandle, Alex Güth, Susanne Pfrommer, Constantin Hanenberg, Elisabeth Schröder, Lena Schäfer, Fabian Slezak, Alex Schmidt, Pia Kramer, Theresa Fabel, Friedrich Geise, Julian Vahl, Fabian Wille, Paul Bullinger, Laura Goldschmitt und Kevin Dorling.

Zuletzt möchte ich mich bei meiner Familie, meinen Eltern und meinem Bruder, sowie meinen Großeltern bedanken, die immer für mich da sind. Auch wenn das was ich mache schwer zu greifen ist, so lassen sie mich immer meinen Weg gehen und unterstützen mich. Danke!

Table of Contents

List of Tables

List of Figures

Summary

This dissertation addresses challenges in modern time series econometrics and provides applications in macroeconomics and finance. In all projects, we propose a new estimation strategy for the respective object of interest; the first project is associated with high-dimensional data and a macroeconomic application, the second project is concerned with the evaluation of a risk measure in financial econometrics, and in the third and fourth project we estimate highly parametrized macro-finance models.

One challenge in modern time series econometrics motivating the study in Chapter 1 arises from the existence of large data sets which can be used to investigate economic channels more thoroughly. We make use of the large amount of available macroeconomic and financial time series and extract the relevant information using a sparse factor model. This enables us to combine a dimension reduction strategy provided by factor models with economically interpretable factors obtained through the introduction of sparsity. These factors are useful for structural analysis. We introduce a regularized factor-augmented vector autoregressive model to investigate the impact of monetary policy shocks on many observable economic and financial time series.

The recent introduction of the risk measure Expected Shortfall as primary market risk measure poses the challenge of how to evaluate and compare forecasts of the Expected Shortfall. This issue is addressed in Chapter 2 in which we propose forecast encompassing tests for this risk measure. Towards this purpose, we introduce an M-estimator to estimate forecast combination weights and three Wald-type tests to test for forecast encompassing.

In macro-finance and structural asset pricing the estimation and assessment of highly parametrized structural models is a challenging task. These models are targeted to describe some economic channels appropriately while also providing a complete model of the technology process and the household decision. The focus in asset pricing is typically placed on the specification of the investor's or household's decision as opposed to the business cycle dynamics. The latter often only represent a caricature of reality and are

openly misspecified. The studies in Chapter 3 and 4 are associated with the estimation of the structural parameters of interest of structural asset pricing models to provide a critical assessment of the underlying models. The proposed econometric strategies are built on the concept of partial indirect inference that allows for the consistent estimation of these parameters in the presence of model misspecification. Whereas the third chapter focuses on the critical assessment of a recently introduced asset pricing model, the fourth chapter emphasizes on the proposed estimation strategy and the impact accounting for model misspecification has on the estimation uncertainty of the structural parameters.

The four chapters in this cumulative dissertation consist of separate research papers which I have written during my doctoral studies at the University of Konstanz. The first chapter, *A Regularized Factor-Augmented Vector Autoregressive Model*, is joint work with Dr. Maurizio Daniele. The second chapter, *Forecast Encompassing Tests for the Expected Shortfall*, is a joint project with Dr. Timo Dimitriadis. The third chapter, *Empirical Asset Pricing in a DSGE Framework: Reconciling Calibration and Econometrics using Partial Indirect Inference*, is a collaboration with Prof. Dr. Joachim Grammig and Dalia Elshiaty from the Eberhard Karls Universität Tübingen. The fourth chapter, *Econometric Analysis of Partial Equilibrium Asset Pricing Models: A Two-Step Partial Indirect Inference Approach*, is entirely my own work. For consistency, the authors or the author of the projects are referred to as "we" throughout the dissertation.

In the first chapter, *A Regularized Factor-Augmented Vector Autoregressive Model*, we analyze the dynamic effects of structural monetary policy shocks on the economy using the informational content of many time series. Conventionally, this is investigated using small scale vector autoregressive (VAR) models which only incorporate few variables that often result in misleading impulse responses that stem from informational deficiencies. The factor-augmented vector autoregressive (FAVAR) model pioneered by Bernanke et al. (2005) allows the use of many economic time series. It decomposes the co-movement in the observed time series into a common and an idiosyncratic component. The common component consists of factors and the associated factor loadings that describe how the factors are affected by the underlying time series. The combined number of factors is considerably smaller than the number of observed time series. Hence, the factor model condenses the information contained in the observed series into the factors. These factors are subsequently used in a VAR model to allow for autoregressive dynamics. We introduce a regularized version of the FAVAR model and impose shrinkage on the factor loadings. This allows to retrieve factors that only load on a subset of time series which can be associated with economic sectors. Hence, these factors can be interpreted economically which enables us to use these factors for structural analysis. Using these factors in a vector autoregressive model enables

to investigate the effects of a structural shock on the included time series by computing impulse response functions. To summarize, we propose a regularized factor-augmented vector autoregressive (RFAVAR) model that allows for sparsity in the factor loadings. In this framework, factors may load only on a subset of variables which simplifies the factor identification and their economic interpretation. The effects of structural shocks can be investigated on economically interpretable factors and on all observed time series included in the RFAVAR model. We prove consistency for the estimators of the factor loadings, the factors, the covariance matrix of the idiosyncratic component, as well as the autoregressive parameters in the dynamic model. In an empirical application, we investigate the effects of a monetary policy shock on a broad range of economically relevant variables. We identify this shock using a joint identification of the factor model and the structural innovations in the VAR model. We find impulse response functions that are in line with economic rationale, both on the factor aggregates and the observed time series level.

In the second chapter, *Forecast Encompassing Tests for the Expected Shortfall*, we introduce three forecast encompassing tests for the risk measure Expected Shortfall (ES). In recent years, the ES has been introduced in the Basel III Accords (Basel Committee, 2016, 2017) as the primary market risk measure for the international banking regulation. This fosters the need for methods to evaluate the predictive ability of competing ES forecasts. The ES at probability level α is defined as the conditional expectation of returns that are smaller than the respective α-quantile. We use encompassing tests to compare ES forecasts which test the null hypothesis whether one forecast performs as good as any (linear) combination of the ES forecasts. The test is carried out by testing whether the combination parameter of the second forecast is significantly different from zero. It, hence, tests whether the second ES forecast provides any additional information to the first forecast. We utilize joint loss functions for the pair ES and Value at Risk introduced by Fissler and Ziegel (2016) to set up three ES encompassing test variants. The tests are built on misspecification robust asymptotic theory, and we investigate the finite sample properties of the tests in an extensive simulation study. We use the encompassing tests to illustrate the potential of forecast combination methods for different financial assets. This paper is accepted for publication in the *International Journal of Forecasting* (Dimitriadis and Schnaitmann, 2020).

The remaining chapters are concerned with the consistent estimation and critical assessment of structural asset pricing and macro-finance models. These models are highly parametrized and try to capture asset pricing fluctuations in financial markets. These models do not claim to offer a description of the entire economy or an appropriate description of the technology processes and are, thus, openly misspecified in this dimension. The third and fourth chapters deal with the estimation and critical assessment of these models. Towards

this purpose, we apply the partial indirect inference method (PII) introduced by Dridi et al. (2007). Thereby, we acknowledge that parts of the model are misspecified, while other parts have the claim of capturing (financial) economic reality. Adapting the PII philosophy to the asset pricing models, we use binding functions that facilitate the consistent estimation of some structural model parameters of interest, while treating others as nuisance parameters.

In the third chapter, *Empirical Asset Pricing in a DSGE Framework: Reconciling Calibration and Econometrics using Partial Indirect Inference*, we provide a critical assessment of the most recent iteration of DGSE asset pricing models introduced by Chen (2017). The class of DSGE models considered in this project connects asset prices with an exogenous technology process and an endogenously generated consumption process. In this model economy, households try to smooth their consumption stream in response to technology shocks, and the fluctuations in consumption determine asset prices. Capital adjustment costs induce that the supply of new capital via household investment and consumption decisions is not perfectly flexible. We pursue a matching approach that acknowledges that parts of the DSGE asset pricing model are evidently misspecified, while others claim to reflect economic reality. Specifically, we focus our estimation efforts on those parameters and parts of the DSGE asset pricing model that are assumed to reflect economic reality: its asset pricing implications. We thereby build a bridge between the data generated by the DSGE asset pricing model and the real-world data. The results of our empirical analysis are not unfavorable for the DSGE asset pricing approach, but they also indicate that the very positive interpretation of calibration results, in particular regarding the resolution of asset pricing puzzles, should be taken with a grain of salt.

In the fourth chapter, *Econometric Analysis of Partial Equilibrium Asset Pricing Models: A Two-Step Partial Indirect Inference Approach*, we propose a two-step estimation strategy for partial equilibrium asset pricing models. In partial equilibrium asset pricing models, the technology processes are modeled to provide consumption and dividend dynamics in an exogenous manner. This part of the model is independent of the specification of the representative agent's intertemporal decision. The consumption dynamics are then used in the asset pricing part of the model whose focus lies on specifying the representative agent's preferences given by the utility function specification. The two-step approach makes use of this structure. The technology parameters are estimated in a first step using macroeconomic time series. In a second step, the asset pricing part is estimated using financial data and taking the technology parameters from the first step as given. Using the two-step PII strategy, we focus on the consistent estimation of model parameters that are economically of interest. The other parameters that stem from parts of the model that are potentially misspecified are secondary and, under certain circumstances, we may choose to calibrate them. In case of

partial equilibrium models, the asset pricing part of the model yields testable restrictions. Hence, we are interested in the consistent estimation of the preference parameters that pertain to this part of the model. We estimate and assess the habit persistence model of Campbell and Cochrane (1999) and the long-run risk model proposed by Bansal and Yaron (2004). We propose a new estimation strategy for partial equilibrium asset pricing models and provide asymptotic results for the proposed strategy. It combines a two-step estimation approach based on the model structure with partial indirect inference. The two-step partial indirect inference approach enables an evaluation of the models that reflects both the model structure via the two-step approach and the modeling focus by using the PII methodology.

References

Bansal, R. and A. Yaron (2004). "Risks for the long run: A potential resolution of asset pricing puzzles". *The Journal of Finance* 59, 1481–1509.

Basel Committee (2016). *Minimum capital requirements for Market Risk*. Tech. rep. Available at http://www.bis.org/bcbs/publ/d352.pdf. Bank for International Settlements.

— (2017). *Pillar 3 disclosure requirements – consolidated and enhanced framework*. Tech. rep. Available at http://www.bis.org/bcbs/publ/d400.pdf. Basel Committee on Banking Supervision.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005). "Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach". *The Quarterly Journal of Economics* 120. (1), 387–422.

Campbell, J. Y. and J. H. Cochrane (1999). "By force of habit: A consumption-based explanation of aggregate stock market behavior". *The Journal of Political Economy* 107, 205–251.

Chen, A. Y. (2017). "External habit in a production economy: A model of asset prices and consumption volatility risk". *The Review of Financial Studies* 30. (8), 2890–2932.

Dimitriadis, T. and J. Schnaitmann (2020). "Forecast Encompassing Tests for the Expected Shortfall". *International Journal of Forecasting (forthcoming)*.

Dridi, R., A. Guay, and E. Renault (2007). "Indirect inference and calibration of dynamic stochastic general equilibrium models". *Journal of Econometrics* 136. (2), 397–430.

Fissler, T. and J. F. Ziegel (2016). "Higher order elicitability and Osband's principle". *Annals of Statistics* 44. (4), 1680–1707.

Zusammenfassung

Diese Dissertation beschäftigt sich mit Herausforderungen in der modernen Zeitreihenökonometrie. In den verschiedenen Kapiteln entwickeln wir neue Schätzmethoden für Forschungsfragen unterschiedlicher Bereiche: Im ersten Projekt verwenden wir hochdimensionelle Daten in einer makroökonomischen Anwendung. Im zweiten Projekt evaluieren wir Prognosen des Risikomaßes Expected Shortfall im finanzökonometrischen Kontext. Das dritte und vierte Projekt beschäftigen sich mit der Schätzung von makroökonomischen Finanzmodellen, die durch viele strukturelle Parameter beschrieben werden.

Heutzutage stehen große Datenmengen zur Verfügung, mit deren Hilfe ökonomische Zusammenhänge besser untersucht werden können. Im ersten Projekt betrachten wir zahlreiche makroökonomische und finanzwirtschaftliche Zeitreihen. Mithilfe eines „*sparse factor model*“ extrahieren wir die wesentlichen Informationen und erhalten ökonomisch interpretierbare Faktoren anhand der „*sparsity*“-Struktur. Diese Faktoren verwenden wir für strukturelle Analysen: mit unserem regularisierten „*factor-augmented vector autoregressive*“ (RFAVAR)-Modell untersuchen wir den Einfluss eines geldpolitischen Schocks auf eine Vielzahl makro- und finanzökonomischer Zeitreihen.

Im finanzwissenschaftlichen Kontext hat die Aufnahme des Expected Shortfalls als primäres Marktrisikomaß in der internationalen Bankenregulierung neue Forschungsfelder geschaffen. Das zweite Projekt setzt sich mit der Bewertung, dem Vergleich und der Kombination verschiedener Expected Shortfall Prognosen auseinander. Dafür entwickeln wir einen M-Schätzer für die Prognosekombinationsgewichte und führen drei „*forecast encompassing*“ Tests ein, die auf dem Wald-Testprinzip basieren.

Das dritte und vierte Projekt befassen sich mit der Schätzung von Modellen aus der strukturellen „*asset pricing*“ und „*macro-finance*“ Literatur. Ziel dieser Modelle ist die Abbildung eines bestimmten ökonomischen Mechanismus, der durch viele strukturelle Parameter beschrieben wird. Die Modelle beinhalten eine Beschreibung der Technologieprozesse und eine Modellierung der Haushaltsentscheidungen. Der Fokus in der „*asset*

pricing“ Literatur liegt auf der Spezifikation der intertemporalen Entscheidung der Haushalte. Diese Entscheidung wirkt sich über den Preismechanismus auf Finanzmärkten aus. Die Beschreibung des realen Geschäftszyklus und die damit verbundenen Schwankungen in der Produktion sind sekundär und häufig fehlspezifiert. Sie beschreiben in diesen Modellen die Realität nur unzureichend und sind als Karikatur der Realität zu verstehen. Das dritte und vierte Projekt beschäftigen sich mit der Schätzung der strukturellen Parameter, die für die Beschreibung des ökonomischen Kernmechanismus - des Preisbildungsmechanismus - relevant sind. Unter Berücksichtigung der Schätzungenauigkeit der Parameter erfolgt eine kritische Bewertung der Modelle. Die Schätzmethoden basieren auf dem Konzept des „*partial indirect inference*“, das uns die konsistente Schätzung der relevanten Parameter ermöglicht, obgleich sie aus einem fehlspezifizierten Modell stammen. Im dritten Projekt steht die kritische Bewertung eines kürzlich vorgeschlagenen „*asset pricing*“-Modells im Vordergrund. Das vierte Projekt legt den Fokus auf die vorgeschlagene Schätzmethode und untersucht welchen Einfluss die Berücksichtigung der Modellfehlspezifikation auf die Schätzungenauigkeit der strukturellen Parameter hat.

Die vier Kapitel der kumulativen Thesis enthalten die Ergebnisse aus vier separaten Forschungspapieren, die ich im Rahmen meiner Doktorarbeit an der Universität Konstanz durchgeführt habe. Das erste Kapitel, *A Regularized Factor-Augmented Vector Autoregressive Model*, ist ein gemeinsames Projekt mit Dr. Maurizio Daniele. Das zweite Kapitel, *Forecast Encompassing Tests for the Expected Shortfall*, ist in Zusammenarbeit mit Dr. Timo Dimitradis entstanden. Das dritte Kapitel, *Empirical Asset Pricing in a DSGE Framework: Reconciling Calibration and Econometrics using Partial Indirect Inference*, ist eine Kooperation mit Prof. Dr. Joachim Grammig und Dalia Elshiaty von der Eberhard Karls Universität Tübingen. Das vierte Kapitel, *Econometric Analysis of Partial Equilibrium Asset Pricing Models: A Two-Step Partial Indirect Inference Approach*, ist das Ergebnis meiner alleinigen Forschung.

Im ersten Kapitel, *A Regularized Factor-Augmented Vector Autoregressive Model*, analysieren wir den dynamischen Einfluss eines strukturellen geldpolitischen Schocks auf die Wirtschaft. In der klassischen Literatur werden solche Effekte mithilfe von vektorautoregressiven Modellen untersucht, die nur wenige Variablen beinhalten. Dadurch entstehen oft irreführende Impuls-Antwort-Funktionen, die durch einen Mangel an Informationen herbeigeführt werden. Das von Bernanke u. a. (2005) eingeführte „*factor-augmented vector autoregressive*“ (FAVAR)-Modell ermöglicht eine erschöpfendere Nutzung des Informationsgehalts vieler ökonomischer Zeitreihen. Es zerlegt die Zeitreihen in zwei Komponenten, eine gemeinsame und eine zeitreihenspezifische Komponente. Die gemeinsame Komponente besteht aus Faktoren und der damit verbundenen Gewichtung. Die Gewichtung beschreibt, wie groß der Einfluss der Zeitreihe auf die Faktoren ist. Das Faktormodell komprimiert

den Informationsgehalt der vielen Zeitreihen zu wenigen Faktoren. Diese Faktoren werden anstatt einiger ausgewählter Zeitreihen im vektorautoregressiven Kontext verwendet, um dynamische Strukturen abzubilden. Wir entwickeln eine regularisierte Version des FAVAR-Modells, welche eine „*sparsity*"-Struktur in den Faktorgewichtungsmatrizen erlaubt. Das regularisierte FAVAR-Modell ermöglicht uns, Faktoren zu extrahieren, die nur von einer Teilmenge der zugrundeliegenden Zeitreihen beeinflusst werden. Diese Teilmengen können mit ökonomischen Sektoren assoziiert werden, wodurch eine ökonomische Interpretation der Faktoren ermöglicht wird. Die Verwendung dieser Faktoren im vektorautoregressiven Kontext erlaubt uns, den Einfluss von strukturellen Schocks durch Impulsantworten auf diese Faktoren und alle zugrundeliegenden Zeitreihen zu untersuchen. Das Kapitel lässt sich folgendermaßen zusammenfassen: wir schlagen ein „*regularized factor-augmented vector autoregressive*" (RFAVAR)-Modell vor, bei dem die Gewichtungsmatrix der Faktoren eine „*sparsity*"-Struktur hat. Diese Struktur impliziert, dass die Faktoren durch eine Teilmenge der unterliegenden Zeitreihen formiert werden. Dies ermöglicht die Identifikation der Faktoren und erlaubt deren ökonomische Interpretation. Mithilfe dieser Methode können wir den Einfluss von strukturellen makroökonomischen Schocks auf ökonomisch interpretierbare Faktoren und alle zugrundeliegenden Zeitreihen untersuchen. Wir beweisen die Konsistenz der Schätzer für die Faktorengewichtungsmatrix, die Faktoren, die Kovarianzmatrix der zeitreihenspezifischen Einflüsse der Zeitreihen und aller autoregressiven Parameter des dynamischen Modells. In einer empirischen Anwendung untersuchen wir den Einfluss von geldpolitischen Schocks auf eine breite Masse von ökonomisch relevanten Zeitreihen. Wir identifizieren diesen Schock mit einem gemeinsamen Identifikationsschema für das Faktormodell und die strukturellen Innovationen des vektorautoregressiven Modells. Die resultierenden Impuls-Antwort-Funktionen sind für die extrahierten Faktoren und die beobachtbaren Zeitreihen ökonomisch plausibel.

Im zweiten Kapitel, *Forecast Encompassing Tests for the Expected Shortfall*, entwickeln wir drei „*forecast encompassing*" Tests für das Risikomaß Expected Shortfall (ES). Das Risikomaß ES wurde in das Basel III Abkommen (Basel Committee, 2016, 2017) als primäres Marktrisikomaß in die internationale Bankenregulierung aufgenommen. Dadurch entstand die Notwendigkeit, die Prognosefähigkeit konkurrierender ES-Prognosen zu evaluieren. Das Risikomaß ES, evaluiert am Signifikanzniveau α, ist definiert als der bedingte Erwartungswert der Renditen, die kleiner als das damit verbundene α-Quantil (VaR) sind. Wir vergleichen ES-Prognosen mithilfe von „*encompassing*"-Tests. Ein „*encompassing*"-Test basiert auf der Nullhypothese, dass eine einzelne ES Prognose so gut ist, wie eine beliebige (lineare) Kombination der ES Prognosen. Dabei untersucht man, ob der Gewichtungsfaktor der zweiten Prognose statistisch von null verschieden ist. Ist dieser statistisch von null verschieden, so

liefert die zweite Prognose zusätzliche Informationen, die in der ersten nicht enthalten sind. Wir verwenden eine von Fissler und Ziegel (2016) eingeführte gemeinsame Verlustfunktion für das Paar ES und VaR und führen drei ES-encompassing-Testvarianten ein. Diese Tests bauen auf einer asymptotischen Theorie auf, die Fehlspezifikationen der Quantilsgleichung korrigiert. Desweiteren untersuchen wir mit Simulationsstudien die Eigenschaften unserer Tests bei endlicher Stichprobengröße. Zuletzt verwenden wir die „*encompassing*"-Tests, um das Potenzial von kombinierten Prognosen für verschiedene Finanzprodukte hervorzuheben. Dieses Forschungspapier ist im *International Journal of Forecasting* zur Publikation angenommen (Dimitriadis und Schnaitmann, 2020).

Die verbleibenden Kapitel befassen sich mit der konsistenten Schätzung und kritischen Beurteilung von strukturellen „*asset pricing*"- und „*macro-finance*"-Modellen. Diese Modelle untersuchen die Schwankungen in der Bepreisung von Finanzprodukten und erheben nicht den Anspruch, eine allgemeingültige Erklärung aller ökonomischen Prozesse abbilden zu können. So müssen nicht zentrale Betrachtungsgegenstände, wie die Technologieprozesse, nicht detailgetreu abgebildet werden. Sie können fehlspezifiziert sein. Im dritten und vierten Kapitel befassen wir uns mit der Schätzung der Modellparameter, mit dem Ziel einer kritischen Beurteilung dieser Modelle. Um die Fehlspezifikationen in den Modellen zu berücksichtigen wenden wir den von Dridi u. a. (2007) vorgeschlagenen „*partial indirect inference*" (PII)-Ansatz an, eine simulationsbasierte Methode, und adaptieren ihn für unsere Zwecke. Dies erlaubt uns zu berücksichtigen, dass Teile eines Modells fehlspezifiziert sind und andere den Anspruch erheben, finanzökonomische Zusammenhänge korrekt abzubilden. Wir bringen die PII-Strategie zur Anwendung, indem wir Bindefunktionen erzeugen, die eine konsistente Schätzung der elementaren strukturellen Parameter ermöglichen.

Im dritten Kapitel, *Empirical Asset Pricing in a DSGE Framework: Reconciling Calibration and Econometrics using Partial Indirect Inference*, evaluieren wir das „*dynamic stochastic general equilibrium (DSGE) asset pricing*"-Modell, das von Chen (2017) in die Literatur eingeführt wurde. Diese kritische Bewertung erfolgt, indem wir die elementaren strukturellen Parameter mit PII schätzen. Die hier behandelte Klasse der DSGE Modelle verbindet die Preisbildung von Finanzprodukten mit einem exogenen Technologieprozess und einem endogen durch das Modell erzeugten Konsumprozess. Der Produktionsprozess und dadurch auch der Konsumprozess werden durch Technologieschocks beeinflusst. In dieser Modellökonomie versuchen Haushalte ihren Konsum über die Zeit zu glätten, indem sie konsumieren oder sparen. Dadurch beeinflussen Fluktuationen im Konsum die Preise von Finanzprodukten. Kapitaltransaktionskosten führen dazu, dass das Angebot von neuem Kapital durch Investitionen der Haushalte und die Konsumentscheidung nicht perfekt flexibel sind. Der für die Parameterschätzung verwendete „*matching*"-Ansatz erlaubt es, dass ein

Teil des Modells die ökonomische Realität abbildet, während ein anderer Teil des Modells fehlspezifiziert ist. Der Fokus der Schätzung liegt auf den Parametern, die Teile des Modells beschreiben, welche ökonomische Realität abbilden: der Preisbildungsmechanismus des Modells. Dadurch bauen wir eine Brücke zwischen den Daten, die aus dem DSGE asset pricing Modell generiert sind und den beobachtbaren Zeitreihen. Die Ergebnisse der empirischen Analyse widerlegen den „*DSGE asset pricing*"-Ansatz nicht. Allerdings zeigen sie auch, dass eine vorsichtigere Interpretation der Kalibrierungsergebnisse notwendig ist, insbesondere was ihr Potenzial angeht alle *asset pricing puzzles* zu lösen.

Im vierten Kapitel, *Econometric Analysis of Partial Equilibrium Asset Pricing Models: A Two-Step Partial Indirect Inference Approach*, schlagen wir eine zweistufige Schätzmethode für „*partial equilibrium asset pricing*"-Modelle vor. In „*partial equilibrium asset pricing*"-Modellen wird die Konsum- und Dividendendynamik exogen modelliert. Diese Beschreibung der Technologie ist unabhängig von der intertemporalen Entscheidung des repräsentativen Agenten. Die Schwankungen im Konsum finden im finanzwissenschaftlichen Teil des Modells Anwendung. Dieser legt den Modellierungsfokus auf die Präferenzen des repräsentativen Agenten, die in der Nutzenspezifikation Ausdruck findet. Der zweistufige Schätzansatz basiert auf dieser Struktur: In einem ersten Schritt werden die Technologieparameter mithilfe von makroökonomischen Zeitreihen geschätzt. Im zweiten Schritt werden die so gewonnenen Makroparameter als gegeben hingenommen und die Parameter des finanzwissenschaftlichen Teil des Modells unter Zuhilfenahme von Finanzdaten geschätzt. Unser zweistufiger PII-Ansatz fokussiert sich auf die konsistente Schätzung von ökonomisch relevanten strukturellen Parametern. Die restlichen Modellparameter beschreiben Teile des Modells, die potenziell fehlspezifiziert sind. Unter bestimmten Voraussetzungen können diese kalibriert werden ohne die konsistente Schätzung der anderen Parameter zu beeinflussen. Bei „*partial equilibrium asset pricing*"-Modellen vertreten wir die Ansicht, dass der finanzwissenschaftliche Teil des Modells korrekt spezifiziert ist und testbare Hypothesen liefert. Deshalb sind wir an der Schätzung der Präferenzparameter interessiert, die diesen Teil des Modells beschreiben. Wir schätzen und evaluieren das *habit persistence model*, das von Campbell und Cochrane (1999) eingeführt wurde und das *long-run risk model*, das von Bansal und Yaron (2004) vorgeschlagen wurde. Wir führen eine neue Schätzstrategie für partielle Gleichgewichtsmodelle im „*asset pricing*"-Kontext ein und leiten asymptotische Theorie für diese her. Die Schätzstrategie kombiniert einen zweistufigen Schätzansatz basierend auf der Modellstruktur mit dem PII-Ansatz. Der PII-Ansatz teilt die Modellparameter danach auf, ob sie ein Teil des Modells darstellen, der die ökonomische Realität abbildet oder ob sie aus Teilen des Modells stammen, die nur notwendig sind, um Daten aus dem Modell zu generieren, aber nicht die ökonomische Wahrheit reflektiert. Diese Aufteilung der

Parameter beeinflusst die asymptotische Verteilung der geschätzten Parameter und deren Konfidenzintervalle. Der zweistufige „*partial indirect inference*“-Ansatz ermöglicht bei der Evaluierung der Modelle sowohl die Modellstruktur als auch den Fokus des Modells zu berücksichtigen.

Literatur

Bansal, R. and A. Yaron (2004). “Risks for the long run: A potential resolution of asset pricing puzzles”. *The Journal of Finance* 59, 1481–1509.

Basel Committee (2016). *Minimum capital requirements for Market Risk*. Tech. rep. Available at http://www.bis.org/bcbs/publ/d352.pdf. Bank for International Settlements.

— (2017). *Pillar 3 disclosure requirements – consolidated and enhanced framework*. Tech. rep. Available at http://www.bis.org/bcbs/publ/d400.pdf. Basel Committee on Banking Supervision.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005). “Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach”. *The Quarterly Journal of Economics* 120. (1), 387–422.

Campbell, J. Y. and J. H. Cochrane (1999). “By force of habit: A consumption-based explanation of aggregate stock market behavior”. *The Journal of Political Economy* 107, 205–251.

Chen, A. Y. (2017). “External habit in a production economy: A model of asset prices and consumption volatility risk”. *The Review of Financial Studies* 30. (8), 2890–2932.

Dimitriadis, T. and J. Schnaitmann (2020). “Forecast Encompassing Tests for the Expected Shortfall”. *International Journal of Forecasting (forthcoming)*.

Dridi, R., A. Guay, and E. Renault (2007). “Indirect inference and calibration of dynamic stochastic general equilibrium models”. *Journal of Econometrics* 136. (2), 397–430.

Fissler, T. and J. F. Ziegel (2016). “Higher order elicitability and Osband’s principle”. *Annals of Statistics* 44. (4), 1680–1707.

Chapter 1

A Regularized Factor-Augmented Vector Autoregressive Model

This project is joint work with Maurizio Daniele. We thank Ralf Brüggemann, Mehmet Caner, and the participants at the $(EC)^2$ 2018 in Rome, the QFFE 2019 in Marseille and the IAAE 2019 in Nicosia for useful comments on earlier versions of the paper. Financial support by the Graduate School of Decision Sciences (GSDS) and the German Science Foundation (DFG) grant number BR 2941/3-1 is gratefully acknowledged.

1.1. Introduction

The dynamic effects of structural shocks on the economy are conventionally analyzed using small scale vector autoregressive (VAR) models. These models typically incorporate only few variables which often leads to informational deficiencies that may result in misleading impulse response patterns. To resolve this problem, Bernanke et al. (2005) introduce the factor-augmented vector autoregressive (FAVAR) model which enables the use of the informational content of many time series for structural analysis.[1] A FAVAR model decomposes the co-movement of the observed time series into a common and idiosyncratic component. The common component is composed of latent and observed factors that affect the underlying time series according to weights represented by the factor loadings matrix. The idiosyncratic innovations allow for weak cross-sectional and serial correlations in the spirit of Chamberlain and Rothschild (1983). As the number of latent and observed factors is much smaller than the number of time series included, the information contained in a large panel of variables is condensed into a small number of factors. To account for dynamics, the FAVAR model incorporates autoregressive structures for both types of factors.

There is an identification problem associated with the FAVAR approach. In contrast to the observed factors that can be interpreted, restrictions have to be imposed on the FAVAR model to identify the latent factors and factor loadings. In the context of structural analysis, it is crucial to consider an identification scheme that allows an economic interpretation of the factors.

In this paper, we propose a regularized factor-augmented vector autoregressive (RFAVAR) model. The regularization is imposed on the factor loadings and allows to retrieve factors which may load only on a subset of the underlying time series. This approach yields factors which can be identified by the corresponding subsets. In the context of economic data, our approach provides factors which can be associated with different sectors of the economy such that they can be interpreted economically. The RFAVAR approach extends a vector autoregressive (VAR) model by these factors. The resulting model can be used to investigate the effects of a structural shock in the presence of many observed time series using impulse response functions. We show that the sparsity in the factor loadings yields the necessary restrictions to identify the RFAVAR model, as well as some structural innovations in the dynamic part.

Our paper contributes to the literature on the estimation and identification of the FAVAR model. Frequentist estimation approaches commonly used in the literature are either based

[1] Dynamic factor models (DFM) introduced and investigated by e.g. Geweke (1977), Sargent, Sims, et al. (1977), Forni et al. (2000), Stock and Watson, (2002, 2005, 2011) and Forni and Gambetti (2010) constitute an alternative to FAVAR models.

on two-step principal component analysis (PCA) (Bernanke et al. (2005)), iterated PCA and least squares (Boivin et al. (2009)) or a likelihood-based two-step method (Bai et al. (2016)). Our RFAVAR model builds on a likelihood-based approach. The analysis of structural shocks exploits the economic interpretability of the latent factors. Usually, a named factor scheme is used to identify the latent factors. This scheme associates each latent factor with a unique observed time series. However, these time series may not represent an entire economic sector appropriately. Furthermore, the naming time series potentially impose implausible relations between the factors and the remaining observed time series that affect the structure of the factor model and its dynamics. Examples of named factor identification schemes can be found in Bernanke et al. (2005), Stock and Watson (2016) or Bai et al. (2016). The RFAVAR approach proposes an alternative identification strategy that relies on the imposed regularization on the factor loadings.

Macroeconomic and financial data sets are typically composed of several subpanels that collect time series from different economic sectors (see e.g. Hallin and Liška (2011)). Hence, these large panels of time series implicate a block structure and imposing sparsity allows us to retrieve factors that represent the underlying structure in the economic data set.

In this paper, we propose a regularized FAVAR model that allows for sparsity in the factor loadings matrices of the latent and observed factors. Factors may only load on a subset of variables which represent economic sectors. This allows to identify and interpret the factors economically without imposing restrictive identification restrictions through a named factor scheme. Our estimation procedure relies on a penalized quasi-maximum likelihood approach and is based on L_1-norm regularization of the factor loadings. The estimation of the loadings for the latent and observed factors is conducted in a single step, allowing for a similar degree of shrinkage for both types of factors, which is necessary to retain their structural interpretation.

We prove consistency for the estimators of the factor loadings, the factors and the covariance matrix of the idiosyncratic component under the average Frobenius norm. Moreover, the autoregressive parameters are consistently estimated as well. The asymptotic results hold for pervasive and weak factors. We also study the finite sample properties of the proposed estimator. Using data generating processes with dense and sparse factors, we assess the correct detection of zero loadings, evaluate the accuracy of the factor loadings estimates and measure the goodness-of-fit of the coefficients of the dynamic equation. The simulation results show that the RFAVAR model precisely estimates the true underlying factor loadings structure across the simulation designs.

Using the RFAVAR approach, the impact of structural shocks is investigated on economically interpretable factors and observed time series using impulse response functions. We

propose a joint identification scheme of the factor model and the structural innovations in the VAR model. More specifically, the scheme focuses on the identification of shocks to the observed factors.

In the empirical literature, FAVAR models have been used to investigate structural monetary policy, fiscal or oil price shocks (see e.g. Bernanke et al. (2005), Boivin and Giannoni (2007), Mumtaz and Surico (2009), Kilian and Lewis (2011) and Stock and Watson (2016)). Using monthly US macroeconomic data, we apply our framework to investigate the effects of a monetary policy shock on the factors and the underlying time series. We focus on a shock to the structural innovation of the Federal Funds rate (FFR). We treat the FFR as an observed factor (c.f. Bernanke et al. (2005) or Boivin et al. (2009)). Using the RFAVAR approach, we identify the factors in a data-driven manner and extract five latent factors that relate to the labor market, prices, industrial production, the stock market and credit spreads. The resulting impulse response functions are in line with economic rationale, both on the factor aggregates and the observed time series level. In particular, following a tightening of monetary policy, prices and industrial production fall, credit conditions deteriorate and the level of employment decreases. Most notably, the prize puzzle typically found in small scale VAR models (e.g. Sims (1992) and Ramey (2016)) is no longer present. These results are similar to those obtained by Forni and Gambetti (2010) in a structural dynamic factor model framework. However, the latter is very sensitive to changes in the model specification in comparison to our regularized FAVAR model.

The remainder of this paper is organized as follows. In Section 1.2, we describe the proposed regularized FAVAR approach and the estimation procedure, where we impose shrinkage on the factor loadings of both, the observed and latent factors. We further provide an identification scheme that jointly identifies the factor model and the structural innovations. In Section 1.3, the asymptotic properties of the sparse factor loadings, the latent factor estimator, the covariance matrix of the idiosyncratic innovations, as well as the coefficients of the dynamic equation are provided. We study the finite sample properties of our regularized FAVAR estimator in Section 1.4. Section 1.5 presents the results of an empirical application that investigates the effects of a monetary policy shock on the US economy. Finally, Section 1.6 concludes.

1.2. Econometric framework

In this section, we introduce the regularized FAVAR approach, discuss its estimation procedure and provide a scheme for the model identification.

1.2.1. A regularized factor-augmented vector autoregressive model

We define the factor-augmented vector autoregressive model as

$$x_t = \Lambda^f f_t + \Lambda^g g_t + e_t, \tag{1.2.1}$$

$$\begin{bmatrix} f_t \\ g_t \end{bmatrix} = \begin{bmatrix} \Phi^{ff}(L) & \Phi^{fg}(L) \\ \Phi^{gf}(L) & \Phi^{gg}(L) \end{bmatrix} \begin{bmatrix} f_{t-1} \\ g_{t-1} \end{bmatrix} + \begin{bmatrix} \eta_t^f \\ \eta_t^g \end{bmatrix}, \quad t = 1, \ldots, T, \tag{1.2.2}$$

where x_t is a $(N \times 1)$ vector of the observable time series, f_t is a $(r_1 \times 1)$ vector of latent factors, g_t is a $(r_2 \times 1)$ vector of observed factors, Λ^f is a $(N \times r_1)$ matrix of factor loadings of the latent factors, and Λ^g is a $(N \times r_2)$ matrix of factor loadings of the observed factors. e_t is a zero mean $(N \times 1)$ vector of idiosyncratic innovations which may be cross-sectionally and serially correlated, as in the approximate factor model framework. Their covariance matrix is given by $\Sigma_e = \mathrm{E}\left[e_t e_t'\right]$. Moreover, η_t^f denotes a $(r_1 \times 1)$ and η_t^g a $(r_2 \times 1)$ vector of factor innovations associated with the latent and observed factors, respectively, and $\Phi^{ff}(L)$, $\Phi^{fg}(L)$, $\Phi^{gf}(L)$ and $\Phi^{gg}(L)$ are lag polynomials of order p. We define the $(r \times 1)$ vector of latent and observed factors as $h_t = \left[f_t', g_t'\right]'$, where $r = r_1 + r_2$.

In the FAVAR specification the latent factors are not identified without further restrictions. The factor interpretability hinges on the identification restrictions imposed on the factor model. For an exact identification of the model, $r_1^2 + r_1 r_2$ restrictions have to be imposed, see e.g. Bernanke et al. (2005) or Bai et al. (2016). Among alternative identification schemes, Bai et al. (2016) analyze the named factor scheme which yields estimated factors that can be interpreted economically. Intuitively, the idea is to re-order the underlying observed time series such that the first r_1 time series define the latent factors. It is assumed that these time series only load onto one factor.[2] Hence, the ordering of the time series is crucial for the structure of the factor model.

We propose a regularized FAVAR model which shrinks single elements of the factor loadings matrices to zero. The resulting sparsity pattern allows us to retrieve the underlying sectors in economic data sets. This leads to a factor loadings matrix which is economically interpretable and enables us to attribute economic meaning to the estimated factors. Our RFAVAR model circumvents any a priori selection of the factors compared to the named factor scheme and is independent of the ordering of the data. Moreover, by introducing sparsity, we can incorporate weak factors which only load on a subset of the time series. As shown by Onatski (2012), it is crucial to account for weakly influential factors in the

[2]The resulting factor loadings matrix is given by $\Lambda_{NF}^f = \left[I_{r_1}, \Lambda^{f\cdot\prime}\right]'$, where I_{r_1} is a $(r_1 \times r_1)$ identity matrix and $\Lambda^{f\cdot} = \Lambda_2^f \left(\Lambda_1^f\right)^{-1}$, with $\Lambda^f = \left[\Lambda_1^{f\prime}, \Lambda_2^{f\prime}\right]'$ and $\Lambda_1^{f\prime}$ denotes the upper $(r_1 \times r_1)$ block.

estimation procedure as conventional methods (e.g. principal component analysis (PCA)) estimate these factors with large errors.

A second issue that we target is the estimation of the regularized FAVAR model. The complication in this model arises because it consists of latent and observable factors. This implies that the conventionally used methodologies to extract the latent factors do not work in presence of the observed factors. The effect of the observed factors in explaining the covariance structure of the data has to be accounted for. The approaches introduced in the literature have tackled this problem in various ways. Bernanke et al. (2005) propose a two-step principal components approach to estimate the FAVAR model which does not take into account that the observed factors also contribute to the factor space used to estimate the latent factors. Bai et al. (2016) address this issue by introducing a two-step quasi maximum likelihood (QML) approach to estimate the FAVAR model. In a first step, they extract the impact of the observed factors by linearly projection and estimate the latent factors, $\tilde{f}_t$ and their factor loadings, $\tilde{\Lambda}^f$ using QML. In a second step, they obtain the factor loadings of the observed factors, $\tilde{\Lambda}^g$ by regressing $(x_t - \tilde{\Lambda}^f \tilde{f}_t)$ on g_t.

We propose a procedure for the joint estimation of the factors loadings of both types of factors. Hence, the informational content explained by the observed factors is directly accounted when estimating the latent factors. This allows us to jointly impose shrinkage on the factor loadings estimates which leads to a similar degree of shrinkage for both types of factors.

1.2.2. Estimation of the regularized FAVAR model

The factors and factor loadings in equation (1.2.1) can be estimated by either PCA[3] or quasi maximum likelihood (QML) estimation under normality.[4] In the following we pursue estimating the factor model by QML, which allows us to introduce sparsity in the factor loadings by penalizing the likelihood function.

Intuitively, we can estimate the factor loadings of the observed and latent factors, Λ^f and Λ^g, jointly using the covariance matrix of the observed series, x_t. In matrix notation, the observation equation of the model is defined as

$$X = \Lambda^f F' + \Lambda^g G' + e = \begin{bmatrix} \Lambda^f & \Lambda^g \end{bmatrix} \begin{bmatrix} F' \\ G' \end{bmatrix} + e = \Lambda H + e, \tag{1.2.3}$$

[3] See e.g. Bai and Ng (2002) or Stock and Watson (2002c) for a detailed treatment of the PCA, in approximate factor models.

[4] Bai and Li (2012) deal with the consistent estimation of the strict factor model, whereas Bai and Li (2016) analyze the approximate factor model estimation by QML.

where $X = [x_1, \ldots, x_T]$ and $e = [e_1, \ldots, e_T]$ are $(N \times T)$ matrices, $F = \left[f_1, \ldots, f_{r_1}\right]$ and $G = \left[g_1, \ldots, g_{r_2}\right]$ are $(T \times r_1)$ and $(T \times r_2)$ matrices, respectively. The covariance matrix of the observed time series X based on the factor model in (1.2.3) can be written as $\Sigma = \Lambda \Sigma_H \Lambda' + \Sigma_e$, where Σ_H is the composite covariance matrix of the observed and latent factors and Σ_e denotes the covariance matrix of the idiosyncratic component.

Using the previous result, we obtain the following expression for the negative quasi log-likelihood function for the covariance matrix of the data in the FAVAR model

$$\mathcal{L}(\Lambda, \Sigma_H, \Sigma_e) = \log |\Lambda \Sigma_H \Lambda' + \Sigma_e| + \operatorname{tr}\left[S_x \left(\Lambda \Sigma_H \Lambda' + \Sigma_e\right)^{-1}\right], \tag{1.2.4}$$

where $S_x = \frac{1}{T} \sum_{t=1}^{T} x_t x_t'$ is the sample covariance matrix of the observed data. In the first estimation step, we treat Σ_e as a diagonal matrix and define $\Phi_e = \operatorname{diag}(\Sigma_e)$ denoting a diagonal matrix that contains only the main diagonal elements of Σ_e to reduce the number of estimated parameters. Thus, the unpenalized objective function reduces to

$$\mathcal{L}(\Lambda, \Sigma_H, \Phi_e) = \log |\Lambda \Sigma_H \Lambda' + \Phi_e| + \operatorname{tr}\left[S_x \left(\Lambda \Sigma_H \Lambda' + \Phi_e\right)^{-1}\right]. \tag{1.2.5}$$

As Σ_e incorporates correlations of general form, equation (1.2.5) may be seen as a quasi log-likelihood function, since it imposes the innovation term structure of a strict factor model. However, Bai and Li (2016) show in the approximate factor model framework that imposing this restrictions on Σ_e does not affect the consistency of the QML estimator. The diagonality assumption on Σ_e is relaxed in a second step based on the soft-thresholding estimator presented in Section 1.2.3.

In order to introduce sparsity in the factor loadings matrix Λ, we shrink each element of Λ towards zero. This is incorporated based on a penalized maximum likelihood estimation of the objective function in (1.2.5) by separate L_1-norm penalties on the factor loadings associated with the observed and latent factors, respectively. More specifically, we focus on the penalized optimization problem

$$\begin{aligned} \min_{\{\Lambda, \Phi_e\}} \; & \log |\Lambda \Sigma_H \Lambda' + \Phi_e| + \operatorname{tr}\left[S_x \left(\Lambda \Sigma_H \Lambda' + \Phi_e\right)^{-1}\right] \\ & + \mu_1 \sum_{i=1}^{N} \sum_{k=1}^{r_1} \left|\lambda_{ik}^{f}\right| + \mu_2 \sum_{i=1}^{N} \sum_{l=1}^{r_2} \left|\lambda_{il}^{g}\right|, \end{aligned} \tag{1.2.6}$$

where μ_1 and μ_2 determine the degree of penalization of the factor loadings corresponding to the latent and observed factors, respectively. A clear separation of both sets of factor loadings has the advantage of offering a more flexible treatment of both components.

The latent factors f_t can be estimated by generalized least squares (GLS)

$$\tilde{f}_t = \left(\tilde{\Lambda}^{f\prime}\tilde{\Phi}_e^{-1}\tilde{\Lambda}^f\right)^{-1}\tilde{\Lambda}^{f\prime}\tilde{\Phi}_e^{-1}x_t\,, \tag{1.2.7}$$

where the estimates $\tilde{\Lambda}^f$ and $\tilde{\Phi}_e$ are obtained from the optimization of the objective function in (1.2.6).

1.2.3. Estimation of the idiosyncratic component covariance matrix

As the diagonality assumption on Σ_e imposed in the first step of our estimation is restrictive, we introduce a procedure that relaxes this assumption and allows for the estimation of a possibly sparse idiosyncratic innovation covariance matrix. More specifically, we re-estimate Σ_e by means of the principal orthogonal complement thresholding (POET) estimator introduced by Fan et al. (2013). The POET estimator allows for sparsity in the idiosyncratic error covariance matrix by shrinking the off-diagonal elements of the sample covariance matrix of the residuals obtained from the estimation of our regularized FAVAR model towards zero using soft-thresholding. The estimated idiosyncratic error covariance matrix $\tilde{\Sigma}_e^\tau$ based on the POET method is defined as

$$\tilde{\Sigma}_e^\tau = s_{ij}^\tau, \quad s_{ij}^\tau = \begin{cases} s_{e,ii}, & i = j \\ \mathcal{S}(s_{e,ij},\tau), & i \neq j \end{cases},$$

where $s_{e,ij}$ is the ij-th element of the sample covariance matrix $S_e = \frac{1}{T}\sum_{t=1}^T(x_t - \tilde{\Lambda}\tilde{h}_t)(x_t - \tilde{\Lambda}\tilde{h}_t)'$ of the estimated factor model residuals, $\tilde{h}_t = \left(\tilde{f}_t', g_t'\right)'$ are the estimated factors, $\tau = \frac{1}{\sqrt{N}} + \sqrt{\frac{\log N}{T}}$ is a threshold and $\mathcal{S}(\cdot)$ denotes the soft-thresholding operator defined as

$$\mathcal{S}(s_{e,ij},\tau) = \text{sign}(s_{e,ij})(|s_{e,ij}| - \tau)_+\,. \tag{1.2.8}$$

1.2.4. Identification restrictions

The FAVAR model in (1.2.1) and (1.2.2) is not identified, as it can be expressed as

$$\begin{aligned} x_t &= \Lambda A^{-1}\cdot A\,h_t + e_t \\ Ah_t &= A\Phi_1A^{-1}\cdot Ah_{t-1} + \cdots + A\Phi_pA^{-1}\cdot Ah_{t-p} + A\eta_t, \end{aligned}$$

where $h_t = \left(f_t', g_t'\right)'$ and A denotes an invertible $[(r_1 + r_2)\times(r_1 + r_2)]$ matrix. Moreover, $\Lambda^* = \Lambda A^{-1}$, $h_{t-j}^* = Ah_{t-j}$ for all $j = 0,\ldots,p$, $\Phi_i^* = A\Phi_iA^{-1}$ for all $i = 1,\ldots,p$ and $\eta_t^* = A\eta_t$ are the uniquely identified quantities given a specific choice of A. Bai et al. (2016) show

that $r_1^2 + r_1 r_2$ restrictions are necessary to uniquely identify the FAVAR model. Hence, we impose restrictions on the factor model and the dynamics of the factors.

The covariance matrix of the (structural) VAR innovations $\eta_t^* = \left(\eta_t^{*f\prime}, \eta_t^{*g\prime}\right)'$ is given by

$$\Omega^* = \begin{bmatrix} \mathrm{E}\left[\eta_t^{*f}\eta_t^{*f\prime}\right] & \mathrm{E}\left[\eta_t^{*f}\eta_t^{*g\prime}\right] \\ \mathrm{E}\left[\eta_t^{*g}\eta_t^{*f\prime}\right] & \mathrm{E}\left[\eta_t^{*g}\eta_t^{*g\prime}\right] \end{bmatrix} = \begin{bmatrix} \Omega_{ff}^* & \Omega_{fg}^* \\ \Omega_{gf}^* & \Omega_{gg}^* \end{bmatrix}. \tag{1.2.9}$$

We consider the following sets of identification restrictions on the covariance matrix of the factor innovations and the factor model.[5]

IRa: $\Omega_{fg}^* = 0$.

IRb: $\Omega_{fg}^* = 0$ and $\Lambda_1 = I_{r_1}$, where Λ_1 is the upper $r_1 \times r_1$ submatrix of Λ^f.

Both identification schemes share the restriction that $\Omega_{fg}^* = 0$. This restriction assures that the observed factors g_t are not rotated. Moreover, as this assumption is imposed on the covariance matrix of the structural innovations this implies that the structural shocks associated with the observed factors are contemporaneously uncorrelated with those of the latent ones.

The first set of identifying restrictions *IRa* is used for our regularized FAVAR model and the condition in *IRa* contributes to $r_1 r_2$ restrictions. Hence, r_1^2 additional restrictions are necessary to fully identify our regularized FAVAR model. This rotational indeterminacy is resolved as long as a sufficient amount of sparsity is imposed on the factor loadings Λ that leads to a local identification of the model. Numerically we can verify, whether enough sparsity is imposed on the factor loadings matrix to achieve identification, by evaluating the rank condition for local identification proposed by Bekker (1986). In fact, if the missing restrictions are covered by zero restrictions in the factor loading matrix, our regularized FAVAR model is identified up to a unitary generalized permutation matrix, by the combination of *IRa* and the L_1-norm penalties on Λ. Hence, the regularized FAVAR model is fully identified, by ordering the columns of the estimated factor loadings matrix $\tilde{\Lambda}$ according to their sparsity and by fixing the sign of each column. We choose the sign of the estimated factors such that they align to the corresponding observable time series.

It is important to note that even for arbitrarily small penalty parameters μ_1, μ_2, the optimization problem in (1.2.6) leads to estimated parameters that minimize the L_1-norm of the factor loadings. However, if the selected penalty parameters are too small, it is likely that the estimated factor loadings are shrunken towards zero, but are not set exactly to zero.

[5] Similar identification restrictions are common in the factor model literature, see e.g. Bai and Ng (2013) or Bai and Wang (2014).

In this special case, the L_1-norm penalty does not resolve the rotational indeterminacy and additional restrictions are necessary to identify the regularized FAVAR model.[6] Moreover, there exist cases in which there is a sufficient amount of sparsity (i.e. at least r_1^2 zeros) but the model is not locally identified. Consider the following special cases in which sparsity does not provide identification power.

(a) $\Lambda^f = \begin{bmatrix} \Lambda^a \\ \mathbf{0} \end{bmatrix}$ where Λ^a is a dense $(N_1 \times r)$ matrix and $\mathbf{0}$ is a $(N - N_1 \times r)$ matrix of zeros, where $(N - N_1) \geq r_1$.

(b) $\Lambda^f = \begin{bmatrix} \Lambda^b & \mathbf{0} \end{bmatrix}$ where Λ^b is a dense $(N \times r - r^*)$ matrix and $\mathbf{0}$ is a $(N \times r^*)$ matrix of zeros, where $1 \leq r^* < r$.

(c) If there exist multiple pervasive factors and weak factors, then the pervasive factors are not locally identified by the sparsity.

A possible solution to the rotational indeterminacy in the cases (a) and (b) is to remove the zero rows or columns and re-estimate the RFAVAR model on the reduced data set or with a smaller number of factors, respectively. If there is no sparsity imposed on the factor loadings matrix (i.e. $\mu_1 = \mu_2 = 0$) or the local identification condition is not satisfied, we need restrictions in addition to *IRa* to identify the FAVAR model. For this case, we use the normalization $\frac{1}{N}\Lambda'\Sigma_e^{-1}\Lambda = Q$, where the diagonal entries of Q are assumed to be distinct and arranged in a decreasing order and $\Sigma_F = \mathbb{E}[f_t f_t'] = I_{r_1}$. These restrictions are usually imposed in the maximum likelihood framework for the approximate factor model (see e.g. Lawley and Maxwell (1971)).[7] The second restriction *IRb* is conventionally[8] referred to as named factor identification. The first r_1 time series offer a direct identification for the latent factors and are assumed to only load on the specific factors, respectively. However, as a consequence the ordering of the observed time series matters crucially for this set of restrictions.

Based on the different sets of identification schemes the resulting estimates change. In scheme *IRa*, the identification restrictions are chosen such that the factor loadings corresponding to the latent factors $\tilde{\Lambda}^f$ are not rotated after their estimation. Hence, the estimated sparsity structure and the corresponding interpretation of $\hat{\Lambda}^f$ is not distorted

[6] The simulation studies illustrated in Section 1.4 confirm that the regularized FAVAR model consistently detects the sparsity structure in the true factor loadings, irrespective of the estimate of Λ, which initializes our estimation procedure outlined in Section 1.C.

[7] Alternatively, we could also impose the restriction $\Omega^*_{ff} = I_{r_1}, \Omega^*_{fg} = 0$ and $\frac{1}{N}\Lambda'\Sigma_e^{-1}\Lambda = Q$ (see e.g. identification restriction *IRa* in Bai et al. (2016)). This imposes an orthogonality restriction on the covariance matrix of the dynamic factor innovations. However, this version implies that the factor loadings of the latent factors are rotated.

[8] Restriction *IRb* is also employed by Bai et al. (2016).

by any rotation. More specifically, the factor loadings estimates are unchanged $\hat{\Lambda}^f = \tilde{\Lambda}^f$. The factor loadings associated with the observed factors are transformed according to $\hat{\Lambda}^g = \tilde{\Lambda}^g + \tilde{\Lambda}^f \tilde{\Omega}_{fg} \tilde{\Omega}_{gg}^{-1}$. Moreover, we rotate the estimated factors by $\hat{F} = \tilde{F} - \tilde{\Omega}_{fg} \tilde{\Omega}_{gg}^{-1} G$ and the autoregressive parameters by $\hat{\Phi}_i = \tilde{A} \tilde{\Phi}_i \tilde{A}^{-1}$, for $i = 1, \ldots, p$, with the rotation matrix $\tilde{A} = \begin{bmatrix} I_{r_1} & -\tilde{\Omega}_{fg} \tilde{\Omega}_{gg}^{-1} \\ 0 & I_{r_2} \end{bmatrix}$. In case of *IRb*, let $\tilde{\Lambda}_1$ be the first $(r_1 \times r_1)$ block of $\tilde{\Lambda}$. The resulting estimated factor loadings are $\hat{\Lambda}^f = \tilde{\Lambda} \tilde{\Lambda}_1^{-1}$ and $\hat{\Lambda}^g = \tilde{\Lambda}^g + \tilde{\Lambda}^f \tilde{\Omega}_{fg} \tilde{\Omega}_{gg}^{-1}$. The estimated factors are given by $\hat{F} = \left(\tilde{F} - \tilde{\Omega}_{fg} \tilde{\Omega}_{gg}^{-1} G \right) \tilde{\Lambda}_1'$ and the autoregressive parameters are obtained as $\hat{\Phi}_i = \tilde{A} \tilde{\Phi}_i \tilde{A}^{-1}$, for $i = 1, \ldots, p$, where we use the rotation matrix $\tilde{A} = \begin{bmatrix} \tilde{\Lambda}_1 & -\tilde{\Lambda}_1 \tilde{\Omega}_{fg} \tilde{\Omega}_{gg}^{-1} \\ 0 & I_{r_2} \end{bmatrix}$. Note that quantities with a '^' are estimates of the identified quantities denoted by a '*'.

1.2.5. Impulse responses

Starting from the VAR representation of the factors in equation (1.2.2), we obtain the vector moving average representation of the model and the impulse response functions. By rewriting the dynamic equation of the factors as $\left(I_r - \Phi_1 L - \cdots - \Phi_p L^p \right) h_t = \eta_t$ and using the inverted lag polynomial to write the factors h_t as a function of their innovations η_t, we obtain

$$h_t = \left(I_r - \Phi_1 L - \cdots - \Phi_p L^p \right)^{-1} \eta_t = \Psi(L) \eta_t,$$

where $\Psi(L) = I_r + \Psi_1 L + \Psi_2 L^2 + \cdots$ and the Ψ_i are moving average coefficients. Hence, $\Psi(L)$ is the conventional inverted lag polynomial and the Ψ_i can be interpreted as the matrices of responses to the innovations η_t. Furthermore, we accumulate the moving average parameters to analyze the effect of a shock in the factor innovations on the level of the factors, when the underlying observed time series x_t are given in first-differences.

In the factor-augmented VAR model, we can also calculate the responses of the observed time series by employing

$$x_t = \Lambda \Psi(L) \eta_t.$$

Impulse responses on the individual time series in levels are obtained by pre-multiplying the moving average coefficients Ψ_i by the factor loadings matrix Λ and accumulating them, when necessary.

Shocks to the factor innovations η_t cannot be interpreted as structural shocks as they are contemporaneously correlated. However, as we identify the factor model and the structural innovations jointly by the schemes described in Section 1.2.4, we obtain innovations with a block diagonal structure. The structural innovations are given by pre-multiplying the

factors and factor innovations by the rotation matrix. The contemporaneous impact matrix is obtained by inverting the rotation matrix and is given by

$$A^{-1} = \begin{bmatrix} I_{r_1} & \Omega_{fg}\Omega_{gg}^{-1} \\ 0 & I_{r_2} \end{bmatrix}, \tag{1.2.10}$$

where Ω_{fg} and Ω_{gg} are the corresponding elements of the covariance matrix of the factor innovations η_t, which is denoted as $\Omega = \begin{bmatrix} \Omega_{ff} & \Omega_{fg} \\ \Omega_{gf} & \Omega_{gg} \end{bmatrix}$. The covariance matrix of the structural innovations η_t^* is given by the block diagonal matrix $\Omega^* = \begin{bmatrix} \Omega_{f\cdot g} & 0 \\ 0 & \Omega_{gg} \end{bmatrix}$, where $\Omega_{f\cdot g} = \Omega_{ff} - \Omega_{fg}\Omega_{gg}^{-1}\Omega_{gf}$. This implies that a structural shock in the innovation of the observed factors are contemporaneously uncorrelated with those of the latent factors. Shocks to the structural innovation of g_t can impact all latent factors contemporaneously. We do not impose zero restrictions on the last r_2 columns of the impact matrix A^{-1} in (1.2.10). For the named factor identification scheme (*IRb*), we obtain a block diagonal covariance matrix of the following form $\Omega^* = \begin{bmatrix} \Lambda_1\Omega_{f\cdot g}\Lambda_1' & 0 \\ 0 & \Omega_{gg} \end{bmatrix}$.

In case of only one observable factor, i.e. $r_2 = 1$, the covariance of the structural innovations Ω_{gg} is a scalar. Hence, in the FAVAR context the structural innovations of the observed factors are identified by the factor model identification irrespective of the identification scheme used.

For more than one observable factor, i.e. $r_2 > 1$, we can generalize our model. The inverse of the contemporaneous impact matrix can be written as $A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$ and the structural innovations of the observed factors are given by $\eta_{g,t}^* = A_{22}\eta_{g,t}$. Hence, their covariance matrix is expressed as

$$\Omega_{gg}^* = \mathbb{E}\left[\eta_{g,t}^*\eta_{g,t}^{*\prime}\right] = A_{22}\Omega_{gg}A_{22}'. \tag{1.2.11}$$

In this setting, we can impose restrictions on A_{22} to achieve identification of the structural innovations with respect to the observable factors g_t. For example, a recursive structure within the observable factors could be used.

Moreover, we derive the impact matrix for the observed data x_t. It describes the contemporaneous impacts of an exogenous structural shock on the observed time series. Starting from the implied structural moving average representation for the observed data in lag operator notation, $x_t = \Lambda\Psi(L)A^{-1}\eta_t^*$, the model in period t reads $x_t = \Lambda A^{-1}\eta_t^* +$

$\Lambda\Psi_1 A^{-1}\eta^*_{t-1} + \Lambda\Psi_2 A^{-1}\eta^*_{t-2} + \cdots$. The impact matrix is denoted as $B_0 = \Lambda^* = \Lambda A^{-1}$. Using the estimated quantities, the impact matrix is given by

$$\hat{B}_0 = \begin{bmatrix} \tilde{\Lambda}^f & \tilde{\Lambda}^f \tilde{\Omega}_{fg} \tilde{\Omega}_{gg}^{-1} + \tilde{\Lambda}^g \end{bmatrix} = \hat{\Lambda}. \tag{1.2.12}$$

Hence, the sparsity in $\hat{\Lambda}$ provides structure to the contemporaneous impact matrix. The zeros in the factor loadings yield timing restrictions which are data-driven but can be explained economically. Moreover, the strength and direction of the effect is guided by the regularized factor loadings.

Inference on the impulse response functions of the regularized FAVAR model is conducted using a residual-bootstrap based on the dynamic factor equation (1.2.2). As shown in the following section, all estimators based on the regularized FAVAR are consistent. Hence, the estimated factors are assumed to be known in the VAR bootstrap. This approach is in the spirit of the factor augmented regression literature as in Bai and Ng (2006) and Gonçalves and Perron (2014).

The details on the implementation of the regularized FAVAR model are depicted in Section 1.C in the Appendix.

1.3. Asymptotic properties

In this section we establish the consistency of the regularized FAVAR model estimator introduced in equation (1.2.6). The following notation is used for the asymptotic analysis: $\pi_{\max}(A)$ and $\pi_{\min}(A)$ are the maximum and minimum eigenvalue of a square matrix A. Further, $\|A\|$ and $\|A\|_F$ denote the spectral and Frobenius norm, respectively. They are defined as $\|A\| = \sqrt{\pi_{\max}(A'A)}$ and $\|A\|_F = \sqrt{\text{tr}\,(A'A)}$. For some constant $c > 0$ and a non-random sequence b_N, we use the notation $b_N = \mathcal{O}(N)$, if $N^{-1}b_N \to c$, for $N \to \infty$. Moreover, $b_N = o(N)$, if $N^{-1}b_N \to 0$, for $N \to \infty$. Similarly, for a random sequence d_N, we say $d_N = \mathcal{O}_p(N)$, if $N^{-1}d_N \xrightarrow{p} c$, for $N \to \infty$ and $d_N = o_p(N)$, if $N^{-1}d_N \xrightarrow{p} 0$, for $N \to \infty$, where $\xrightarrow{p}$ denotes convergence in probability. Moreover, for the sequences a_N and b_N, we write $a_N \asymp b_N$, if $a_N = o(b_N)$ and $b_N = o(a_N)$.

Subsequently, we state the necessary assumptions.

Assumption 1.3.1 (Data generating process).

(i) $\{e_t, h_t\}_{t\geq 1}$ are strictly stationary, for $h_t = (f_t', g_t')'$. In addition, $\mathrm{E}[e_{it}] = \mathrm{E}[e_{it}h_{kt}] = 0$, for all $i \leq N$, $k \leq r$ and $t \leq T$.

(ii) There exists a constant $c > 0$ such that, for all N,

$$c^{-1} < \pi_{\min}\left(\frac{\Lambda'\Lambda}{N^{\beta}}\right) \leq \pi_{\max}\left(\frac{\Lambda'\Lambda}{N^{\beta}}\right) < c, \text{ where } 1/2 \leq \beta \leq 1.^{9}$$

(iii) There exists $b_1, b_2 > 0$ and $c_1, c_2 > 0$, such that for any $s > 0$, $i \leq N$ and $k \leq r$,

$$\mathbb{P}\left(|e_{it}| > s\right) \leq \exp(-(s/b_1)^{c_1}), \quad \mathbb{P}\left(|h_{kt}| > s\right) \leq \exp(-(s/b_2)^{c_2})$$

(iv) Define the mixing coefficient:

$$\alpha(T) = \sup_{A \in \mathcal{F}_{-\infty}^{0}, B \in \mathcal{F}_{T}^{\infty}} \left|\mathbb{P}(A)\,\mathbb{P}(B) - \mathbb{P}(AB)\right|,$$

where $\mathcal{F}_{-\infty}^{0}$ and $\mathcal{F}_{T}^{\infty}$ denote the σ-algebras generated by $\{(h_t, e_t) : -\infty \leq t \leq 0\}$ and $\{(h_t, e_t) : T \leq t \leq \infty\}$
Strong mixing: There exist $c_3 > 0$ and $C > 0$ satisfying: for all $T \in \mathcal{Z}^{+}$,

$$\alpha(T) \leq \exp(-CT^{c_3})$$

(v) There exist constants $d_1, d_2, d_3, d_4, d_5 > 0$ such that $d_2 < \pi_{\min}(\Sigma_e) \leq \pi_{\max}(\Sigma_e) < d_1$, $d_4 < \pi_{\min}(\Sigma_H) \leq \pi_{\max}(\Sigma_H) < d_3$ and $\max_{i \leq N, k \leq r} |\lambda_{ik}| < d_5$.

(vi) The factors $h_t = (f_t', g_t')'$ follow the VAR representation in (1.2.2), where η_t is an iid process with $\mathrm{E}[\eta_t] = 0$, $\mathrm{E}\left[\eta_t \eta_t'\right] = \Omega$, where Ω is positive semi-definite and all the roots of the polynomial $\Phi(L) = (I_r - \Phi_1 L - \cdots - \Phi_p L^p) = 0$ are outside the unit circle. Moreover, η_s and e_{it} are independent for all i, t, s.

The assumptions in 1.3.1 impose regularity conditions on the data generating process and are similar to those imposed by Bai and Liao (2016). Condition (i) induces strict stationarity for e_t and h_t and requires that both terms are uncorrelated. Condition (ii) relaxes the pervasiveness assumption commonly imposed in the approximate factor model literature and allows for weak factors that may affect only a subset of the entire data.[10] Condition (iii) requires exponential-type tails, which allows to use large deviation theory for $\frac{1}{T}\sum_{t=1}^{T} e_{it}e_{jt} - \sigma_{e,ij}$ and $\frac{1}{T}\sum_{t=1}^{T} h_{jt}e_{it}$. Condition (iv) imposes a strong mixing condition to

[9] The lower limit 1/2 for β is necessary for a consistent estimation of the factors. See Lemma 1.D.6 in Section 1.D in the Appendix.

[10] Note that Condition (ii) can be generalized such that the parameter β is different for each of the eigenvalues of $\Lambda'\Lambda$. Hence, each eigenvalue can diverge at a different rate N^{β_i}, for $i = 1, \ldots, r$ and $\beta_1 \geq \cdots \geq \beta_i \geq \cdots \geq \beta_r$. To simplify the notation in the theoretical parts, we denote $\beta_1 = \beta$. Thus, the largest eigenvalue of $\Lambda'\Lambda$ diverges with the rate $\mathcal{O}\left(N^{\beta}\right)$.

allow for weak serial dependence. Further, Condition (v) implies bounded eigenvalues of the idiosyncratic error covariance matrix and the covariance matrix of the factors, which is a common identifying assumption in the factor model framework. Condition (vi) imposes common VAR assumptions on the factor processes.

To control the sparsity in both Λ and Σ_e, we impose the following sparsity assumptions

Assumption 1.3.2 (Sparsity).

(i) $L_N = \sum_{k=1}^{r} \sum_{i=1}^{N} \mathbb{1}\{\lambda_{ik} \neq 0\} \asymp N^{\beta}$,

(ii) $S_N = \max_{i \leq N} \sum_{j=1}^{N} \mathbb{1}\{\sigma_{e,ij} \neq 0\}$, $S_N^2 \frac{N d_T}{L_N} = o_p(1)$ and $S_N \max(\mu_1, \mu_2) = o(1)$,

where $\mathbb{1}\{\cdot\}$ defines an indicator function that is equal to one if the boolean argument in braces is true, $\sigma_{e,ij}$ is the ij-th element of Σ_e and $d_T = \frac{\log N^{\beta}}{N} + \frac{1}{N^{\beta}} \frac{\log N}{T}$, for $1/2 \leq \beta \leq 1$.

Condition (i) defines the quantity L_N that represents the number of non-zero elements in the factor loadings matrix Λ. As the number of factors r are assumed to be constant, (i) upper bounds the number of non-zero elements in each column of Λ by N^{β}. Hence, for $\beta < 1$, we allow for zero elements in Λ. On the other hand, Condition (i) allows for a dense factor loadings matrix as well, for $\beta = 1$.[11] Condition (ii) specifies S_N that corresponds to the maximum number of non-zero elements in each row of Σ_e, following the definition in Fan et al. (2013). Moreover, the assumption $S_N^2 \frac{N d_T}{L_N} = o_p(1)$ limits the amount of correlations allowed between the idiosyncratic components and implies that Σ_e does not have an excessive dense structure. Hence, even for increasing sample sizes Σ_e only comprises of a weak amount of correlations, which is in accordance with the approximate factor model framework.

Theorem 1.3.1 (Consistency of the estimators of the regularized FAVAR model before rotation).
Under Assumptions 1.3.1 and 1.3.2 the regularized FAVAR model in (1.2.6) satisfies for T and $N \to \infty$, the following properties

$$\frac{1}{N} \left\| \tilde{\Lambda} - \Lambda \right\|_F^2 = \mathcal{O}_p \left(\frac{N d_T}{L_N} + \max(\mu_1, \mu_2) \right),$$
$$\frac{1}{N} \left\| \tilde{\Phi}_e - \Phi_e \right\|_F^2 = \mathcal{O}_p \left(\frac{\log N^{\beta}}{N} + \frac{\log N}{T} \right),$$

[11] It is important to note that the assumption on a constant number of factors r can be easily relaxed in our regularized FAVAR framework. In fact, r can slowly diverge with N. The analysis of this setting is left for future research.

where $d_T = \frac{\log N^\beta}{N} + \frac{1}{N^\beta}\frac{\log N}{T}$, for $1/2 \leq \beta \leq 1$.
Hence, for $\log(N) = o(T)$ and the regularization parameters $\mu_1 = o(1)$, $\mu_2 = o(1)$, we obtain

$$\frac{1}{N}\left\|\tilde{\Lambda} - \Lambda\right\|_F^2 = o_p(1), \qquad \frac{1}{N}\left\|\tilde{\Phi}_e - \Phi_e\right\|_F^2 = o_p(1).$$

Furthermore, for all $t \leq T$:

$$\left\|\tilde{f}_t - f_t\right\| = \mathcal{O}_p\left(N^{-\beta/2} + \frac{\sqrt{L_N}}{N^\beta}\left(\mu_1 + \sqrt{\frac{N d_T}{L_N} + \mu_2}\right)\right) = o_p(1).$$

For the second step estimator of the idiosyncratic error covariance matrix, specified in Section 1.2.3, we get

$$\left\|\tilde{\Sigma}_e^\tau - \Sigma_e\right\| = \mathcal{O}_p\left(S_N\sqrt{\frac{N d_T}{L_N} + \max(\mu_1, \mu_2)}\right).$$

Moreover, for the autoregressive matrices in the dynamic equation (1.2.2), we have

$$\tilde{\Phi}_i - \Phi_i = \left(\sum_{t=\bar{p}}^{T} \eta_t \psi_t'\right)\left(\sum_{t=\bar{p}}^{T} \psi_t \psi_t'\right)^{-1} (\iota_i \otimes I_r) + o_p(1),$$

where ι_i is the i-th column of the $r \times r$ identity matrix, $\bar{p} = p + 1$ and $\psi_t = \left(h_{t-1}', \ldots, h_{t-p}'\right)'$.

The proof of Theorem 1.3.1 is given in Section 1.D in the Appendix. Under the imposed regularity conditions, Theorem 1.3.1 establishes the average consistency in the Frobenius norm of the estimators of the factor loadings matrix and idiosyncratic error covariance matrix based on our regularized FAVAR model before rotation. More specifically, we can see that Λ and Φ are estimated consistently, even though we impose the strict factor model structure on Σ_e in the first step of our estimation procedure. Consequently, the latent factors f_t estimated based on GLS are consistent as long as $1/2 \leq \beta \leq 1$. Intuitively, the lower limit on β ensures that the factors are not too weak such that there is still a clear distinction between the common and idiosyncratic component. Furthermore, the second step estimator for the idiosyncratic error covariance matrix introduced in Section 1.2.3 is consistent under the spectral norm as long as $S_N^2 d_T = o_p(1)$ and $S_N \max(\mu_1, \mu_2) = o(1)$. Finally, the autoregressive parameter matrices in the dynamic VAR equation in (1.2.2) are consistently estimated under the Frobenius norm as well.

The following theorem summarizes the consistency results for the complete set of estimators based on our regularized FAVAR model after rotating them with matrix $\tilde{A} = \begin{bmatrix} I_{r_1} & -\tilde{\Omega}_{fg}\tilde{\Omega}_{gg}^{-1} \\ 0 & I_{r_2} \end{bmatrix}$.

Theorem 1.3.2 (Consistency of the rotated estimators based on the regularized FAVAR model). Under Assumptions 1.3.1 and 1.3.2 the regularized FAVAR model in (1.2.6) satisfies for T and $N \to \infty$, the following

$$\begin{aligned}
\max_{i \leq N} \left\| \hat{\lambda}_i^g - \lambda_i^{*g} \right\| &= \mathcal{O}_p\left(\sqrt{\mu_1} + \sqrt{\mu_2} + \sqrt{\frac{N d_T}{L_N}} \right) = o_p(1), \\
\left\| \hat{f}_t - f_t^* \right\| &= \mathcal{O}_p\left(N^{-\beta/2} + \frac{\sqrt{L_N}}{N^\beta} \left(\mu_1 + \sqrt{\frac{N d_T}{L_N} + \mu_2} \right) \right) = o_p(1), \\
\hat{\Phi}_i - \Phi_i^* &= \left(\sum_{t=\bar{p}}^{T} \eta_t \psi_t' \right) \left(\sum_{t=\bar{p}}^{T} \psi_t \psi_t' \right)^{-1} (\iota_i \otimes I_r) \\
&\quad + \mathcal{O}_p\left(\frac{\sqrt{L_N}}{N^\beta} \left(\mu_1 + \sqrt{\frac{N d_T}{L_N} + \mu_2} \right) \right),
\end{aligned}$$

where ι_i is the i-th column of the $r \times r$ identity matrix, $\bar{p} = p + 1$, $\psi_t = \left(h_{t-1}', \ldots, h_{t-p}' \right)'$ and $1/2 \leq \beta \leq 1$.

The proof of Theorem 1.3.2 is given in Section 1.D in the Appendix. Theorem 1.3.2 shows that the estimators of the latent factors, the factor loadings of the observed factors and the coefficients of the dynamic equation in (1.2.2) are consistently estimated under the Euclidean and Frobenius norm after rotating them based on matrix $\tilde{A}$. Note that an analysis of the estimated factor loadings of the latent factors $\tilde{\lambda}_i^f$ is redundant, as they are unaffected by the rotation.

1.4. Monte Carlo evidence

This section analyzes the finite sample properties of our regularized FAVAR estimator.

1.4.1. Monte Carlo designs

We focus on a similar data generating process as in Bai et al. (2016). The data x_t are generated according to the following equation

$$x_t = \Lambda^f f_t + \Lambda^g g_t + e_t, \tag{1.4.1}$$

where the factors are assumed to comply with the following VAR(1) process

$$h_t = \Phi h_{t-1} + \eta_t, \tag{1.4.2}$$

with $h_t = (f_t', g_t')'$ and η_t is an iid $N(0, \Omega^*)$ process. The matrix Ω^*, defined in (1.2.9) is constructed such that the identification restrictions IR_a, which are applied to our regularized FAVAR model are fulfilled. Hence, $\Omega^*_{fg} = \Omega^*_{gf} = 0$ and Ω^*_{ff} and Ω^*_{gg} are both symmetric positive definite matrices. The matrix Ω^* is generated according to $\Omega^* = M(M'M)^{-1/2}D(M'M)^{-1/2}M'$, where M is a $(r \times r)$-dimensional standard normal matrix and D is a diagonal matrix, where its diagonal entries are drawn from $[1 + U(0,1)]^2$. Moreover, we consider two specifications for the parameters of the dynamic equation Φ. For the first specification Φ_1 is fixed to $0.7I_r$, whereas for the second specification Φ is set to

$$\Phi_2 = \begin{bmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{bmatrix}.$$

In all simulations we set the number of latent and observed factors r_1 and r_2, to two and one, respectively, such that the total number of factors is $r = 3$. For the true factor loadings we consider five different specifications. In the first specification, Λ^f_1 and Λ^g_1 are both dense. The structures of the factor loadings matrices in the second specification are given by

$$\Lambda^f_2 = \begin{bmatrix} \lambda^f_{1:n_1,1} & \lambda^f_{1:n_1,2} \\ \lambda^f_{(N+1)-n_2:N,1} & 0 \end{bmatrix}, \quad \Lambda^g_2 = \begin{bmatrix} 0 \\ \lambda^g_{(N+1)-n_2:N,1} \end{bmatrix}, \tag{1.4.3}$$

where $n_1 = n_2 = 0.6 \cdot N$. In the third specification, Λ^g_3 is dense and Λ^f_3 follows

$$\Lambda^f_3 = \begin{bmatrix} \lambda^f_{1:n_1,1} & 0 \\ 0 & \lambda^f_{(N+1)-n_2:N,2} \end{bmatrix}. \tag{1.4.4}$$

The columns of the fourth loadings matrix corresponding to the latent factors Λ^f_4 have the following cardinalities: $\left|\lambda^f_1\right| = N$, $\left|\lambda^f_2\right| = N^{0.9}$, whereas the cardinality of the loadings matrix corresponding to the observed factor Λ^g_4 is $\left|\lambda^g_1\right| = N^{0.8}$. Finally, the cardinalities of the loadings in the fifth factor loadings specification Λ^f_5 and Λ^g_5 are set to: $\left|\lambda^f_1\right| = N^{0.95}$, $\left|\lambda^f_2\right| = N^{0.85}$ and $\left|\lambda^g_1\right| = N^{0.8}$. The positions of the zero loadings in the fourth and fifth factor loadings specifications are drawn randomly.[12] The non-zero factor loadings in all five

[12] All cardinalities are rounded to the larger integer values.

settings are drawn from $1 + N(0,1)$ and the idiosyncratic components $e_t = (e_{1t}, \ldots, e_{Nt})$ are independently drawn from $e_{it} \sim N(0, \sigma_i^2)$, where $\sigma_i^2 \sim 1 + U(0,1)$.

Similar as in Bai and Li (2016) or Doz et al. (2011), we evaluate the accuracy of the factor loadings estimates based on the Trace-Ratio criterion (TR), which is given by

$$TR(\hat{\Lambda}) = \text{tr}\left[(\Lambda'\hat{\Lambda})(\hat{\Lambda}'\hat{\Lambda})^{-1}(\hat{\Lambda}'\Lambda)\right] / \text{tr}\left(\Lambda'\Lambda\right). \tag{1.4.5}$$

To measure the goodness-of-fit of the parameters of the dynamic equation, we use the average Frobenius norm of the difference between the estimated parameters $\hat{\Phi}$ and the true ones Φ, according to

$$FN(\hat{\Phi}) = \frac{1}{r}\left\|\hat{\Phi} - \Phi\right\|_F^2 = \frac{1}{r}\text{tr}\left[\left(\hat{\Phi} - \Phi\right)'\left(\hat{\Phi} - \Phi\right)\right]. \tag{1.4.6}$$

1.4.2. Simulation results

Table 1.4.1 reports the trace ratios for the estimated factor loading matrices, as well as the average Frobenius norms of the difference between the estimated parameters of the dynamic equation $\hat{\Phi}$ and the true ones, for the first factor loadings specification. This specification only incorporates strong latent and observed factors, as the loadings matrices corresponding to both types of factors are dense. The first column of Table 1.4.1 considers different combinations of N and T, where we allow $N, T = 50, 100$ and 200. The second and third columns report the results for first and second specification of the true parameter matrix of the dynamic equation (Φ_1, Φ_2), respectively. Overall, Table 1.4.1 shows that the trace ratios for the factor loadings estimated by our regularized FAVAR model approach one for both specifications of Φ, as N and T increase. Hence our RFAVAR model consistently estimates the true factor loadings for $N, T \to \infty$. At the same time, we observe that the average Frobenius norm of the difference between $\hat{\Phi}$ and Φ_i, for $i = 1, 2$, tend to zero as $N, T \to \infty$. Hence, the parameters of the dynamic equation are consistently estimated by the regularized FAVAR model as well. These results are in line with the large sample properties established in Section 1.3. In addition, we can depict that the convergence rates of the estimators decrease, if the dynamics in the latent and observed factors evolve according to Φ_2 compared to specification Φ_1. This result is anticipated as Φ_2 allows for interactions between the factors, which generally complicates the estimation.

The trace ratios for $\hat{\Lambda}$ and Frobenius norm differences for $\hat{\Phi}$ for the second to fifth factor loadings specifications $(\Lambda_2^f, \Lambda_2^g - \Lambda_5^f, \Lambda_5^g)$ are reported in Table 1.4.2, as well as in Tables 1.A.1 - 1.A.3 in the Appendix 1.A. These factor loadings specifications incorporate different sparsity structures in the loadings of the latent or observed factors and hence allow for weakly

Table 1.4.1: Simulation results - First factor loading specification $(\Lambda_1^f, \Lambda_1^g)$

		Φ_1		Φ_2	
N	T	$TR(\hat{\Lambda})$	$FN(\hat{\Phi})$	$TR(\hat{\Lambda})$	$FN(\hat{\Phi})$
50	50	0.9555	0.0580	0.9312	0.1178
50	100	0.9748	0.0201	0.9451	0.0908
100	50	0.9843	0.0344	0.9799	0.0775
100	100	0.9927	0.0121	0.9860	0.0637
200	200	0.9977	0.0059	0.9941	0.0566

Note: The table reports the trace ratios and the differences in average Frobenius norms for the estimators of our RFAVAR model given in equations (1.4.5) and (1.4.6), for the first factor loadings specification. The second and third column illustrates the results for the first and second specification of the true parameters Φ_1, Φ_2, respectively. The results are based on 1000 repetitions.

Table 1.4.2: Simulation results - Second factor loading specification $(\Lambda_2^f, \Lambda_2^g)$

		Φ_1		Φ_2	
N	T	$TR(\hat{\Lambda})$	$FN(\hat{\Phi})$	$TR(\hat{\Lambda})$	$FN(\hat{\Phi})$
50	50	0.9520	0.0314	0.9302	0.0757
50	100	0.9708	0.0138	0.9507	0.0608
100	50	0.9809	0.0299	0.9682	0.0711
100	100	0.9884	0.0128	0.9778	0.0562
200	200	0.9964	0.0055	0.9938	0.0501

Note: The table reports the trace ratios and the differences in average Frobenius norms for the estimators of our RFAVAR model given in equations (1.4.5) and (1.4.6), for the second factor loadings specification. The second and third column illustrates the results for the first and second specification of the true parameters Φ_1, Φ_2, respectively. The results are based on 1000 repetitions.

influential factors. The results in terms of $TR(\hat{\Lambda})$ and $FN(\hat{\Phi})$ are quite similar compared to the first loadings specification, across the different sparsity patterns in the true factor loadings matrices. More specifically, the regularized FAVAR model consistently estimates the factor loadings and parameters of the dynamic equation as N,T increase. Hence, the RFAVAR model is able to consistently detect the true sparsity pattern in Λ and therefore is capable of consistently estimating weak factors.

1.5. Empirical application: The effects of monetary policy shocks

In this section, we use the regularized FAVAR model to investigate the effects of a monetary policy shock on economic variables. We use the Federal Funds rate (FFR) as the monetary policy instrument. The monetary policy shock is defined as a shock to the innovation of the FFR. In our FAVAR setting, we treat the FFR as an observable factor.

1.5.1. Data description

The data set x_t consists of 126 macroeconomic and financial time series on monthly frequency. It spans the period from January 1985 until December 2016 which results in 384 monthly observations. The data set is mainly based on the FRED-MD database by McCracken and Ng (2016). To represent a broad range of economic sectors we augment the data set by manufacturing data, stock market data and various short-term interest rate spreads. The resulting data set is comparable to the one used by Forni and Gambetti (2010), Stock and Watson (2016) and Kerssenfischer (2019). A detailed overview can be found in Table 1.E.1

in Appendix 1.E. We transform the data such that x_t contains only stationary time series.[13] Furthermore, without loss of generality we standardize x_t.

1.5.2. Overview of models

This section provides an overview of the models that are compared in the empirical analysis. The regularized FAVAR (**RFAVAR**) leads to a data driven identification of the latent factors. We obtain an economic structure by shrinking single elements in the factor loadings matrix to zero. Hence, the estimated factors only load on a subset of the observed time series that correspond to different sectors of the economy. For the identification of our model we use scheme *IRa* in Section 1.2.4 that relies on a data-driven identification of the FAVAR model that is invariant to the ordering of the data. Hence, we circumvent imposing a restrictive a priori assumption on the structure of the model. We initialize our model with eight latent factors according to the IC_1 selection criterion by Bai and Ng (2002) and are left with five latent factors after the regularization.

We set the lag order for our regularized FAVAR model to $p = 12$. In a dynamic factor model setting, this lag order captures the dynamics in the observed and latent factors sufficiently. Concerning the choice of the regularization parameters we use the procedure described in Section 1.C.2.[14]

The commonly used scheme in the literature to obtain economically identified factors is based on the named factor identification. This corresponds to identification scheme *IRb* in Section 1.2.4, which implies that the first r_1 time series drive the dynamics of the latent factors. In our analysis we denote this model as **FAVAR-NF** and the following time series are used as naming variables: civilian unemployment rate, consumer price index, industrial production index, S&P 500 composite index and Baa corporate bond yield (see Table 1.E.2). This specific choice of time series is guided by the selection of our regularized FAVAR model. More specifically, these are the time series with the highest loadings in absolute value on the estimated factors. Moreover, the selection is also economically sensible as relevant sectors of the economy are represented by those time series. We use $p = 12$ lags in the named factor FAVAR model.

[13] The specific transformations can be found in Table 1.E.1 in Appendix 1.E.

[14] Different robustness checks in our empirical application reveal that the structure of the regularized FAVAR model does not change much for different values of μ_2. Hence, to reduce the computational time, we can fix μ_2 to a value in the interval [0.7, 1.3] and optimize only for values of μ_1.

In the following, we introduce two additional models that are restricted versions of the general dynamic factor model proposed by Forni et al. (2000). Closely related models are studied by Stock and Watson (2002, 2002, 2005). More specifically, the model is given by

$$x_t = \Lambda^f f_t + e_t, \tag{1.5.1}$$

$$f_t^{\ddagger} = \Theta(L) f_{t-1}^{\ddagger} + u_t, \tag{1.5.2}$$

where f_t in Equation (1.5.1) is a $(q \times 1)$-vector of latent static factors and $\Theta(L)$ is a p-th order lag polynomial. The first model we consider is the pure dynamic factor model (**DFM**) by Forni and Gambetti (2010). In this model $f_t^{\ddagger} = f_t$, $u_t = R\varepsilon_t$, R is a $(q \times q_1)$-dimensional matrix and ε_t is $(q_1 \times 1)$-vector of primitive shocks.

In the second specification, we augment the factors in equation (1.5.2) by observable variables y_t. Hence, in this setting $f_t^{\ddagger} = [f_t', y_t']'$. We denote this model as **VAR-F**. In comparison to the FAVAR model introduced in Equations (1.2.1) and (1.2.2), the model specification in Equation (1.5.1) only includes latent variables on the right hand side.

For the model specification of the DFM, we follow Forni and Gambetti (2010) by setting $q = 16$, $q_1 = 4$ and $p = 12$.[15] The identification of the structural shocks relies on a recursive Cholesky scheme, which includes industrial production (IP), consumer prices (CPI), the FFR and the excess bond premium (EBP). A tight monetary policy shock increases the FFR and has no contemporaneous effect on IP and CPI. The FFR can only be affected by industrial production, consumer prices and itself contemporaneously, whereas EBP reacts to shocks to the three others. The model settings for the VAR-F are comparable to Kerssenfischer (2019) and are set to $q_1 = 9$ and $p = 12$. As observable variables, we include IP, CPI and the FFR. For the identification of the structural shocks corresponding to the observable variables we use a recursive Cholesky scheme. As above, IP and CPI are not affected by a contractionary monetary policy shock on impact, whereas the FFR increases.

1.5.3. Results for the factor model analysis

To get some insights on the data, we start with an unregularized factor model estimated by PCA. The number of included factors is determined by the Bai and Ng (2002) IC_1 criterion, where we set the maximum number of allowed factors ($r_{\max}$) to ten.

Figure 1.B.1 shows the R^2 of univariate regressions of the latent factors on the observed time series. Of the eight factors that are selected by the criterion, only five factors have blocks with high explanatory power. It is well documented in the literature that the information criteria of Bai and Ng (2002) tend to overestimate the true number of factors when there is

[15] The selection criterion by Bai and Ng (2007) also suggests 4 dynamic factors for our dataset.

remaining correlations in the idiosyncratic component (see e.g. Ahn and Horenstein (2013) and Caner and Han (2014)). Furthermore, the standard factor model is only identified statistically and the estimated factors may not be economically meaningful. For the structural analysis we are interested in the dynamics of a model that is economically interpretable. Hence, it is crucial to economically identify the latent factors.

In the regularized FAVAR model we enhance the economic interpretability by shrinking elements of the factor loadings matrix to zero. We illustrate the R^2 results of univariate regressions associated with a sparse factor structure in Figure 1.B.2. The separation of the latent factors is notably more distinct compared to the factors estimated by PCA. More specifically, we obtain a block structure in the factor loadings which leads to factors that are linked to different sectors in the economy. Furthermore, the RFAVAR model estimates five latent factors. Hence, it shrinks the factors that do not add additional explanatory power to zero. The economic groups associated to the latent factors are: the labor market, prices, industrial production, the stock market and credit spreads. In our context, the labor market factor is mostly linked to employment time series. The corresponding time series plots are shown in Figure 1.E.1 in Appendix 1.E. The obtained factors are closely aligned with the underlying economic time series which is due to the fact that the informational content of various time series is used to construct the latent factor estimates. Table 1.E.1 in Appendix 1.E provides an overview of which factor drives each variable. A comparable number of factors has been found by Stock and Watson (2016).

The same analysis is repeated for the named factor FAVAR (FAVAR-NF) model. The resulting R^2 are plotted in Figures 1.B.3. The choice of the naming variables in the FAVAR-NF model leads to factors that have a comparable allocation to economic sectors as the RFAVAR model. This result is anticipated as the selection of the naming variables is guided by the RFAVAR. However, the obtained factor time series in Figures 1.E.2 in Appendix 1.E are less aligned to the observed time series in comparison to the RFAVAR. This effect is more pronounced for naming variables that do not represent the entire sector. In these cases the resulting factors are distorted by sector unrelated variables, as in the labor market and credit spread factors.

It is important to note that the model identification crucially depends on the ordering of the observed time series. For example, Stock and Watson (2016) use the following naming scheme in an oil price application: Real personal consumption expenditures, industrial production index, civilian employment rate, S&P 500 composite index, trade weighted US dollar index major currencies and producer price index.[16] The resulting R^2 are plotted in Figure 1.B.4. Even though the naming variables are chosen based on economic reasoning

[16]As Stock and Watson (2016) work on quarterly frequency they can use data that is not available on monthly frequency. We deviate by using a different proxy for employment and by omitting government spending.

the selection is rather arbitrary and leads to very different results. Hence, the selection of the naming variables constitutes a restrictive assumption on the structural model.

We omit this analysis for the DFM and VAR-F models because these frameworks are not concerned with the interpretability of the factors.

1.5.4. Results from the impulse response analysis

In the following, we analyze the effects of a monetary policy shock on the observed time series based on our model specification (RFAVAR). Additionally, we compare the results to the ones obtained for the FAVAR-NF, DFM and VAR-F. Our focus lies on the dynamic responses to a shock in the monetary policy instrument (FFR). This corresponds to a partial identification of the structural model, where we identify the monetary policy shock. More specifically, the innovations corresponding to the FFR are contemporaneously uncorrelated with the latent factor innovations, whereas the latter can be contemporaneously correlated. Our RFAVAR framework allows for the structural analysis of the dynamics for both the factors and the underlying observed time series as pointed out in Section 1.2.5.

In a first step, we elaborate on the effect on the estimated factors for the RFAVAR and the FAVAR-NF model. In both settings, the obtained factors are economically identified and serve as proxies of economic aggregates. The contemporaneous structural effect on the factors are depicted in Table 1.A.4, where the last column in both panels is associated with the monetary policy shock.

For our RFAVAR model, the strength of the impact is given in Panel A of Table 1.A.4. The sign of the contemporaneous impact is in line with economic reasoning. More specifically, the price, industrial production and the labor market factors react negatively on impact in response to a tight monetary policy shock.

The impulse responses of the factors to a 100bp shock in the innovation of the Federal Funds rate are illustrated in Figure 1.5.1. The responses of all factors are transitory and stabilize to a new level after one or two years. More precisely, the labor market reacts significantly negative on impact and no longer reacts after 12 months. Moreover, as expected by economic rationale, the price level, the level of industrial production and the credit spread factor are impacted negatively by an exogenous increase in the monetary policy rate. The stock market factor does not react significantly to a monetary policy shock for most of the periods. It slightly increases after five months for about one month.

For the FAVAR-NF model, the impact matrix is reported in Panel B in Table 1.A.4.[17]

[17]Note that we only provide standard errors for the estimates in the last column of Panel B. This follows from the structure of the rotation matrix $\tilde{A}$ given at the end of Section 1.2.4. As the first $(r_1 \times r_1)$ block of $\tilde{A}$ depends on $\tilde{\Lambda}_1$, which is kept fixed in the residual-bootstrap based on the dynamic factor equation (1.2.2), we cannot compute bootstrap-based standard errors for the estimated quantities in the first $(r_1 \times r_1)$ block of the impact

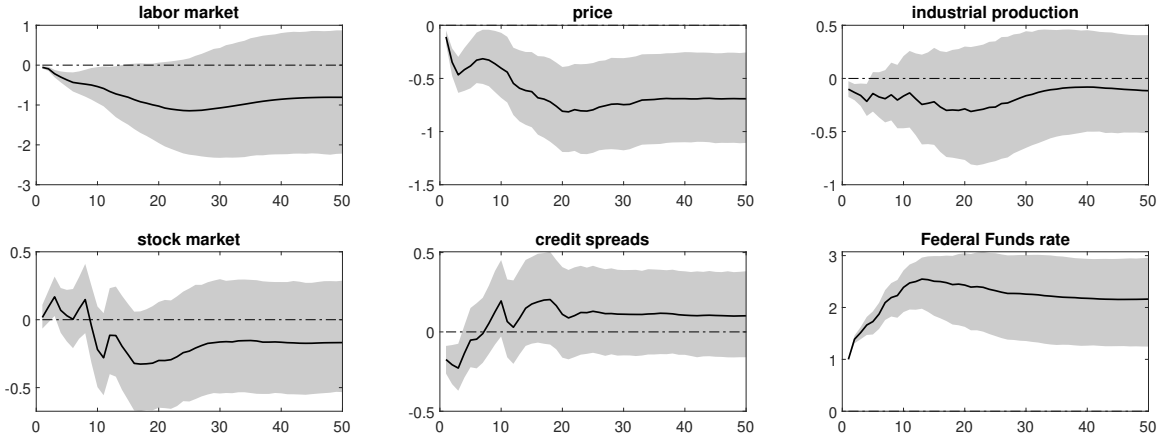


Figure 1.5.1: Accumulated impulse responses to a monetary policy shock on the factors for the regularized FAVAR.
The graph shows accumulated impulse responses to a 100bp shock in the innovation of the FFR. The dashed lines correspond to 68% bootstrap confidence intervals.

Furthermore, in Figure 1.B.5, we plot the impulse response functions associated with the named factor scheme. Both, the contemporaneous impacts and the resulting impulse responses on the estimated factors are not always in line with the economic theory. For example, unemployment reacts negatively in response to a tight monetary policy shock, whereas industrial production increases for about ten months. Moreover, we obtain relatively large confidence intervals for the remaining factors leading to statistically insignificant impulse responses. Even though the choice of the naming variables is economically motivated, the implied structural dynamics fail to span the monetary policy shock.

In a second step, we investigate the impulse responses on the observable variables x_t for all models. For the RFAVAR and FAVAR-NF models the contemporaneous impact matrix on x_t is given by $\hat{\Lambda}$ corresponding to the rotated factor loadings estimate, as outlined in Section 1.2.5. The impact matrix for our regularized FAVAR model is given in Figure 1.5.2. The contemporaneous impact of a tightening monetary policy shock is in accordance with economic reasoning. Following an exogenous increase in the FFR, the short term interest rates go up on impact. Further, price and IP as well as employment time series react negatively, whereas unemployment series rise. The stock market plummets and credit conditions deteriorate.

The contemporaneous impact matrices for the FAVAR-NF model is depicted in Figure 1.B.6. The economic implications are implausible, as the sign of the effects is often incorrect. For example, the labor market, industrial production and prices react positively on impact for a contractionary monetary policy shock. In contrast to our RFAVAR model, the factor

matrix. However, this is not harmful for the upcoming analysis, as we are interested in the dynamic effects of a monetary policy shock, whose contemporaneous impacts are given by the last column of the impact matrix.

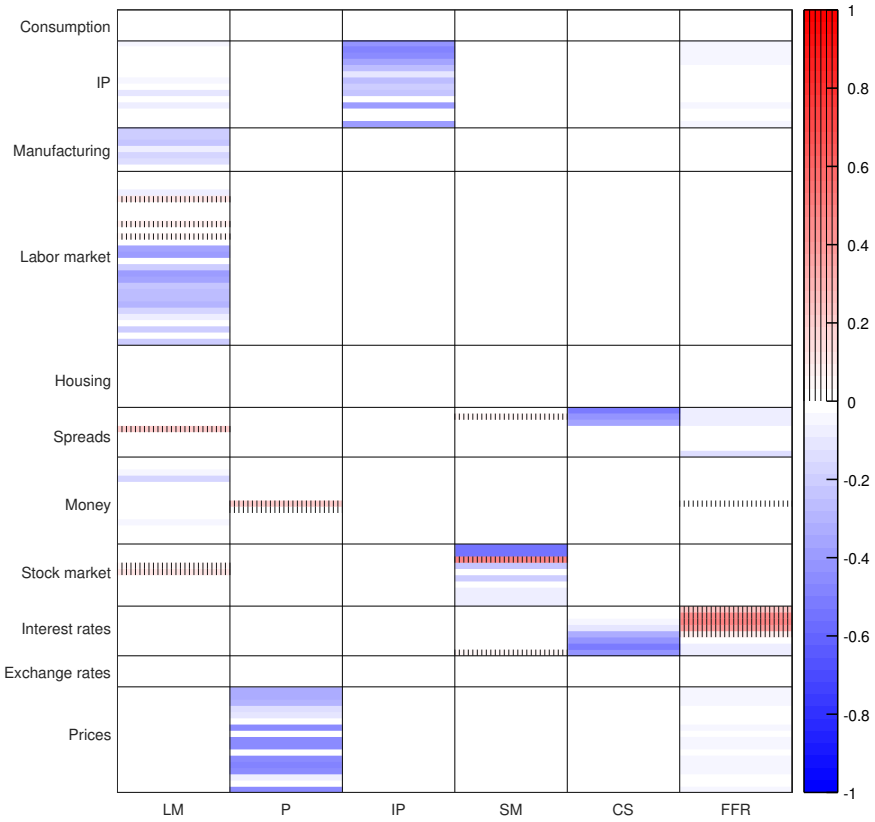


Figure 1.5.2: Impact matrix on the observed variables x_t for the regularized FAVAR.
This graph shows the contemporaneous impact matrix of a 100bp shock in the factor innovations to the observed time series for the regularized FAVAR model. The factors are abbreviated as follows: 'LM' labor market, 'P' price, 'IP' industrial production, 'SM' stock market, 'CS' credit spread and 'FFR' Federal Funds rate.

loadings matrix of the FAVAR-NF model is more dense and an economic sector association is not possible. Hence, even though the R^2 plot of univariate regressions of the named factors on the observed time series in Figure 1.B.4 shows a high explanatory power of the latent factors, the implied rotation scheme yields a factor loadings matrix which we can no longer explain economically.

In the following, we analyze the impulse responses for selected time series, i.e. we concentrate on the effects of a tight monetary policy shock on the consumer price index (CPI), the civilian unemployment rate, the IP index and the three months Treasury bill. Figure 1.5.3 shows the level effects for our RFAVAR model on the specific variables. CPI and the IP index series react negatively for around two years, respectively and stabilize to a new level for the remaining periods. The unemployment rate increases for approximately 24 months, whereas the short-term interest rate reacts positively on impact for 12 months.

The impulse responses for the FAVAR-NF are provided in Figure 1.B.7. Overall, the results show contradicting effects contemporaneously as well as over time for the unemployment rate and the IP index.

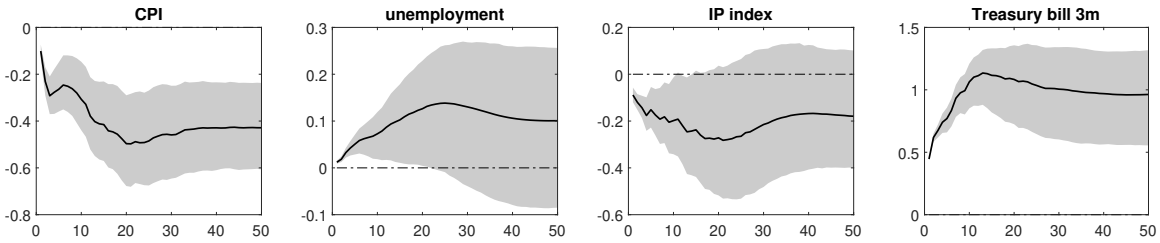


Figure 1.5.3: Accumulated impulse responses to a monetary policy shock on the observed variables x_t for the regularized FAVAR.
The graph shows accumulated impulse responses to a 100bp shock in the innovation of the FFR. The dashed lines correspond to 68% bootstrap confidence intervals.

For the DFM, the impulse responses for CPI, unemployment, industrial production and the three month Treasury bill rate are illustrated in Figure 1.B.8. The point estimates for CPI, the IP index and unemployment are economically not reasonable. Moreover, the DFM is sensitive to the number of included latent factors and lags, which leads to a high degree of estimation noise.

The impulse responses for the VAR-F model are given in Figure 1.B.9. The responses of unemployment and IP are in line with economic rationale. Moreover, the impulse response of CPI is very volatile and statistically insignificant. The three month Treasury bill rate reacts negatively over all horizons. The results are very sensitive to the number of included factors and lags. If we do not include enough factors or lags, the impulse response patterns are hardly interpretable (e.g. a price puzzle is obtained).

1.6. Conclusion

In this paper, we propose a regularized factor-augmented vector autoregressive model that enables the identification of the factors and their economic interpretation. Our estimation procedure relies on a penalized quasi-maximum likelihood approach and is based on a L_1-norm regularization of the factor loadings matrix. The named factor identification scheme conventionally used in the FAVAR literature imposes specific relations between the factors and observed time series. More precisely, the ordering of the variables determines the structure of the model. The sparse factor loadings structure in our regularized FAVAR model allows for a direct factor identification. Hence, we are able to identify the latent factors in a data-driven manner without implicitly assuming their form a priori in our identification scheme. In this framework, the effects of structural shocks can be investigated on economically meaningful estimated factors and on all observed time series included in the model.

We prove consistency under the Frobenius norm for the estimators of the factor loadings, the latent factors and the covariance matrix of the idiosyncratic component based on the regularized FAVAR model. The factors estimated based on GLS are shown to be consistent. The autoregressive parameters in the dynamic equation are consistently estimated as well. We analyze the finite sample properties of the RFAVAR in various simulation settings. The results show that the model detects the true structure of the factor loadings and the autoregressive parameters precisely.

In an empirical study, we investigate the effects of monetary policy shocks on a broad range of economically relevant variables. We identify this shock using a joint identification of the factor model and the structural innovations in the vector autoregressive model. We extract five latent factors that relate to the labor market, prices, industrial production, the stock market and credit spreads. Furthermore, the Federal Funds rate is used as an observed factor. We find impulse response functions which are in line with economic rationale both on the factor aggregates and the observed time series level. More specifically, we do not observe a price puzzle and the obtained impulse response patterns are economically plausible. In particular, following a tight monetary policy, industrial production falls, credit conditions deteriorate and the level of employment decreases.

References

Ahn, S. C. and A. R. Horenstein (2013). "Eigenvalue ratio test for the number of factors". *Econometrica* 81. (3), 1203–1227.

Bai, J. and K. Li (Feb. 2012). "Statistical analysis of factor models of high dimension". *The Annals of Statistics* 40. (1), 436–465.

— (2016). "Maximum likelihood estimation and inference for approximate factor models of high dimension". *Review of Economics and Statistics* 98. (2), 298–309.

Bai, J., K. Li, and L. Lu (2016). "Estimation and inference of FAVAR models". *Journal of Business & Economic Statistics* 34. (4), 620–641.

Bai, J. and Y. Liao (2016). "Efficient estimation of approximate factor models via penalized maximum likelihood". *Journal of Econometrics* 191. (1), 1–18.

Bai, J. and S. Ng (2002). "Determining the number of factors in approximate factor models". *Econometrica* 70. (1), 191–221.

— (2006). "Confidence intervals for diffusion index forecasts and inference for factor-Augmented regressions". *Econometrica* 74. (4), 1133–1150.

— (2007). "Determining the number of primitive shocks in factor models". *Journal of Business & Economic Statistics* 25. (1), 52–60.

— (2013). "Principal components estimation and identification of static factors". *Journal of Econometrics* 176. (1), 18–29.

Bai, J. and P. Wang (2014). "Identification theory for high dimensional static and dynamic factor models". *Journal of Econometrics* 178. (2), 794–804.

Bekker, P. A. (1986). "A note on the identification of restricted factor loading matrices". *Psychometrika* 51. (4), 607–611.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005). "Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach". *The Quarterly Journal of Economics* 120. (1), 387–422.

Bien, J. and R. J. Tibshirani (2011). "Sparse estimation of a covariance matrix". *Biometrika* 98. (4), 807–820.

Boivin, J. and M. P. Giannoni (2007). "Global forces and monetary policy effectiveness". In: *International Dimensions of Monetary Policy*. University of Chicago Press, 429–478.

Boivin, J., M. P. Giannoni, and I. Mihov (2009). "Sticky prices and monetary policy: Evidence from disaggregated US data". *American Economic Review* 99. (1), 350–384.

Caner, M. and X. Han (2014). "Selecting the correct number of factors in approximate factor models: The large panel case with group bridge estimators". *Journal of Business & Economic Statistics* 32. (3), 359–374.

Chamberlain, G. and M. Rothschild (1983). "Arbitrage, factor structure, and mean-variance analysis on large asset markets". *Econometrica* 51. (5), 1281–1304.

Daniele, M., W. Pohlmeier, and A. Zagidullina (2018). *Sparse approximate factor estimation for high-dimensional covariance matrices*. Working Paper. University of Konstanz.

Doz, C., D. Giannone, and L. Reichlin (2011). "A two-step estimator for large approximate dynamic factor models based on Kalman filtering". *Journal of Econometrics* 164. (1), 188–205.

Fan, J., Y. Liao, and M. Mincheva (2011). "High dimensional covariance matrix estimation in approximate factor models". *Annals of Statistics* 39. (6), 3320–3356.

— (2013). "Large covariance estimation by thresholding principal orthogonal complements". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75. (4), 603–680.

Forni, M. and L. Gambetti (2010). "The dynamic effects of monetary policy: A structural factor model approach". *Journal of Monetary Economics* 57. (2), 203–216.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). "The generalized dynamic-factor model: Identification and estimation". *Review of Economics and Statistics* 82. (4), 540–554.

Geweke, J. (1977). "The dynamic factor analysis of economic time series". *Latent variables in socio-economic models*.

Gonçalves, S. and B. Perron (2014). "Bootstrapping factor-augmented regression models". *Journal of Econometrics* 182. (1), 156–173.

Hallin, M. and R. Liška (2011). "Dynamic factors in the presence of blocks". *Journal of Econometrics* 163. (1), 29–41.

Kerssenfischer, M. (2019). "The puzzling effects of monetary policy in VARs: Invalid identification or missing information?" *Journal of Applied Econometrics* 34. (1), 18–25.

Kilian, L. and L. T. Lewis (2011). "Does the Fed respond to oil price shocks?" *The Economic Journal* 121. (555), 1047–1072.

Lawley, D. and A. Maxwell (1971). *Factor analysis as a statistical method, second ed.* Butterworths, London.

McCracken, M. W. and S. Ng (2016). "FRED-MD: A monthly database for macroeconomic research". *Journal of Business & Economic Statistics* 34. (4), 574–589.

Mumtaz, H. and P. Surico (2009). "The transmission of international shocks: A factor-augmented VAR approach". *Journal of Money, Credit and Banking* 41. (s1), 71–100.

Onatski, A. (2012). "Asymptotics of the principal components estimator of large factor models with weakly influential factors". *Journal of Econometrics* 168. (2), 244–258.

Ramey, V. A. (2016). "Macroeconomic shocks and their propagation". In: *Handbook of Macroeconomics*. Vol. 2. Elsevier, 71–162.

Sargent, T. J., C. A. Sims, et al. (1977). "Business cycle modeling without pretending to have too much a priori economic theory". *New methods in business cycle research* 1, 145–168.

Sims, C. A. (1992). "Interpreting the macroeconomic time series facts: The effects of monetary policy". *European Economic Review* 36. (5), 975–1000.

Stock, J. H. and M. Watson (2011). "Dynamic factor models". *Oxford Handbook on Economic Forecasting*.

Stock, J. H. and M. W. Watson (2002a). "Forecasting using principal components from a large number of predictors". *Journal of the American Statistical Association* 97. (460), 1167–1179.

— (2002b). "Forecasting using principal components from a large number of predictors". *Journal of the American Statistical Association* 97. (460), 1167–1179.

— (2002c). "Macroeconomic forecasting using diffusion indexes". *Journal of Business & Economic Statistics* 20. (2), 147–162.

— (2005). *Implications of dynamic factor models for VAR analysis*. Tech. rep. w11467. National Bureau of Economic Research.

— (2016). "Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics". In: *Handbook of Macroeconomics*. Vol. 2. Elsevier, 415–525.

Wu, J. C. and F. D. Xia (2016). "Measuring the macroeconomic impact of monetary policy at the zero lower bound". *Journal of Money, Credit and Banking* 48. (2-3), 253–291.

Appendix 1.A Tables

Table 1.A.1: Simulation results - Third factor loading specification $(\Lambda_3^f, \Lambda_3^g)$

		Φ_1		Φ_2	
N	T	$TR(\hat{\Lambda})$	$FN(\hat{\Phi})$	$TR(\hat{\Lambda})$	$FN(\hat{\Phi})$
50	50	0.9554	0.0447	0.9246	0.1009
50	100	0.9706	0.0154	0.9478	0.0691
100	50	0.9795	0.0325	0.9680	0.0807
100	100	0.9896	0.0130	0.9770	0.0638
200	200	0.9966	0.0060	0.9940	0.0525

Note: The table reports the trace ratios and the differences in average Frobenius norms for the estimators of our RFAVAR model given in equations (1.4.5) and (1.4.6), for the third factor loadings specification. The second and third column illustrates the results for the first and second specification of the true parameters Φ_1, Φ_2, respectively. The results are based on 1000 repetitions.

Table 1.A.2: Simulation results - Fourth factor loading specification $(\Lambda_4^f, \Lambda_4^g)$

		Φ_1		Φ_2	
N	T	$TR(\hat{\Lambda})$	$FN(\hat{\Phi})$	$TR(\hat{\Lambda})$	$FN(\hat{\Phi})$
50	50	0.9275	0.0349	0.9059	0.0797
50	100	0.9576	0.0128	0.9344	0.0630
100	50	0.9745	0.0336	0.9568	0.0770
100	100	0.9867	0.0124	0.9772	0.0583
200	200	0.9954	0.0050	0.9925	0.0505

Note: The table reports the trace ratios and the differences in average Frobenius norms for the estimators of our RFAVAR model given in equations (1.4.5) and (1.4.6), for the fourth factor loadings specification. The second and third column illustrates the results for the first and second specification of the true parameters Φ_1, Φ_2, respectively. The results are based on 1000 repetitions.

Table 1.A.3: Simulation results - Fifth factor loading specification $(\Lambda_5^f, \Lambda_5^g)$

		Φ_1		Φ_2	
N	T	$TR(\hat{\Lambda})$	$FN(\hat{\Phi})$	$TR(\hat{\Lambda})$	$FN(\hat{\Phi})$
50	50	0.9196	0.0308	0.8762	0.0739
50	100	0.9480	0.0119	0.9127	0.0605
100	50	0.9593	0.0292	0.9432	0.0653
100	100	0.9789	0.0115	0.9678	0.0546
200	200	0.9935	0.0051	0.9892	0.0424

Note: The table reports the trace ratios and the differences in average Frobenius norms for the estimators of our RFAVAR model given in equations (1.4.5) and (1.4.6), for the fifth factor loadings specification. The second and third column illustrates the results for the first and second specification of the true parameters Φ_1, Φ_2, respectively. The results are based on 1000 repetitions.

Table 1.A.4: Contemporaneous impact matrices

Panel A: regularized FAVAR model - RFAVAR						
LM	1	0	0	0	0	-0.0555 (0.0316)
P	0	1	0	0	0	-0.1093 (0.0676)
IP	0	0	1	0	0	-0.1014 (0.0721)
SM	0	0	0	1	0	0.0169 (0.0895)
CS	0	0	0	0	1	-0.1742 (0.0836)
FFR	0	0	0	0	0	1
Panel B: named factors scheme - FAVAR-NF						
LM	-1.2426	0.3996	0.2934	0.2029	0.0824	-0.4235 (0.0449)
P	-0.8293	-0.9389	0.0315	-0.2053	0.0276	-0.1201 (0.0738)
IP	-0.6111	-0.0810	-0.2619	-0.1330	-1.1376	0.0337 (0.0870)
SM	1.7036	-0.3395	0.9813	1.1365	0.3018	0.1589 (0.0752)
CS	-1.1194	-0.1157	-0.9908	0.7985	0.5888	-0.1626 (0.0836)
FFR	0	0	0	0	0	1

Note: This tables shows the contemporaneous impact matrices associated with the different identification schemes. The factors are abbreviated as follows: 'LM' labor market, 'P' price, 'IP' industrial production, 'SM' stock market, 'CS' credit spread and 'FFR' Federal Funds rate. Panel A shows the impact matrix for our structural regularized FAVAR model (*IRa*). Panel B gives the impact matrix associated with the named factor identification scheme (*IRb*). Standard errors are given in brackets.

Appendix 1.B Figures

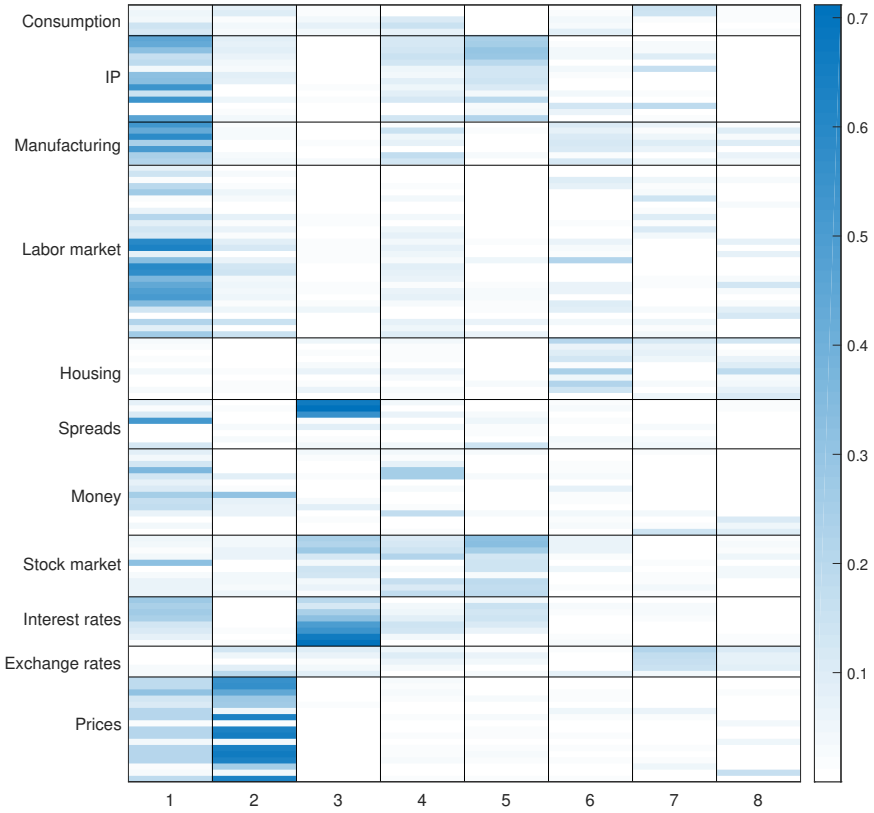


Figure 1.B.1: R^2 plot for a factor model based on principal components analysis
This graph shows R^2 for univariate regressions of the observed variables on each of the eight factors. The number of factors is chosen based on the Bai and Ng (2002) IC_1 criterion.

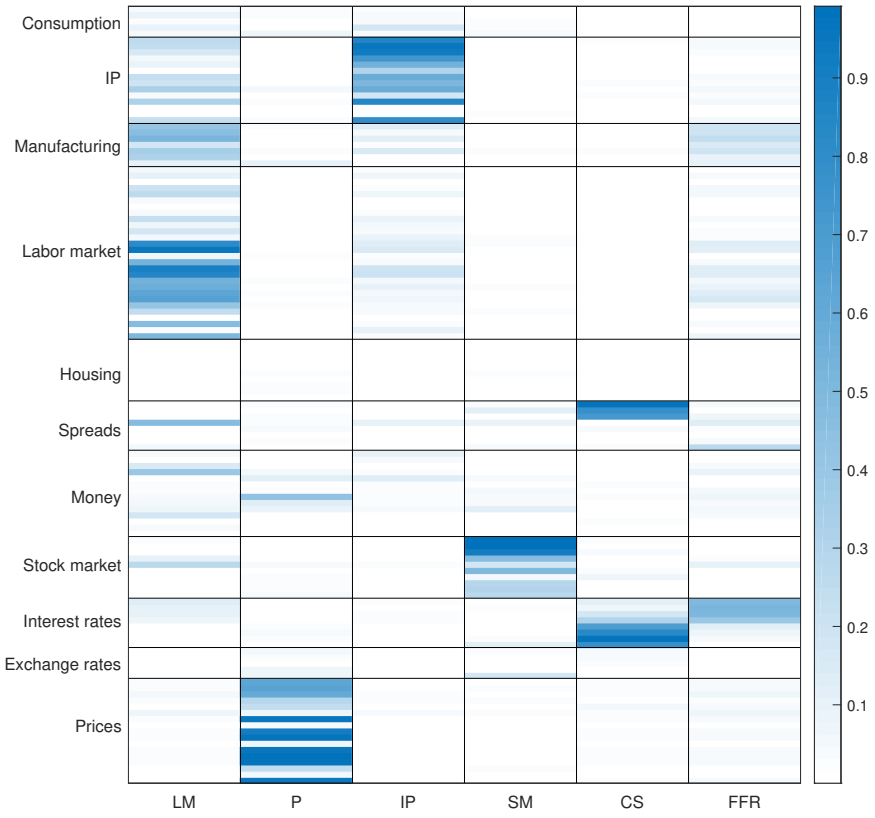


Figure 1.B.2: R^2 plot for the regularized FAVAR
This graph shows R^2 for univariate regressions of the observed variables on each of the factors. The factors are abbreviated as follows: 'LM' labor market, 'P' price, 'IP' industrial production, 'SM' stock market, 'CS' credit spread and 'FFR' Federal Funds rate.

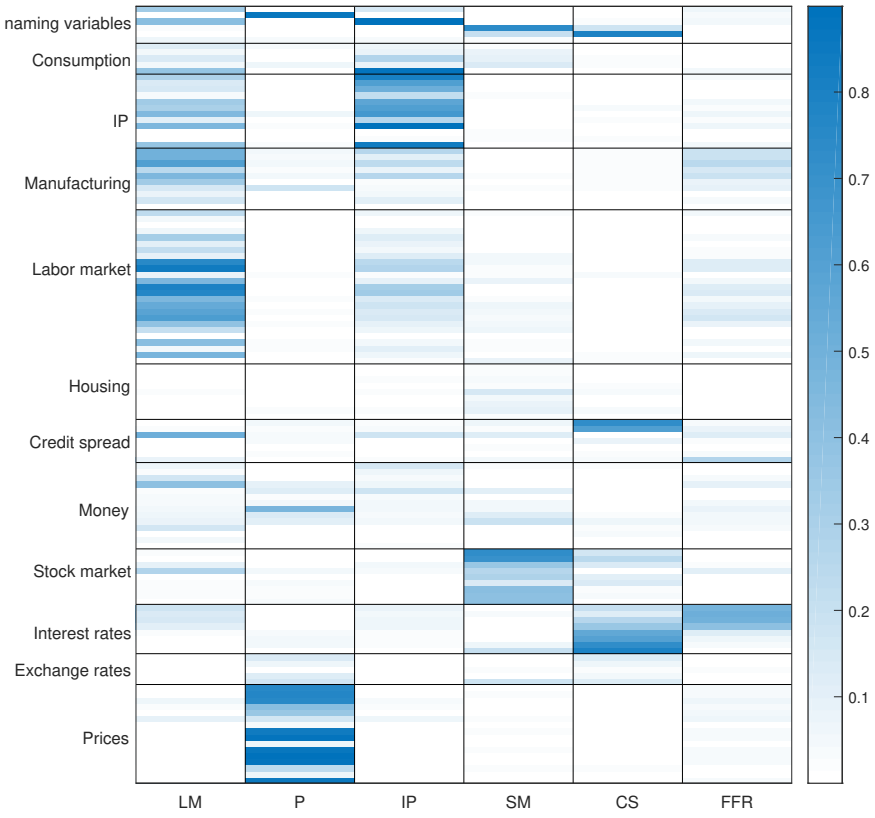


Figure 1.B.3: R^2 plot for the named factor FAVAR
This graph shows R^2 for univariate regressions of the observed variables on each of the factors. The factors are abbreviated as follows: 'LM' labor market, 'P' price, 'IP' industrial production, 'SM' stock market, 'CS' credit spread and 'FFR' Federal Funds rate. The naming variables are given in Table 1.E.2 in Appendix 1.E.

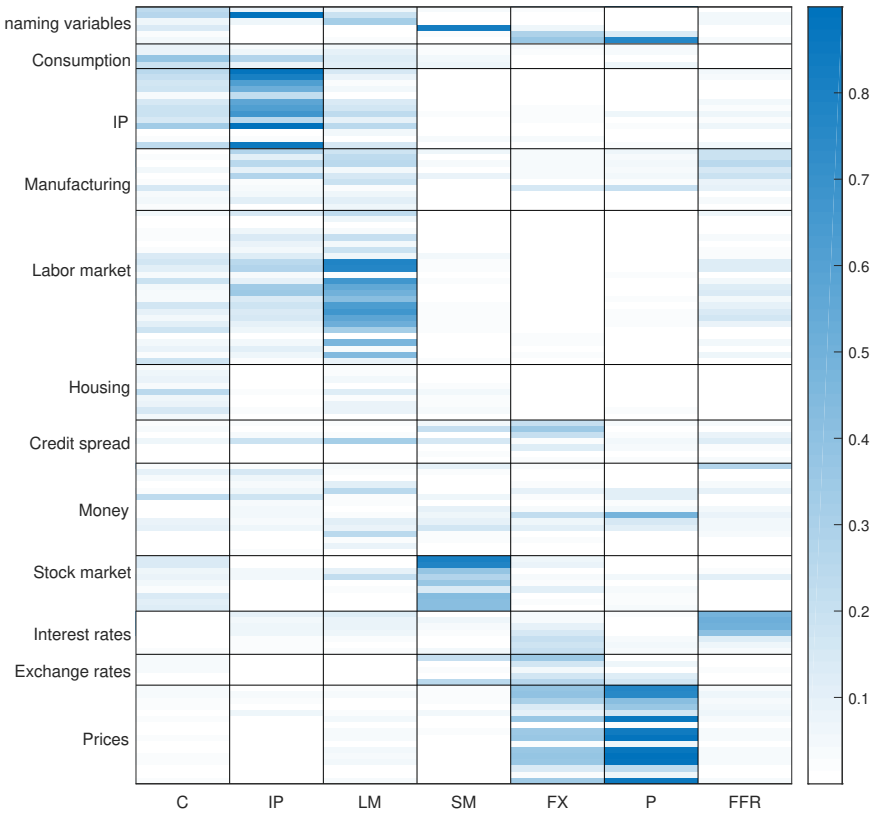


Figure 1.B.4: R^2 plot for the named factor FAVAR: oil price application
This graph shows R^2 for univariate regressions of the observed variables on each of the factors. The factors are abbreviated as follows: 'LM' labor market, 'P' price, 'IP' industrial production, 'SM' stock market, 'C' consumption, 'FX' is a trade weighted currency index and 'FFR' Federal Funds rate. The naming variables are based on Stock and Watson (2016).

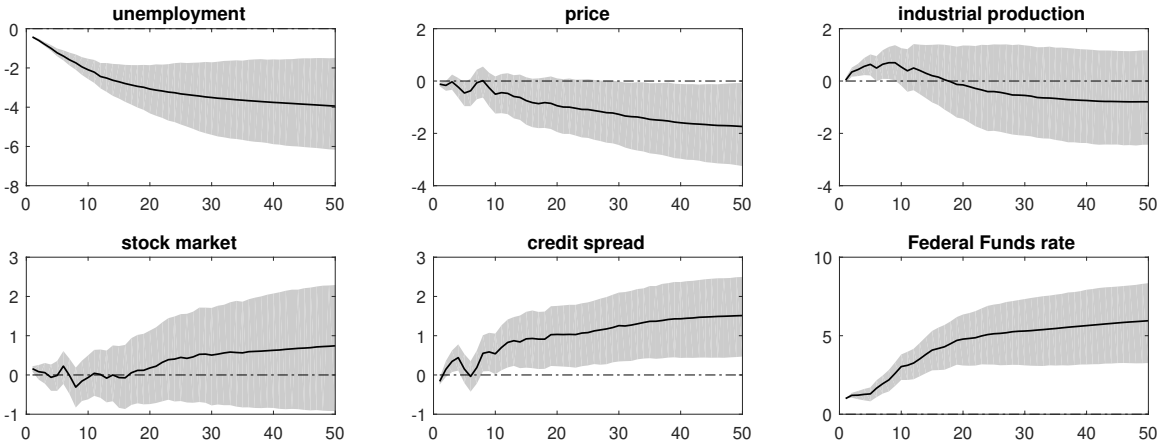


Figure 1.B.5: Accumulated impulse responses to a monetary policy shock on the factors for the named factor FAVAR
The graph shows accumulated impulse responses to a 100bp shock in the innovation of the FFR. The dashed lines correspond to 68% bootstrap confidence intervals.

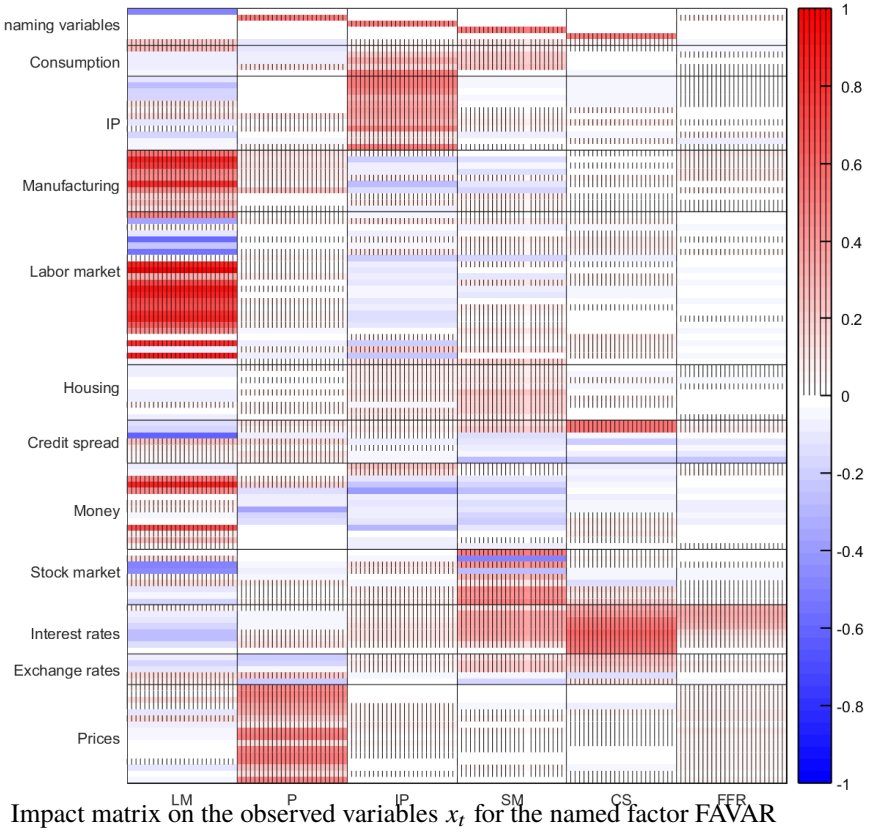


Figure 1.B.6: Impact matrix on the observed variables x_t for the named factor FAVAR
This graph shows the contemporaneous impact matrix of a 100bp shock in the factor innovations to the observed time series for a factor model with named factor identification. The factors are abbreviated as follows: 'LM' labor market, 'P' price, 'IP' industrial production, 'SM' stock market, 'CS' credit spread and 'FFR' Federal Funds rate. The naming variables for the first scheme are given in Table 1.E.2 in Appendix 1.E.

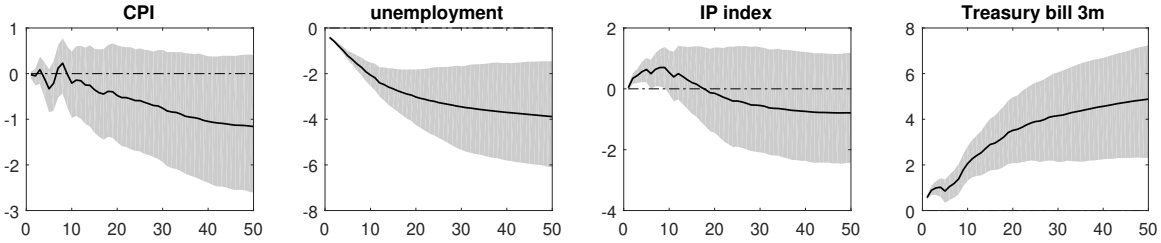


Figure 1.B.7: Accumulated impulse responses to a monetary policy shock on the observed variables x_t for the named factor FAVAR
The graph shows accumulated impulse responses to a 100bp shock in the innovation of the FFR. The dashed lines correspond to 68% bootstrap confidence intervals.

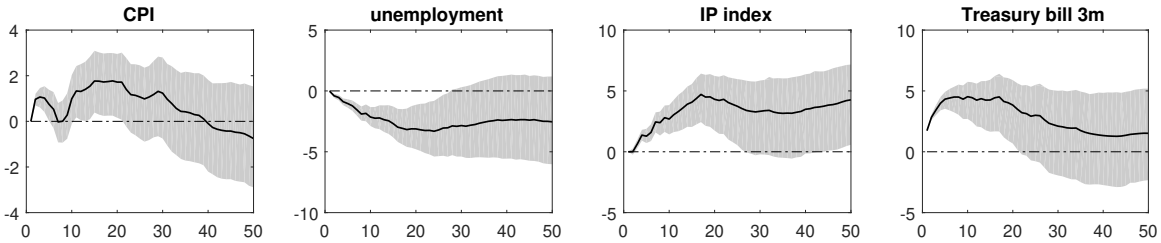


Figure 1.B.8: Accumulated impulse responses on the observed variables x_t for the dynamic factor model
The graph shows accumulated impulse responses to a 100bp shock in the innovation of the FFR. The dashed lines correspond to 68% bootstrap confidence intervals.

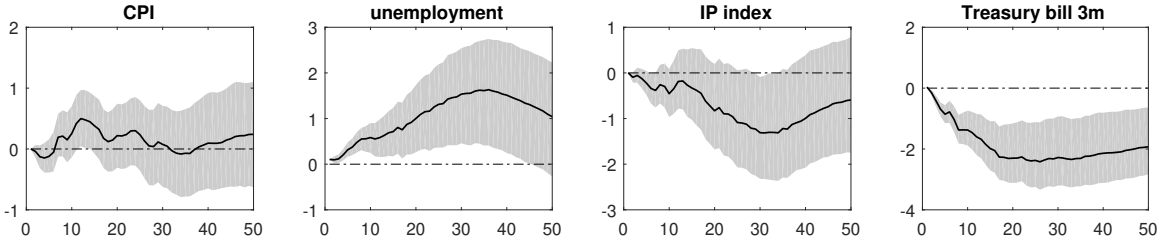


Figure 1.B.9: Accumulated impulse responses on the observed variables x_t for the VAR-F model
The graph shows impulse responses to a 100bp shock in the innovation of the FFR. The dashed lines correspond to 68% bootstrap confidence intervals.

Appendix 1.C Implementation of the regularized FAVAR estimator

1.C.1 Majorize-minimize expectation maximization algorithm

For the implementation of the regularized FAVAR model, we employ the majorize-minimize expectation maximization (EM) algorithm by Bien and Tibshirani (2011). The idea of this algorithm is to replace the optimization of the nonconvex objective function in equation (1.2.4) by a sequence of convex problems that are numerically easy to solve by algorithms

for convex optimization. In that respect, we majorize the log-likelihood function in (1.2.4) by the tangent plane of the concave part $\log\left|\tilde{\Sigma}\right|$, which leads to the following expression

$$\mathcal{L}_m^* = \log\left|\tilde{\Sigma}_m\right| + \operatorname{tr}\left[\left(2\tilde{\Sigma}_{H,m}\tilde{\Lambda}_m'\right)\tilde{\Sigma}_m^{-1}\left(\Lambda - \tilde{\Lambda}_m\right)\right] + \operatorname{tr}\left[S_x\Sigma^{-1}\right], \tag{1.C.1}$$

where $\tilde{\Sigma}_m = \tilde{\Lambda}_m\tilde{\Sigma}_{H,m}\tilde{\Lambda}_m' + \hat{\Phi}_{e,m}$ and $\Sigma = \Lambda\tilde{\Sigma}_{H,m}\Lambda' + \hat{\Phi}_{e,m}$ with $\tilde{\Sigma}_{H,m} = \begin{bmatrix}\tilde{\Sigma}_{F,m} & 0\\ 0 & \tilde{\Sigma}_{G,m}\end{bmatrix}$. $\tilde{F}$ denotes a initial estimate for the latent factors that can be obtained by unpenalized maximum likelihood and the subscript m is the m-th step in the iterative procedure. In the following, we augment the majorized log-likelihood function in (1.C.1) by L_1-penalty terms for the factor loadings Λ^f and Λ^g, which leads to the following optimization problem for our regularized FAVAR model

$$\begin{aligned}\min_{\{\Lambda\}}\ &\log\left|\tilde{\Sigma}_m\right| + \operatorname{tr}\left[\left(2\tilde{\Lambda}_m'\right)\tilde{\Sigma}_m^{-1}\left(\Lambda - \tilde{\Lambda}_m\right)\right] + \operatorname{tr}\left[S_x\Sigma^{-1}\right]\\ &+ \mu_1\sum_{i=1}^{N}\sum_{k=1}^{r_1}\left|\lambda_{ik}^f\right| + \mu_2\sum_{i=1}^{N}\sum_{l=1}^{r_2}\left|\lambda_{il}^g\right|.\end{aligned} \tag{1.C.2}$$

As the optimization problem in equation (1.C.2) has the appealing property of being entirely convex it can be easily solved by a convex optimizer. In this respect, we rely on an efficient projected gradient descent algorithm and optimize the following minimization problem

$$\min_{\{\Lambda\}}\frac{1}{2c}\sum_{i=1}^{N}\sum_{j=1}^{r}\left(\lambda_{ij} - \tilde{\lambda}_{ij,m} + c\cdot\tilde{D}_{ij,m}\right)^2 + \mu_1\sum_{i=1}^{N}\sum_{k=1}^{r_1}\left|\lambda_{ik}^f\right| + \mu_2\sum_{i=1}^{N}\sum_{l=1}^{r_2}\left|\lambda_{il}^g\right|, \tag{1.C.3}$$

where c determines the depth of the projection in the gradient decent algorithm[18] and

$$\tilde{D}_m = \left[\tilde{\Sigma}_m^{-1} - \tilde{\Sigma}_m^{-1}S_x\tilde{\Sigma}_m^{-1}\right]\left(2\tilde{\Lambda}_m'\right),$$

which corresponds to the first-derivative of $\mathcal{L}_m^*$ with respect to each element of Λ. The minimization of the objective function (1.C.3) leads to the following first order conditions with respect to each element of the quantities Λ^f and Λ^g

$$\begin{aligned}\frac{\partial}{\partial\lambda_{ik}^f} &= \sum_{i=1}^{N}\sum_{k=1}^{r_1}\left(\lambda_{ik} - \tilde{\lambda}_{ik} + c\cdot\tilde{D}_{ik,m}\right) + c\cdot\mu_1\sum_{k=1}^{r_1}\sum_{i=1}^{N}\nu_{ik}^f \overset{!}{=} 0,\\ \frac{\partial}{\partial\lambda_{il}^g} &= \sum_{i=1}^{N}\sum_{l=r_1+1}^{r}\left(\lambda_{il} - \tilde{\lambda}_{il} + c\cdot\tilde{D}_{il,m}\right) + c\cdot\mu_2\sum_{l=r_1+1}^{r}\sum_{i=1}^{N}\nu_{il}^g \overset{!}{=} 0,\end{aligned}$$

[18] In all our applications we set $c = 0.01$.

where ν_{ik}^f and ν_{il}^g are the subgradients of $\left|\lambda_{ik}^f\right|$ and $\left|\lambda_{il}^g\right|$, respectively. Hence, solving for a specific $\tilde{\lambda}_{ik}^f$ or $\tilde{\lambda}_{ik}^g$ leads to the following updating formulas for the factor loadings estimates

$$\tilde{\lambda}_{ik,m+1}^f = \mathcal{S}\left(\tilde{\lambda}_{ik,m}^f - c \cdot \tilde{D}_{ik,m}, c \cdot \mu_1\right), \quad \text{for } i = 1,\dots,N;\ k = 1,\dots,r_1 \tag{1.C.4}$$

$$\tilde{\lambda}_{il,m+1}^g = \mathcal{S}\left(\tilde{\lambda}_{il,m}^g - c \cdot \tilde{D}_{il,m}, c \cdot \mu_2\right), \quad \text{for } i = 1,\dots,N;\ l = 1,\dots,r_2, \tag{1.C.5}$$

where $\mathcal{S}(\cdot)$ is the soft-thresholding function defined in (1.2.8).

To obtain an update for the estimate of the covariance matrix of the idiosyncratic error Φ_e, we use the formula in the EM algorithm suggested by Bai and Li (2012)

$$\tilde{\Phi}_{e,m+1} = \operatorname{diag}\left[S_x - \tilde{\Lambda}_{m+1}\tilde{\Lambda}_m'\left(\tilde{\Lambda}_m\tilde{\Lambda}_m' + \tilde{\Phi}_{e,m}\right)^{-1} S_x\right].$$

Our iterative estimation procedure for the regularized FAVAR model is therefore described by the following steps:

Starting from the FAVAR model $X = \Lambda^f F' + \Lambda^g G' + e$

Step 1: Transform the data by linearly projecting the observed factors G from the observed data. This is achieved by post-multiplying the FAVAR model by $M = I_T - G(G'G)^{-1}G'$. We obtain the transformed data $XM = \Lambda^f F'M + eM$ or $\dot{X} = \Lambda^f \dot{F}' + \dot{e}$.

Step 2: Set $m = 1$ and obtain an initial estimate for the factor loading matrix $\tilde{\Lambda}_m^f$, factors $\tilde{F}_m$ and the diagonal idiosyncratic error covariance matrix $\tilde{\Phi}_{e,m}$, e.g. by using unpenalized MLE on $\dot{X}$. Get an initial estimate of the factor loadings of the observed factor by $\tilde{\Lambda}_m^g = \left(X - \tilde{\Lambda}_m^f \tilde{F}_m'\right) G\,(G'G)^{-1}$.

Step 3: Update $\tilde{\lambda}_{ik,m}^f$ by $\tilde{\lambda}_{ik,m+1}^f = \mathcal{S}\left(\tilde{\lambda}_{ik,m}^f - c \cdot \tilde{D}_{ik,m},\, c \cdot \mu_1\right)$, for $i = 1,\dots,N;\ k = 1,\dots,r_1$ and $\tilde{\lambda}_{il,m}^g$ by $\tilde{\lambda}_{il,m+1}^g = \mathcal{S}\left(\tilde{\lambda}_{il,m} - c \cdot \tilde{D}_{il,m}, c \cdot \mu_2\right)$, for $i = 1,\dots,N;\ l = 1,\dots,r_2$.

Step 4: Update $\tilde{\Phi}_e$ using the EM algorithm in Bai and Li (2012), according to $\tilde{\Phi}_{e,m+1} = \operatorname{diag}\left[S_x - \tilde{\Lambda}_{m+1}\tilde{\Lambda}_m'\left(\tilde{\Lambda}_m\tilde{\Lambda}_m' + \tilde{\Phi}_{e,m}\right)^{-1} S_x\right]$

Step 5: If $\left\|\tilde{\Lambda}_{m+1}^f - \tilde{\Lambda}_m^f\right\|$ and $\left\|\tilde{\Phi}_{e,m+1} - \tilde{\Phi}_{e,m}\right\|$ are sufficiently small, stop the procedure, otherwise set $m = m + 1$ and return to *Step 3*.

Step 6: Estimate the latent factors by $\tilde{f}_t = \left(\tilde{\Lambda}^{f\prime}\tilde{\Phi}_e^{-1}\tilde{\Lambda}^f\right)^{-1}\tilde{\Lambda}^{f\prime}\tilde{\Phi}_e^{-1}x_t$, where $\tilde{\Lambda}^f$ and $\tilde{\Phi}_e$ are the parameter estimates after convergence.

Step 7: Re-estimate the covariance matrix of the idiosyncratic errors based on the procedure introduced in Section 1.2.3.

Step 8: Based on the estimated and observed factors, $\tilde{h}_t = \left(\tilde{f}_t', g_t'\right)'$, we estimate the following VAR regression: $\tilde{h}_t = \Phi_1 \tilde{h}_{t-1} + \Phi_2 \tilde{h}_{t-2} + \cdots + \Phi_p \tilde{h}_{t-p} + \eta_t$, we obtain the residuals $\tilde{\eta}_t$ from the previous regression and calculate $\tilde{\Omega} = \frac{1}{T-p} \sum_{t=p+1}^{T} \tilde{\eta}_t \tilde{\eta}_t'$.

Step 9: For the identification of our regularized FAVAR model, we use the identification restriction *IRa*. Thus, the factor loadings estimates for the unobserved factors are unchanged $\hat{\Lambda}^f = \tilde{\Lambda}^f$. Further, we estimate $\hat{\Lambda}^g = \tilde{\Lambda}^g + \tilde{\Lambda}^f \tilde{\Omega}_{fg} \tilde{\Omega}_{gg}^{-1}$, $\hat{F} = \tilde{F} - G\tilde{\Omega}_{gg}^{-1} \tilde{\Omega}_{gf}$ and the autoregressive parameters by $\hat{\Phi}_i = \tilde{A} \tilde{\Phi}_i \tilde{A}^{-1}$, for $i = 1, \ldots, p$, with the rotation matrix $\tilde{A} = \begin{bmatrix} I_{r_1} & -\tilde{\Omega}_{fg} \tilde{\Omega}_{gg}^{-1} \\ 0 & I_{r_2} \end{bmatrix}$.

1.C.2 Selection of the regularization parameters

We provide a selection criterion to estimate the regularization parameters μ_1 and μ_2. The information criterion is based on an adaptation of the Bayesian information criterion, comparable to the criteria in Bai and Ng (2002), and takes the following form

$$IC(\mu_1, \mu_2) = \mathcal{L}\left(\tilde{\Lambda}, S_{\tilde{H}}, \tilde{\Sigma}_e^{\tau}\right) + \kappa_{\mu_1,\mu_2} \sqrt{\frac{\log(2N)}{N} + \frac{\log N}{NT}}, \tag{1.C.6}$$

where κ_{μ_1,μ_2} denotes the number of non-zero elements in the factor loadings matrix $\tilde{\Lambda}$ for specific μ_1 and μ_2. $\mathcal{L}\left(\tilde{\Lambda}, S_{\tilde{H}}, \tilde{\Sigma}_e^{\tau}\right)$ is the value of the log-likelihood function in equation (1.2.4) evaluated at the estimates of the factor loadings $\tilde{\Lambda}$, the sample covariance matrix of the factors $S_{\tilde{H}}$ and the covariance matrix of the idiosyncratic component $\tilde{\Sigma}_e^{\tau}$. The penalty function in (1.C.6) relates to the convergence rate of the estimator $\tilde{\Lambda}$ and vanishes for $N, T \to \infty$. In order to select the optimal penalty parameters, we calculate the information criterion in (1.C.6) for a grid of different values for μ_1 and μ_2 and select the ones that minimize the criterion. The grids for the regularization parameters are set to $\mu_1 = \left[0, \mu_{1,\max}\right]$ and $\mu_2 = \left[0, \mu_{2,\max}\right]$, where $\mu_{1,\max}$ and $\mu_{2,\max}$ denote the highest values of the penalty parameters such that we do not select an empty loadings matrix and the set of identifying conditions *IRa* in Section 1.2.4 are still fulfilled.

Appendix 1.D Consistency of the regularized FAVAR estimator

Proof of Theorem 1.3.1 (Consistency of the estimators of the regularized FAVAR model before rotation).

To establish the consistency of the regularized FAVAR model, we proceed in a similar fashion as in Daniele et al. (2018). Initially, we define the following penalized log-likelihood

$$\mathcal{L}_p(\bar{\Lambda}, \bar{\Sigma}_H, \bar{\Sigma}_e) = Q_1(\bar{\Lambda}, \bar{\Sigma}_e) + Q_2(\bar{\Lambda}, \bar{\Sigma}_e) + Q_3(\bar{\Lambda}, \bar{\Sigma}_H, \bar{\Sigma}_e), \tag{1.D.1}$$

where

$$\begin{aligned}
Q_1(\bar{\Lambda}, \bar{\Sigma}_e) &= \frac{1}{N}\log\left|\bar{\Sigma}_e\right| + \frac{1}{N}\mathrm{tr}\left(S_e\bar{\Sigma}_e^{-1}\right) - \frac{1}{N}\log|\Sigma_e| - \frac{1}{N}\mathrm{tr}\left(S_e\Sigma_e^{-1}\right) \\
&\quad + \frac{1}{N}\mu_1\sum_{i=1}^{N}\sum_{k=1}^{r_1}\left(\left|\bar{\lambda}_{ik}^f\right| - \left|\lambda_{ik}^f\right|\right) + \frac{1}{N}\mu_2\sum_{i=1}^{N}\sum_{l=1}^{r_2}\left(\left|\bar{\lambda}_{il}^g\right| - \left|\lambda_{il}^g\right|\right), \\
Q_2(\bar{\Lambda}, \bar{\Sigma}_e) &= \frac{1}{N}\mathrm{tr}\left[\left(\bar{\Lambda}-\Lambda\right)'\bar{\Sigma}_e^{-1}\left(\bar{\Lambda}-\Lambda\right)S_H - \left(\bar{\Lambda}-\Lambda\right)'\bar{\Sigma}_e^{-1}\bar{\Lambda}\left(\bar{\Lambda}'\bar{\Sigma}_e^{-1}\bar{\Lambda}\right)^{-1}\bar{\Lambda}'\bar{\Sigma}_e^{-1}\left(\bar{\Lambda}-\Lambda\right)S_H\right], \\
Q_3(\bar{\Lambda}, \bar{\Sigma}_H, \bar{\Sigma}_e) &= \frac{1}{N}\log\left|\bar{\Lambda}\bar{\Sigma}_H\bar{\Lambda}' + \bar{\Sigma}_e\right| + \frac{1}{N}\mathrm{tr}\left[S_x\left(\bar{\Lambda}\bar{\Sigma}_H\bar{\Lambda}' + \bar{\Sigma}_e\right)^{-1}\right] - Q_2(\bar{\Lambda}, \bar{\Sigma}_e) \\
&\quad - \frac{1}{N}\log\left|\bar{\Sigma}_e\right| - \frac{1}{N}\mathrm{tr}\left(S_e\bar{\Sigma}_e^{-1}\right),
\end{aligned}$$

and $\Lambda = \left[\Lambda^f \ \Lambda^g\right]$. Hence, the penalized log-likelihood in (1.D.1) can be written as

$$\begin{aligned}
\mathcal{L}_p(\bar{\Lambda}, \bar{\Sigma}_H, \bar{\Sigma}_e) &= \frac{1}{N}\log\left|\bar{\Lambda}\bar{\Sigma}_H\bar{\Lambda}' + \bar{\Sigma}_e\right| + \frac{1}{N}\mathrm{tr}\left[S_x\left(\bar{\Lambda}\bar{\Sigma}_H\bar{\Lambda}' + \bar{\Sigma}_e\right)^{-1}\right] \\
&\quad - \frac{1}{N}\log|\Sigma_e| - \frac{1}{N}\mathrm{tr}\left(S_e\Sigma_e^{-1}\right) \\
&\quad + \frac{1}{N}\mu_1\sum_{i=1}^{N}\sum_{k=1}^{r_1}\left(\left|\bar{\lambda}_{ik}^f\right| - \left|\lambda_{ik}^f\right|\right) + \frac{1}{N}\mu_2\sum_{i=1}^{N}\sum_{l=1}^{r_2}\left(\left|\bar{\lambda}_{il}^g\right| - \left|\lambda_{il}^g\right|\right).
\end{aligned} \tag{1.D.2}$$

Given the constants $c, d_1, d_2, d_3, d_4 > 0$, consider the following set

$$\begin{aligned}
\Psi_\delta = \{(\Lambda, \Sigma_H, \Sigma_e) :\ & c^{-1} < \pi_{\min}\left(\frac{\Lambda'\Lambda}{N^\beta}\right) \leq \pi_{\max}\left(\frac{\Lambda'\Lambda}{N^\beta}\right) < c, \\
& c_2 < \pi_{\min}(\Sigma_H) \leq \pi_{\max}(\Sigma_H) < c_1, \\
& c_4 < \pi_{\min}(\Sigma_e) \leq \pi_{\max}(\Sigma_e) < c_3\}, \quad \text{for } 1/2 \leq \beta \leq 1.
\end{aligned}$$

Further, we define $\Phi_e = \text{diag}(\Sigma_e)$, which corresponds to a covariance matrix that contains only the diagonal elements of Σ_e on its main diagonal. To control the sparsity in both Λ and Σ_e, we impose the following sparsity assumptions as in Assumption 1.3.2

$$L_N = \sum_{k=1}^{r} \sum_{i=1}^{N} \mathbb{1}\{\lambda_{ik} \neq 0\} \asymp N^{\beta}, \text{ for } 1/2 \leq \beta \leq 1,$$
$$S_N = \max_{i \leq N} \sum_{j=1}^{N} \mathbb{1}\left\{\sigma_{e,ij} \neq 0\right\},$$

where $\mathbb{1}\{\cdot\}$ denotes the indicator function that is equal to one if the boolean argument in braces is true. Hence, L_N is the number of non-zero elements in the factor loadings matrix Λ and S_N denotes the maximum number of non-zero elements in each row of Σ_e.

The following lemma will be useful for the forthcoming derivations.

Lemma 1.D.1.

(i) $\max_{i,j \leq N} \left|\frac{1}{T}\sum_{t=1}^{T} e_{it}e_{jt} - \mathrm{E}\left[e_{it}e_{jt}\right]\right| = \mathcal{O}_p\left(\sqrt{(\log N)/T}\right)$,

(ii) $\max_{i,j \leq r} \left|\frac{1}{T}\sum_{t=1}^{T} f_{it}f_{jt} - \mathrm{E}\left[f_{it}f_{jt}\right]\right| = \mathcal{O}_p\left(\sqrt{(\log T)/T}\right)$,

(iii) $\max_{i \leq r, j \leq N} \left|\frac{1}{T}\sum_{t=1}^{T} h_{it}e_{jt}\right| = \mathcal{O}_p\left(\sqrt{(\log N)/T}\right)$.

Proof. See *Lemmas A.3 and B.1* in Fan et al. (2011).

□

Lemma 1.D.2.

$$\sup_{(\Lambda,\Sigma_H,\Sigma_e)\in\Psi_\delta} \left|Q_3(\tilde{\Lambda},\tilde{\Sigma}_H,\tilde{\Sigma}_e)\right| = \mathcal{O}_p\left(\frac{\log N^{\beta}}{N} + \frac{1}{N^{\beta}}\frac{\log N}{T}\right).$$

Proof. The proof can be conducted in the a similar fashion as in Daniele et al. (2018). We start with the unpenalized log-likelihood, which is given by

$$\mathcal{L}(\bar{\Lambda},\bar{\Sigma}_H,\bar{\Sigma}_e) = \frac{1}{N}\log\left|\bar{\Lambda}\bar{\Sigma}_H\bar{\Lambda}' + \bar{\Sigma}_e\right| + \frac{1}{N}\text{tr}\left[S_x\left(\bar{\Lambda}\bar{\Sigma}_H\bar{\Lambda}' + \bar{\Sigma}_e\right)^{-1}\right]. \tag{1.D.3}$$

The first term on the right-hand side of equation (1.D.3) can be expressed as

$$\frac{1}{N}\log\left|\bar{\Lambda}\bar{\Sigma}_H\bar{\Lambda}' + \bar{\Sigma}_e\right| = \frac{1}{N}\log\left|\bar{\Sigma}_e\right| + \frac{1}{N}\log\left|\bar{\Sigma}_H\right| + \frac{1}{N}\log\left|\bar{\Sigma}_H^{-1} + \bar{\Lambda}'\bar{\Sigma}_e^{-1}\bar{\Lambda}\right|$$

Hence, we obtain

$$\begin{aligned}\frac{1}{N}\log\left|\bar{\Lambda}\bar{\Sigma}_H\bar{\Lambda}'+\bar{\Sigma}_e\right| &= \frac{1}{N}\log\left|\bar{\Sigma}_e\right| + \mathcal{O}\left(\frac{r}{N}\right) + \mathcal{O}\left(r\frac{\log N^\beta}{N}\right)\\ &= \frac{1}{N}\log\left|\bar{\Sigma}_e\right| + \mathcal{O}\left(\frac{\log N^\beta}{N}\right).\end{aligned} \tag{1.D.4}$$

We continue with analyzing the second term on the right-hand side of (1.D.3). The variable S_x is given by

$$S_x = \frac{1}{T}\sum_{t=1}^{T} x_t x_t' = \Lambda S_H \Lambda' + S_e + \Lambda\frac{1}{T}\sum_{t=1}^{T} h_t e_t' + \left(\Lambda\frac{1}{T}\sum_{t=1}^{T} h_t e_t'\right)',$$

where $S_H = \frac{1}{T}\sum_{t=1}^{T} h_t h_t'$ and $S_e = \frac{1}{T}\sum_{t=1}^{T} e_t e_t'$. Moreover, based on the matrix inversion formula we have

$$\left(\bar{\Lambda}\bar{\Sigma}_H\bar{\Lambda}'+\bar{\Sigma}_e\right)^{-1} = \bar{\Sigma}_e^{-1} - \bar{\Sigma}_e^{-1}\bar{\Lambda}\left(\bar{\Sigma}_H^{-1}+\bar{\Lambda}'\bar{\Sigma}_e^{-1}\bar{\Lambda}\right)^{-1}\bar{\Lambda}'\bar{\Sigma}_e^{-1}.$$

Hence, we obtain

$$\begin{aligned}\frac{1}{N}\mathrm{tr}\left[S_x\left(\bar{\Lambda}\bar{\Sigma}_H\bar{\Lambda}'+\bar{\Sigma}_e\right)^{-1}\right] &= \frac{1}{N}\mathrm{tr}\left(\Lambda'\bar{\Sigma}_e^{-1}\Lambda S_H\right) + \frac{1}{N}\mathrm{tr}\left(S_e\bar{\Sigma}_e^{-1}\right)\\ &\quad - A_1 + A_2 + A_3 - A_4,\end{aligned} \tag{1.D.5}$$

where $A_1 = \frac{1}{N}\mathrm{tr}\left[\Lambda S_H\Lambda'\bar{\Sigma}_e^{-1}\bar{\Lambda}\left(\bar{\Sigma}_H^{-1}+\bar{\Lambda}'\bar{\Sigma}_e^{-1}\bar{\Lambda}\right)^{-1}\bar{\Lambda}'\bar{\Sigma}_e^{-1}\right]$, $A_2 = \frac{1}{N}\mathrm{tr}\left[\Lambda\frac{1}{T}\sum_{t=1}^{T} h_t e_t'\left(\bar{\Lambda}\bar{\Sigma}_H\bar{\Lambda}'+\bar{\Sigma}_e\right)^{-1}\right]$, $A_3 = \frac{1}{N}\mathrm{tr}\left[\frac{1}{T}\sum_{t=1}^{T} e_t h_t'\Lambda'\left(\bar{\Lambda}\bar{\Sigma}_H\bar{\Lambda}'+\bar{\Sigma}_e\right)^{-1}\right]$ and $A_4 = \frac{1}{N}\mathrm{tr}\left[S_e\bar{\Sigma}_e^{-1}\bar{\Lambda}\left(\bar{\Sigma}_H^{-1}+\bar{\Lambda}'\bar{\Sigma}_e^{-1}\bar{\Lambda}\right)^{-1}\bar{\Lambda}'\bar{\Sigma}_e^{-1}\right]$.

In the following, we bound each term $A_1 - A_4$, separately. By the matrix inversion formula, we can write A_1 as follows

$$\begin{aligned}A_1 &= \frac{1}{N}\mathrm{tr}\left[\Lambda'\bar{\Sigma}_e^{-1}\bar{\Lambda}\left(\bar{\Lambda}'\bar{\Sigma}_e^{-1}\bar{\Lambda}\right)^{-1}\bar{\Lambda}'\bar{\Sigma}_e^{-1}\Lambda S_H\right]\\ &\quad - \frac{1}{N}\mathrm{tr}\left[\Lambda'\bar{\Sigma}_e^{-1}\bar{\Lambda}\left(\bar{\Lambda}'\bar{\Sigma}_e^{-1}\bar{\Lambda}\right)^{-1}\bar{\Sigma}_H^{-1}\left(\bar{\Sigma}_H^{-1}+\bar{\Lambda}'\bar{\Sigma}_e^{-1}\bar{\Lambda}\right)^{-1}\bar{\Lambda}'\bar{\Sigma}_e^{-1}\Lambda S_H\right].\end{aligned}$$

The second term can be bounded using Lemma 1.D.1., $\pi_{\max}\left[\left(\bar{\Sigma}_H^{-1}+\bar{\Lambda}'\bar{\Sigma}_e^{-1}\bar{\Lambda}\right)^{-1}\right] \leq \pi_{\max}\left[\left(\bar{\Lambda}'\bar{\Sigma}_e^{-1}\bar{\Lambda}\right)^{-1}\right] = \mathcal{O}_p\left(N^{-\beta}\right)$ and $\pi_{\max}(S_H) \leq \pi_{\max}(\Sigma_H) + \|S_H-\Sigma_H\|_F \leq \mathcal{O}(1) + r\|S_H-\Sigma_H\|_{\max} = \mathcal{O}_p\left(1+\sqrt{\frac{\log T}{T}}\right)$ by

$$\begin{aligned}&\frac{1}{N}\text{tr}\left[\Lambda'\bar{\Sigma}_e^{-1}\bar{\Lambda}\left(\bar{\Lambda}'\bar{\Sigma}_e^{-1}\bar{\Lambda}\right)^{-1}\bar{\Sigma}_H^{-1}\left(\bar{\Sigma}_H^{-1}+\bar{\Lambda}'\bar{\Sigma}_e^{-1}\bar{\Lambda}\right)^{-1}\bar{\Lambda}'\bar{\Sigma}_e^{-1}\Lambda S_H\right]\\ &\quad\leq \frac{1}{N}\left\|\Lambda'\bar{\Sigma}_e^{-1}\bar{\Lambda}\right\|_F^2\mathcal{O}\left(N^{-2\beta}\right)\mathcal{O}_p\left(1+\sqrt{\frac{\log T}{T}}\right)=\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{N}\sqrt{\frac{\log T}{T}}\right).\end{aligned}$$

Thus,

$$A_1 = \frac{1}{N}\text{tr}\left[\Lambda'\bar{\Sigma}_e^{-1}\bar{\Lambda}\left(\bar{\Lambda}'\bar{\Sigma}_e^{-1}\bar{\Lambda}\right)^{-1}\bar{\Lambda}'\bar{\Sigma}_e^{-1}\Lambda S_H\right]+\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{N}\sqrt{\frac{\log T}{T}}\right).$$

Since, $\pi_{\max}\left[\left(\bar{\Lambda}\bar{\Sigma}_H\bar{\Lambda}'+\bar{\Sigma}_e\right)^{-1}\right] \leq \pi_{\max}\left[\left(\bar{\Lambda}\bar{\Lambda}'+\bar{\Sigma}_e\right)^{-1}\right] = \mathcal{O}\left(N^{-\beta}\right)$, we can bound the terms A_2-A_4 in a similar fashion as in *Lemma S.1.2.* in Daniele et al. (2018). Hence, $A_2 = \mathcal{O}_p\left(\frac{1}{N}+\frac{1}{N^\beta}\frac{\log N}{T}\right), A_3 = \mathcal{O}_p\left(\frac{1}{N}+\frac{1}{N^\beta}\frac{\log N}{T}\right)$ and $A_4 = \frac{1}{N}\text{tr}\left(S_e\bar{\Sigma}_e^{-1}\right)\mathcal{O}(1)$.

Based on the previous results, we can bound the unpenalized likelihood by

$$\begin{aligned}\mathcal{L}(\bar{\Lambda},\bar{\Sigma}_H,\bar{\Sigma}_e) &= \frac{1}{N}\log\left|\bar{\Sigma}_e\right|+\frac{1}{N}\text{tr}\left(S_e\bar{\Sigma}_e^{-1}\right)+\frac{1}{N}\text{tr}\left(\Lambda'\bar{\Sigma}_e^{-1}\Lambda S_H\right)\\ &\quad-\frac{1}{N}\text{tr}\left[\Lambda'\bar{\Sigma}_e^{-1}\bar{\Lambda}\left(\bar{\Lambda}'\bar{\Sigma}_e^{-1}\bar{\Lambda}\right)^{-1}\bar{\Lambda}'\bar{\Sigma}_e^{-1}\Lambda S_H\right]+\mathcal{O}_p\left(\frac{\log N^\beta}{N}+\frac{1}{N^\beta}\frac{\log N}{T}\right)\\ &= \frac{1}{N}\log\left|\bar{\Sigma}_e\right|+\frac{1}{N}\text{tr}\left(S_e\bar{\Sigma}_e^{-1}\right)+\frac{1}{N}\text{tr}\left[\left(\bar{\Lambda}-\Lambda\right)'\bar{\Sigma}_e^{-1}\left(\bar{\Lambda}-\Lambda\right)S_H\right]\\ &\quad-\frac{1}{N}\text{tr}\left[\left(\bar{\Lambda}-\Lambda\right)'\bar{\Sigma}_e^{-1}\bar{\Lambda}\left(\bar{\Lambda}'\bar{\Sigma}_e^{-1}\bar{\Lambda}\right)^{-1}\bar{\Lambda}'\bar{\Sigma}_e^{-1}\left(\bar{\Lambda}-\Lambda\right)S_H\right]+\mathcal{O}_p\left(\frac{\log N^\beta}{N}+\frac{1}{N^\beta}\frac{\log N}{T}\right)\\ &= \frac{1}{N}\log\left|\bar{\Sigma}_e\right|+\frac{1}{N}\text{tr}\left(S_e\bar{\Sigma}_e^{-1}\right)+Q_2(\bar{\Lambda},\bar{\Sigma}_e)+\mathcal{O}_p\left(\frac{\log N^\beta}{N}+\frac{1}{N^\beta}\frac{\log N}{T}\right).\end{aligned}$$

The result follows from the definition of $Q_3(\bar{\Lambda},\bar{\Sigma}_H,\bar{\Sigma}_e)$.

□

Furthermore, by *Lemma S.1.3.* in Daniele et al. (2018), we obtain

$$Q_1\left(\tilde{\Lambda},\tilde{\Sigma}_e\right)+Q_2\left(\tilde{\Lambda},\tilde{\Sigma}_e\right) \leq d_T, \tag{1.D.6}$$

where $d_T = \frac{\log N^\beta}{N}+\frac{1}{N^\beta}\frac{\log N}{T}$.

In the following, we establish the consistency results for the diagonal idiosyncratic error covariance matrix estimator $\tilde{\Phi}_e$ and the factor loadings estimator $\tilde{\Lambda}$.

Lemma 1.D.3.

$$\frac{1}{N}\left\|\tilde{\Phi}_e - \Phi_e\right\|_F^2 = \mathcal{O}_p\left(\frac{\log N}{T} + d_T\right) = o_p(1).$$

Proof. By the definition of $Q_1\left(\tilde{\Lambda}, \tilde{\Sigma}_e\right)$ and $Q_2\left(\tilde{\Lambda}, \tilde{\Sigma}_e\right)$ and equation (1.D.6) we define

$$B_1 + B_2 \leq d_T, \tag{1.D.7}$$

where

$$\begin{aligned}
B_1 &= \frac{1}{N}\log\left|\tilde{\Sigma}_e\right| + \frac{1}{N}\mathrm{tr}\left(S_e\tilde{\Sigma}_e^{-1}\right) - \frac{1}{N}\log\left|\Sigma_e\right| - \frac{1}{N}\mathrm{tr}\left(S_e\Sigma_e^{-1}\right), \\
B_2 &= \frac{1}{N}\mathrm{tr}\left[\left(\tilde{\Lambda}-\Lambda\right)'\tilde{\Sigma}_e^{-1}\left(\tilde{\Lambda}-\Lambda\right)S_H - \left(\tilde{\Lambda}-\Lambda\right)'\tilde{\Sigma}_e^{-1}\tilde{\Lambda}\left(\tilde{\Lambda}'\tilde{\Sigma}_e^{-1}\tilde{\Lambda}\right)^{-1}\tilde{\Lambda}'\tilde{\Sigma}_e^{-1}\left(\tilde{\Lambda}-\Lambda\right)S_H\right] \\
&\quad + \frac{1}{N}\mu_1\sum_{i=1}^{N}\sum_{k=1}^{r_1}\left(\left|\tilde{\lambda}_{ik}^f\right| - \left|\lambda_{ik}^f\right|\right) + \frac{1}{N}\mu_2\sum_{i=1}^{N}\sum_{l=1}^{r_2}\left(\left|\tilde{\lambda}_{il}^g\right| - \left|\lambda_{il}^g\right|\right).
\end{aligned}$$

The result follows by the same argument as in *Lemma S.1.4.* in Daniele et al. (2018). □

To establish the consistency of the factor loadings matrix $\tilde{\Lambda}$, we analyze both sets of factor loadings corresponding to the latent and observed factors separately. Initially, we lower bound the first term in B_2 and obtain the following

Lemma 1.D.4.

$$\begin{aligned}
&\frac{1}{N}\mathrm{tr}\left[\left(\tilde{\Lambda}-\Lambda\right)'\tilde{\Sigma}_e^{-1}\left(\tilde{\Lambda}-\Lambda\right)S_H - \left(\tilde{\Lambda}-\Lambda\right)'\tilde{\Sigma}_e^{-1}\tilde{\Lambda}\left(\tilde{\Lambda}'\tilde{\Sigma}_e^{-1}\tilde{\Lambda}\right)^{-1}\tilde{\Lambda}'\tilde{\Sigma}_e^{-1}\left(\tilde{\Lambda}-\Lambda\right)S_H\right] \\
&\geq \mathcal{O}_p\left(\frac{1-L_N}{N}\right)\max_{i\leq N}\left\|\tilde{\lambda}_i - \lambda_i\right\|^2
\end{aligned}$$

Proof. Since $\pi_{\max}(S_H) = \mathcal{O}_p\left(1 + \sqrt{\frac{\log T}{T}}\right)$, we have

$$\begin{aligned}
&\frac{1}{N}\mathrm{tr}\left[\left(\tilde{\Lambda}-\Lambda\right)'\tilde{\Sigma}_e^{-1}\left(\tilde{\Lambda}-\Lambda\right)S_H - \left(\tilde{\Lambda}-\Lambda\right)'\tilde{\Sigma}_e^{-1}\tilde{\Lambda}\left(\tilde{\Lambda}'\tilde{\Sigma}_e^{-1}\tilde{\Lambda}\right)^{-1}\tilde{\Lambda}'\tilde{\Sigma}_e^{-1}\left(\tilde{\Lambda}-\Lambda\right)S_H\right] \\
&\geq \frac{1}{N}\mathrm{tr}\left[\left(\tilde{\Lambda}-\Lambda\right)'\left(\tilde{\Lambda}-\Lambda\right)\right]\pi_{\min}\left(\tilde{\Sigma}_e^{-1}\right)\pi_{\min}(S_H) \\
&\quad - \frac{1}{N}\mathrm{tr}\left[\left(\tilde{\Lambda}-\Lambda\right)'\left(\tilde{\Lambda}-\Lambda\right)\right]\pi_{\max}\left[\tilde{\Sigma}_e^{-1}\tilde{\Lambda}\left(\tilde{\Lambda}'\tilde{\Sigma}_e^{-1}\tilde{\Lambda}\right)^{-1}\tilde{\Lambda}'\tilde{\Sigma}_e^{-1}\right]\pi_{\max}(S_H) \\
&\geq \mathcal{O}_p\left(\frac{1}{N}\right)\max_{i\leq N}\left\|\tilde{\lambda}_i-\lambda_i\right\|^2 - \mathcal{O}_p\left(\frac{L_N}{N}\right)\max_{i\leq N}\left\|\tilde{\lambda}_i-\lambda_i\right\|^2 = \mathcal{O}_p\left(\frac{1-L_N}{N}\right)\max_{i\leq N}\left\|\tilde{\lambda}_i-\lambda_i\right\|^2
\end{aligned}$$

□

Furthermore, the consistency results for $\tilde{\Lambda}^f$ and $\tilde{\Lambda}^g$ are summarized in the following lemma.

Lemma 1.D.5.

$$\max_{i\leq N}\left\|\tilde{\lambda}_i-\lambda_i\right\| = \mathcal{O}_p\left(\sqrt{\frac{Nd_T}{L_N}+\max(\mu_1,\mu_2)}\right),$$

where $d_T = \frac{\log N^\beta}{N} + \frac{1}{N^\beta}\frac{\log N}{T}$.

Proof. By the definition of B_2, and equations (1.D.7) and Lemma 1.D.4. we obtain

$$\mathcal{O}_p\left(\frac{1-L_N}{N}\right)\max_{i\leq N}\left\|\tilde{\lambda}_i-\lambda_i\right\|^2 + \frac{1}{N}\mu_1\sum_{i=1}^{N}\sum_{k=1}^{r_1}\left(\left|\tilde{\lambda}_{ik}^f\right|-\left|\lambda_{ik}^f\right|\right) + \frac{1}{N}\mu_2\sum_{i=1}^{N}\sum_{l=1}^{r_2}\left(\left|\tilde{\lambda}_{il}^g\right|-\left|\lambda_{il}^g\right|\right) \leq d_T. \tag{1.D.8}$$

We start with analyzing $\tilde{\Lambda}^f$. The left hand side of (1.D.8) can be further lower bounded by

$$\begin{aligned}
\mathcal{O}_p\left(\frac{1-L_N}{N}\right)\max_{i\leq N}\left\|\tilde{\lambda}_i^f-\lambda_i^f\right\|^2 - \frac{1}{N}\mu_1\sum_{i=1}^{N}\sum_{k=1}^{r_1}\left(\left|\lambda_{ik}^f\right|-\left|\tilde{\lambda}_{ik}^f\right|\right) - \frac{1}{N}\mu_2\sum_{i=1}^{N}\sum_{l=1}^{r_2}\left|\lambda_{il}^g\right| \leq d_T \\
\mathcal{O}_p\left(\frac{1-L_N}{N}\right)\max_{i\leq N}\left\|\tilde{\lambda}_i^f-\lambda_i^f\right\|^2 - \frac{1}{N}\mu_1\sum_{i=1}^{N}\sum_{k=1}^{r_1}\left|\tilde{\lambda}_{ik}^f-\lambda_{ik}^f\right| \leq d_T + \mathcal{O}\left(\frac{L_N}{N}\right)\mu_2 r_2 \\
\mathcal{O}_p\left(\frac{1-L_N}{N}\right)\max_{i\leq N}\left\|\tilde{\lambda}_i^f-\lambda_i^f\right\|^2 - \mathcal{O}\left(\frac{L_N}{N}\right)\mu_1\sqrt{r_1}\sqrt{\max_{i\leq N}\left\|\tilde{\lambda}_i^f-\lambda_i^f\right\|^2} \leq d_T + \mathcal{O}\left(\frac{L_N}{N}\right)\mu_2 r_2
\end{aligned}$$

Solving for $\max_{i\leq N}\left\|\tilde{\lambda}_i^f - \lambda_i^f\right\|$ yields

$$\begin{aligned}\max_{i\leq N}\left\|\tilde{\lambda}_i^f - \lambda_i^f\right\| &\leq \mathcal{O}_p\left(\frac{L_N}{1-L_N}\right)\mu_1 + \sqrt{\mathcal{O}_p\left(\left(\frac{L_N}{1-L_N}\right)^2\right)\mu_1^2 + \mathcal{O}_p\left(\frac{Nd_T}{1-L_N} + \frac{L_N}{1-L_N}\mu_2\right)} \\ &\leq \mathcal{O}_p\left(\frac{L_N}{1-L_N}\mu_1 + \sqrt{\frac{Nd_T}{1-L_N} + \frac{L_N}{1-L_N}\mu_2}\right).\end{aligned}$$

Since, $\lim_{N\to\infty}\frac{L_N}{1-L_N} \to -1$, we obtain

$$\max_{i\leq N}\left\|\tilde{\lambda}_i^f - \lambda_i^f\right\| \leq \mathcal{O}_p\left(\mu_1 + \sqrt{\frac{Nd_T}{L_N} + \mu_2}\right)$$

Equivalently, by the same argument as above for $\tilde{\Lambda}^g$, we obtain

$$\max_{i\leq N}\left\|\tilde{\lambda}_i^g - \lambda_i^g\right\| = \mathcal{O}_p\left(\mu_2 + \sqrt{\frac{Nd_T}{L_N} + \mu_1}\right),$$

which completes the proof.

□

The consistency of the latent factor estimator is established in the following lemma.

Lemma 1.D.6.

$$\left\|\tilde{f}_t - f_t\right\| = \mathcal{O}_p\left(N^{-\beta/2} + \frac{\sqrt{L_N}}{N^\beta}\left(\mu_1 + \sqrt{\frac{Nd_T}{L_N} + \mu_2}\right)\right) = o_p(1).$$

Proof. The latent factor estimator in equation (1.2.7) yields

$$\tilde{f}_t - f_t = -\left(\tilde{\Lambda}^{f\prime}\tilde{\Phi}_e^{-1}\tilde{\Lambda}^f\right)^{-1}\tilde{\Lambda}^{f\prime}\tilde{\Phi}_e^{-1}\left(\tilde{\Lambda}^f - \Lambda^f\right)f_t + \left(\tilde{\Lambda}^{f\prime}\tilde{\Phi}_e^{-1}\tilde{\Lambda}^f\right)^{-1}\tilde{\Lambda}^{f\prime}\tilde{\Phi}_e^{-1}e_t. \quad (1.D.9)$$

As $L_N \asymp N^\beta$, for $1/2 \leq \beta \leq 1$, the first term on the right-hand side is upper bounded by

$$\begin{aligned}
&\mathcal{O}_p\left(N^{-\beta}\right)\sqrt{\sum_{i=1}^{N}\left\|\left(\tilde{\Lambda}^{f\prime}\tilde{\Phi}_e^{-1}\right)_i\left(\tilde{\lambda}_i^f-\lambda_i^f\right)\right\|^2}\,\|f_t\| \\
&\leq \mathcal{O}_p\left(N^{-\beta}\right)\sqrt{\mathcal{O}_p\left(\sum_{i=1}^{N}\left\|\tilde{\lambda}_i^f-\lambda_i^f\right\|^2\right)} \\
&\leq \mathcal{O}_p\left(N^{-\beta}\right)\sqrt{\mathcal{O}_p\left(L_N\max_{i\leq N}\left\|\tilde{\lambda}_i^f-\lambda_i^f\right\|^2\right)} \\
&= \mathcal{O}_p\left(\frac{\sqrt{L_N}}{N^\beta}\right)\mathcal{O}_p\left(\mu_1+\sqrt{\frac{Nd_T}{L_N}+\mu_2}\right). \qquad (1.D.10)
\end{aligned}$$

In the following, we bound the second term on the right-hand side of (1.D.9). For this we analyze the term $\tilde{\Lambda}^{f\prime}\tilde{\Phi}_e^{-1}e_t$.

$$\begin{aligned}
&\mathcal{O}_p\left(N^{-\beta}\right)\left\|\left(\tilde{\Lambda}^{f\prime}\tilde{\Phi}_e^{-1}-\Lambda^{f\prime}\Phi_e^{-1}\right)e_t\right\|_F \\
&\quad\leq \mathcal{O}_p\left(N^{-\beta}\right)\left\|\left(\tilde{\Lambda}^{f}-\Lambda^{f}\right)'\tilde{\Phi}_e^{-1}e_t\right\|_F+\mathcal{O}_p\left(N^{-\beta}\right)\left\|\Lambda^{f\prime}\left(\tilde{\Phi}_e^{-1}-\Phi_e^{-1}\right)e_t\right\|_F.
\end{aligned}$$

By Lemma 1.D.5., the first term is bounded by

$$\begin{aligned}
&\mathcal{O}_p\left(N^{-\beta}\right)\sqrt{\sum_{i=1}^{N}\left\|\left(\tilde{\lambda}_i^f-\lambda_i^f\right)\left(\tilde{\Phi}_e^{-1}e_t\right)_i\right\|^2} \\
&\quad\leq \mathcal{O}_p\left(N^{-\beta}\right)\sqrt{L_N\max_{i\leq N}\left\|\tilde{\lambda}_i^f-\lambda_i^f\right\|^2\mathcal{O}_p(1)} \\
&\quad= \mathcal{O}_p\left(\frac{\sqrt{L_N}}{N^\beta}\right)o_p(1)=o_p(1). \qquad (1.D.11)
\end{aligned}$$

The second term can be bounded using Lemma 1.D.3. according to

$$\begin{aligned}
\mathcal{O}_p\left(N^{-\beta}\right)\left\|\Lambda^{f\prime}\left(\tilde{\Phi}_e^{-1}-\Phi_e^{-1}\right)e_t\right\|_F &= \mathcal{O}_p\left(N^{-\beta}\right)\sqrt{\sum_{i=1}^{N}\left\|\left(\Lambda^{f\prime}\Phi_e^{-1}\right)_i\left(\phi_{ie}-\tilde{\phi}_{ie}\right)\left(\tilde{\Phi}_e^{-1}e_t\right)_i\right\|^2} \\
&= \mathcal{O}_p\left(\frac{\log N}{N^\beta}\left\|\tilde{\Phi}_e-\Phi_e\right\|_F\right)=o_p(1). \qquad (1.D.12)
\end{aligned}$$

Hence, equations (1.D.10), (1.D.11) and (1.D.12) yield

$$\begin{aligned}\left\|\tilde{f}_t - f_t\right\| &= \mathcal{O}_p\left(N^{-\beta}\right)\sum_{i=1}^{N}\left\|\left(\Lambda^{f'}\Phi_e^{-1}\right)_i e_{it}\right\| + \mathcal{O}_p\left(\frac{\sqrt{L_N}}{N^{\beta}}\left(\mu_1 + \sqrt{\frac{Nd_T}{L_N} + \mu_2}\right)\right)\\ &= \mathcal{O}_p\left(N^{-\beta/2} + \frac{\sqrt{L_N}}{N^{\beta}}\left(\mu_1 + \sqrt{\frac{Nd_T}{L_N} + \mu_2}\right)\right) = o_p(1).\end{aligned}$$

□

To establish the consistency of the second step estimator of the idiosyncratic error covariance matrix $\tilde{\Sigma}_e^{\tau}$, we first compute the convergence rate of idiosyncratic errors e_{it}.

Lemma 1.D.7.

$$\max_{i\leq N}\frac{1}{T}\sum_{t=1}^{T}|\tilde{e}_{it} - e_{it}|^2 = \mathcal{O}_p\left(\frac{Nd_T}{L_N} + \max(\mu_1,\mu_2)\right).$$

Proof. As $\tilde{e}_{it} - e_{it} = \left(\tilde{\lambda}_i - \lambda_i\right)\tilde{h}_t' + \lambda_i\left(\tilde{h}_t - h_t\right)'$, we obtain by Lemma 1.D.5. and Lemma 1.D.6.

$$\begin{aligned}\max_{i\leq N}\frac{1}{T}\sum_{t=1}^{T}|\tilde{e}_{it} - e_{it}|^2 &\leq 2\max_{i\leq N}\left\|\tilde{\lambda}_i - \lambda_i\right\|^2\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{h}_t\right\|^2 + 2\max_{i\leq N}\|\lambda_i\|^2\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{f}_t - f_t\right\|^2\\ &\leq \mathcal{O}_p\left(\max_{i\leq N}\left\|\tilde{\lambda}_i - \lambda_i\right\|^2\right) + \mathcal{O}_p\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{f}_t - f_t\right\|^2\right)\\ &= \mathcal{O}_p\left(\frac{Nd_T}{L_N} + \max(\mu_1,\mu_2)\right).\end{aligned}$$

□

By Lemma 1.D.7 we have shown that $\max_{i\leq N}\frac{1}{T}\sum_{t=1}^{T}|\tilde{e}_{it} - e_{it}|^2 = o_p(1)$. Hence, by a similar argument as in *Lemma S.1.9.* in Daniele et al. (2018), we have

$$\max_{i,j\leq N}\left|\tilde{\sigma}_{ij} - \sigma_{ij}\right| = \mathcal{O}_p\left(\sqrt{\frac{Nd_T}{L_N} + \max(\mu_1,\mu_2)}\right). \tag{1.D.13}$$

In what follows, we are going to determine the convergence rate of the idiosyncratic error covariance matrix estimator based on second step soft-thresholding estimator introduced in Section 1.2.3.

Lemma 1.D.8.

$$\left\|\tilde{\Sigma}_e^{\tau} - \Sigma_e\right\| = \mathcal{O}_p\left(S_N\sqrt{\frac{Nd_T}{L_N} + \max(\mu_1,\mu_2)}\right).$$

Proof. The result follows from equation (1.D.13) and Theorem A.1. of Fan et al. (2013). □

In the following, we derive the convergence rates for the autoregressive matrices Φ_i, for $i = 1, \dots p$, in the dynamic equation (1.2.2). We proceed in the same fashion as in *Proposition A.2.* in Bai et al. (2016). In order to improve the readability of the upcoming technicalities, we define $\bar{p} = p + 1$ and $\bar{T} = T - p - 1$.

Lemma 1.D.9. For $\frac{L_N}{N^{2\beta}} \frac{N \log N}{T} = o(1)$, we have

(a) $\frac{1}{\bar{T}} \sum_{t=\bar{p}}^{T} \tilde{h}_{t-q} \tilde{h}'_{t-s} - \frac{1}{\bar{T}} \sum_{t=\bar{p}}^{T} h_{t-q} h'_{t-s} = \mathcal{O}_p \left(\frac{\sqrt{L_N}}{N^{\beta}} \left(\mu_1 + \sqrt{\frac{N d_T}{L_N} + \mu_2} \right) \right)$, for $q, s = 0, \dots, p$.

(b) $\frac{1}{\bar{T}} \sum_{t=\bar{p}}^{T} \left(\tilde{h}_{t-q} - h_{t-q} \right) \tilde{h}'_{t-s} = \mathcal{O}_p \left(\frac{\sqrt{L_N}}{N^{\beta}} \left(\mu_1 + \sqrt{\frac{N d_T}{L_N} + \mu_2} \right) \right)$, for $q, s = 0, \dots, p$.

(c) $\frac{1}{\bar{T}} \sum_{t=\bar{p}}^{T} \eta_t \tilde{h}'_{t-q} - \frac{1}{\bar{T}} \sum_{t=\bar{p}}^{T} \eta_t h'_{t-q} = \mathcal{O}_p \left(\frac{\sqrt{L_N}}{N^{\beta}} \left(\mu_1 + \sqrt{\frac{N d_T}{L_N} + \mu_2} \right) \right)$, for $q = 0, \dots, p$.

Proof. We start with expression (a). As $\tilde{h}_t = \left[\tilde{f}'_t, g'_t \right]'$ and $h_t = \left[f'_t, g'_t \right]'$, the left hand side of (a) can be expressed as $\begin{bmatrix} W_{11} & W_{12} \\ W_{21} & 0 \end{bmatrix}$, where $W_{11} = \frac{1}{\bar{T}} \sum_{t=\bar{p}}^{T} \left(\tilde{f}_{t-q} - f_{t-q} \right) \left(\tilde{f}_{t-s} - f_{t-s} \right)' + \frac{1}{\bar{T}} \sum_{t=\bar{p}}^{T} \left(\tilde{f}_{t-q} - f_{t-q} \right) f'_{t-s} + \frac{1}{\bar{T}} \sum_{t=\bar{p}}^{T} f_{t-q} \left(\tilde{f}_{t-s} - f_{t-s} \right)'$, $W_{12} = \frac{1}{\bar{T}} \sum_{t=\bar{p}}^{T} \left(\tilde{f}_{t-q} - f_{t-q} \right) g'_{t-s}$ and $W_{21} = \frac{1}{\bar{T}} \sum_{t=\bar{p}}^{T} g_{t-q} \left(\tilde{f}_{t-s} - f_{t-s} \right)'$. Now, we analyse each of the three quantities separately. Based on Lemma 1.D.6, we can see that the first term on the right hand side of W_{11} is $\mathcal{O}_p \left(\frac{\mu_1^2 L_N + N d_T + L_N \mu_2)}{N^{2\beta}} \right)$.

Using equation (1.D.9), we obtain the following expression for the second term in W_{11}

$$\begin{aligned} \frac{1}{\bar{T}} \sum_{t=\bar{p}}^{T} \left(\tilde{f}_{t-q} - f_{t-q} \right) f'_{t-s} = & - \left(\tilde{\Lambda}^{f\prime} \tilde{\Phi}_e^{-1} \tilde{\Lambda}^{f} \right)^{-1} \tilde{\Lambda}^{f\prime} \tilde{\Phi}_e^{-1} \left(\tilde{\Lambda}^{f} - \Lambda^{f} \right) \frac{1}{\bar{T}} \sum_{t=\bar{p}}^{T} f_{t-q} f'_{t-s} \\ & + \left(\tilde{\Lambda}^{f\prime} \tilde{\Phi}_e^{-1} \tilde{\Lambda}^{f} \right)^{-1} \tilde{\Lambda}^{f\prime} \tilde{\Phi}_e^{-1} \frac{1}{\bar{T}} \sum_{t=\bar{p}}^{T} e_{t-q} f'_{t-s}. \end{aligned} \tag{1.D.14}$$

The first term on the right hand side of equation (1.D.14) is bounded by the expression $\mathcal{O}_p \left(\frac{\sqrt{L_N}}{N^{\beta}} \left(\mu_1 + \sqrt{\frac{N d_T}{L_N} + \mu_2} \right) \right)$, similarly as in Lemma 1.D.6, as $\frac{1}{\bar{T}} \sum_{t=\bar{p}}^{T} f_{t-q} f'_{t-s} =$

$\mathcal{O}_p(1)$. To bound the second term in equation (1.D.14), we first analyze the expression $\tilde{\Lambda}^{f\prime}\tilde{\Phi}_e^{-1}\frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T} e_{t-q}f'_{t-s}$.

$$\begin{aligned}
&\mathcal{O}_p\left(N^{-\beta}\right)\left\|\left(\tilde{\Lambda}^{f\prime}\tilde{\Phi}_e^{-1}-\Lambda^{f\prime}\Phi_e^{-1}\right)\frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T} e_{t-q}f'_{t-s}\right\|_F \\
&\leq \mathcal{O}_p\left(N^{-\beta}\right)\left[\left\|\left(\tilde{\Lambda}^{f}-\Lambda^{f}\right)'\tilde{\Phi}_e^{-1}\frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T} e_{t-q}f'_{t-s}\right\|_F+\left\|\Lambda^{f\prime}\left(\tilde{\Phi}_e^{-1}-\Phi_e^{-1}\right)\frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T} e_{t-q}f'_{t-s}\right\|_F\right].
\end{aligned} \tag{1.D.15}$$

Based on Lemma 1.D.1 and Lemma 1.D.5, the first term in (1.D.15) is upper bounded by

$$\mathcal{O}_p\left(\frac{\sqrt{L_N}}{N^{\beta}}\right)\sqrt{\max_{i\leq N}\left\|\tilde{\lambda}_i^f-\lambda_i^f\right\|^2}\,\mathcal{O}_p\left(\sqrt{\frac{N\log N}{T}}\right).$$

Hence, as $\frac{L_N}{N^{2\beta}}\frac{N\log N}{T}=o(1)$, the first term of equation (1.D.15) is $o_p(1)$. Using Lemma 1.D.3, the second term in (1.D.15) is upper bounded by

$$\begin{aligned}
&\mathcal{O}_p\left(N^{-\beta}\right)\left\|\tilde{\Phi}_e-\Phi_e\right\|_F\left\|\frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T} e_{t-q}f'_{t-s}\right\|_F \\
&\leq o_p(1)\sqrt{\frac{N\log N}{T}}=o_p(1).
\end{aligned}$$

Finally, the second term on the right hand side of equation (1.D.14) is bounded by

$$\begin{aligned}
&\mathcal{O}_p\left(N^{-\beta}\right)\sqrt{\max_{i\leq N}\left(\Lambda^{f\prime}\Phi_e^{-1}\right)_i}\left\|\frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T} e_{t-q}f'_{t-s}\right\|_F \\
&\leq \mathcal{O}_p\left(N^{-\beta}\right)\sqrt{\frac{N\log N}{T}}=o_p(1).
\end{aligned}$$

Thus, the second term in W_{11} is $\mathcal{O}_p\left(\frac{\sqrt{L_N}}{N^{\beta}}\left(\mu_1+\sqrt{\frac{Nd_T}{L_N}}+\mu_2\right)\right)$. It can be shown that the last term in W_{11} is of the same order. By summarizing these results, we have that $W_{11}=\mathcal{O}_p\left(\frac{\sqrt{L_N}}{N^{\beta}}\left(\mu_1+\sqrt{\frac{Nd_T}{L_N}}+\mu_2\right)\right)$. By similar arguments as for W_{11}, it can be shown that W_{12} and W_{21} are $\mathcal{O}_p\left(\frac{\sqrt{L_N}}{N^{\beta}}\left(\mu_1+\sqrt{\frac{Nd_T}{L_N}}+\mu_2\right)\right)$ as well. Hence, (a) follows by this.

Similar as in Bai et al. (2016), expression (b) can be written as

$$\begin{bmatrix}\frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T}\left(\tilde{f}_{t-q}-f_{t-q}\right)\tilde{f}'_{t-s} & \frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T}\left(\tilde{f}_{t-q}-f_{t-q}\right)g'_{t-s}\\ 0 & 0\end{bmatrix}.$$

Both terms above are of order $\mathcal{O}_p\left(\frac{\sqrt{L_N}}{N^\beta}\left(\mu_1+\sqrt{\frac{Nd_T}{L_N}+\mu_2}\right)\right)$, as shown in (a). (b) follows by this result.

The left hand side of part (c) can be expressed as

$$\begin{bmatrix} \frac{1}{T}\sum_{t=\bar{p}}^{T}\eta_t\left(\tilde{f}_{t-q}-f_{t-q}\right)' \\ 0 \end{bmatrix}.$$

By using equation (1.D.9), $\frac{1}{T}\sum_{t=\bar{p}}^{T}\eta_t\left(\tilde{f}_{t-q}-f_{t-q}\right)'$ can be written as

$$\begin{aligned}\frac{1}{T}\sum_{t=\bar{p}}^{T}\eta_t\left(\tilde{f}_{t-q}-f_{t-q}\right)' = &-\frac{1}{T}\sum_{t=\bar{p}}^{T}\eta_t f_{t-q}'\left(\tilde{\Lambda}^f-\Lambda^f\right)'\tilde{\Phi}_e^{-1}\tilde{\Lambda}^f\left(\tilde{\Lambda}^{f\prime}\tilde{\Phi}_e^{-1}\tilde{\Lambda}^f\right)^{-1}\\ &+\frac{1}{T}\sum_{t=\bar{p}}^{T}\eta_t e_{t-q}'\tilde{\Phi}_e^{-1}\tilde{\Lambda}^f\left(\tilde{\Lambda}^{f\prime}\tilde{\Phi}_e^{-1}\tilde{\Lambda}^f\right)^{-1},\end{aligned}$$

which is bounded by

$$\begin{aligned}&\left\|\frac{1}{T}\sum_{t=\bar{p}}^{T}\eta_t f_{t-q}'\left(\tilde{\Lambda}^f-\Lambda^f\right)'\tilde{\Phi}_e^{-1}\tilde{\Lambda}^f\left(\tilde{\Lambda}^{f\prime}\tilde{\Phi}_e^{-1}\tilde{\Lambda}^f\right)^{-1}\right\|_F\\ &\leq \mathcal{O}_p\left(N^{-\beta}\right)\sqrt{\sum_{i=1}^{N}\left\|\left(\tilde{\Lambda}_i^f-\Lambda_i^f\right)'\right\|}=\mathcal{O}_p\left(\frac{\sqrt{L_N}}{N^\beta}\left(\mu_1+\sqrt{\frac{Nd_T}{L_N}+\mu_2}\right)\right)=o_p(1)\end{aligned}$$

Hence, (c) follows from this result.

□

Lemma 1.D.10.

$$\tilde{\Phi}_i-\Phi_i=\left(\sum_{t=\bar{p}}^{T}\eta_t\psi_t'\right)\left(\sum_{t=\bar{p}}^{T}\psi_t\psi_t'\right)^{-1}(\iota_i\otimes I_r)+\mathcal{O}_p\left(\frac{\sqrt{L_N}}{N^\beta}\left(\mu_1+\sqrt{\frac{Nd_T}{L_N}+\mu_2}\right)\right),$$

where $\Phi=\left(\Phi_1,\ldots,\Phi_p\right)$ and $\psi_t=\left(h_{t-1}',\ldots,h_{t-p}'\right)'$.

Proof. We denote $\tilde{\Phi}$ the estimator of Φ, which is obtained by estimating the regression

$$\tilde{h}_t=\Phi_1\tilde{h}_{t-1}+\cdots+\Phi_p\tilde{h}_{t-p}+error,$$

which yields the estimator

$$\tilde{\Phi}=\left(\sum_{t=\bar{p}}^{T}\tilde{h}_t\tilde{\psi}_t'\right)\left(\sum_{t=\bar{p}}^{T}\tilde{\psi}_t\tilde{\psi}_t'\right)^{-1},$$

with $\psi_t = \left(h'_{t-1}, \ldots, h'_{t-p}\right)'$ and $h_t = \Phi\psi_t + \eta_t$.

By Lemma 1.D.9 and the same arguments as in the proof of *Proposition A.2.* in Bai et al. (2016), we have that

$$\tilde{\Phi} - \Phi = \left(\sum_{t=\bar{p}}^{T} \eta_t \psi'_t\right)\left(\sum_{t=\bar{p}}^{T} \psi_t \psi'_t\right)^{-1} + \mathcal{O}_p\left(\frac{\sqrt{L_N}}{N^{\beta}}\left(\mu_1 + \sqrt{\frac{N d_T}{L_N} + \mu_2}\right)\right).$$

The result follows by post-multiplying $\iota_i \otimes I_r$ on both sides.

□

Proof of Theorem 1.3.2 (Consistency of the rotated estimators based on the regularized FAVAR model).

In the following, we will focus on the consistency of the parameter estimates after rotation. As the rotation matrix A in Section 1.2.4 depends on the covariance matrix Ω of the VAR innovations η_t in equation (1.2.2), we first concentrate on the consistency of Ω. For this, we introduce the following two lemmas that are similarly established by Bai et al. (2016).

Lemma 1.D.11. For any two compatible matrices $\mathcal{B}$ and $\mathcal{C}$ and the corresponding estimates $\tilde{\mathcal{B}}$ and $\tilde{\mathcal{C}}$, we have

$$\tilde{\mathcal{B}}\tilde{\mathcal{C}}^{-1}\tilde{\mathcal{B}}' - \mathcal{B}\mathcal{C}^{-1}\mathcal{B}' = \left(\tilde{\mathcal{B}} - \mathcal{B}\right)\mathcal{C}^{-1}\mathcal{B}' + \mathcal{B}\mathcal{C}^{-1}\left(\tilde{\mathcal{B}} - \mathcal{B}\right)' - \mathcal{B}\mathcal{C}^{-1}\left(\tilde{\mathcal{C}} - \mathcal{C}\right)\mathcal{C}^{-1}\mathcal{B}' + \mathcal{R},$$

where

$$\begin{aligned}\mathcal{R} = &- \left(\tilde{\mathcal{B}} - \mathcal{B}\right)\tilde{\mathcal{C}}^{-1}\left(\tilde{\mathcal{C}} - \mathcal{C}\right)\mathcal{C}^{-1}\mathcal{B}' + \left(\tilde{\mathcal{B}} - \mathcal{B}\right)\tilde{\mathcal{C}}^{-1}\left(\tilde{\mathcal{B}} - \mathcal{B}\right)' \\ &+ \mathcal{B}\tilde{\mathcal{C}}^{-1}\left(\tilde{\mathcal{C}} - \mathcal{C}\right)\mathcal{C}^{-1}\left(\tilde{\mathcal{C}} - \mathcal{C}\right)\mathcal{C}^{-1}\mathcal{B}' - \mathcal{B}\tilde{\mathcal{C}}^{-1}\left(\tilde{\mathcal{C}} - \mathcal{C}\right)\mathcal{C}^{-1}\left(\tilde{\mathcal{B}} - \mathcal{B}\right)'.\end{aligned}$$

Proof. See *Lemma B.1.* in Bai et al. (2016).

□

Lemma 1.D.12.

$$\frac{1}{\bar{T}}\tilde{H}' M_{\tilde{\Phi}} \tilde{H} - \frac{1}{\bar{T}} H' M_{\Phi} H = \mathcal{O}_p\left(\frac{\sqrt{L_N}}{N^{\beta}}\left(\mu_1 + \sqrt{\frac{N d_T}{L_N} + \mu_2}\right)\right),$$

where

$$\frac{1}{\bar{T}}\tilde{H}' M_{\tilde{\Phi}} \tilde{H} = \frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T} \tilde{h}_t \tilde{h}'_t - \left(\frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T} \tilde{h}_t \tilde{\psi}'_t\right)\left(\frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T} \tilde{\psi}_t \tilde{\psi}'_t\right)^{-1}\left(\frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T} \tilde{\psi}_t \tilde{h}'_t\right),$$

and $\frac{1}{\bar{T}}H'M_{\Phi}H$ is defined similarly.

Proof. The proof is conducted as in Bai et al. (2016). Hereby, we consider the following three expressions that are bounded using Lemma 1.D.9 (a).

$$\frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T}\tilde{h}_t\tilde{h}_t' - \frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T}h_t h_t' = \mathcal{O}_p\left(\frac{\sqrt{L_N}}{N^{\beta}}\left(\mu_1 + \sqrt{\frac{Nd_T}{L_N} + \mu_2}\right)\right),$$
$$\frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T}\tilde{h}_t\tilde{\psi}_t' - \frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T}h_t\psi_t' = \mathcal{O}_p\left(\frac{\sqrt{L_N}}{N^{\beta}}\left(\mu_1 + \sqrt{\frac{Nd_T}{L_N} + \mu_2}\right)\right),$$
$$\frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T}\tilde{\psi}_t\tilde{\psi}_t' - \frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T}\psi_t\psi_t' = \mathcal{O}_p\left(\frac{\sqrt{L_N}}{N^{\beta}}\left(\mu_1 + \sqrt{\frac{Nd_T}{L_N} + \mu_2}\right)\right).$$

The result follows based on the above expressions and Lemma 1.D.11.

□

The covariance matrix estimator of the innovations η_t can be bounded according to the following lemma.

Lemma 1.D.13.

$$\left\|\tilde{\Omega} - \Omega\right\|_F = \mathcal{O}_p\left(\frac{\sqrt{L_N}}{N^{\beta}}\left(\mu_1 + \sqrt{\frac{Nd_T}{L_N} + \mu_2}\right)\right).$$

Proof. The estimator of $\tilde{\eta}_t$ is defined as

$$\tilde{\Omega} = \frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T}\tilde{\eta}_t\tilde{\eta}_t',$$

where $\tilde{\eta}_t$ are the residuals of the regression $\tilde{h}_t = \Phi_1\tilde{h}_{t-1} + \cdots + \Phi_p\tilde{h}_{t-p} + error$. Hence,

$$\tilde{\eta}_t = \tilde{h}_t - \left(\sum_{t=\bar{p}}^{T}\tilde{h}_t\tilde{\psi}_t'\right)\left(\sum_{t=\bar{p}}^{T}\tilde{\psi}_t\tilde{\psi}_t'\right)^{-1}\tilde{\psi}_t,$$

where $\tilde{\psi}_t = \left(\tilde{h}_{t-1}', \ldots, \tilde{h}_{t-p}'\right)'$. Based on the previous result we have

$$\tilde{\Omega} = \frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T}\tilde{h}_t\tilde{h}_t' - \left(\frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T}\tilde{h}_t\tilde{\psi}_t'\right)\left(\frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T}\tilde{\psi}_t\tilde{\psi}_t'\right)^{-1}\left(\frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T}\tilde{\psi}_t\tilde{h}_t'\right). \tag{1.D.16}$$

Hence, given (1.D.16) we obtain

$$\begin{aligned}\tilde{\Omega} - \Omega = &\frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T}\tilde{h}_t\tilde{h}_t' - \left(\frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T}\tilde{h}_t\tilde{\psi}_t'\right)\left(\frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T}\tilde{\psi}_t\tilde{\psi}_t'\right)^{-1}\left(\frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T}\tilde{\psi}_t\tilde{h}_t'\right)\\ &- \frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T}h_t h_t' + \left(\frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T}h_t\psi_t'\right)\left(\frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T}\psi_t\psi_t'\right)^{-1}\left(\frac{1}{\bar{T}}\sum_{t=\bar{p}}^{T}\psi_t h_t'\right).\end{aligned} \tag{1.D.17}$$

The expression (1.D.17) is bounded by Lemma 1.D.12. These results yield

$$\left\|\tilde{\Omega} - \Omega\right\|_F = \mathcal{O}_p\left(\frac{\sqrt{L_N}}{N^{\beta}}\left(\mu_1 + \sqrt{\frac{Nd_T}{L_N}} + \mu_2\right)\right).$$

□

The estimated rotation matrix $\tilde{A} = \begin{bmatrix} I_{r_1} & -\tilde{\Omega}_{fg}\tilde{\Omega}_{gg}^{-1} \\ 0 & I_{r_2} \end{bmatrix}$ can be bounded using the following lemma.

Lemma 1.D.14.

$$\left\|\tilde{A}_{12} - A_{12}\right\|_F = \mathcal{O}_p\left(\frac{\sqrt{L_N}}{N^{\beta}}\left(\mu_1 + \sqrt{\frac{Nd_T}{L_N}} + \mu_2\right)\right),$$

where $\tilde{A}_{12} = -\tilde{\Omega}_{fg}\tilde{\Omega}_{gg}^{-1}$ and $A_{12} = -\Omega_{fg}\Omega_{gg}^{-1}$.

Proof. Given the definitions of $\tilde{A}_{12}$ and A_{12}, we obtain

$$\begin{aligned}\tilde{A}_{12} - A_{12} &= \Omega_{fg}\Omega_{gg}^{-1} - \tilde{\Omega}_{fg}\tilde{\Omega}_{gg}^{-1}\\ &= -\left(\tilde{\Omega}_{fg} - \Omega_{fg}\right)\left(\tilde{\Omega}_{gg}^{-1} - \Omega_{gg}^{-1}\right) - \Omega_{fg}\left(\tilde{\Omega}_{gg}^{-1} - \Omega_{gg}^{-1}\right) - \left(\tilde{\Omega}_{fg} - \Omega_{fg}\right)\Omega_{gg}^{-1}.\end{aligned} \tag{1.D.18}$$

Based on Lemma 1.D.13, we can upper bound the expression (1.D.18) by

$$\begin{aligned}\left\|\tilde{A}_{12} - A_{12}\right\|_F &\leq -\left\|\tilde{\Omega}_{fg} - \Omega_{fg}\right\|_F\left\|\tilde{\Omega}_{gg}^{-1} - \Omega_{gg}^{-1}\right\|_F - \left\|\Omega_{fg}\right\|\left\|\tilde{\Omega}_{gg}^{-1} - \Omega_{gg}^{-1}\right\|_F - \left\|\tilde{\Omega}_{fg} - \Omega_{fg}\right\|_F\left\|\Omega_{gg}^{-1}\right\|\\ &= \mathcal{O}_p\left(\frac{\sqrt{L_N}}{N^{\beta}}\left(\mu_1 + \sqrt{\frac{Nd_T}{L_N}} + \mu_2\right)\right).\end{aligned}$$

□

Lemma 1.D.15.

$$\max_{i\leq N}\left\|\hat{\lambda}_i^g-\lambda_i^{*g}\right\|=\mathcal{O}_p\left(\sqrt{\mu_1}+\sqrt{\mu_2}+\sqrt{\frac{Nd_T}{L_N}}\right).$$

Proof. The rotated factor loadings of the observed factors are defined as $\hat{\lambda}_i^g=\tilde{\lambda}_i^g-\tilde{\lambda}_i^f\tilde{A}_{12}$ and $\lambda_i^{*g}=\lambda_i^g-\lambda_i^fA_{12}$. Hence, their difference can be expressed as

$$\hat{\lambda}_i^g-\lambda_i^{*g}=\tilde{\lambda}_i^g-\lambda_i^g-\left(\tilde{\lambda}_i^f-\lambda_i^f\right)A_{12}-\lambda_i^f\left(\tilde{A}_{12}-A_{12}\right)-\left(\tilde{\lambda}_i^f-\lambda_i^f\right)\left(\tilde{A}_{12}-A_{12}\right). \tag{1.D.19}$$

The above expression can be bounded using Lemma 1.D.5 and Lemma 1.D.14 according to

$$\begin{aligned}\max_{i\leq N}\left\|\hat{\lambda}_i^g-\lambda_i^{*g}\right\| &\leq \max_{i\leq N}\left\|\tilde{\lambda}_i^g-\lambda_i^g\right\|-\max_{i\leq N}\left\|\tilde{\lambda}_i^f-\lambda_i^f\right\|\|A_{12}\|_F\\ &\quad-\max_{i\leq N}\left\|\lambda_i^f\right\|\left\|\tilde{A}_{12}-A_{12}\right\|_F-\max_{i\leq N}\left\|\tilde{\lambda}_i^f-\lambda_i^f\right\|\left\|\tilde{A}_{12}-A_{12}\right\|_F\\ &=\mathcal{O}_p\left(\mu_1+\mu_2+\sqrt{\frac{Nd_T}{L_N}+\mu_1}+\sqrt{\frac{Nd_T}{L_N}+\mu_2}\right)\\ &\leq\mathcal{O}_p\left(\sqrt{\mu_1}+\sqrt{\mu_2}+\sqrt{\frac{Nd_T}{L_N}}\right).\end{aligned}$$

□

Lemma 1.D.16.

$$\left\|\hat{f}_t-f_t^*\right\|=\mathcal{O}_p\left(N^{-\beta/2}+\frac{\sqrt{L_N}}{N^\beta}\left(\mu_1+\sqrt{\frac{Nd_T}{L_N}+\mu_2}\right)\right).$$

Proof. The rotated unobserved factors are defined as $\hat{f}_t=\tilde{f}_t+\tilde{A}_{12}g_t$ and $f_t^*=f_t+A_{12}g_t$. Thus, the difference of both terms is given by

$$\hat{f}_t-f_t^*=\tilde{f}_t-f_t+\left(\tilde{A}_{12}-A_{12}\right)g_t. \tag{1.D.20}$$

Using the euclidean norm, we determine the upper bound of (1.D.20), using Lemma 1.D.6 and Lemma 1.D.14, by

$$\begin{aligned}\left\|\hat{f}_t - f_t^*\right\| &\leq \left\|\tilde{f}_t - f_t\right\| + \left\|\tilde{A}_{12} - A_{12}\right\|_F g_t. \\ &= \mathcal{O}_p\left(N^{-\beta/2}\right) + \mathcal{O}_p\left(\frac{\sqrt{L_N}}{N^{\beta}}\left(\mu_1 + \sqrt{\frac{Nd_T}{L_N} + \mu_2}\right)\right) = o_p(1).\end{aligned}$$

□

Finally, for the rotated coefficient matrices of the dynamic equation in (1.2.2) we establish the following lemma.

Lemma 1.D.17.

$$\hat{\Phi}_i - \Phi_i^* = \left(\sum_{t=\vec{p}}^{T} \eta_t \psi_t'\right)\left(\sum_{t=\vec{p}}^{T} \psi_t \psi_t'\right)^{-1}(\iota_i \otimes I_r) + \mathcal{O}_p\left(\frac{\sqrt{L_N}}{N^{\beta}}\left(\mu_1 + \sqrt{\frac{Nd_T}{L_N} + \mu_2}\right)\right),$$

where ι_i is the i-th column of the $r \times r$ identity matrix.

Proof. Given the definitions $\hat{\Phi}_i = \tilde{A}\tilde{\Phi}_i\tilde{A}^{-1}$ and $\Phi_i^* = A\Phi_i A^{-1}$, we obtain

$$\begin{aligned}\hat{\Phi} - \Phi^* = &\left(\tilde{A} - A\right)\left(\tilde{\Phi} - \Phi\right)\left(\tilde{A}^{-1} - A^{-1}\right) + A\left(\tilde{\Phi} - \Phi\right)\left(\tilde{A}^{-1} - A^{-1}\right) \\ &+ \left(\tilde{A} - A\right)\left(\tilde{\Phi} - \Phi\right)A^{-1} + A\left(\tilde{\Phi} - \Phi\right)A^{-1} + \left(\tilde{A} - A\right)\Phi A^{-1} \\ &+ \left(\tilde{A} - A\right)\Phi\left(\tilde{A}^{-1} - A^{-1}\right) + A\Phi\left(\tilde{A}^{-1} - A^{-1}\right)\end{aligned} \tag{1.D.21}$$

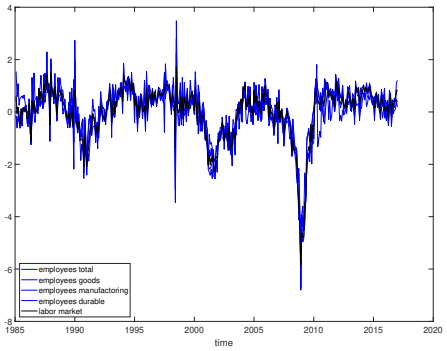
Based on Lemma 1.D.10 and Lemma 1.D.14, we get the expression

$$\hat{\Phi} - \Phi^* = \left(\sum_{t=\vec{p}}^{T} \eta_t \psi_t'\right)\left(\sum_{t=\vec{p}}^{T} \psi_t \psi_t'\right)^{-1} + \mathcal{O}_p\left(\frac{\sqrt{L_N}}{N^{\beta}}\left(\mu_1 + \sqrt{\frac{Nd_T}{L_N} + \mu_2}\right)\right).$$

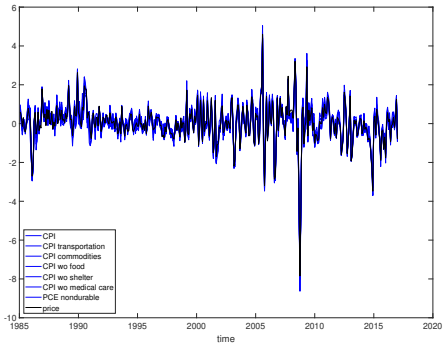
The result follows by post-multiplying $\iota_i \otimes I_r$ on both sides.

□

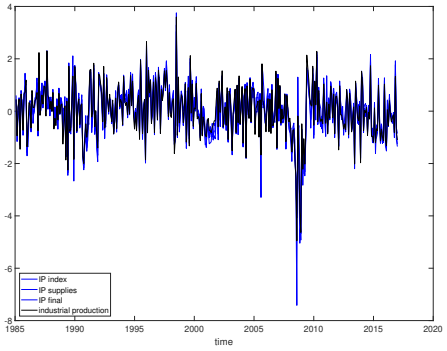
Appendix 1.E Additional figures and tables



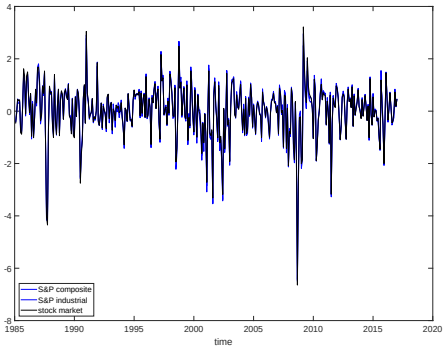
(a) Labor market



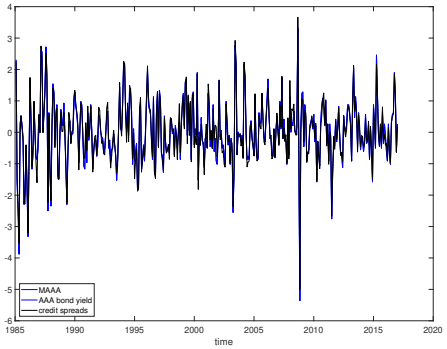
(b) Price



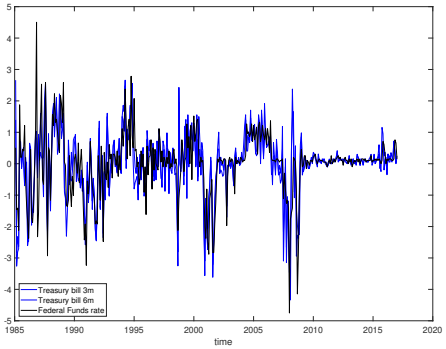
(c) Industrial production



(d) Stock market

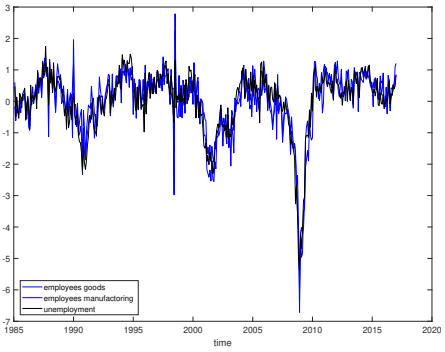


(e) Credit spread

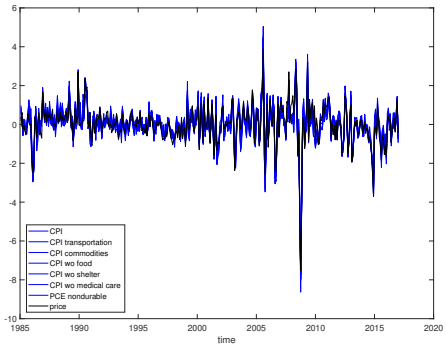


(f) Federal funds rate

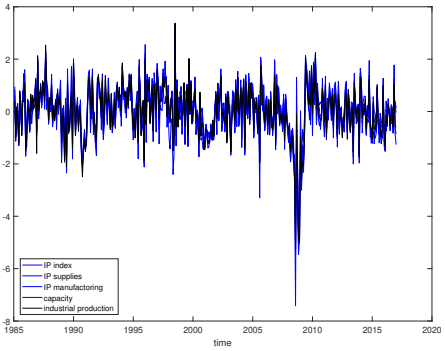
Figure 1.E.1: Factor plots for the regularized FAVAR



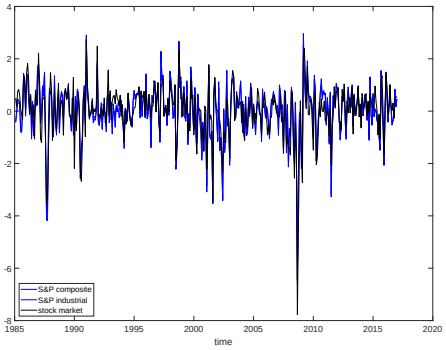
(a) Labor market



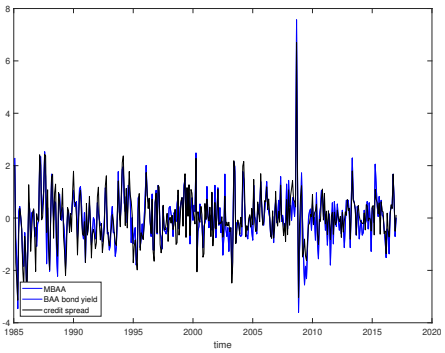
(b) Price



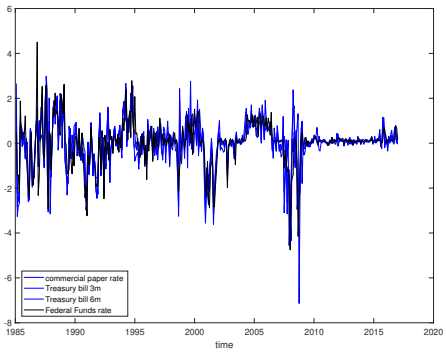
(c) Industrial production



(d) Stock market



(e) Credit spread



(f) Federal funds rate

Figure 1.E.2: Factor plots for the named factor FAVAR

Table 1.E.1: Data and transformations

	Description of the variables	Transformation code	Latent factor	FFR
1	Real personal income	5		0
2	Real personal income ex transfer receipts	5		0
3	Real personal consumption expenditures	5		0
4	Real manufacturing and trade industries sales	5	5	-0.0025
5	Retail and food services sales	5		0
6	IP index	5	1,5	-0.0444
7	IP final products and nonindustrial supplies	5	1,5	-0.0494
8	IP final products	5	5	-0.0473
9	IP consumer goods	5	5	-0.0362
10	IP durable consumer goods	5	5	-0.0267
11	IP nondurable consumer goods	5	5	-0.0103
12	IP business equipment	5	1,5	-0.0283
13	IP materials	5	1,5	-0.0230
14	IP durable materials	5	1,5	-0.0300
15	IP nondurable materials	5	5	-0.0026
16	IP manufacturing	5	1,5	-0.0445
17	IP residuential utilities	5		0
18	IP fuels	5		0
19	Capacity utilization	2	1,5	-0.0402
20	US ISM Manufacturers survey: production index	1	1	-0.0110
21	US ISM Manufacturers survey: employment index	1	1	-0.0119
22	US ISM Purchasing Managers Index	1	1	-0.0138
23	US ISM Manufacturers survey: supplier delivery index	1	1	-0.0038
24	US ISM Manufacturers survey: new orders index	1	1	-0.0096
25	US ISM Manufacturers survey: inventories index	1	1	-0.0075
26	US ISM Manufacturers survey: prices paid index	1	1,2	-0.0021
27	Help wanted index	2		0
28	Ratio of help wanted/number unemployed	2	1	-0.0005
29	Civilian labor force	5		0
30	Civilian employment	5	1	-0.0041
31	Civilian unemployment rate	2	1	0.0060
32	Average duration of unemployment	2		0
33	Civilians unemployed - less than 5 weeks	5		0
34	Civilians unemployed for 5-14 weeks	5		0
35	Civilians unemployed - 15 weeks and over	5	1	0.0050
36	Civilians unemployed for 15-26 weeks	5		0
37	Civilians unemployed for 27 weeks and over	5	1	0.0030
38	Initial claims	5		0
39	All employees: total nonfarm	5	1	-0.0194
40	All employees: goods-producing industries	5	1	-0.0222
41	All employees: mining	5		0
42	All employees: construction	5	1	-0.0120
43	All employees: manufacturing	5	1,5	-0.0217
44	All employees: durable goods	5	1,5	-0.0219
45	All employees: nondurable goods	5	1	-0.0130
46	All employees: service-providing industries	5	1	-0.0139

Continued on next page

Table 1.E.1 – *cont.*

	Description of the variables	Transformation code	Latent factor	FFR
47	All employees: trade, transportation and utilities	5	1	-0.0152
48	All employees: wholesale trade	5	1	-0.0159
49	All employees: retail trade	5	1	-0.0098
50	All employees: financial activities	5	1	-0.0044
51	All employees: government	5		0
52	Average hourly earnings: goods-producing	1	1	-0.0110
53	Average weekly overtime hours: manufacturing	2		0
54	Average weekly hours: manufacturing	1	1	-0.0121
55	Housing starts: total new privately owed	5		0
56	Housing starts: NE	5		0
57	Housing starts: MW	5		0
58	Housing starts: S	5		0
59	Housing starts: W	5		0
60	New private housing permits	5		0
61	New private housing permits: NE	5		0
62	New private housing permits: MW	5		0
63	New private housing permits: S	5		0
64	New private housing permits: W	5		0
65	Moodys Seasoned Aaa Corporate Bond Yield, Percent	1	3	-0.0912
66	Moodys Seasoned Baa Corporate Bond Yield, Percent	2	3,4	-0.0722
67	30-Year Fixed Rate Mortgage Average in the United States, Percent	2	3	-0.0644
68	Excess Bond Premium	1	1	0.0119
69	Spread UK-US	2		0
70	Spread CAN-US	2		0
71	Spread SW-US	2		0
72	Spread JPN-US	2		-0.1383
73	New orders for durable goods	5		0
74	New orders for nondefense capital goods	5		0
75	Unfilled orders for durable goods	5	1	-0.0019
76	Business Inventories	5	1	-0.0100
77	Inventories to sales ratio	2		0
78	M1 money stock	5		0
79	M2 money stock	5		0
80	Real M2 money stock	5	2	0.0232
81	St. Louis adjusted monetary base	5	2	0.0023
82	Total reserves of depository institutions	5		0
83	Commercial and industrial loans	5	1	-0.0029
84	Real estate loans at all commercial banks	5		0
85	Total nonrevolving credit	5		0
86	Nonrevolving comsumer credit to personal income	2		0
87	S&P 500 composite	5	4	0.0092
88	S&P industrials	5	4	0.0091
89	S&P common stock dividend yields	2	4	-0.0083
90	S&P common stock price-earnings ratio	5	1,4	0.0066
91	VXO	1	1,4	0.0063

Continued on next page

Table 1.E.1 – *cont.*

	Description of the variables	Transformation code	Latent factor	FFR
92	Dow Jones Industrial	5	4	0.0034
93	Dow Jones Utilities	5		0
94	Nasdaq Composite	5	4	0.0015
95	Dow Jones	5	4	0.0015
96	Nasdaq Industrial	5	4	0.0013
97	3 month financial commercial paper rate	2	1	0.2361
98	3 month Treasury bill	2		0.4455
99	6 month Treasury bill	2	1,3	0.4731
100	1 year Treasury rate	2	3	0.4023
101	5 year Treasury rate	2	3	0.0871
102	10 year Treasury rate	2	3	-0.0207
103	Moody AAA corporate bond yield	2	3	-0.0912
104	Moody BAA corporate bond yield	2	3,4	-0.0722
105	Trade weighted US Dollar index major currencies	5		0
106	Switzerland US foreign exchange rate	5		0
107	Japan US foreign exchange rate	5		0
108	US UK foreign exchange rate	5		0
109	Canada US foreign exchange rate	5		0
110	PPI: finished goods	5	2	-0.0349
111	PPI: finished consumer goods	5	2	-0.0353
112	PPI: intermediate materials	5	2	-0.0332
113	PPI: crude materials	5	2	-0.0145
114	crude oil	5	2	-0.0113
115	PPI: metals	5		0
116	CPI: all items	5	2	-0.0505
117	CPI: apparel	5		0
118	CPI: transportation	5	2	-0.0480
119	CPI: commodities	5	2	-0.0508
120	CPI: durables	5		0
121	CPI: all items less food	5	2	-0.0501
122	CPI: all items less shelter	5	2	-0.0516
123	CPI: all items less medical care	5	2	-0.0509
124	Personal consumption expenditure: chain index	6	2	-0.0090
125	Personal consumption expenditure: durable goods	5		0
126	Personal consumption expenditure: nondurable goods	5	2	-0.0491

Note: This table shows the different observed variables on monthly frequency for a sampling period January 1985 until December 2016. The data is retrieved from the McCracken and Ng (2016) FRED-MD data base and datastream. The transformation codes are labeled as follows: 1 = no transformation, 2 = Δx_t, 3 = $\Delta^2 x_t$, 4 = $\log(x_t)$, 5 = $\Delta\log(x_t)$, 6 = $\Delta^2\log(x_t)$. The last two columns are associated with our regularized FAVAR model. The column "Latent factor" shows the latent factors the specific time series load on. The latent factors are denotes as 1 = labor market, 2 = price, 3 = industrial production, 4 = stock market, 5 = credit spread. The column "FFR" denotes the estimated loadings onto the observable factor.

Table 1.E.2: Naming variables scheme

Index	Description of the variables
6	Industrial production (IP) index
31	Civilian unemployment rate
87	S&P 500 composite
66	Moody's Seasoned Baa corporate bond yield
116	Consumer price index (CPI): all items

Note: This table lists the observed time series that are used as naming variables for the named factor identification case (*IRb*).

Appendix 1.F Robustness checks

We evaluate the robustness of our findings in various ways: First, we use the shadow rate provided by Wu and Xia (2016) instead of the Federal Funds rate when the policy rate is at the zero lower bound between December 16, 2008 and December 15, 2015.[19] Figure 1.G.1 shows both, the shadow rate and the Federal Funds rate, in one graph. The replacement of the policy rate does not affect the results qualitatively for our regularized FAVAR model. In Figure 1.G.2 we plot the impulse responses of the observed variables to a shock in the innovation of the shadow rate. The impulse response patterns are very similar to those in Figure 1.5.3. For the named factor scheme, the positive impact on the IP index is not significant in Figure 1.G.3 and the other three impulse responses are not substantially different. The impulse response patterns of the DFM model in Figure 1.G.4 are similar to those obtained in the previous section. Finally, for the VAR-F model in Figure 1.G.5, the impulse responses of CPI, the unemployment rate and the IP index remain economically plausible when we use the shadow rate instead of the FFR. More precisely, CPI reacts significantly negative for the first 6 months, whereas the IP index decreases significantly for approximately three years. Overall, using the shadow rate in periods where the FFR is at the zero lower bound improves the results of our competing models slightly but does not affect our RFAVAR model qualitatively.

Second, we augment the maximum number of latent factors used in the initial steps to $r_{\max} = 20$. Figure 1.G.6 contains the R^2 of univariate regressions of the principal components factors on the observed time series. Qualitatively, it is similar to the one obtained for $r^* = 10$ factors depicted in Figure 1.B.1. In both settings, there are five factors which have separate block-wise explanatory power. Once we employ shrinkage onto the factor loadings, we obtain five latent factors for both initial number of factors. Our method is robust against

[19]The shadow rate is retrieved from Jing Cynthia Wu's website https://sites.google.com/view/jingcynthiawu/shadow-rates.

alternative initial number of factors. This can be seen in Figures 1.B.2 and 1.G.7 which are based on different starting points for the shrinkage but yield similar latent factors and factor loadings estimates.

Third, we alter the lag order of the regularized FAVAR model to allow for different dynamics in the impulse responses. The resulting response functions of the factors and observed time series of a RFAVAR(2) model can be found in Figures 1.G.8 and 1.G.9. Moreover, the results for a RFAVAR(3) and RFAVAR(6) model are depicted in Figures 1.G.10, 1.G.11 and 1.G.12, 1.G.13, respectively. The shape of the impulse responses and the evolution over the horizons does not change qualitatively.

Forth, we shorten the time span to the period prior to the global financial crisis. Hence, the data set spans a period for January 1985 to December 2006. We set the initial number of factors to $r_{\max} = 10$. Based on the Bai and Ng (2002) IC_1 criterion, we initialize the factor model with $r^* = 9$ factors. The R^2 of univariate regressions of the PCA based factors onto the observed time series can be found in Figure 1.G.14. For the pre-crisis period, the block-wise dependence is pronounced for four or five factors. When we impose sparsity onto the factor loadings, we retrieve four latent factors: prices, credit spreads, real activity, and the stock market. Figure 1.G.15 shows the R^2 of univariate regressions of the regularized factors on the observed time series and Figure 1.G.16 shows the time series of the latent factors. A block-structure is still present, however, the real activity sector is represented by the industrial production sector. The labor market sector is no longer a separate latent factor in the pre-crisis period because the two real activity proxies covary heavily. For the entire sample, the labor market factor reacts more sluggish in comparison to industrial production which leads to two separate real activity factors (compare Figure 1.E.1). Our regularized factor model yields economically sensible latent factors and factor loadings estimates in the pre-crisis period. Figure 1.G.17 contains the contemporaneous effects of a 100bp structural shock to the FFR onto the observed time series. The direction of the impact is economically plausible for all cases. Lastly, Figure 1.G.18 shows impulse responses of four observed time series: the consumer price index and the industrial production index react negatively in response to a tight monetary policy shock, the level of employment in the manufacturing sector plummets, whereas the three months Treasury bill increases.

Appendix 1.G Figures for the robustness checks

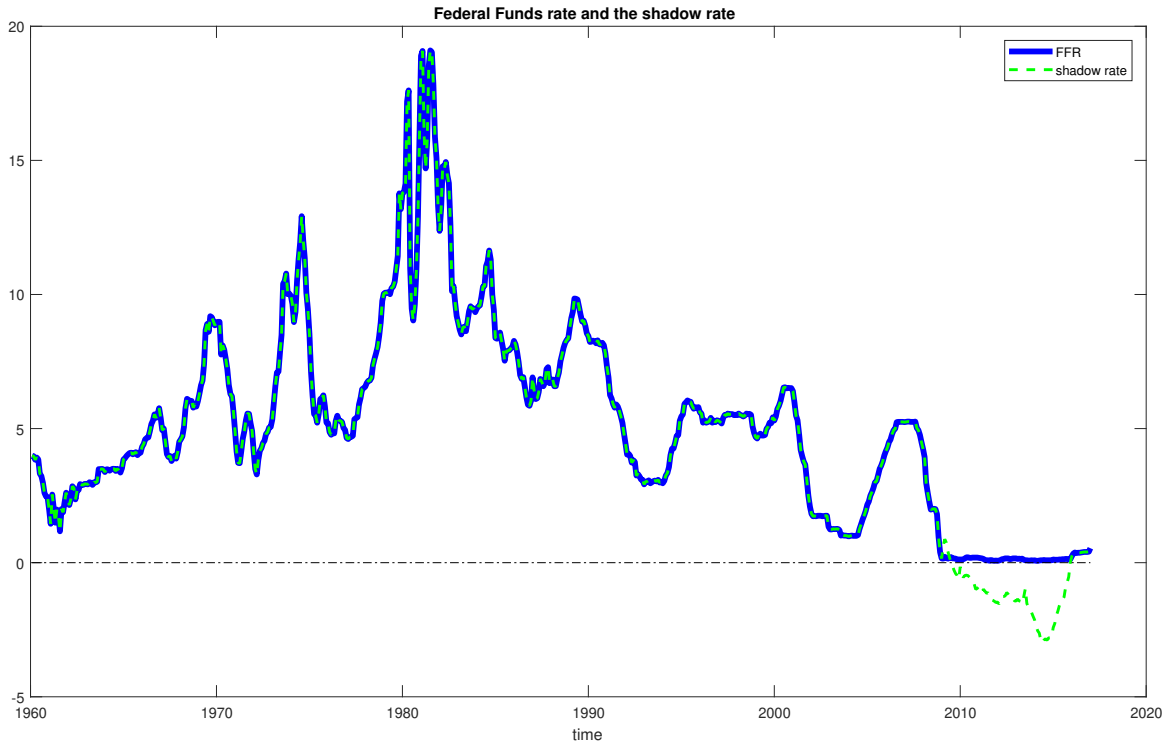


Figure 1.G.1: Shadow and Federal funds rate
This graph shows the shadow rate and the Federal Funds rate. The shadow rate takes effect whenever the FFR is at the zero lower bound, i.e. when it is below 25bp.

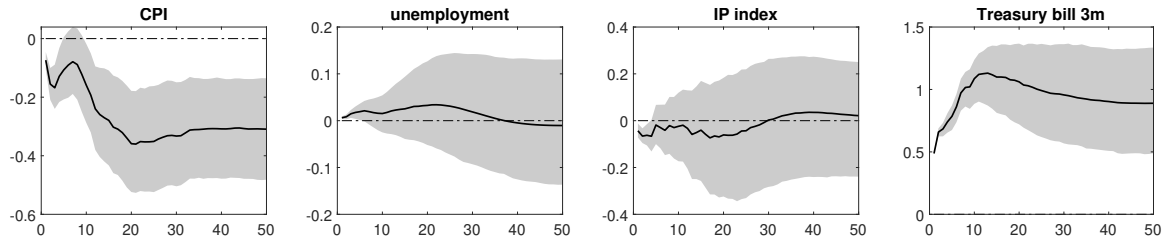


Figure 1.G.2: Accumulated impulse responses to a monetary policy shock on the observed variables x_t for the regularized FAVAR
The graph shows accumulated impulse responses to a 100bp shock in the innovation of the shadow rate. The dashed lines correspond to 68% bootstrap confidence intervals. The impulse responses are based on a RFAVAR(12) model.

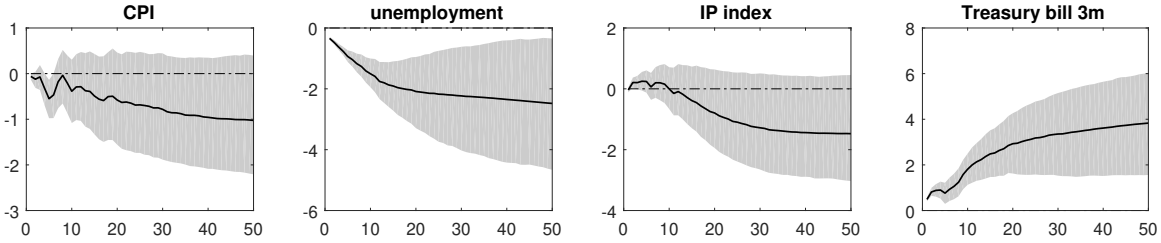


Figure 1.G.3: Accumulated impulse responses to a monetary policy shock on the observed variables x_t for the named factor FAVAR
The graph shows accumulated impulse responses to a 100bp shock in the innovation of the shadow rate. The dashed lines correspond to 68% bootstrap confidence intervals.

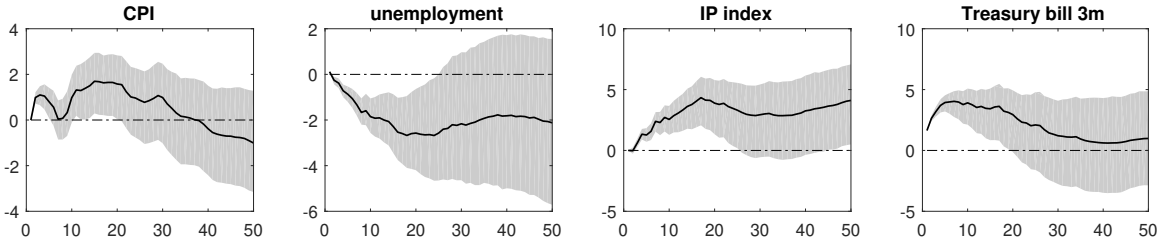


Figure 1.G.4: Accumulated impulse responses on the observed variables x_t for the dynamic factor model
The graph shows accumulated impulse responses to a 100bp shock in the innovation of the shadow rate. The dashed lines correspond to 68% bootstrap confidence intervals.

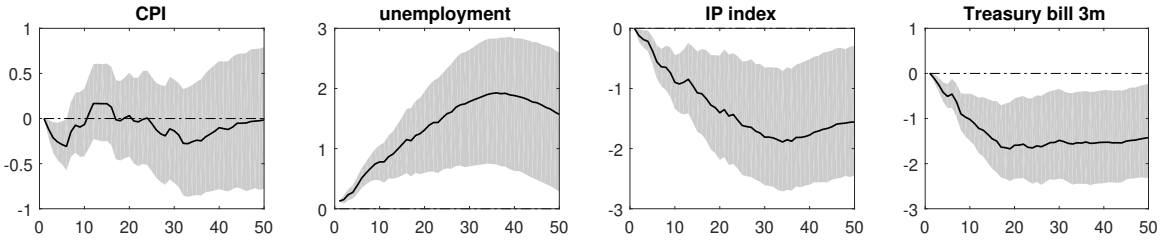


Figure 1.G.5: Accumulated impulse responses on the observed variables x_t for the VAR-F model
The graph shows impulse responses to a 100bp shock in the innovation of the shadow rate. The dashed lines correspond to 68% bootstrap confidence intervals.

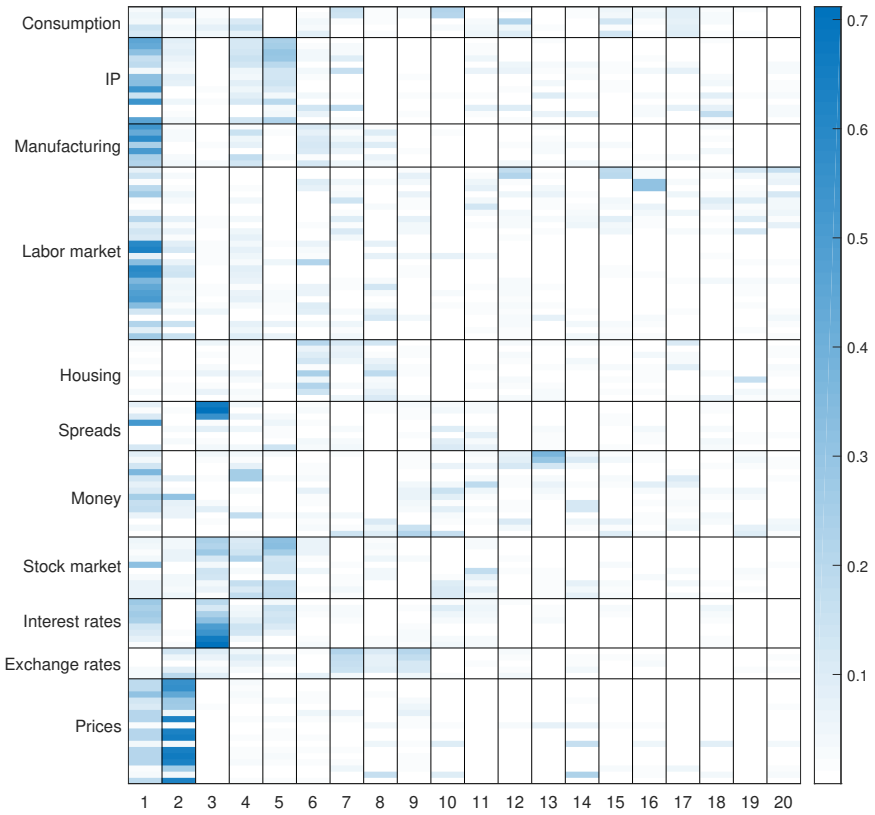


Figure 1.G.6: R^2 plot for a factor model based on principal components analysis
This graph shows R^2 for univariate regressions of the observed variables on each of the 20 factors.

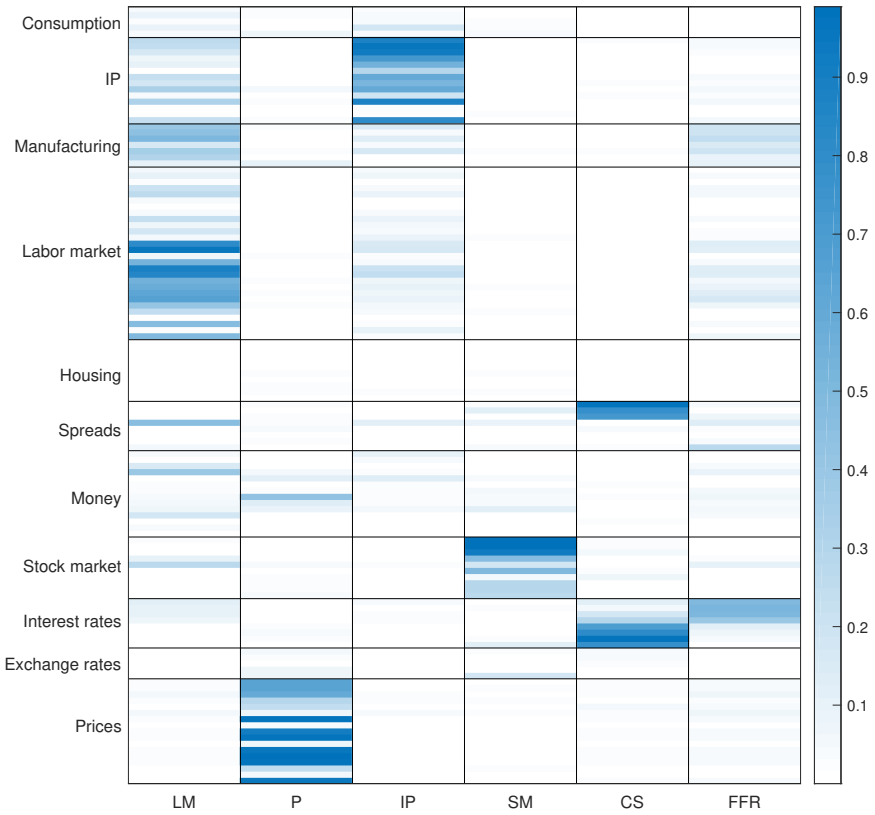


Figure 1.G.7: R^2 plot for the regularized FAVAR
This graph shows R^2 for univariate regressions of the observed variables on each of the factors. The factors are abbreviated as follows: 'LM' labor market, 'P' price, 'IP' industrial production, 'SM' stock market, 'CS' credit spread and 'FFR' Federal Funds rate.

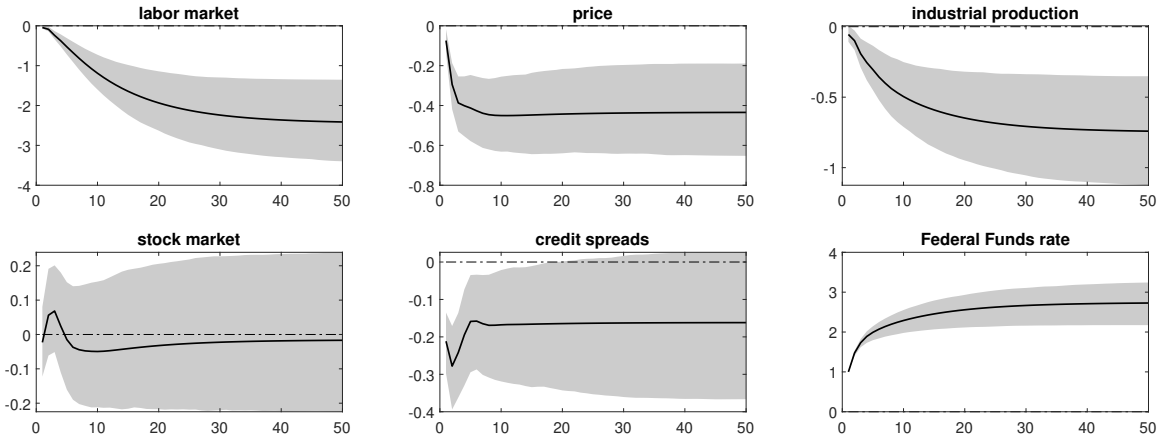


Figure 1.G.8: Accumulated impulse responses to a monetary policy shock on the factors for the regularized FAVAR
The graph shows accumulated impulse responses to a 100bp shock in the innovation of the FFR. The dashed lines correspond to 68% bootstrap confidence intervals. The impulse responses are based on a RFAVAR(2) model.

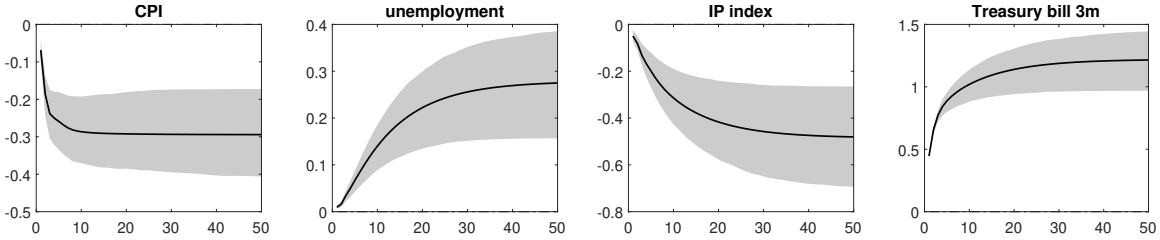


Figure 1.G.9: Accumulated impulse responses to a monetary policy shock on the observed variables x_t for the regularized FAVAR
The graph shows accumulated impulse responses to a 100bp shock in the innovation of the FFR. The dashed lines correspond to 68% bootstrap confidence intervals. The impulse responses are based on a RFAVAR(2) model.

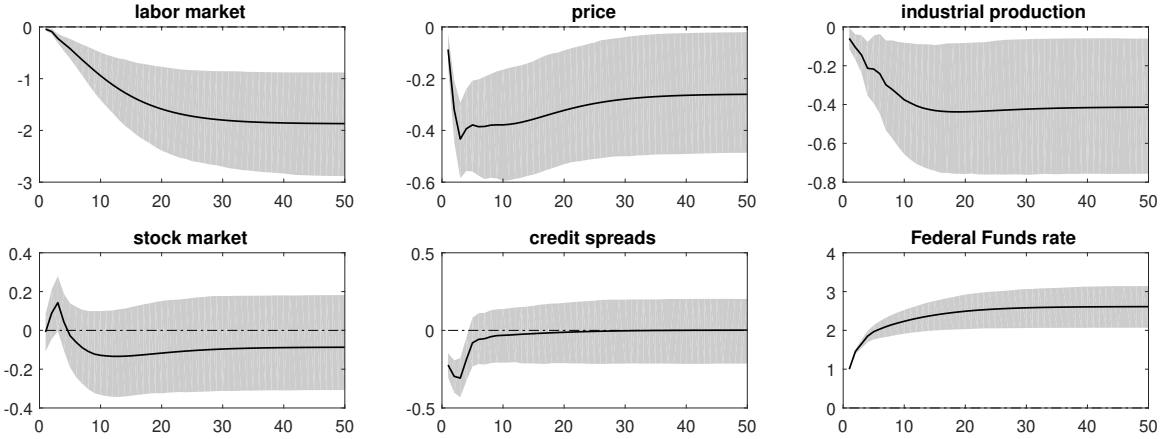


Figure 1.G.10: Accumulated impulse responses to a monetary policy shock on the factors for the regularized FAVAR
The graph shows accumulated impulse responses to a 100bp shock in the innovation of the FFR. The dashed lines correspond to 68% bootstrap confidence intervals. The impulse responses are based on a RFAVAR(3) model.

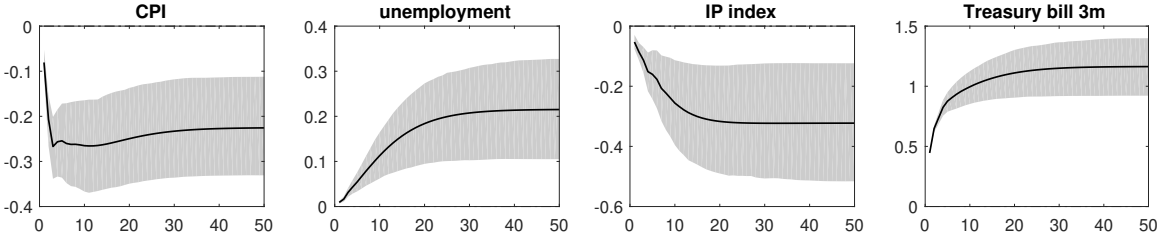


Figure 1.G.11: Accumulated impulse responses to a monetary policy shock on the observed variables x_t for the regularized FAVAR
The graph shows accumulated impulse responses to a 100bp shock in the innovation of the FFR. The dashed lines correspond to 68% bootstrap confidence intervals. The impulse responses are based on a RFAVAR(3) model.

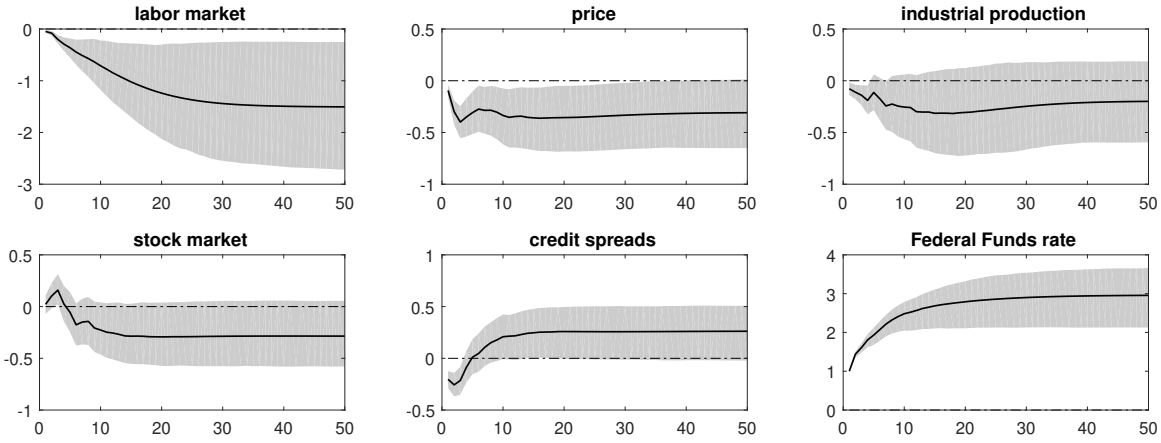


Figure 1.G.12: Accumulated impulse responses to a monetary policy shock on the factors for the regularized FAVAR
The graph shows accumulated impulse responses to a 100bp shock in the innovation of the FFR. The dashed lines correspond to 68% bootstrap confidence intervals. The impulse responses are based on a RFAVAR(6) model.

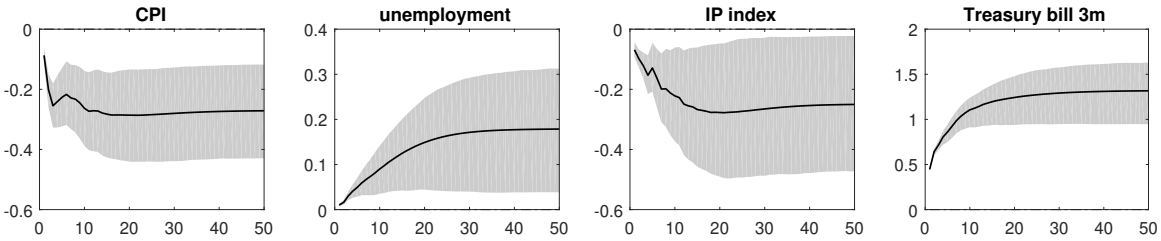


Figure 1.G.13: Accumulated impulse responses to a monetary policy shock on the observed variables x_t for the regularized FAVAR
The graph shows accumulated impulse responses to a 100bp shock in the innovation of the FFR. The dashed lines correspond to 68% bootstrap confidence intervals. The impulse responses are based on a RFAVAR(6) model.

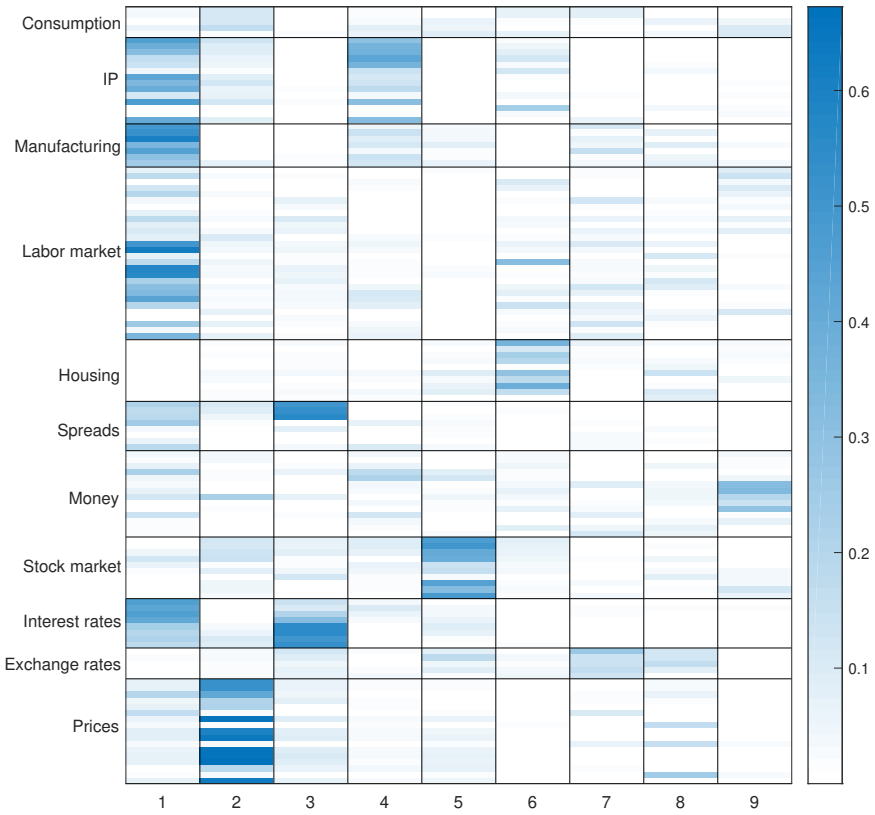


Figure 1.G.14: R^2 plot for the FAVAR based on principal components analysis
This graph shows R^2 for univariate regressions of the observed variables on each of the nine factors. The number of factors is chosen based on the Bai and Ng (2002) IC_1 criterion. The sample spans the pre-crisis period from January 1985 to December 2006.

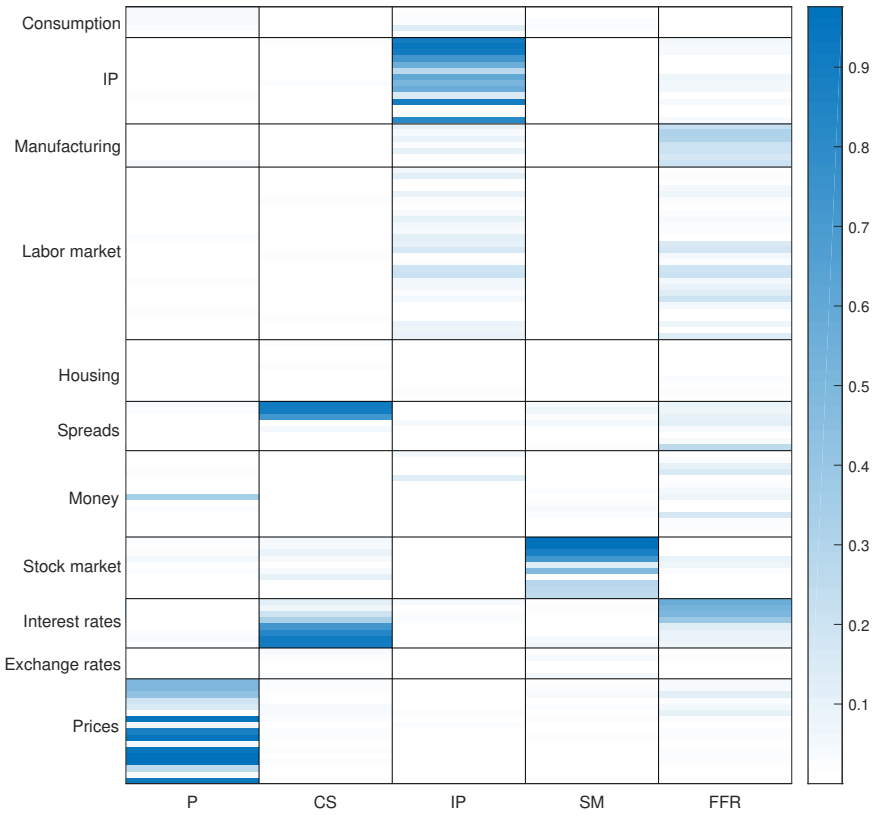
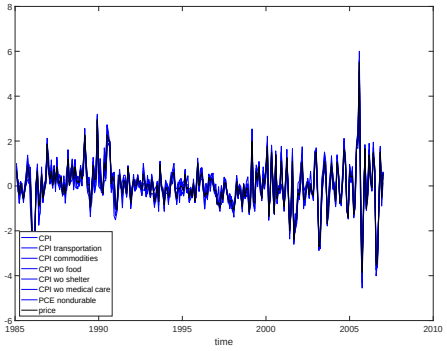
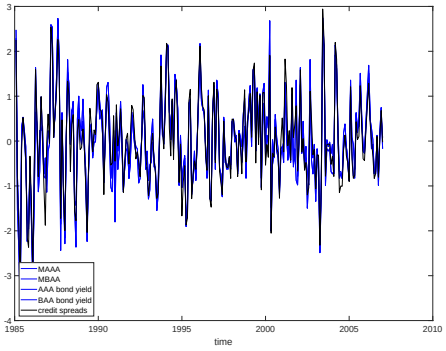


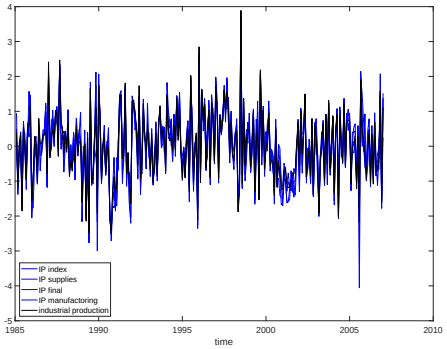
Figure 1.G.15: R^2 plot for the regularized FAVAR
This graph shows R^2 for univariate regressions of the observed variables on each of the factors. The factors are abbreviated as follows: 'P' price, 'CS' credit spreads, 'IP' industrial production, 'SM' stock market and 'FFR' Federal Funds rate. The sample spans the pre-crisis period from January 1985 to December 2006.



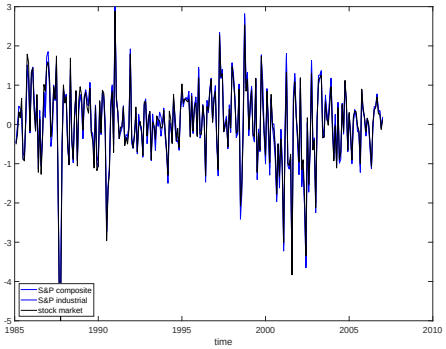
(a) Price



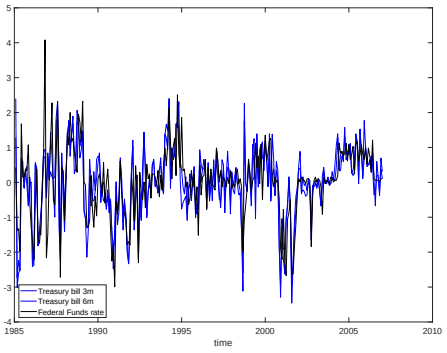
(b) Credit Spread



(c) Industrial Production



(d) Stock Market



(e) Federal Funds Rate

Figure 1.G.16: Factor plots for the regularized FAVAR

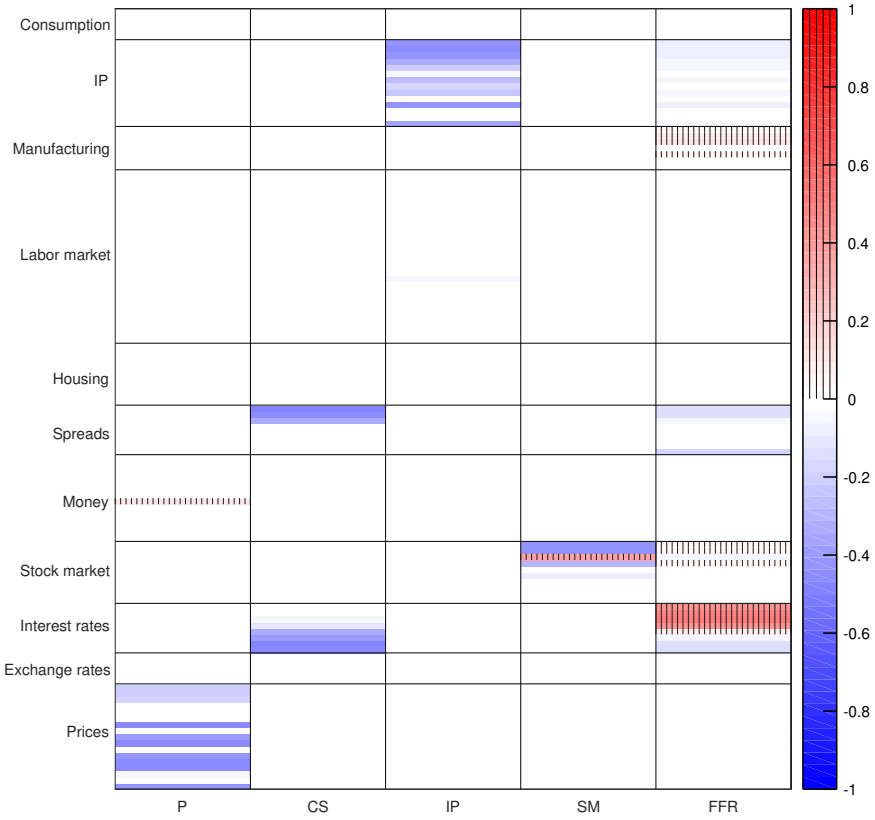


Figure 1.G.17: Impact matrix on the observed variables x_t for the regularized FAVAR
This graph shows the contemporaneous impact matrix of a 100bp shock in the factor innovations to the observed time series for the regularized FAVAR model. The factors are abbreviated as follows: 'P' price, 'CS' credit spreads, 'IP' industrial production, 'SM' stock market and 'FFR' Federal Funds rate. The sample spans the pre-crisis period from January 1985 to December 2006.

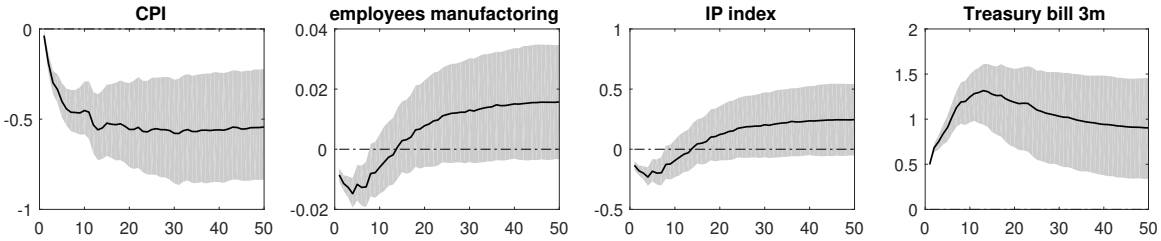


Figure 1.G.18: Accumulated impulse responses to a monetary policy shock on the observed variables x_t for the regularized FAVAR
The graph shows accumulated impulse responses to a 100bp shock in the innovation of the FFR. The dashed lines correspond to 68% bootstrap confidence intervals. The impulse responses are based on a RFAVAR(12) model. The sample spans the pre-crisis period from January 1985 to December 2006.

Chapter 2

Forecast Encompassing Tests for the Expected Shortfall

This project is joint work with Timo Dimitriadis. For helpful comments, we thank the editor (Michael McCracken), the associate editor and two referees, as well as Sander Barendse, Sebastian Bayer, Ralf Brüggemann, Joachim Grammig, Alastair Hall, Andrew Patton, James Taylor and the seminar participants at Universität Konstanz, Duke University, Universität Hohenheim, the 2019 QFFE conference in Marseille, the 2019 IAAE conference in Nicosia, the 2019 ESEM conference in Manchester, the 2019 Statistische Woche in Trier and the 2019 Financial Econometrics Workshop in Örebro. Financial support by the Klaus Tschira Foundation, the Universität Hohenheim, the Graduate School of Decision Sciences (GSDS) and a travel fund of the IAAE is gratefully acknowledged. This paper is accepted for publication in the *International Journal of Forecasting* (Dimitriadis and Schnaitmann, 2020) and is available here: https://doi.org/10.1016/j.ijforecast.2020.07.008 .

2.1. Introduction

Through the recent introduction of Expected Shortfall (ES) as the primary market risk measure for the international banking regulation in the Basel III Accords (Basel Committee, 2016, 2017), there is a great demand for reliable methods for evaluating and comparing the predictive ability of competing ES forecasts. The ES at probability level $\alpha \in (0, 1)$ is defined as the expectation of the returns smaller than the respective α-quantile (the Value at Risk, VaR), where α is usually chosen to be 2.5% as proposed by the Basel Accords. The ES is replacing the VaR in the banking regulation as it overcomes several shortcomings of the latter such as being not coherent and its inability to capture tail risks beyond the α-quantile (Artzner et al., 1999; Basel Committee, 2013; Danielsson et al., 2001). While the empirical properties favor the ES over the VaR as a risk measure, the ES lacks elicitability, which implies that no strictly consistent loss functions exist. The non-elicitability of the ES is overcome by considering the pair VaR and ES which are *jointly elicitable*, i.e. there exist joint loss functions for the VaR and the ES (Fissler and Ziegel, 2016). This discovery triggered a rapidly growing branch of literature in developing forecasting methods and forecast evaluation techniques for the ES, see Patton et al. (2019), Dimitriadis and Bayer (2019), Bayer and Dimitriadis (2020), Taylor (2019), Barendse (2020), Fissler et al. (2016) and Nolde and Ziegel (2017) among others.

A desirable tool for the comparison of ES forecasts are encompassing tests, which however build upon the existence of strictly consistent loss functions. Given two competing forecasts A and B, forecast encompassing tests the null hypothesis that forecast A *performs not worse* than any (linear) combination of these forecasts. This is carried out by testing whether the optimal combination weight of forecast B deviates significantly from zero.[1] This null hypothesis allows for the convenient interpretation that forecast B does not add any information to forecast A and thus, forecast A is superior to forecast B. The existence of appropriate loss functions is inevitable for encompassing tests for two reasons. First, the *superior performance* of competing forecasts is defined in the statistical sense by using strictly consistent loss functions. Second, loss and identification functions are crucial for M- or GMM-estimation of the optimal forecast combination weights through an appropriate regression framework for the risk measure under consideration.

In this paper, we introduce novel encompassing tests for the ES based on the joint loss functions for the VaR and ES developed in Fissler and Ziegel (2016). We introduce the following three test variants for the ES. First, we propose to jointly test forecast encompassing

[1]For the classical theory on forecast encompassing see Hendry and Richard (1982), Mizon and Richard (1986), Diebold (1989), Ericsson (1993), Harvey et al. (1998), Clark and McCracken (2001), Giacomini and Komunjer (2005), Newbold and Harvey (2007) and Clements and Harvey (2009) among others.

for the VaR and ES, henceforth denoted the *joint VaR and ES encompassing test*. We introduce a second test variant, denoted the *auxiliary ES encompassing test*, which estimates the optimal combination weights for the vector of the VaR and ES, however, only tests the parameters associated with the ES. While incorporating both, VaR and ES forecasts, this variant only tests encompassing of the ES forecasts. The third variant overcomes the tests' dependence on VaR forecasts and tests encompassing of competing ES forecasts stand-alone, which comes at the cost of a potential model misspecification. We henceforth call this test the *strict ES encompassing test*. This variant is particularly relevant due to the current set of rules established by the Basel Committee of Banking Supervision, which only imposes the financial institutions to report ES forecasts (Basel Committee, 2016, 2017). Only this test variant can be applied in situations where the person evaluating the forecasts merely has forecasts for the ES at hand. However, in situations where both, the VaR and ES forecasts (stemming from the same model or forecasting procedure) are available, application of the joint or auxiliary tests is generally recommended.

We implement the encompassing tests through M-estimation of the optimal combination weights (Dimitriadis and Bayer, 2019; Patton et al., 2019) and in an environment with asymptotically non-vanishing estimation uncertainty of the forecasting procedures (Giacomini and Komunjer, 2005; Giacomini and White, 2006). As the strict ES encompassing test is potentially subject to model misspecification, we derive the asymptotic distribution of the test statistics in a general setting which allows for misspecified models. This generalizes the asymptotic theory of Patton et al. (2019), Dimitriadis and Bayer (2019) and Bayer and Dimitriadis (2020) to potentially misspecified (and nonlinear) models. We base the Wald test statistics of the encompassing tests on a misspecification-robust covariance estimator. Our implementation further introduces a *link* or *combination function* which captures the different linear and nonlinear forecast combination methods in the existing encompassing testing literature, see Clements and Harvey (2009) and Clements and Harvey (2010) among others.

We analyze the finite sample behavior of our encompassing tests and the effect of the potential model misspecification in an extensive simulation study using models from various model classes associated with the ES. For this, we consider classical GARCH models, the GAS (generalized auto-regressive score) models with time-varying higher moments of Creal et al. (2013), the GAS models for the VaR and ES of Patton et al. (2019) and the ES-CAViaR models of Taylor (2019). Data stemming from the latter three model classes induces some model misspecification for the strict ES encompassing test, which allows us to evaluate the effect the misspecification has on our tests. We find that all tests exhibit approximately correct size and good power properties for all considered simulations. This also holds for

the strict ES encompassing test which demonstrates that this test is robust to the degree of model misspecification we usually encounter in financial applications.

Tests for forecast encompassing are commonly used to establish a theoretical basis for forecast combinations in cases when encompassing is rejected for both forecasts (Clements and Harvey, 2009; Giacomini and Komunjer, 2005; Newbold and Harvey, 2007). This implies that neither of the forecasts stand-alone performs as good as an optimal forecast combination, which indicates that a forecast combination incorporates more information than the individual forecasts. Giacomini and Komunjer (2005), Timmermann (2006), Halbleib and Pohlmeier (2012) and Taylor (2020) advocate general forecast combination methods for multiple reasons and particularly for risk measures with small probability levels, as it is customary for the VaR and the ES.

We apply our encompassing tests to ES forecasts from classical GARCH and GAS models, but also from the recently developed dynamic ES models of Taylor (2019) and Patton et al. (2019) for daily returns of the IBM stock, the S&P 500 and the DAX 30 indices. The test results imply that for the IBM stock, forecast combination methods outperform the stand-alone forecasting models in many instances. In comparison, this pattern seems to be less pronounced for the S&P 500 and the DAX 30 indices, which are already well diversified through their versatile composition. Thus, classical diversification gains (Timmermann, 2006) of forecast combination methods might be less pronounced for stock indices. The two ES based test variants exhibits very similar results, which further indicates that the strict ES test is robust against potential misspecifications in financial settings.

The classical idea of forecast encompassing goes back to Hendry and Richard (1982), Chong and Hendry (1986) and Mizon and Richard (1986) and is developed for mean forecasts under the squared loss function. Broad reviews on encompassing testing are provided e.g. by Newbold and Harvey (2007) and Clements and Harvey (2009). Harvey and Newbold (2000) extend the encompassing technique which classically focuses on two competing forecasts to encompassing of multiple forecasts. Giacomini and Komunjer (2005) develop (conditional) encompassing of quantile forecasts and focus on encompassing tests for *methods* instead of *models*. Clements and Harvey (2010) generalize encompassing tests to probabilistic forecasts by relying on strictly consistent scoring rules. Giacomini and Komunjer (2005) and Clements and Harvey (2010) investigate extensions of encompassing to more complicated functionals of the conditional distribution. Our work pursues this path by developing encompassing tests for the ES as a prominent example of higher-order elicitable functionals where only joint loss functions for vector-valued functionals are available. Our testing approach can be adapted to further higher-order elicitable functionals such as the pair mean, variance and the Range Value at Risk (Cont et al., 2010; Embrechts et al., 2018; Fissler and Ziegel, 2019).

The rest of the paper is organized as follows. In Section 2.2, we introduce encompassing tests for the ES and derive the asymptotic distribution of the associated test statistics under model misspecification. Section 2.3 presents an extensive simulation study analyzing the size and power properties of our tests. In Section 2.4, we apply the testing procedure to daily financial returns of the IBM stock and the S&P 500 and DAX 30 indices and Section 2.5 concludes. All proofs are deferred to Appendix 2.A, additional data generating processes for the simulation study are described in Appendix 2.B, technical details of the proofs can be found in Appendix 2.C and additional tables and figures are provided in Appendices 2.D and 2.E.

2.2. Theory

We consider a stochastic process $Z = \left\{Z_t : \Omega \to \mathbb{R}^{l+1}, l \in \mathbb{N}, t = 1, \ldots, T\right\}$, which is defined on some common and complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F} = \{\mathcal{F}_t, t = 1, \ldots, T\}$ and $\mathcal{F}_t = \sigma\{Z_s, s \leq t\}$. We partition the stochastic process as $Z_t = (Y_t, X_t)$, where $Y_t : \Omega \to \mathbb{R}$ is an absolutely continuous random variable of interest and $X_t : \Omega \to \mathbb{R}^l$ is a vector of explanatory variables. We denote the conditional distribution of Y_{t+1} given the information set $\mathcal{F}_t$ by F_t. Accordingly, $\mathbb{E}_t$, Var_t and h_t denote the expectation, variance and density corresponding to F_t. Following Giacomini and Komunjer (2005), we consider ($\mathcal{F}_t$-measurable) one-step ahead forecasts, henceforth denoted by $\hat{f}_t, \hat{q}_t$ and $\hat{e}_t$, which are generated by a function $f\left(\gamma_{t,m}, Z_t, Z_{t-1}, \ldots\right)$, which is fixed over time. For this, $\gamma_{t,m}$ denotes the (estimated) model parameters at time t or alternatively the semi- or non-parametric estimator used in the construction of the forecasts. This construction allows for both, fixed forecasting schemes, where the model parameters $\gamma_{t,m}$ are only estimated once, and rolling window forecasting schemes, where the parameters $\gamma_{t,m}$ are re-estimated in each step. We denote general competing forecasts by $\hat{\boldsymbol{f}}_t = (\hat{f}_{1,t}, \hat{f}_{2,t})$, specific VaR (quantile) forecasts by $\hat{\boldsymbol{q}}_t = (\hat{q}_{1,t}, \hat{q}_{2,t})$ and ES forecasts by $\hat{\boldsymbol{e}}_t = (\hat{e}_{1,t}, \hat{e}_{2,t})$.

In the context of evaluating point forecasts, an important property of risk measures (or more general statistical functionals) is *elicitability* (Gneiting, 2011). Elicitability means that there exist strictly consistent loss functions, i.e. loss functions $\rho(Y, f)$ depending on the random variable $Y \sim F$ and the issued forecast f, whose expectation $\mathbb{E}\left[\rho(Y, \cdot)\right]$ is uniquely minimized by the true risk measure $\Gamma(F)$. Using such a loss function, one can assess the quality of issued forecasts by comparing their average losses induced by the realizations of the predicted variable. Evaluating forecasts through strictly consistent loss functions has the desired impact that it incentivizes financial institutions to truthfully report their correct forecasts (Fissler et al., 2016; Gneiting, 2011). As a direct consequence, the literature on tests for forecast comparison and forecast rationality evolves around the associated loss

functions, see Mizon and Richard (1986), Diebold and Mariano (1995), Elliott et al. (2005), Giacomini and Komunjer (2005), Giacomini and White (2006), Patton and Timmermann (2007), Clements and Harvey (2010), Gneiting (2011) and Patton (2011) among many others.

Many important statistical functionals such as the variance, the ES, the minimum, the maximum and the mode are not elicitable, i.e. no strictly consistent loss functions exist (Fissler and Ziegel, 2016; Gneiting, 2011; Heinrich, 2014). This deficiency calls for generalized approaches in many academic disciplines. We built our test procedure for the ES on such an approach, which considers multiple functionals stacked as vectors and considers joint elicitability. Fissler and Ziegel (2016) show that the ES is jointly elicitable with the VaR by constructing strictly consistent joint loss functions for this pair, which we utilize in our encompassing approach.

In the following section, we formally introduce the concept of forecast encompassing in the classical case of one-dimensional, real-valued and elicitable functionals. Subsequently, we make use of the higher-order elicitability of the ES and generalize the encompassing approach to ES forecasts in Section 2.2.2.

2.2.1. The encompassing principle

Following e.g. Hendry and Richard (1982), Mizon and Richard (1986), Diebold (1989) and Giacomini and Komunjer (2005), we formally introduce the classical concept of linear forecast encompassing for one-dimensional, real-valued and elicitable functionals. We assume that two competing forecasters predict the variable of interest Y_{t+1} and issue one-step ahead point forecasts $\hat{\boldsymbol{f}}_t = \left(\hat{f}_{1,t}, \hat{f}_{2,t}\right)$ for a given functional $\Gamma(F_t)$.[2] In order to conduct the forecast evaluation in an out-of-sample fashion, we divide the sample size T in an in-sample part of size m and an out-of-sample part of size n such that $T = m + n$. The in-sample period is used to generate the forecasts $\hat{f}_{1,t}$ and $\hat{f}_{2,t}$ as described in the beginning of Section 2.2, while the out-of-sample period is used for the evaluation of the forecasts. This procedure poses little restrictions on how to generate the forecasts and allows for parametric, semiparametric or nonparametric techniques and for nested and non-nested forecasting procedures (Giacomini and Komunjer, 2005).

Let $\rho\left(Y_{t+1}, \hat{f}_t\right)$ be a strictly consistent loss function for $\Gamma(\cdot)$. Then, we say that forecast $\hat{f}_{1,t}$ encompasses $\hat{f}_{2,t}$ at time t, if

$$\mathbb{E}\left[\rho\left(Y_{t+1}, \hat{f}_{1,t}\right)\right] \leq \mathbb{E}\left[\rho\left(Y_{t+1}, \theta_1 \hat{f}_{1,t} + \theta_2 \hat{f}_{2,t}\right)\right], \tag{2.2.1}$$

[2] While we focus our approach on one-step ahead forecasts, extensions to multi-step ahead forecasts are straight-forward by employing a HAC-type estimator for the asymptotic covariance.

for all $(\theta_1, \theta_2) \in \Theta \subseteq \mathbb{R}^2$. Equation (2.2.1) implies that, in terms of the loss induced by ρ, the forecast $\hat{f}_{1,t}$ is at least as good as any (linear) combination of $\hat{f}_{1,t}$ and $\hat{f}_{2,t}$. Hence, forecast $\hat{f}_{2,t}$ does not add any information on Y_{t+1} which is not already incorporated in $\hat{f}_{1,t}$. We define (θ_1^*, θ_2^*) as the optimal combination parameters which minimize the expected loss,

$$(\theta_1^*, \theta_2^*) = \underset{(\theta_1,\theta_2)\in\Theta}{\arg\min}\, \mathbb{E}\left[\rho\big(Y_{t+1}, \theta_1 \hat{f}_{1,t} + \theta_2 \hat{f}_{2,t}\big)\right]. \tag{2.2.2}$$

By definition, it holds that $\mathbb{E}\left[\rho\big(Y_{t+1}, \theta_1 \hat{f}_{1,t} + \theta_2 \hat{f}_{2,t}\big)\right] \geq \mathbb{E}\left[\rho\big(Y_{t+1}, \theta_1^* \hat{f}_{1,t} + \theta_2^* \hat{f}_{2,t}\big)\right]$ for all $(\theta_1, \theta_2) \in \Theta$. In particular, this implies that

$$\mathbb{E}\left[\rho\big(Y_{t+1}, \hat{f}_{1,t}\big)\right] \geq \mathbb{E}\left[\rho\big(Y_{t+1}, \theta_1^* \hat{f}_{1,t} + \theta_2^* \hat{f}_{2,t}\big)\right]. \tag{2.2.3}$$

Combining (2.2.1) and (2.2.3) yields the following definition of forecast encompassing.

Definition 2.2.1 (Linear Forecast Encompassing for Elicitable Functionals). We say that forecast $\hat{f}_{1,t}$ encompasses $\hat{f}_{2,t}$ at time t with respect to the loss function ρ if and only if

$$\mathbb{E}\left[\rho\big(Y_{t+1}, \hat{f}_{1,t}\big)\right] = \mathbb{E}\left[\rho\big(Y_{t+1}, \theta_1^* \hat{f}_{1,t} + \theta_2^* \hat{f}_{2,t}\big)\right], \tag{2.2.4}$$

which is equivalent to $(\theta_1^*, \theta_2^*) = (1, 0)$.

Tests for forecast encompassing are carried out through the following steps. First, we regress the realizations Y_{t+1} onto the forecasts $\hat{f}_{1,t}$ and $\hat{f}_{2,t}$ using an appropriate regression technique for the functional under consideration in order to obtain the estimated combination (or encompassing) parameters $\hat{\theta}_n$ and their asymptotic distribution. Then, we test whether these parameters equal one and zero respectively.

As discussed e.g. in Clements and Harvey (2009) and Clements and Harvey (2010), there exist several different testing specifications available for the encompassing principle, which differ in terms of the admissible specifications of the linear (or nonlinear) forecast combination formula. We generalize and unify these approaches by introducing a general *link* or *combination function*,

$$g : \mathfrak{F} \times \Theta \to \mathbb{R}, \qquad (\hat{\boldsymbol{f}}_t, \theta) \mapsto g(\hat{\boldsymbol{f}}_t, \theta), \tag{2.2.5}$$

which maps the forecasts and the respective parameters onto a linear or nonlinear forecast combination and where $\mathfrak{F}$ denotes the random space of the issued forecasts. For this, the function g and the parameter space Θ have to be chosen such that there exists a $\theta_0 \in \Theta$, such that $g(\hat{\boldsymbol{f}}_t, \theta_0) = \hat{f}_{1,t}$ almost surely, which enables testing whether $\hat{f}_{1,t}$ alone captures the full

information provided by any forecast combination through testing the parametric restriction $\theta^* = \theta_0$.

Definition 2.2.2 (General Forecast Encompassing for Elicitable Functionals). We say that forecast $\hat{f}_{1,t}$ encompasses $\hat{f}_{2,t}$ at time t with respect to the loss function ρ and with respect to the link function g if and only if

$$\mathbb{E}\left[\rho\big(Y_{t+1}, \hat{f}_{1,t}\big)\right] = \mathbb{E}\left[\rho\big(Y_{t+1}, g(\hat{\boldsymbol{f}}_t, \theta^*)\big)\right], \tag{2.2.6}$$

which is equivalent to $\theta^* = \theta_0$.

This general definition unifies the following existing specifications of forecast encompassing, but also allows for more general linear and nonlinear specifications, see e.g. Ericsson (1993), Clements and Harvey (2009) and Clements and Harvey (2010).

Example 2.2.1. Prominent examples for linear and nonlinear forecast encompassing are the following link functions and associated null hypotheses,

(1) $g(\hat{\boldsymbol{f}}_t, \theta) = \theta_1 + \theta_2 \hat{f}_{1,t} + \theta_3 \hat{f}_{2,t}$ and $\mathbb{H}_0 : (\theta_2^*, \theta_3^*) = (1,0)$ or $\mathbb{H}_0 : (\theta_1^*, \theta_2^*, \theta_3^*) = (0,1,0)$,

(2) $g(\hat{\boldsymbol{f}}_t, \theta) = \theta_1 + \theta_2 \hat{f}_{1,t} + (1-\theta_2) \hat{f}_{2,t}$ and $\mathbb{H}_0 : \theta_2^* = 1$ or $\mathbb{H}_0 : (\theta_1^*, \theta_2^*) = (0,1)$,

(3) $g(\hat{\boldsymbol{f}}_t, \theta) = \theta_1 + \hat{f}_{1,t} + \theta_2 \hat{f}_{2,t}$ and $\mathbb{H}_0 : \theta_2^* = 0$ or $\mathbb{H}_0 : (\theta_1^*, \theta_2^*) = (0,0)$,

(4) $g(\hat{\boldsymbol{f}}_t, \theta) = \theta_1 \hat{f}_{1,t} + \theta_2 \hat{f}_{2,t}$ and $\mathbb{H}_0 : (\theta_1^*, \theta_2^*) = (1,0)$,

(5) $g(\hat{\boldsymbol{f}}_t, \theta) = \theta_1 \hat{f}_{1,t} + (1-\theta_1) \hat{f}_{2,t}$ and $\mathbb{H}_0 : \theta_1^* = 1$,

(6) $g(\hat{\boldsymbol{f}}_t, \theta) = \hat{f}_{1,t} + \theta_1 \hat{f}_{2,t}$ and $\mathbb{H}_0 : \theta_1^* = 0$,

(7) $g(\hat{\boldsymbol{f}}_t, \theta) = \theta_1 \pm \exp\left(\theta_2 \log(\pm \hat{f}_{1,t}) + \theta_3 \log(\pm \hat{f}_{2,t})\right)$ and $\mathbb{H}_0 : (\theta_2^*, \theta_3^*) = (1,0)$.

2.2.2. Forecast encompassing for the Expected Shortfall

In this section, we consider encompassing tests for the ES. For absolutely continuous distributions F_t, the ES is formally defined as

$$ES_{t,\alpha}(Y_{t+1}) = \mathbb{E}_t\left[Y_{t+1} | Y_{t+1} \leq Q_{t,\alpha}(Y_{t+1})\right], \tag{2.2.7}$$

where $Q_{t,\alpha}(Y_{t+1})$ denotes the conditional α-quantile of Y_{t+1} given $\mathcal{F}_t$. As discussed in the previous section, the main ingredient of forecast encompassing tests is the specification of the underlying loss function, which has to be associated with the risk measures we consider forecasts for. As such loss functions do not exist for the ES stand-alone, we utilize a strictly

consistent joint loss function for the pair consisting of the ES and the VaR, given by Fissler and Ziegel (2016) as

$$\rho(Y, q_\alpha, e_\alpha) = -\frac{1}{e_\alpha}\left(e_\alpha - q_\alpha + \frac{(q_\alpha - Y)\mathbb{1}_{\{Y \leq q_\alpha\}}}{\alpha}\right) + \log(-e_\alpha), \tag{2.2.8}$$

where the arguments Y, q_α and e_α denote the return realization, the quantile and the ES respectively. As this loss function exhibits the desirable property of having loss differences which are homogeneous of order zero, it is often denoted as the FZ0-loss function, see e.g. Patton et al. (2019). While there exist infinitely many strictly consistent loss functions for the pair VaR and ES, the recent literature seems to agree upon this choice: Dimitriadis and Bayer (2019) find that it exhibits a stable numerical performance in M-estimation and empirically yields relatively efficient parameter estimates. Nolde and Ziegel (2017) discuss the desirable property of homogeneity of these loss functions and Patton et al. (2019), Bayer and Dimitriadis (2020) and Taylor (2019) use this loss function to estimate dynamic ES models.

Following the specification of a link function in (2.2.5), we introduce the quantile- and ES-specific link functions

$$g^q : \mathfrak{Q} \times \Theta^\beta \to \mathbb{R}, \qquad (\hat{\boldsymbol{q}}_t, \beta) \mapsto g^q(\hat{\boldsymbol{q}}_t, \beta), \tag{2.2.9}$$

$$g^e : \mathfrak{E} \times \Theta^\eta \to \mathbb{R}\,, \qquad (\hat{\boldsymbol{e}}_t, \eta) \mapsto g^e(\hat{\boldsymbol{e}}_t, \eta), \tag{2.2.10}$$

where $\mathfrak{Q}$ and $\mathfrak{E}$ denote the random spaces of the VaR and ES forecasts, $\Theta^\beta \subseteq \mathbb{R}^{k_\beta}$ and $\Theta^\eta \subseteq \mathbb{R}^{k_\eta}$ such that $\Theta = \Theta^\beta \times \Theta^\eta$ and $k_\beta + k_\eta = k \in \mathbb{N}$. We assume that the functions g^q, g^e and the parameter space Θ are chosen such that there exist values $\beta_0 \in \Theta^\beta$ and $\eta_0 \in \Theta^\eta$, such that $g^q(\hat{\boldsymbol{q}}_t, \beta_0) = \hat{q}_{1,t}$ and $g^e(\hat{\boldsymbol{e}}_t, \eta_0) = \hat{e}_{1,t}$ almost surely.

In the following, we introduce the concept of joint forecast encompassing for the pair consisting of the VaR and the ES. Analogously to (2.2.2), we define the optimal combination parameters for the VaR and ES as

$$\theta^* = (\beta^*, \eta^*) = \underset{(\beta,\eta)\in\Theta}{\arg\min}\, \mathbb{E}\left[\rho\big(Y_{t+1}, g^q(\hat{\boldsymbol{q}}_t, \beta), g^e(\hat{\boldsymbol{e}}_t, \eta)\big)\right]. \tag{2.2.11}$$

Definition 2.2.3 (Joint VaR and ES Forecast Encompassing). Let $\left(\hat{q}_{1,t}, \hat{e}_{1,t}\right)$ and $\left(\hat{q}_{2,t}, \hat{e}_{2,t}\right)$ denote pair-wise competing forecasts for the pair consisting of the conditional quantile and

ES of F_t. We say that $(\hat{q}_{1,t}, \hat{e}_{1,t})$ encompasses $(\hat{q}_{2,t}, \hat{e}_{2,t})$ at time t with respect to the link functions g^q and g^e if and only if

$$\mathbb{E}\left[\rho\big(Y_{t+1}, \hat{q}_{1,t}, \hat{e}_{1,t}\big)\right] = \mathbb{E}\left[\rho\big(Y_{t+1}, g^q(\hat{\boldsymbol{q}}_t, \beta^*), g^e(\hat{\boldsymbol{e}}_t, \eta^*)\big)\right], \tag{2.2.12}$$

where the loss function ρ is given in (2.2.8). This holds if and only if $(\beta^*, \eta^*) = (\beta_0, \eta_0)$.

We test whether the sequence of joint quantile and ES forecasts $(\hat{q}_{1,t}, \hat{e}_{1,t})$ encompasses the sequence $(\hat{q}_{2,t}, \hat{e}_{2,t})$ for all $t = m, \ldots, T-1$ by estimating the parameters of the following semiparametric regression,

$$Y_{t+1} = g^q(\hat{\boldsymbol{q}}_t, \beta) + u^q_{t+1}, \qquad \text{and} \qquad Y_{t+1} = g^e(\hat{\boldsymbol{e}}_t, \eta) + u^e_{t+1}, \tag{2.2.13}$$

where $Q_\alpha(u^q_{t+1}|\mathcal{F}_t) = 0$ and $ES_\alpha(u^e_{t+1}|\mathcal{F}_t) = 0$ almost surely for all $t = m, \ldots, T-1$ by using the M-estimation technique introduced in Patton et al. (2019) and Dimitriadis and Bayer (2019). We then test for $(\beta^*, \eta^*) = (\beta_0, \eta_0)$ using a Wald type test statistic.

Definition 2.2.3 develops a *joint* encompassing test for the VaR and ES, which is reasonable given the joint elicitability property of the VaR and ES. However, a further objective of this paper is to construct encompassing tests for the ES stand-alone, which we do in the following.

Definition 2.2.4 (Auxiliary ES Forecast Encompassing). Let $(\hat{q}_{1,t}, \hat{e}_{1,t})$ and $(\hat{q}_{2,t}, \hat{e}_{2,t})$ denote competing forecasts for the pair consisting of the conditional quantile and ES of F_t. We say that $\hat{e}_{1,t}$ *auxiliarily encompasses* $\hat{e}_{2,t}$ at time t with respect to the link functions g^q and g^e if and only if

$$\mathbb{E}\left[\rho\big(Y_{t+1}, g^q(\hat{\boldsymbol{q}}_t, \beta^*), \hat{e}_{1,t}\big)\right] = \mathbb{E}\left[\rho\big(Y_{t+1}, g^q(\hat{\boldsymbol{q}}_t, \beta^*), g^e(\hat{\boldsymbol{e}}_t, \eta^*)\big)\right], \tag{2.2.14}$$

that is, if and only if $\eta^* = \eta_0$.[3]

This parameter restriction is tested using a Wald type test statistic based on the estimates of the regression setup given in (2.2.13). As we do not test the quantile specific parameters β^*, we do not impose that the underlying quantile forecast also encompasses its competitor under this null hypothesis. Hence, even though this test is based on the joint regression, it only tests encompassing of the ES forecasts. We call this test *auxiliary ES encompassing test* as it still depends on the auxiliary quantile forecasts which are used for the estimation of the optimal combination parameters.

[3]It is important to notice that the optimal combination parameter β^* on the left-hand side of (2.2.14) is given by (2.2.11) and not in the sense of an optimal combination parameter of a *restricted* model.

Given that both, the VaR and ES forecasts are available, application of either the joint or auxiliary test is the most plausible approach given their joint elicitability. However, even though the emphasis of the auxiliary encompassing test is on the ES, it still requires quantile forecasts for the implementation of the parameter estimation. This can be problematic for two reasons. First, the quantile forecasts are still used in the estimation procedure and thus have an indirect effect on the parameter estimates of the ES specific parameters. E.g., the previous tests are not applicable for ES forecasts which are based on the same VaR forecasts, as this implies perfect collinearity of the quantile regressors. Second, the auxiliary test is only applicable in the setup where the person applying the test has access to the quantile forecasts. In the current implementation of the regulatory framework of the Basel Committee (Basel Committee, 2016, 2017), the banks are only obligated to report their ES forecasts (at probability level 2.5%), but not the corresponding VaR forecasts. Thus, the accompanying VaR forecasts, which the ES forecasts are internally based on, are in general not available to the regulator who has to decide on an adequate risk management of the financial institution at hand.

In order to account for these scenarios, we further introduce the *strict ES encompassing test*, which only requires ES forecasts in the following. For this, we slightly modify the definition of (2.2.11) by replacing $g^q(\hat{\boldsymbol{q}}_t, \beta)$ through $g^q(\hat{\boldsymbol{e}}_t, \beta)$,[4]

$$\theta^* = (\beta^*, \eta^*) = \underset{(\beta,\eta)\in\Theta}{\arg\min}\, \mathbb{E}\left[\rho\big(Y_{t+1}, g^q(\hat{\boldsymbol{e}}_t, \beta), g^e(\hat{\boldsymbol{e}}_t, \eta)\big)\right]. \tag{2.2.15}$$

Definition 2.2.5 (Strict ES Forecast Encompassing). Let $\hat{e}_{1,t}$ and $\hat{e}_{2,t}$ denote competing ES forecasts of the underlying predictive distribution F_t. We say that $\hat{e}_{1,t}$ strictly encompasses $\hat{e}_{2,t}$ at time t with respect to the link functions g^q and g^e if and only if

$$\mathbb{E}\left[\rho\big(Y_{t+1}, g^q(\hat{\boldsymbol{e}}_t, \beta^*), \hat{e}_{1,t}\big)\right] = \mathbb{E}\left[\rho\big(Y_{t+1}, g^q(\hat{\boldsymbol{e}}_t, \beta^*), g^e(\hat{\boldsymbol{e}}_t, \eta^*)\big)\right], \tag{2.2.16}$$

that is, if and only if $\eta^* = \eta_0$.

We test whether $\hat{e}_{1,t}$ strictly encompasses $\hat{e}_{2,t}$ for all $t = m, \dots, T-1$ by setting up the slightly transformed regression

$$Y_{t+1} = g^q(\hat{\boldsymbol{e}}_t, \beta) + u^q_{t+1}, \qquad \text{and} \qquad Y_{t+1} = g^e(\hat{\boldsymbol{e}}_t, \eta) + u^e_{t+1}, \tag{2.2.17}$$

where $Q_\alpha(u^q_{t+1}|\mathcal{F}_t) = 0$ and $ES_\alpha(u^e_{t+1}|\mathcal{F}_t) = 0$ almost surely for all $t = m, \dots, T-1$. The crucial difference between this test and the joint and auxiliary encompassing tests is that instead of using the quantile forecasts $\hat{\boldsymbol{q}}_t$ in the quantile link function g^q, we use the ES

[4]Note that the parameters denoted θ^* in (2.2.11) and in (2.2.15) can generally differ.

forecasts $\hat{\boldsymbol{e}}_t$ for both, the quantile and ES link functions g^q and g^e. We argue that this can be seen as a best feasible solution due to the lack of loss functions for the ES stand-alone together with the necessity of developing forecast evaluation methods for the ES stand-alone due to the current setup of the Basel III regulatory framework (Basel Committee, 2016, 2017).

The underlying idea of this test is mainly motivated by pure scale models, i.e. $Y_{t+1} = \sigma_t u_{t+1}$, $u_{t+1} \sim F(0,1)$, which is still the most frequently used class of models for risk management with the GARCH and stochastic volatility models as prime examples. For this model class, the VaR and ES forecasts are perfectly colinear, $\hat{e}_t = \frac{\xi_\alpha}{z_\alpha}\hat{q}_t$, where z_α and ξ_α are the α-quantile and α-ES of the distribution $F(0,1)$. Hence, the quantile model $g^q(\hat{\boldsymbol{e}}_t, \beta) = g^q(\hat{\boldsymbol{q}}_t \xi_\alpha / z_\alpha, \beta) = g^q(\hat{\boldsymbol{q}}_t, \tilde{\beta})$ is correctly specified, but with transformed quantile parameters $\tilde{\beta}$.[5] As we only test on the ES-specific parameters η as described in Definition 2.2.5, our test is invariant to this (often linear) transformation of the parameter β and thus, it is correctly specified for pure scale models.

In the general case, the quantile equation can possibly be misspecified. Thus, we provide asymptotic theory under general model misspecification for the M-estimator in the following section. The potential model misspecification might bias the pseudo-true parameters and challenge the interpretability of the test decision, but we argue that this effect is negligible for this setup. First, the misspecification is only *slight* in the sense that daily financial return data is approximated well by pure scale processes. Second, the misspecification is *indirect* in the sense that while the quantile parameters are potentially misspecified, we only test the ES parameters, which are influenced by the misspecification only indirectly through the joint estimation. Furthermore, we illustrate that the performance of our strict ES encompassing test is not negatively influenced by more general data generating processes in the simulation study in Section 2.3 by considering GAS models with time-varying higher moments of Creal et al. (2013) and the dynamic ES models of Patton et al. (2019) and Taylor (2019).

Tests for equal (superior) predictive ability in the sense of Diebold and Mariano (1995), Clark and McCracken (2001), Giacomini and White (2006), West (2006) and the model confidence set approach of Hansen et al. (2011) can be seen as a general alternative to encompassing tests. As these tests are directly based on the average loss difference, they can only test the predictive ability of the VaR and ES jointly. In contrast, encompassing tests are based on the regression coefficients of the semiparametric quantile and ES models and hence, only indirectly on the respective loss function. This fundamental difference allows for stand-alone encompassing tests for ES forecasts, which constitutes a great advantage for ES encompassing tests.

[5] For the prominent case of linear encompassing link formulas $g^q(\cdot)$, it holds that $\tilde{\beta} = \beta z_\alpha / \xi_\alpha$.

Strictly speaking, strict consistency of loss functions only implies that the *optimal* forecast exhibits the *smallest possible* loss in expectation. In reality however, competing forecasts are often misspecified due to estimation error or misspecified forecasting models. Patton (2020) shows that then, the ranking induced by the loss functions can be sensitive towards the choice of (strictly consistent) loss functions or even misleading. Holzmann and Eulert (2014) show that for competing forecasts which are based on nested information sets and which are correctly specified given their underlying (but usually incomplete) information set (auto-calibrated), applying any strictly consistent loss function results in a correct ranking of the forecasts. In our case of testing forecast encompassing, we indeed build on nested information sets as it obviously holds that $\sigma\{\hat{f}_{1,t}, \hat{f}_{2,t}\} \supseteq \sigma\{\hat{f}_{1,t}\}$. Thus, by further assuming that the issued forecasts are auto-calibrated given the forecaster's information set, we can conclude that the ranking implied by (2.2.1) is indeed the correct one and invariant towards the choice of strictly consistent loss functions.

2.2.3. Asymptotic theory under model misspecification

In the following, we use the short notation $g_t^e(\eta) = g^e(\hat{\boldsymbol{e}}_t, \eta)$ and $g_t^q(\beta) = g^q(\hat{\boldsymbol{q}}_t, \beta)$ (or $g_t^q(\beta) = g^q(\hat{\boldsymbol{e}}_t, \beta)$ in the case of the strict test). We define the M-estimator as

$$\hat{\theta}_n = \underset{\theta\in\Theta}{\arg\min}\, Q_n(\theta), \quad \text{where} \quad Q_n(\theta) = \frac{1}{n}\sum_{t=m}^{T-1} \rho\big(Y_{t+1}, g_t^q(\beta), g_t^e(\eta)\big), \tag{2.2.18}$$

and the pseudo-true parameter as[6]

$$\theta_n^* = \underset{\theta\in\Theta}{\arg\min}\, Q_n^0(\theta), \quad \text{where} \quad Q_n^0(\theta) = \frac{1}{n}\sum_{t=m}^{T-1} \mathbb{E}\left[\rho\big(Y_{t+1}, g_t^q(\beta), g_t^e(\eta)\big)\right]. \tag{2.2.19}$$

When the link (regression) functions $g^q(\cdot)$ and $g^e(\cdot)$ are correctly specified, we get that the pseudo-true parameter θ_n^* equals the classical true regression parameter and it is independent of the sample size n. We further define the corresponding identification functions, which are almost surely the derivative of the loss function ρ with respect to θ,

$$\psi\big(Y_{t+1}, g_t^q(\beta), g_t^e(\eta)\big) = \begin{pmatrix} -\frac{\nabla g_t^q(\beta)}{\alpha g_t^e(\eta)}\left(\mathbb{1}_{\{Y_{t+1}\le g_t^q(\beta)\}} - \alpha\right) \\ \frac{\nabla g_t^e(\eta)}{g_t^e(\eta)^2}\left(g_t^e(\eta) - g_t^q(\beta) + \frac{1}{\alpha}(g_t^q(\beta) - Y_{t+1})\mathbb{1}_{\{Y_{t+1}\le g_t^q(\beta)\}}\right) \end{pmatrix}. \tag{2.2.20}$$

We restrict our attention to processes which satisfy the following conditions.

Assumption 2.2.1. We assume that

[6]The pseudo-true parameter can generally depend on the issued loss function, i.e. in this case on the zero homogeneous choice in (2.2.8).

(a) the process Z_t is strong mixing of size $-r/(r-2)$ for some $r > 2$,

(b) the parameter space $\Theta = \Theta^\beta \times \Theta^\eta \subseteq \mathbb{R}^k$ is compact and non-empty,

(c) the pseudo-true parameter θ_n^* defined in (2.2.19) is in the interior of Θ and is the unique minimizer of the objective function $Q_n^0(\theta)$ and the sequence $\mathbb{E}_t\left[\psi\left(Y_{t+1}, g_t^q(\beta), g_t^e(\eta)\right)\right]$, defined in (2.2.20) is uncorrelated,

(d) the distribution of Y_{t+1} given $\mathcal{F}_t$, denoted by F_t is absolutely continuous with continuous and strictly positive density h_t, which is bounded from above almost surely on the whole support of F_t and Lipschitz continuous,

(e) for all θ in a neighborhood of θ_n^*, it holds that $\left|\frac{1}{g_t^e(\eta)}\right| \leq K < \infty$ for some constant $K > 0$,

(f) the link functions $g_t^q(\beta)$ and $g_t^e(\eta)$ are $\mathcal{F}_t$-measurable, twice continuously differentiable in $\theta = (\beta, \eta)$ on $\text{int}(\Theta)$ almost surely and if $\mathbb{P}\big(g_t^q(\beta_1) = g_t^q(\beta_2) \cap g_t^e(\eta_1) = g_t^e(\eta_2)\big) = 1$, then $\theta_1 = \theta_2$,

(g) the matrices Λ_n and Σ_n, defined in Proposition 2.2.2 are positive definite with a determinant bounded away from zero for all n sufficiently large,

(h) it holds that $g_t^q(\beta) \leq Q$, $\nabla g_t^q(\beta) \leq Q_1$, $H_t^q(\beta) \leq Q_2$, $\nabla H_t^q(\beta) \leq Q_3$, and $g_t^e(\eta) \leq E$, $\nabla g_t^e(\eta) \leq E_1$, $H_t^e(\eta) \leq E_2$, $\nabla H_t^e(\eta) \leq E_3$, for all θ in a neighborhood of θ_n^*, where the random variables $Q, E, Q_1, E_1, Q_2, E_2, Q_3, E_3$ are all $\mathcal{F}_t$-measurable and for some $r > 2$ (from condition (a)), the following moments are bounded (i) $\mathbb{E}[Q_1^{r+1}]$, (ii) $\mathbb{E}[E_1^{r+1}]$, (iii) $\mathbb{E}[Q_2^{(r+1)/2}]$, (iv) $\mathbb{E}[E_2^{(r+1)/2}]$, (v) $\mathbb{E}[E_1 Q_2]$, (vi) $\mathbb{E}[Q_1 Q_2]$, (vii) $\mathbb{E}[Q_1 E_2]$, (viii) $\mathbb{E}[Q_1^2 E_1]$, (ix) $\mathbb{E}[E E_1^3]$, (x) $\mathbb{E}[E E_3]$, (xi) $\mathbb{E}[E E_1 E_2]$, (xii) $\mathbb{E}[Q E_1 E_2]$, (xiii) $\mathbb{E}[Q E_1^3]$, (xiv) $\mathbb{E}[Q_1 Q^r E_1^r]$, (xv) $\mathbb{E}[E_1^{r-1} E_2 | Y_t]^r]$, (xvi) $\mathbb{E}[E_1^{r+1} | Y_t]^r]$, (xvii) $\mathbb{E}[Y_t^{2r}]$,

(i) for any n, the term $\sup_{\beta \in \Theta^\beta} \sum_{t=m}^{T-1} \mathbb{1}_{\{Y_{t+1} = g_t^q(\beta)\}}$ is almost surely bounded from above.

The following propositions show consistency and asymptotic normality of the M-estimator under potential model misspecification.

Proposition 2.2.1. Given the conditions in Assumption 2.2.1, it holds that $\hat{\theta}_n - \theta_n^* \xrightarrow{\mathbb{P}} 0$.

Proposition 2.2.2. Given the conditions in Assumption 2.2.1, it holds that

$$\Omega_n^{-1/2}(\theta_n^*)\sqrt{n}\big(\hat{\theta}_n - \theta_n^*\big) \xrightarrow{d} \mathcal{N}(0, I_k), \tag{2.2.21}$$

with $\Omega_n(\theta_n^*) = \Lambda_n^{-1}(\theta_n^*)\,\Sigma_n(\theta_n^*)\,\Lambda_n^{-1}(\theta_n^*)$, where $\Lambda_n(\theta_n^*) = \begin{pmatrix} \Lambda_{n,qq}(\theta_n^*) & \Lambda_{n,qe}(\theta_n^*) \\ \Lambda_{n,eq}(\theta_n^*) & \Lambda_{n,ee}(\theta_n^*) \end{pmatrix}$, and $\Sigma_n(\theta_n^*) = \frac{1}{n}\sum_{t=m}^{T-1}\mathbb{E}\left[\psi\left(Y_{t+1}, g_t^q(\beta_n^*), g_t^e(\eta_n^*)\right)\cdot\psi\left(Y_{t+1}, g_t^q(\beta_n^*), g_t^e(\eta_n^*)\right)^\top\right]$. Furthermore, the components of $\Lambda_n(\theta_n^*)$ are given by

$$\Lambda_{n,qq}(\theta_n^*) = -\frac{1}{n}\sum_{t=m}^{T-1}\mathbb{E}\left[\frac{H_t^q(\beta_n^*)}{\alpha g_t^e(\eta_n^*)}\left(F_t(g_t^q(\beta_n^*)) - \alpha\right) + \frac{\nabla g_t^q(\beta_n^*)\nabla g_t^q(\beta_n^*)^\top}{\alpha g_t^e(\eta_n^*)} h_t(g_t^q(\beta_n^*))\right], \tag{2.2.22}$$

$$\Lambda_{n,qe}(\theta_n^*) = \Lambda_{n,eq}(\theta_n^*)^\top = \frac{1}{n}\sum_{t=m}^{T-1}\mathbb{E}\left[\frac{\nabla g_t^q(\beta_n^*)\nabla g_t^e(\eta_n^*)^\top}{\alpha g_t^e(\eta_n^*)^2}\left(F_t(g_t^q(\beta_n^*)) - \alpha\right)\right], \tag{2.2.23}$$

$$\Lambda_{n,ee}(\theta_n^*) = \frac{1}{n}\sum_{t=m}^{T-1}\mathbb{E}\left[\frac{\nabla g_t^e(\eta_n^*)\nabla g_t^e(\eta_n^*)^\top}{g_t^e(\eta_n^*)^2} + \left(\frac{H_t^e(\eta_n^*)}{g_t^e(\eta_n^*)^2} - 2\frac{\nabla g_t^e(\eta_n^*)\nabla g_t^e(\eta_n^*)^\top}{g_t^e(\eta_n^*)^3}\right)\times \tag{2.2.24}$$

$$\left(g_t^e(\eta_n^*) - g_t^q(\beta_n^*) + \frac{1}{\alpha}(g_t^q(\beta_n^*) - Y_{t+1})\mathbb{1}_{\{Y_{t+1}\le g_t^q(\beta_n^*)\}}\right)\Bigg], \tag{2.2.25}$$

where $H_t^q(\beta)$ and $H_t^e(\eta)$ are the Hessian matrices of $g_t^q(\beta)$ and $g_t^e(\eta)$ respectively.

The two preceding propositions extend the asymptotic theory of Patton et al. (2019) to the case of possibly misspecified models, and the misspecification theory for linear models of Bayer and Dimitriadis (2020) to nonlinear models. The proofs in Appendix 2.A combine, extend and go along the lines of the ideas of Engle and Manganelli (2004) and Patton et al. (2019). The conditions closely resemble the regularity conditions of Patton et al. (2019). As we further allow for model misspecification, we impose the unique minimization condition (c) and slightly strengthen the moment conditions (h). In the baseline case of linear encompassing link functions g^q and g^e, the required moment conditions simplify to those given in Bayer and Dimitriadis (2020).

For the estimation of the asymptotic covariance matrix $\widehat{\Omega}_n$ under possible model misspecification, we follow the approach of Dimitriadis and Bayer (2019) and Bayer and Dimitriadis (2020). We deal with the three nuisance quantities in $\widehat{\Omega}_n$ as follows. In order to estimate the density quantile function $h_t(g_t^q(\beta_n^*))$, we follow the *nid*-estimator of Hendricks and Koenker (1992). As the degree of misspecification in the investigated financial time series is small (Bayer and Dimitriadis, 2020), we approximate $F_t(g_t^q(\beta_n^*)) \approx \alpha$. For the conditional truncated variance, $\mathrm{Var}_t\left(g_t^q(\beta_n^*) - Y_{t+1}\middle| Y_{t+1} \le g_t^q(\beta_n^*)\right)$, we employ the *scl-sp* estimator of Dimitriadis and Bayer (2019).

We now consider the asymptotic distributions of our three ES encompassing tests proposed in Section 2.2.2 under the null hypotheses and for general link functions, where we test certain s-dimensional ($s \in \mathbb{N}, s \le k$) sub-vectors of θ. For this, let $R \in \mathbb{R}^{k\times s}$ be a selection matrix whose columns consist of k-dimensional Cartesian unit (column)

vectors $e_j \in \mathbb{R}^k$, which are zero apart from a one in dimension j. E.g., when $g_t^q(\beta)$ and $g_t^e(\eta)$ equal the linear link functions with intercept, given in the first point of Example 2.2.1, $\theta = (\beta_1, \beta_2, \beta_3, \eta_1, \eta_2, \eta_3)$. Then, for the strict and auxiliary ES encompassing tests, $R = (e_5, e_6)$ and for the joint test $R = (e_2, e_3, e_5, e_6)$. These choices pick the respective parameters from θ. Then, we define the respective test statistics by

$$Z_{R,n} = n\big(\hat{\theta}_n R - \theta_n^* R\big)\, R^\top \widehat{\Omega}_n^{-1} R \left(\hat{\theta}_n R - \theta_n^* R\right)^\top. \tag{2.2.26}$$

Theorem 2.2.2 (ES Encompassing Tests). Given the conditions of Assumption 2.2.1 and given that $\widehat{\Omega}_n - \Omega_n \xrightarrow{\mathbb{P}} 0$, under the respective null hypotheses given in Definition 2.2.3 - 2.2.5, it holds that

$$Z_{R,n} \xrightarrow{d} \chi_s^2. \tag{2.2.27}$$

For linear link functions, this theorem implies that the limiting χ^2 distribution of the joint test has four degrees of freedom, while the one of the strict and auxiliary tests has two degrees of freedom.

An important application of these ES encompassing tests is in the context of selecting the best-performing forecast, i.e. selecting at time T a superior forecasting method for the future. This is particularly relevant as the ES is recently introduced into the Basel regulations without having proper forecast selection procedures at hand. Following Giacomini and Komunjer (2005), we propose the following decision rule. We test the two encompassing hypotheses $\mathbb{H}_0^{(1)}$: $\hat{e}_{1,t}$ encompasses $\hat{e}_{2,t}$ and $\mathbb{H}_0^{(2)}$: $\hat{e}_{2,t}$ encompasses $\hat{e}_{1,t}$ for $t = m, \ldots, T-1$. Then, there are four possible scenarios: (1) if neither $\mathbb{H}_0^{(1)}$ nor $\mathbb{H}_0^{(2)}$ are rejected, the test is not helpful for forecast selection. (2) If $\mathbb{H}_0^{(1)}$ is rejected while $\mathbb{H}_0^{(2)}$ is not rejected, we can conclude that forecast $\hat{e}_{2,t}$ does add information to forecast $\hat{e}_{1,t}$, while we cannot conclude the reverse. Thus, we decide to use the forecasting method of $\hat{e}_{2,t}$. (3) If $\mathbb{H}_0^{(2)}$ is rejected while $\mathbb{H}_0^{(1)}$ is not rejected, the same logic applies inversely and we use the forecasting method of $\hat{e}_{1,t}$. (4) If both, $\mathbb{H}_0^{(1)}$and $\mathbb{H}_0^{(2)}$are rejected, the test delivers statistical evidence that both forecasts contain exclusive information and that a forecast combination outperforms the stand-alone forecasts. Consequently, we use a combined forecast $\hat{e}_{c,t} = g^e(\hat{\boldsymbol{e}}_t, \hat{\eta}_n)$ where the estimated combination parameters $\hat{\eta}_n$ are obtained from the M-estimator proposed here.

Testing forecast encompassing *conditional* on some information set $\tilde{\mathcal{G}}_t = \sigma\{\boldsymbol{W}_t\}$ based on some $\mathcal{F}_t$-measurable vector of instruments $\boldsymbol{W}_t$ in the sense of Giacomini and Komunjer (2005) can be facilitated through estimating the regression parameters through (overidentified) GMM-estimation instead of M-estimation. However, for the strict ES test, this approach requires asymptotic theory under model misspecification for the overidentified

GMM estimator based on *nonsmooth* objective functions. While such theory is available for smooth moment conditions (see e.g. Hall and Inoue (2003) and Hansen and Lee (2019)), its generalization to nonsmooth objective functions is not straight-forward and thus, we leave *conditional* ES encompassing tests based on misspecified GMM-estimation for future research. The moment conditions of our unconditional approach can be interpreted as *conditional* encompassing with respect to the instruments $\nabla g_t^q(\beta)$ and $\nabla g_t^e(\eta)$. In the classical baseline case of linear forecast encompassing, these instruments simplify to $\hat{\boldsymbol{q}}_t$ and $\hat{\boldsymbol{e}}_t$ and thus, our approach tests *conditional* encompassing with respect to the information set $\mathcal{G}_t = \sigma\{1, \hat{\boldsymbol{q}}_t, \hat{\boldsymbol{e}}_t\} \subseteq \mathcal{F}_t$, which in most cases already captures the most relevant information which is available.

2.3. Simulation study

In this section, we evaluate the size and power properties of our three proposed ES encompassing tests and compare them to the VaR encompassing test of Giacomini and Komunjer (2005). For this, we describe the simulation setup in Section 2.3.1 and we report and discuss the simulation results in Section 2.3.2. Section 2.3.3 considers three extensions of the simulation setup with respect to additional data generating processes (DGPs), loss and link functions.

2.3.1. The simulation setup

We employ the three encompassing tests based on the linear link functions $g^q(\hat{\boldsymbol{f}}_t, \beta) = \beta_1 + \beta_2 \hat{f}_{1,t} + \beta_3 \hat{f}_{2,t}$ and $g^e(\hat{\boldsymbol{e}}_t, \eta) = \eta_1 + \eta_2 \hat{e}_{1,t} + \eta_3 \hat{e}_{2,t}$, where $\hat{\boldsymbol{f}}_t = \hat{\boldsymbol{q}}_t$ for the joint and auxiliary tests and $\hat{\boldsymbol{f}}_t = \hat{\boldsymbol{e}}_t$ for the strict test, together with the parameter space $\Theta = \{\theta = (\beta, \eta) \in \mathbb{R}^6 : ||\theta|| \leq K\}$.[7] For the respective encompassing tests, in each case we test the following two opposing hypotheses:

$$\begin{aligned}
&\text{Joint :} && \mathbb{H}_0^{(1)} : (\beta_2^*, \beta_3^*, \eta_2^*, \eta_3^*) = (1,0,1,0), && \mathbb{H}_0^{(2)} : (\beta_2^*, \beta_3^*, \eta_2^*, \eta_3^*) = (0,1,0,1),\\
&\text{Str \& Aux :} && \mathbb{H}_0^{(1)} : (\eta_2^*, \eta_3^*) = (1,0), && \mathbb{H}_0^{(2)} : (\eta_2^*, \eta_3^*) = (0,1),\\
&\text{VaR :} && \mathbb{H}_0^{(1)} : (\beta_2^*, \beta_3^*) = (1,0), && \mathbb{H}_0^{(2)} : (\beta_2^*, \beta_3^*) = (0,1).
\end{aligned} \tag{2.3.1}$$

In the following, we describe two DGPs where for the first, both forecasting models stem from classical GARCH models while the second considers two joint GAS models for the VaR and ES of Patton et al. (2019). For both model classes, we simulate data as

[7] We choose the constant K large enough such that the parameter estimation is not restricted in realistic settings but the parameter space Θ is indeed convex.

a convex combination of two distinct models with a flexible convex combination weight $\pi \in [0,1]$. This implies that for $\pi = 0$, the first model encompasses the second, while for $\pi = 1$, the inverse holds. For all intermediate parameters $\pi \in (0,1)$, the data stems from a linear combination and both forecast encompassing null hypotheses should be rejected which indicates that a forecast combination method is preferred.

The GARCH DGP

The two GARCH models, which are calibrated to daily IBM returns, are given by $\tilde{Y}_{j,t+1} = \hat{\sigma}_{j,t} u_{t+1}$, for $j = 1,2$, where $u_{t+1} \overset{iid}{\sim} \mathcal{N}(0,1)$ and the two distinct volatility specifications are given by

$$\hat{\sigma}^2_{1,t} = 0.042 + 0.053\tilde{Y}^2_{1,t} + 0.925\hat{\sigma}^2_{1,t-1}, \qquad \text{and} \tag{2.3.2}$$

$$\hat{\sigma}^2_{2,t} = 0.044 + \big(0.024 + 0.058 \cdot \mathbb{1}_{\{\tilde{Y}_{2,t} \leq 0\}}\big)\tilde{Y}^2_{2,t} + 0.923\hat{\sigma}^2_{2,t-1}. \tag{2.3.3}$$

For both models, we obtain VaR and ES forecasts by $\hat{q}_{j,t} = z_\alpha \hat{\sigma}_{j,t}$ and $\hat{e}_{j,t} = \xi_\alpha \hat{\sigma}_{j,t}$, for $j = 1,2$, where z_α and ξ_α are the α-quantile and α-ES of the standard normal distribution. Notice that the time index t on $\hat{\sigma}_{j,t}$ indicates that it is a $\mathcal{F}_t$-measurable forecast for time $t+1$. While the first specification in (2.3.2) is a classical GARCH(1,1) model (Bollerslev, 1986), the second specification in (2.3.3) follows the GJR-GARCH model of Glosten et al. (1993), which allows for a leverage effect. We simulate data from the convex combination of these processes, $Y_{t+1} = \big((1-\pi)\hat{\sigma}_{1,t} + \pi\hat{\sigma}_{2,t}\big)u_{t+1}$ for 21 equally spaced values of $\pi \in [0,1]$, where $u_{t+1} \overset{iid}{\sim} \mathcal{N}(0,1)$.

The VaR/ES GAS DGP

In the second simulation setup, we implement the one-factor (1F) and two-factor (2F) GAS models for the VaR and ES of Patton et al. (2019). The 1F-GAS model evolves as

$$\hat{q}_{1,t} = -1.164\exp(\hat{\kappa}_t) \qquad \text{and} \qquad \hat{e}_{1,t} = -1.757\exp(\hat{\kappa}_t), \qquad \text{where} \tag{2.3.4}$$

$$\hat{\kappa}_t = 0.995\hat{\kappa}_{t-1} + \frac{0.007}{\hat{e}_{1,t-1}}\left(\frac{\tilde{Y}_{1,t}}{\alpha}\mathbb{1}_{\{\tilde{Y}_{1,t} \leq \hat{q}_{1,t-1}\}} - \hat{e}_{1,t-1}\right). \tag{2.3.5}$$

The 2F-GAS model follows the specification

$$\begin{pmatrix}\hat{q}_{2,t}\\ \hat{e}_{2,t}\end{pmatrix} = \begin{pmatrix}-0.009\\ -0.010\end{pmatrix} + \begin{pmatrix}0.993 & 0\\ 0 & 0.994\end{pmatrix}\begin{pmatrix}\hat{q}_{2,t-1}\\ \hat{e}_{2,t-1}\end{pmatrix} + \begin{pmatrix}-0.358 & -0.351\\ -0.003 & -0.003\end{pmatrix}\lambda_t, \tag{2.3.6}$$

where the forcing variable is given by $\lambda_t = \big(\hat{q}_{2,t-1}(\alpha - \mathbb{1}_{\{\tilde{Y}_{2,t} \leq \hat{q}_{2,t-1}\}}), \mathbb{1}_{\{\tilde{Y}_{2,t} \leq \hat{q}_{2,t-1}\}}\tilde{Y}_{2,t}/\alpha - \hat{e}_{2,t-1}\big)^\top$. For both models, $j = 1, 2$, we simulate $\tilde{Y}_{j,t+1} \sim \mathcal{N}\big(\hat{\mu}_{j,t}, \hat{\sigma}^2_{j,t}\big)$, where the conditional mean and standard deviations are given by $\hat{\mu}_{j,t} = \hat{q}_{j,t} - z_\alpha \frac{\hat{e}_{j,t} - \hat{q}_{j,t}}{\xi_\alpha - z_\alpha}$ and $\hat{\sigma}_{j,t} = \frac{\hat{e}_{j,t} - \hat{q}_{j,t}}{\xi_\alpha - z_\alpha}$, such that $Q_\alpha(\tilde{Y}_{j,t+1}|\mathcal{F}_t) = \hat{q}_{j,t}$ and $ES_\alpha(\tilde{Y}_{j,t+1}|\mathcal{F}_t) = \hat{e}_{j,t}$ almost surely. The parameter values for this model are obtained from Table 8 of Patton et al. (2019) and correspond to calibrated parameters to daily S&P 500 returns.

In order to simulate returns which follow a convex combination of these two conditional distributions, we simulate Bernoulli draws $\pi_{t+1} \sim \text{Bern}(\pi)$ for 21 equally spaced values of $\pi \in [0, 1]$, and let $Y_{t+1} = (1 - \pi_{t+1})\tilde{Y}_{1,t+1} + \pi_{t+1}\tilde{Y}_{2,t+1}$. Thus, for $\pi = 0$, Y_{t+1} follows the 1F-GAS model, for $\pi = 1$, Y_{t+1} follows the 2F-GAS model and for $\pi \in (0, 1)$, Y_{t+1} follows some convex combination of these two models.[8]

Both models in the GARCH DGP generate data from a pure scale (volatility) process resulting in perfectly colinear VaR and ES forecasts. In contrast, the more general VaR/ES GAS models in the second DGP generate VaR and ES forecasts which are not colinear and consequently introduce misspecification in the quantile model of the strict ES encompassing test. As the utilized parameters are calibrated to daily financial returns, these models reflect a realistic degree of misspecification encountered in practical risk management.

2.3.2. Simulation results

Table 2.3.1 reports the empirical sizes of the three different ES encompassing tests introduced in Section 2.2 together with the VaR encompassing test of Giacomini and Komunjer (2005) at a 10% nominal significance level based on 2000 Monte Carlo replications. Table 2.D.1 and 2.D.2 present equivalent results for nominal sizes of 1% and 5%. The column panel $\mathbb{H}_0^{(1)}$ indicates that we test whether model 1 encompasses model 2, while the panel $\mathbb{H}_0^{(2)}$ indicates the reverse.

We find that the two ES encompassing tests (the strict and auxiliary test) are well-sized, especially in large samples for both DGPs and for both null hypotheses. While the joint VaR and ES test is slightly oversized, the VaR test exhibits even larger sizes. This behavior is especially remarkable as the ES is considerably further in the tail than the VaR at the same probability level and hence, harder to estimate and test. This pattern can be explained by the fact that the asymptotic covariance of the two tests involving the VaR is subject to estimation of the density quantile function $h_t(g_t^q(\beta_n^*))$, which is hard to estimate for small probability

[8] While generating returns stemming from convex combinations of GARCH-type volatility models is straight-forward by using convex combinations of the conditional volatilities, this is not as simple for the more general GAS models considered in this section. Consequently, we use this more involved approach based on Bernoulli draws in order to generate these convex model combinations.

Table 2.3.1: Empirical sizes of the forecast encompassing tests

	$\mathbb{H}_0^{(1)}$				$\mathbb{H}_0^{(2)}$			
	Str ES	Aux ES	VaR ES	VaR	Str ES	Aux ES	VaR ES	VaR
n	GARCH							
500	15.25	15.20	18.35	22.75	14.40	14.65	18.80	22.50
1000	11.55	11.10	15.60	20.10	12.30	12.70	17.80	22.85
2500	11.45	11.55	16.35	18.80	11.00	11.25	14.60	17.55
5000	10.05	10.25	13.10	15.35	9.75	10.15	13.90	15.75
n	VaR/ES GAS							
500	29.35	29.75	24.15	27.85	21.70	21.15	23.10	27.80
1000	22.75	21.85	19.55	23.95	18.15	18.60	18.15	22.80
2500	16.05	15.80	16.20	18.35	12.65	13.50	15.65	19.65
5000	13.50	13.60	14.05	16.60	10.60	11.35	14.10	17.95

Notes: This table presents the empirical sizes (in %) of our three forecast encompassing tests for the ES together with a VaR encompassing test of Giacomini and Komunjer (2005) for a nominal size of 10%. The results are shown for the two DGPs described in Section 2.3.1 in the horizontal panels, for both tested hypotheses in the vertical panels and for different sample sizes.

levels (Dimitriadis and Bayer, 2019; Giacomini and Komunjer, 2005; Koenker and Bassett, 1978).

We further find that the strict and the auxiliary tests behave almost identically. This also holds for the VaR/ES GAS DGP for which the regression model of the strict ES encompassing test is potentially misspecified.[9] This suggests that the approximation error induced by the misspecification in the strict ES test is negligible for realistic financial settings. Remarkably, in the vast majority of cases, the strict ES test exhibits better size properties than the correctly specified joint VaR and ES and the VaR encompassing tests.

We present power curves (empirical rejection rates) for both DGPs and different sample sizes in the individual plot panels in Figure 2.3.1. In each plot, we depict the respective power curves for our three ES encompassing tests and the VaR encompassing test of Giacomini and Komunjer (2005) for both null hypotheses and for a nominal significance level of 10% based on 2000 Monte Carlo replications. We observe increasing power for both DGPs, both tested null hypotheses and for all four encompassing tests for increasing (decreasing) values of the combination parameter π. We find that while the VaR and joint VaR and ES tests

[9]Figure 2.E.1 plots the ratio of the VaR and ES forecasts for a simulated return series of the 2F-GAS model (and further misspecified models described in Appendix 2.B). Following the discussion after Definition 2.2.5, this ratio mainly governs the degree of misspecification the regression model of the strict ES encompassing test is subject to. For the 2F-GAS model, the ratio of the VaR and ES forecasts fluctuates approximately between 0.7 and 0.85, while it equals 0.84 for the location-scale approaches under normality. This demonstrates that the VaR/ES GAS simulation designs, which are calibrated to real financial data, generate some moderate degree of model misspecification in the regression model.

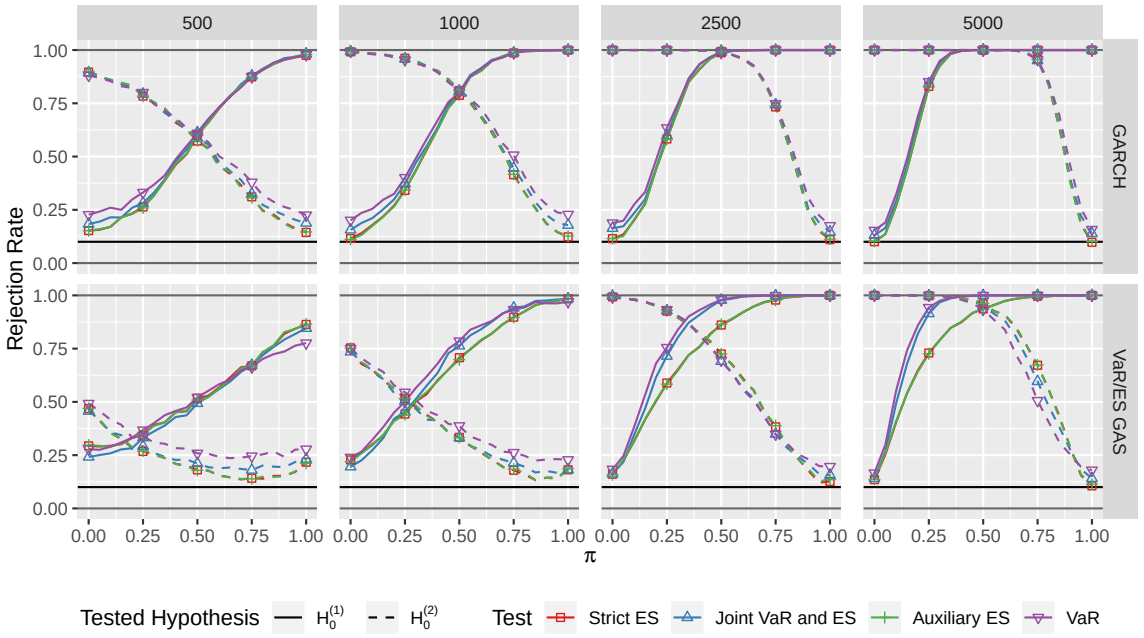


Figure 2.3.1: This figure shows power curves (empirical rejection frequencies) for the encompassing tests with a nominal size of 10% and for the two DGPs described in Section 2.3.1 in the plot rows. The plot columns depict different sample sizes, while the colors indicate the four different encompassing tests and the line types refer to the two tested null hypotheses. An ideal test exhibits a rejection frequency of 10% for $\pi = 0$ and for $\mathbb{H}_0^{(1)}$ (and inversely for $\pi = 1$ and $\mathbb{H}_0^{(2)}$) and as sharply increasing rejection rates as possible for increasing (decreasing) values of π.

are considerably oversized, they produce a similar test power compared to the strict and auxiliary ES encompassing tests, especially for larger (smaller) values of π. Again, the strict and auxiliary ES encompassing tests are almost indistinguishable, which implies that the strict test is robust against the misspecification induced by the VaR/ES GAS models. Interestingly, we find that the power curves for the VaR/ES GAS specifications are slightly asymmetric implying that the tests react differently to the specifications of different numbers of driving factors in the GAS models.

2.3.3. Extensions of the simulation setup

In this section, we consider three extensions of our simulation setup. First, in Section 2.3.3, we present two additional DGPs. Second, in Section 2.3.3, we analyze the behavior of our tests under different strictly consistent loss functions for the pair VaR and ES. Third, in Section 2.3.3, we employ two additional link functions by testing for forecast encompassing for affine and nonlinear forecast combinations.

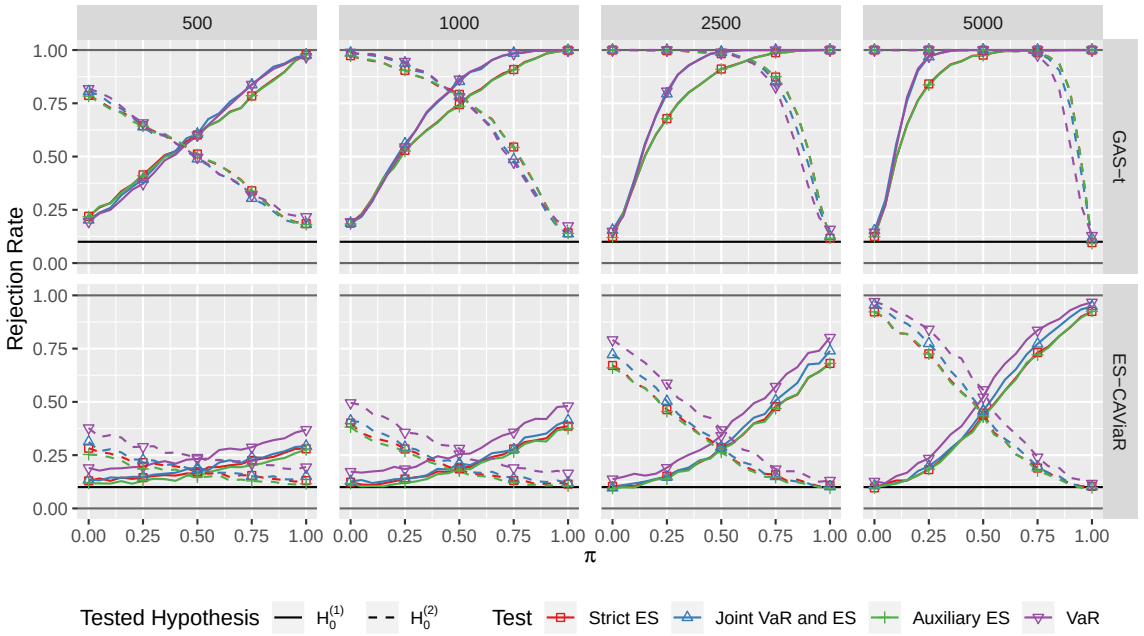


Figure 2.3.2: This figure shows power curves (empirical rejection frequencies) for the encompassing tests with a nominal size of 10%. The plot rows correspond to the GAS-*t* and the ES-CAViaR DGPs described in Section 2.B. The plot columns depict different sample sizes, while the colors indicate the four different encompassing tests and the line types refer to the two tested null hypotheses.

Forecast encompassing under different data generating processes

In this subsection, we consider two additional DGPs, namely the GAS-*t* model of Creal et al. (2013) and the ES-CAViaR models of Taylor (2019), which are described in detail in Appendix 2.B. Both models go beyond the class of pure scale models and consequently generate model misspecification in the quantile regression equation of the strict ES encompassing test. Figure 2.3.2 presents power curves for these two DGPs and Table 2.D.3 reports the corresponding test sizes.

The results of these two DGPs qualitatively confirm the simulation results of Section 2.3.2. The three ES specific tests exhibit accurate empirical test sizes, especially in large samples and the strict and the auxiliary tests generally exhibit better size properties than the joint VaR and ES and the stand-alone VaR encompassing tests. Increasing the sample size results in increasing power for both DGPs and for all considered encompassing tests. Noteworthy, all tests show considerably lower power for the ES-CAViaR DGP compared to the other DGPs. This result is comparable with the power results of Giacomini and Komunjer (2005) as this DGP is a slightly modified version of their DGP. In summary, these two additional DGPs demonstrate that the new ES encompassing tests perform well for a variety of realistic data generating processes.

Forecast encompassing under different loss functions

The three encompassing test specifications for the ES presented in Section 2.2.2 are built on the zero-homogeneous joint loss function for the VaR and ES given in (2.2.8). While this loss function is the most popular choice in the recent literature on semiparametric ES models (Bayer and Dimitriadis, 2020; Patton et al., 2019; Taylor, 2019), there exists an entire class of joint loss functions for the VaR and ES, proposed by Fissler and Ziegel (2016) as

$$\begin{aligned}\rho(Y, q_\alpha, e_\alpha) = &\left(\mathbb{1}_{\{Y \leq q_\alpha\}} - \alpha\right)\mathfrak{g}(q_\alpha) - \mathbb{1}_{\{Y \leq q_\alpha\}}\mathfrak{g}(Y) \\ &+ \phi'(e_\alpha)\left(e_\alpha - q_\alpha + \frac{(q_\alpha - Y)\mathbb{1}_{\{Y \leq q_\alpha\}}}{\alpha}\right) - \phi(e_\alpha) + a(Y),\end{aligned} \tag{2.3.7}$$

where the function $\mathfrak{g}$ is twice continuously differentiable and increasing, ϕ is three times continuously differentiable, strictly increasing and strictly convex, and a and $\mathfrak{g}$ are integrable functions (Fissler and Ziegel, 2016). The loss function in (2.2.8) is a special case of (2.3.7) for the choices $\mathfrak{g}(z) = 0$, $\phi(z) = -\log(-z)$ and $a(z) = 0$.

The ES encompassing tests can generally be set up by using any choice of (2.3.7) (fulfilling certain further weak regularity conditions). We consider two additional specifications in the following. Following the theory of homogeneous loss functions (Nolde and Ziegel, 2017) and the numerical performance in linear regression settings (Dimitriadis and Bayer, 2019), we fix $\mathfrak{g}(z) = 0$ and in addition to $\phi(z) = -\log(-z)$, we employ the choices $\phi(z) = 1/\sqrt{-z}$ and $\phi(z) = -1/z$.[10]

Figure 2.3.3 shows rejection rates for the VaR/ES GAS DGP for different sample sizes and the three encompassing test specifications for the ES, where the different line colors represent the three different loss specifications. Table 2.D.4 reports the corresponding test sizes. We find that the tests based on the three different loss functions perform almost indistinguishably, especially in large samples. This result is not unexpected (especially in large samples where the expectation is well approximated by the sample mean) as Fissler and Ziegel (2016) show that all loss functions in the class in (2.3.7) are uniquely minimized by the true VaR and ES.

For the potentially misspecified strict ES encompassing test, the pseudo-true parameter defined in (2.2.19) may theoretically depend on the underlying loss function and notice that the VaR/ES GAS DGP used in Figure 2.3.3 allows for such a model misspecification. However, we find that neither the rejection rates of the strict ES encompassing test are affected by employing different loss functions, nor are the rejection rates of the two correctly

[10]Strictly speaking, the asymptotic theory in Theorem 2.2.2 only covers the M-estimator based on the loss function in (2.2.8). However, the proofs and the resulting asymptotic covariance matrices are easily extended to the general case by combining the methods of this paper with the extension of Dimitriadis and Bayer (2019) to the case of general loss functions as given in (2.3.7).

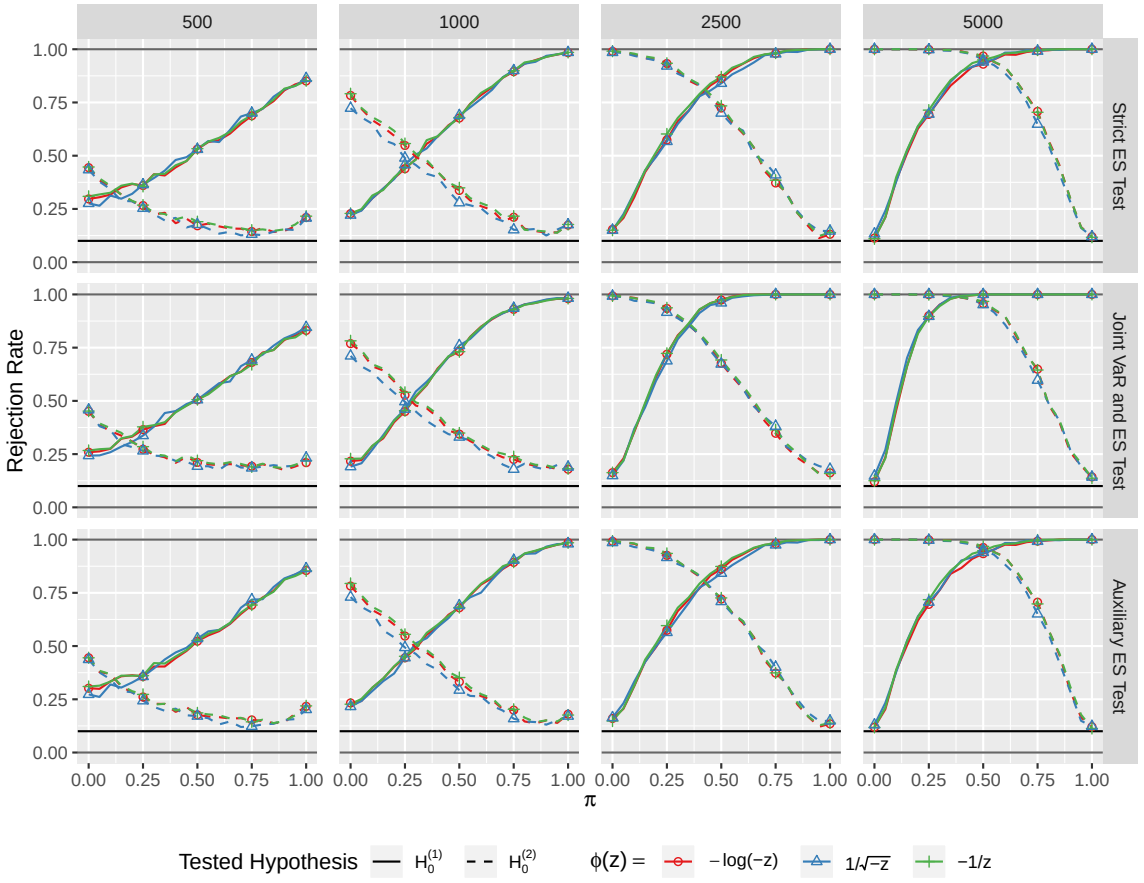


Figure 2.3.3: This figure shows power curves (empirical rejection frequencies) for the three ES specific encompassing tests with a nominal size of 10% for the VaR/ES GAS DGP, where the different colors represent the different loss functions employed in the tests, given in and below (2.3.7). The plot rows depict different sample sizes, the plot columns show the three different ES encompassing tests and the line types refer to the tested null hypotheses.

specified encompassing tests in Figure 2.3.3. This result implies that the potentially different pseudo-true parameters are almost entirely unaffected by the misspecification.

Forecast encompassing under different link functions

In this section, we employ two additional link function specifications. First, we consider an affine combination including an intercept[11]

$$g^q(\hat{\boldsymbol{f}}_t, \beta) = \beta_1 + \beta_2 \hat{f}_{1,t} + (1-\beta_2)\hat{f}_{2,t} \quad \text{and} \quad g^e(\hat{\boldsymbol{e}}_t, \eta) = \eta_1 + \eta_2 \hat{e}_{1,t} + (1-\eta_2)\hat{e}_{2,t}, \tag{2.3.8}$$

[11]We include an intercept as this stabilizes the performance of the associated quantile regression. We do not include classical convex combinations (where $0 \leq \beta_2 \leq 1$) as our theoretical framework does not allow for testing on the boundary (see e.g. Andrews (1999) for details).

where $\hat{\boldsymbol{f}}_t = \hat{\boldsymbol{q}}_t$ for the joint and auxiliary tests and $\hat{\boldsymbol{f}}_t = \hat{\boldsymbol{e}}_t$ for the strict test. For the joint test, the first null hypothesis is given by $\mathbb{H}_0^{(1)} : (\beta_2^*, \eta_2^*) = (1, 1)$ while for the strict and auxiliary tests, it is given by $\mathbb{H}_0^{(1)} : \eta_2^* = 1$. The second, opposing null hypotheses, $\mathbb{H}_0^{(2)}$ are obtained by replacing the ones by zeros. For the affine link functions, we employ the same DGPs as for the encompassing tests based on linear link functions.

Furthermore we employ the nonlinear link functions, where $\hat{\boldsymbol{f}}_t$ is given as in (2.3.8) and

$$\begin{aligned} g^q(\hat{\boldsymbol{f}}_t, \beta) &= \beta_1 - \exp\left(\beta_2 \log(-\hat{f}_{1,t}) + \beta_3 \log(-\hat{f}_{2,t})\right), \quad \text{and} \\ g^e(\hat{\boldsymbol{e}}_t, \eta) &= \eta_1 - \exp\left(\eta_2 \log(-\hat{e}_{1,t}) + \eta_3 \log(-\hat{e}_{2,t})\right), \end{aligned} \tag{2.3.9}$$

and we test the same null hypotheses as for the linear link functions described in (2.3.1). For the nonlinear link function, we employ a slightly modified GARCH DGP. As in Section 2.3.1, let $\hat{\sigma}_{1,t}$ and $\hat{\sigma}_{2,t}$ denote the conditional volatilities of the GARCH and GJR-GARCH models. Then, we simulate data according to

$$Y_{t+1} = \exp\left((1-\pi)\log(\hat{\sigma}_{1,t}) + \pi \log(\hat{\sigma}_{2,t})\right) \cdot u_{t+1} \tag{2.3.10}$$

for an equally spaced grid of 21 values for $\pi \in [0, 1]$ and where $u_{t+1} \overset{iid}{\sim} \mathcal{N}(0, 1)$. This ensures that for $\pi = 0$, $\hat{q}_{1,t}$ and $\hat{e}_{1,t}$ are the correct VaR and ES forecasts, and vice versa for $\pi = 1$. For any $\pi \in (0, 1)$, the true VaR and ES are given by a combination according to the nonlinear link functions in (2.3.9).

Figure 2.3.4 presents power curves for these two additional link functions, where for the affine combination, we consider both, the GARCH and the VaR/ES GAS DGPs. Table 2.D.5 presents the associated test sizes. We find that the power curves for the affine combinations in Figure 2.3.4 are comparable to the linear specifications, as given in Figure 2.3.1 for both considered DGPs. Moreover, the encompassing tests based on the nonlinear link function in the third row of plots perform similar to the linear encompassing tests in the GARCH setting. This extension of the simulation setup shows that our ES encompassing tests can be applied based on a variety of different link functions.

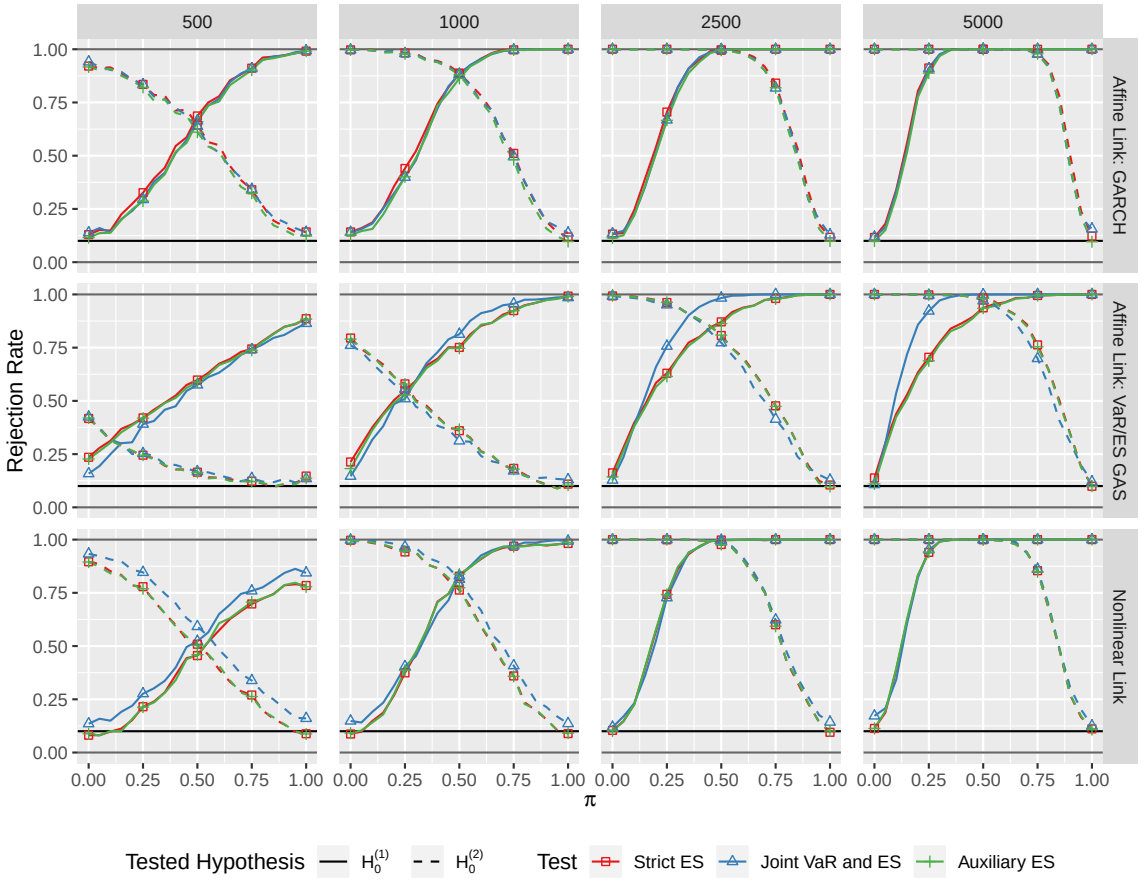


Figure 2.3.4: This figure shows power curves (empirical rejection frequencies) for the encompassing tests with a nominal size of 10% for two additional link functions. The first two rows of plots present results for the affine link functions given in (2.3.8) and for the GARCH and VaR/ES GAS DGPs. The third plot row presents results for the nonlinear link functions given in (2.3.9) and the respective nonlinear GARCH DGP. The plot columns depict different sample sizes, while the colors indicate the three ES specific encompassing tests and the line types refer to the two tested null hypotheses.

2.4. Empirical application

We use close-to-close returns from the IBM stock, the S&P 500 index and the DAX 30 index from June 1st, 2000 until May 31st, 2019, which amounts to a total of $T = 4779$ daily observations. We use a fixed forecasting scheme, i.e. the model parameters are estimated once on the first $m = 2000$ in-sample observations. These parameter estimates are used to generate the VaR and ES forecasts in a rolling-window fashion for the remaining out-of-sample period of $n = 2779$ days. Following the suggestion of the Basel III Accords, we use the probability level $\alpha = 2.5\%$ for the VaR and the ES.

For the analysis, we consider the following competing forecasting models. First, we employ the Historical Simulation (HS) model which generates VaR and ES forecasts by computing the empirical quantile and ES at level α of the past 250 trading days. The second

model is the RiskMetrics (RM) model, which models the conditional volatility as an IGARCH equation with fixed parameter values, $\hat{\sigma}_t^2 = 0.94\hat{\sigma}_{t-1}^2 + 0.06Y_t^2$ and Gaussian residuals. Third, we use the GJR-GARCH(1,1)-t model of Glosten et al. (1993) with Student-t residuals. The fourth model is given by the Student-t-GAS model with time-varying variance and degrees of freedom introduced in Appendix 2.B. The fifth and sixth model are the one and two factor GAS models for the VaR and ES of Patton et al. (2019) set out in Section 2.3.1 and estimated by minimizing the strictly consistent loss function for the VaR and ES given in (2.2.8). The last two models are the two dynamic ES-CAViaR models of Taylor (2019) described in Appendix 2.B. Table 2.D.6 shows the correlations of the respective VaR and ES forecasts of these models. We find that no pair of forecasts is perfectly correlated, which is crucial for the applicability of the encompassing tests as implied by condition (f) of Assumption 2.2.1.

We run pair-wise encompassing tests comparing all eight forecasting methods. Hence, for each model pair, we run encompassing tests for both hypotheses, i.e. that the first forecast encompasses the second, denoted by $\mathbb{H}_0^{(1)}$ and the inverse, denoted by $\mathbb{H}_0^{(2)}$. This results in four possible outcomes of these two tests: (1) *non-rejection* (NR) indicates that none of the null hypotheses is rejected and the tests are not helpful. (2) *encompassed* (E1) denotes the setting where the first model is encompassed by the competitor model but does not encompass it, i.e. $\mathbb{H}_0^{(1)}$ is rejected but $\mathbb{H}_0^{(2)}$ is not, which results in choosing the competitor model. (3) *encompassing* (E2) indicates that the first model encompasses the other but is not encompassed by it, i.e. $\mathbb{H}_0^{(1)}$ is not rejected but $\mathbb{H}_0^{(2)}$ is, which implies that we choose the first model. Finally, (4) *combination* (C) refers to a setting where both null hypotheses are rejected and we consequently opt for a forecast combination.

In Table 2.4.1, we report relative frequencies of test outcomes at the 10% significance level for the different encompassing tests for all three return time series. Tables 2.D.7, 2.D.8 and 2.D.9 report the individual p-values of the encompassing tests. The results can be summarized as follows: first, for the IBM stock returns we find many cases of double rejections and hence empirical evidence for using forecast combinations. This implies that the individual models provide additional and exclusive information and hence, a forecast combination is often superior to the stand-alone forecasting models. This finding supports the theoretical advantages of forecast combinations, presented in a general setting e.g., by Giacomini and Komunjer (2005), Timmermann (2006) and Halbleib and Pohlmeier (2012), and specifically for the pair VaR and ES by Taylor (2020). Second, for the S&P 500 index, and especially for the DAX 30 index, we overall observe less instances of double rejections of the ES encompassing tests. While the decrease in cases where the VaR encompassing test opts for a forecast combination is smaller, these rejections have to be considered carefully given that the VaR encompassing test is oversized in all simulation setups in Section 2.3,

Table 2.4.1: Encompassing test results for the IBM Stock, the DAX 30 and the S&P 500 indices

	Strict ES				Aux ES				Joint VaR ES				VaR			
Model	NR	E1	E2	C	NR	E1	E2	C	NR	E1	E2	C	NR	E1	E2	C
								IBM								
HS		57		43		57		43		43		57		43		57
RM		57		43		57		43		43		57	14	43		43
GJR		57		43		57		43		43		57	14	29		57
GAS		43		57		43		57		29		71		29		71
G1F	14	29	43	14	14	29	43	14		14		86		14		86
G2F	14	29	57		14	29	57			29	57	14		29	43	29
ASES	14		86		14		86		14		57	29	14		57	29
SAVES	14		86		14		86		14		86		14		86	
								S&P 500								
HS		86		14		86		14		71		29		86		14
RM		71	14	14		71	14	14		71	14	14	14	43	14	29
GJR	29		71		29		71				71	29			71	29
GAS	29	29	29	14	14	43	29	14		57	29	14	14	43	14	29
G1F		29		71		29		71		14		86	29	14	29	29
G2F	29	14	57		14	14	71				57	43	14		29	57
ASES	14		71	14	14		71	14			43	57			29	71
SAVES	14	43	29	14	14	43	29	14		43	43	14	14	29	29	29
								DAX 30								
HS		86		14		100				86		14		86		14
RM		43	57			43	57			43	43	14		29	14	57
GJR		57	43			57	43			14	71	14	14	14	43	29
GAS	14	14	71		29		71		14	29	43	14		14	57	29
G1F		71	29			71	29			71	14	14		43	14	43
G2F		86		14		86	14			57		43		43		57
ASES	29		71		29		71				86	14	14		86	
SAVES	14		86		29		71		14		43	43	29	14	29	29

Notes: This table shows results of the pair-wise encompassing test outcomes for the three ES encompassing tests and the VaR encompassing test of Giacomini and Komunjer (2005), the eight considered models estimated on an in-sample period of $m = 2000$ days, a nominal test significance level of 10% and for the IBM stock, the S&P 500 and the DAX 30 indices. The individual columns report the relative frequencies (in %) how often the respective test outcome occurs for a model (given in the rows) when this model generates the *first* forecasts in the encompassing tests. The column "NR" (non-rejection) reports the frequency of cases where neither null hypothesis is rejected, while the column "C" (combination) indicates the frequency that both null hypotheses are rejected. The column "E1" (encompassed) refers to the case that $\mathbb{H}_0^{(1)}$ is rejected while $\mathbb{H}_0^{(2)}$ is not, i.e. it gives the relative frequencies that this model is encompassed by its competitors and the column "E2" (encompassing) refers to the inverse case, i.e. it gives the relative frequencies that this model encompasses its competitors.

even in large samples. This result can be explained by the fact that the S&P 500 and DAX indices are well diversified and the returns fluctuate to a lesser extent and exhibit less extreme outliers than single stock return series. Furthermore, the considered VaR and ES forecasts show larger correlations for the indices than for the single stock in Table 2.D.6, which negatively influences the tests' power. Third, in terms of the frequencies of the cases E1 and

E2, we observe recurring patterns over the different models for both time series. Especially the ES-specific GAS and CAViaR type models seem to exhibit a superior performance, while the HS, RM, GARCH and GAS-t models tend to be encompassed more often. Lastly, the two tests which only focus on testing encompassing of ES forecasts perform almost identically, which supports the conclusion from the simulation study that the potential misspecification does not negatively influence the performance of the strict ES test in realistic financial settings. This is encouraging as the strict ES encompassing test can be applied in cases where one does not have VaR forecasts at hand, such as it is currently imposed by the Basel Committee of Banking Supervision (Basel Committee, 2016, 2017).

Tables 2.D.10, 2.D.11 and 2.D.12 report the joint VaR and ES losses for the zero-homogeneous loss function in (2.2.8) for forecasts stemming from the stand-alone models and the respective forecast combinations with estimated combination weights from the underlying regressions. These results qualitatively confirm the results of the encompassing tests: e.g., for the IBM stock in the first panel of Table 2.4.1, we find that forecast combinations are particularly preferred for the first four models (in the model ordering of the table). The average losses in Table 2.D.10 show a similar pattern throughout all three panels. We observe that the optimal forecast combinations exhibit substantially smaller losses compared to the stand-alone models for the first four models while this decrease is of a considerably smaller magnitude for the last four models.

2.5. Conclusion

With the implementation of the third Basel Accords (Basel Committee, 2016, 2017), risk managers and regulators currently shift attention towards the risk measure Expected Shortfall (ES), which demonstrates the necessity of forecast evaluation and comparison tools for the ES. In this paper, we introduce new forecast encompassing tests for the ES, which are based on a joint loss function and an associated joint regression framework for the ES together with the Value at Risk (Dimitriadis and Bayer, 2019; Fissler and Ziegel, 2016; Patton et al., 2019). We propose three variants of the ES encompassing test, which allow for testing forecast encompassing for flexible parametric forecast combination methods. As one test variant is potentially subject to model misspecification, we extend the existing asymptotic theory of Patton et al. (2019), Dimitriadis and Bayer (2019) and Bayer and Dimitriadis (2020) to cases of potential model misspecification for flexible parametric models. Potential future research on extending the setting presented in this paper includes encompassing tests for convex forecast combinations and forecast combinations which theoretically prevent crossings of the VaR and the ES. Both approaches require non-standard asymptotic theory for tests on the boundary of the parameter space.

Tests for forecast encompassing establish a theoretical foundation for forecast combinations of two competing forecasts when both opposing hypotheses of forecast encompassing are rejected. This situation corresponds to the case when neither forecast encompasses its competitor. Generally, applying forecast combinations can be highly beneficial through the diversification gains stemming from combining different model specifications and underlying information sets. This benefit can be particularly pronounced for extreme risk measures such as the ES (Taylor, 2020), as the stand-alone models are very sensitive to the very little observations in the tails of the return distributions. Thus, combining forecasts can be seen as a robustification of the forecasts. Our application empirically validates this conjecture, especially pronounced for daily returns from the IBM stock.

While we propose encompassing tests suitable for the cases that the ES is reported jointly with and without its accompanying VaR forecast, testing encompassing for the VaR and ES jointly is most natural given their joint elicitability. Furthermore, this joint elicitability property clearly hints towards reporting ES forecasts jointly with their corresponding VaR forecasts by default. In contrast, the auxiliary test can be seen as the first forecast comparison procedure for the ES, which focuses almost entirely on the ES. Theoretically, its application is most reasonable in cases where the primary focus of the researcher is on the ES, even though VaR forecasts are available. Eventually, the strict ES test is specific to scenarios where only competing ES forecasts are available. Furthermore, in cases where competing ES forecasts are built on the same VaR model (forecasts), the strict ES test is applicable while the application of the joint and auxiliary tests is infeasible due to the collinearity of the (identical) VaR forecasts.

References

Andrews, D. W. K. (1999). “Estimation when a parameter is on a boundary”. *Econometrica* 67. (6), 1341–1383.

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). “Coherent measures of risk”. *Mathematical Finance* 9. (3), 203–228.

Barendse, S. (2020). “Efficiently weighted estimation of tail and interquartile expectations”. Working Paper, available at https://drive.google.com/file/d/1nI0QAWbM_VchAZDVg79p2vJcKoCrQB8o/view.

Basel Committee (2013). *Fundamental review of the trading book: A revised market risk framework*. Tech. rep. Available at http://www.bis.org/publ/bcbs265.pdf. Bank for International Settlements.

— (2016). *Minimum capital requirements for Market Risk*. Tech. rep. Available at http://www.bis.org/bcbs/publ/d352.pdf. Bank for International Settlements.

— (2017). *Pillar 3 disclosure requirements – consolidated and enhanced framework*. Tech. rep. Available at http://www.bis.org/bcbs/publ/d400.pdf. Basel Committee on Banking Supervision.

Bayer, S. and T. Dimitriadis (2020). "Regression based Expected Shortfall backtesting". *Journal of Financial Econometrics (forthcoming)*. available at https://arxiv.org/abs/1801.04112, doi: 10.1093/jjfinec/nbaa013.

Bollerslev, T. (1986). "Generalized autoregressive conditional heteroskedasticity". *Journal of Econometrics* 31. (3), 307–327.

Chong, Y. Y. and D. Hendry (1986). "Econometric evaluation of linear macroeconomic models". *Review of Economic Studies* 53. (4), 671–690.

Clark, T. E. and M. W. McCracken (2001). "Tests of equal forecast accuracy and encompassing for nested models". *Journal of Econometrics* 105. (1), 85–110.

Clements, M. and D. Harvey (2009). "Forecast combination and encompassing". In: *Palgrave Handbook of Econometrics: Volume 2: Applied Econometrics*. Ed. by Mills, T. C. and Patterson, K. London: Palgrave Macmillan UK, 169–198.

— (2010). "Forecast encompassing tests and probability forecasts". *Journal of Applied Econometrics* 25. (6), 1028–1062.

Cont, R., R. Deguest, and G. Scandolo (2010). "Robustness and sensitivity analysis of risk measurement procedures". *Quantitative Finance* 10. (6), 593–606.

Creal, D., S. J. Koopman, and A. Lucas (2013). "Generalized autoregressive score models with applications". *Journal of Applied Econometrics* 28. (5), 777–795.

Danielsson, J., P. Embrechts, C. Goodhart, C. Keating, F. Muennich, O. Renault, and H. S. Shin (2001). "An academic response to Basel II". FMG Special Papers. Financial Markets Group Special Papers, available at https://EconPapers.repec.org/RePEc:fmg:fmgsps:sp130.

Diebold, F. X. and R. S. Mariano (1995). "Comparing predictive accuracy". *Journal of Business & Economic Statistics* 13. (3), 253–63.

Diebold, F. X. (1989). "Forecast combination and encompassing: Reconciling two divergent literatures". *International Journal of Forecasting* 5. (4), 589 –592.

Dimitriadis, T. and S. Bayer (2019). "A joint quantile and expected shortfall regression framework". *Electron. J. Statist.* 13. (1), 1823–1871.

Dimitriadis, T. and J. Schnaitmann (2020). "Forecast Encompassing Tests for the Expected Shortfall". *International Journal of Forecasting (forthcoming)*.

Elliott, G., I. Komunjer, and A. Timmermann (2005). "Estimation and testing of forecast rationality under flexible loss". *The Review of Economic Studies* 72. (4), 1107–1125.

Embrechts, P., H. Liu, and R. Wang (2018). "Quantile-based risk sharing". *Operations Research* 66. (4), 936–949.

Engle, R. and S. Manganelli (2004). "CAViaR: Conditional autoregressive value at risk by regression quantiles". *Journal of Business & Economic Statistics* 22. (4), 367–381.

Ericsson, N. R. (1993). "On the limitations of comparing mean square forecast errors: Clarifications and extensions". *Journal of Forecasting* 12. (8), 644–651.

Fissler, T. and J. F. Ziegel (2016). "Higher order elicitability and Osband's principle". *Annals of Statistics* 44. (4), 1680–1707.

— (2019). "Elicitability of range value at risk". arXiv:1902.04489 [math.ST].

Fissler, T., J. F. Ziegel, and T. Gneiting (2016). "Expected Shortfall is jointly elicitable with Value at Risk - Implications for backtesting". *Risk* January, 58–61.

Giacomini, R. and I. Komunjer (2005). "Evaluation and combination of conditional quantile forecasts". *Journal of Business & Economic Statistics* 23. (4), 416–431.

Giacomini, R. and H. White (2006). "Tests of conditional predictive ability". *Econometrica* 74. (6), 1545–1578.

Gikhman, I. I. and A. V. Skorokhod (2004). *The theory of stochastic processes I*. Vol. 210. Classics in Mathematics. Springer Berlin Heidelberg.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). "On the relation between the expected value and the volatility of the nominal excess return on stocks". *The Journal of Finance* 48. (5), 1779–1801.

Gneiting, T. (2011). "Making and evaluating point forecasts". *Journal of the American Statistical Association* 106. (494), 746–762.

Halbleib, R. and W. Pohlmeier (2012). "Improving the value at risk forecasts: Theory and evidence from the financial crisis". *Journal of Economic Dynamics and Control* 36. (8), 1212–1228.

Hall, A. R. and A. Inoue (2003). "The large sample behaviour of the generalized method of moments estimator in misspecified models". *Journal of Econometrics* 114. (2), 361–394.

Hansen, B. E. and S. Lee (2019). "Inference for iterated GMM under misspecification". Working Paper, available at https://www.ssc.wisc.edu/~bhansen/papers/IteratedGMM.html.

Hansen, P. R., A. Lunde, and J. M. Nason (2011). "The model confidence set". *Econometrica* 79. (2), 453–497.

Harvey, D., S. Leybourne, and P. Newbold (1998). "Tests for forecast encompassing". *Journal of Business & Economic Statistics* 16. (2), 254–259.

Harvey, D. and P. Newbold (2000). "Tests for multiple forecast encompassing". *Journal of Applied Econometrics* 15. (5), 471–482.

Heinrich, C. (2014). "The mode functional is not elicitable". *Biometrika* 101. (1), 245–251.

Hendricks, W. and R. Koenker (1992). "Hierarchical spline models for conditional quantiles and the demand for electricity". *Journal of the American Statistical Association* 87. (417), 58–68.

Hendry, D. and J.-F. Richard (1982). "On the formulation of empirical models in dynamic econometrics". *Journal of Econometrics* 20. (1), 3–33.

Holzmann, H. and M. Eulert (Mar. 2014). "The role of the information set for forecasting with applications to risk management". *Ann. Appl. Stat.* 8. (1), 595–621.

Huber, P. J. (1967). "The behavior of maximum likelihood estimates under nonstandard conditions". In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press, 221–233.

Koenker, R. W. and G. Bassett (1978). "Regression quantiles". *Econometrica* 46. (1), 33–50.

Mizon, G. and J.-F. Richard (1986). "The encompassing principle and its application to testing non-nested hypotheses". *Econometrica* 54. (3), 657–78.

Newbold, P. and D. Harvey (2007). "Forecast combination and encompassing". In: *A Companion to Economic Forecasting*. John Wiley and Sons, Ltd. Chap. 12, 268–283.

Newey, W. and D. McFadden (1994). "Large sample estimation and hypothesis testing". In: *Handbook of Econometrics*. Ed. by Engle, R. and McFadden, D. Vol. 4. Elsevier. Chap. 36, 2111–2245.

Nolde, N. and J. F. Ziegel (Dec. 2017). "Elicitability and backtesting: Perspectives for banking regulation". *The Annals of Applied Statistics* 11. (4), 1833–1874.

Patton, A. J. (2011). "Data-based ranking of realised volatility estimators". *Journal of Econometrics* 161. (2), 284–303.
Patton, A. J. (2020). "Comparing possibly misspecified forecasts". *Journal of Business & Economic Statistics* 38. (4), 796–809.
Patton, A. J. and A. Timmermann (2007). "Testing forecast optimality under unknown loss". *Journal of the American Statistical Association* 102. (480), 1172–1184.
Patton, A. J., J. F. Ziegel, and R. Chen (2019). "Dynamic semiparametric models for expected shortfall (and Value-at-Risk)". *Journal of Econometrics* 211. (2), 388 –413.
Taylor, J. W. (2019). "Forecasting value at risk and Expected Shortfall using a semiparametric approach based on the asymmetric Laplace distribution". *Journal of Business & Economic Statistics* 37. (1), 121–133.
— (2020). "Forecast combinations for value at risk and expected shortfall". *International Journal of Forecasting* 36. (2), 428–441.
Timmermann, A. (2006). "Forecast combinations". In: *Handbook of Economic Forecasting*. Ed. by Elliott, G., Granger, C., and Timmermann, A. 1st ed. Vol. 1. Elsevier. Chap. 04, 135–196.
Weiss, A. A. (1991). "Estimating nonlinear dynamic models using least absolute error estimation". *Econometric Theory* 7. (1), 46–68.
West, K. (2006). "Forecast evaluation". In: *Handbook of Economic Forecasting*. Ed. by Elliott, G., Granger, C., and Timmermann, A. 1st ed. Vol. 1. Elsevier. Chap. 03, 99–134.
White, H. (1994). *Estimation, inference and specification analysis*. Econometric Society Monographs. Cambridge University Press.
— (2001). *Asymptotic theory for econometricians*. San Diego: Academic Press.

Appendix 2.A Proofs

Proof. (Proof of Proposition 2.2.1) We check that the necessary conditions (i) - (iv) of the basic consistency theorem, given in Theorem 2.1 in Newey and McFadden (1994), p. 2121 hold, where we consider the objective functions $Q_n(\theta)$ and $Q_n^0(\theta)$ as defined in (2.2.18) and (2.2.19). First, notice that condition (ii) holds by imposing condition (b). The unique identification condition (i) holds by assumption (c). Next, we verify the uniform convergence condition (iv) by applying the uniform weak law of large numbers given in Theorem A.2.5. in White (1994). For that, we have to show that

1. the map $\theta \mapsto \rho\big(Y_{t+1}, g_t^q(\beta), g_t^e(\eta)\big)$ is Lipschitz-L_1 on Θ,[12]

[12]See Definition A.2.3 in White (1994) for a definition of Lipschitz-L_1. Notice that we do not have a double index and thus we suppress the n in the notation of White (1994). Furthermore, we apply the definition by using the identify function for a_t^o.

2. For all $\theta^o \in \Theta$, there exists $\delta^o > 0$, such that for all $\delta, 0 < \delta \leq \delta^o$, the sequences

$$\bar{\rho}_t(\theta^o, \delta) := \sup_{\theta \in \Theta} \left\{ \rho\big(Y_{t+1}, g_t^q(\beta), g_t^e(\eta)\big) \middle| \, ||\theta - \theta^o|| < \delta \right\} \quad \text{and} \tag{2.A.1}$$

$$\underline{\rho}_t(\theta^o, \delta) := \inf_{\theta \in \Theta} \left\{ \rho\big(Y_{t+1}, g_t^q(\beta), g_t^e(\eta)\big) \middle| \, ||\theta - \theta^o|| < \delta \right\} \tag{2.A.2}$$

obey a weak law of large numbers.

Condition 1 follows directly from Lemma 2.C.1 and we turn to condition 2. As the process Z_t is strong mixing of size $-r/(r-2)$ for some $r > 2$ by condition (a) and as the functions $\rho\big(Y_{t+1}, g_t^q(\beta), g_t^e(\eta)\big)$ and the supremum/infimum functions are $\mathcal{F}_t$-measurable for all $t \in \mathbb{N}$, we can conclude that the sequences $\bar{\rho}_t(\theta^o, \delta)$ and $\underline{\rho}_t(\theta^o, \delta)$ are also strong mixing of the same size by applying the same theorem.

Furthermore, for $\tilde{r} > 1$ and for some $\delta > 0$ sufficiently small, $r \geq \tilde{r} + \delta$ and thus $\mathbb{E}\left[|\bar{\rho}_t(\theta^o, \delta)|^{\tilde{r}+\delta}\right] \leq \sup_{1 \leq t \leq T} \mathbb{E}\left[\sup_{\theta \in \Theta} \left|\rho\big(Y_{t+1}, g_t^q(\beta), g_t^e(\eta)\big)\right|^r\right]$ for all $t, 1 \leq t \leq T, T \geq 1$. As Θ is compact, there exists some $c > 0$ such that $\sup_{\theta \in \Theta} ||\theta|| \leq c$ and thus, for all $t = 1, \ldots, T$, it holds that

$$\mathbb{E}\left[\sup_{\theta \in \Theta} \left|\rho\big(Y_{t+1}, g_t^q(\beta), g_t^e(\eta)\big)\right|^r\right] \tag{2.A.3}$$

$$\leq 4^{r-1} \left\{ 1 + \left(\frac{c}{K} \left(1 + \frac{1}{\alpha} \right) \right) \mathbb{E}||g_t^q(\beta)||^r + \frac{1}{\alpha K} \mathbb{E}|Y_{t+1}|^r + \sup_{\theta \in \Theta} \mathbb{E}||\log(g_t^e(\eta))||^r \right\}, \tag{2.A.4}$$

which is bounded by condition (h) and as $\log(z) \leq z$ for z sufficiently large. The same inequality holds for $|\underline{\rho}_t(\theta^o, \delta)|$. Thus, we can apply the weak law of large numbers for strong mixing sequences in Corollary 3.48 in White (2001), p. 49 in order to conclude that for all $\theta^o \in \Theta$ such that $||\theta^o - \theta|| \leq \delta$, it holds that $\frac{1}{n} \sum_{t=m}^{T-1} \left(\bar{\rho}_t(\theta^o, \delta) - \mathbb{E}\left[\bar{\rho}_t(\theta^o, \delta)\right] \right) \xrightarrow{\mathbb{P}} 0$ and $\frac{1}{n} \sum_{t=m}^{T-1} \left(\underline{\rho}_t(\theta^o, \delta) - \mathbb{E}\left[\underline{\rho}_t(\theta^o, \delta)\right] \right) \xrightarrow{\mathbb{P}} 0$, which shows condition 2. Consequently, the uniform convergence condition (iv) holds by applying the uniform weak law of large numbers given in Theorem A.2.5. in White (1994).

As we have shown that the map $\theta \mapsto \rho\big(Y_{t+1}, g_t^q(\beta), g_t^e(\eta)\big)$ is Lipschitz-L_1 in Lemma 2.C.1, the map $\theta \mapsto Q_n^0 = \frac{1}{n} \sum_{t=m}^{T-1} \mathbb{E}\left[\rho\big(Y_{t+1}, g_t^q(\beta), g_t^e(\eta)\big)\right]$ is also continuous which shows condition (iii). Thus, we can apply Theorem 2.1. of Newey and McFadden (1994) which concludes the proof of this proposition. □

Proof. (Proof of Proposition 2.2.2) We define $\Psi_n(\theta) = \frac{1}{n} \sum_{t=m}^{T-1} \psi\big(Y_{t+1}, g_t^q(\beta), g_t^e(\eta)\big)$ and $\Psi_n^0(\theta) = \mathbb{E}[\Psi_n(\theta)]$. From the proof of Lemma 2.C.2, we get the mean value expansion (for $\hat{\theta}_n$ close to θ_n^*),

$$\Psi_n^0(\hat{\theta}_n) - \Psi_n^0(\theta_n^*) = \Delta_n(\tilde{\theta}_1, \ldots, \tilde{\theta}_k)\big(\hat{\theta}_n - \theta_n^*\big), \tag{2.A.5}$$

for (possibly different) values $\tilde{\theta}_1, \ldots, \tilde{\theta}_k$ somewhere on the line between $\hat{\theta}_n$ and θ_n^*, where the components of $\Delta_n(\tilde{\theta}_1, \ldots, \tilde{\theta}_k)$ are given in Lemma 2.C.2, and where $\Psi_n^0(\theta_n^*) = 0$.[13]

Furthermore, it holds that $\Delta_n(\theta_n^*, \ldots, \theta_n^*) = \Lambda_n(\theta_n^*)$ and $\Delta_n(\tilde{\theta}_1, \ldots, \tilde{\theta}_k)$ is a continuous function in its arguments $\tilde{\theta}_1, \ldots, \tilde{\theta}_k$. Using that $\Lambda_n(\theta_n^*)$ has Eigenvalues bounded away from zero (for n large enough), we also get that $\Delta_n(\tilde{\theta}_1, \ldots, \tilde{\theta}_k)$ is non-singular in a neighborhood around θ_n^* (for all arguments) for n large enough as the map which maps the matrix onto its Eigenvalues is continuous. As we further know that $\hat{\theta}_n - \theta_n^* \xrightarrow{\mathbb{P}} 0$ and $||\tilde{\theta}_j - \theta_n^*|| \leq ||\hat{\theta}_n - \theta_n^*||$ for all $j = 1, \ldots, k$, we get from the continuous mapping theorem that

$$\Delta_n^{-1}(\tilde{\theta}_1, \ldots, \tilde{\theta}_k) - \Lambda_n^{-1}(\theta_n^*) \xrightarrow{\mathbb{P}} 0. \tag{2.A.6}$$

In the following, we apply Lemma A.1 in Weiss (1991) (by verifying its assumptions), which extends the iid results of Huber (1967) to strong mixing sequences. Assumption (N1) of Lemma A.1 in Weiss (1991) is satisfied as every almost surely continuous stochastic process is separable in the sense of Doob (Gikhman and Skorokhod, 2004) and the functions $\psi\big(Y_{t+1}, g_t^q(\beta), g_t^e(\eta)\big)$ are almost surely continuous for all $t \in \mathbb{N}$. Assumption (N2) is satisfied as shown in the proof of Proposition 2.2.1. Assumption (N3)(i) is shown in Lemma 2.C.2. The technical Assumptions (N3)(ii) and (N3)(iii) follow from Lemma 4 and Lemma 5 in the supplemental appendix of Patton et al. (2019). For this, notice that the moment conditions in Assumption 2 (C) and (D) of Patton et al. (2019) are implied by the condition (h) in Assumption 2.2.1. Assumption (N4) follows from the moment conditions (h) in Assumption 2.2.1 and Assumption (N5) from the strong mixing condition (a). Furthermore, Lemma 2 in the supplemental appendix of Patton et al. (2019) implies that $\sqrt{n}\Psi_n(\hat{\theta}_n) \xrightarrow{\mathbb{P}} 0$. Thus, we can apply Lemma A.1 in Weiss (1991) and get that

$$\sqrt{n}\Psi_n^0(\hat{\theta}_n) - \sqrt{n}\Psi_n(\theta_n^*) \xrightarrow{\mathbb{P}} 0. \tag{2.A.7}$$

Combining (2.A.5), (2.A.6) and (2.A.7), we get that

$$\sqrt{n}\big(\hat{\theta}_n - \theta_n^*\big) = -\Delta_n(\tilde{\theta}_1, \ldots, \tilde{\theta}_k)^{-1}\sqrt{n}\,\Psi_n^0(\hat{\theta}_n) \tag{2.A.8}$$

$$= -\Big(\Lambda_n^{-1}(\theta_n^*) + o_p(1)\Big) \cdot \Big(\sqrt{n}\Psi_n(\theta_n^*) + o_p(1)\Big) = -\Lambda_n^{-1}(\theta_n^*) \cdot \sqrt{n}\Psi_n(\theta_n^*) + o_p(1). \tag{2.A.9}$$

Furthermore, $\Sigma_n^{-1/2}(\theta_n^*)\sqrt{n}\Psi_n(\theta_n^*) \xrightarrow{d} \mathcal{N}\big(0, I_k\big)$ by Lemma 2.C.3 and thus, $\Sigma_n^{-1/2}(\theta_n^*)\Lambda_n(\theta_n^*)\,\sqrt{n}\big(\hat{\theta}_n - \theta_n^*\big) \xrightarrow{d} \mathcal{N}\big(0, I_k\big)$, which concludes the proof of this proposition. □

[13] The mean-value theorem cannot be generalized in a straight-forward fashion to vector-valued functions. Thus, we have to consider the mean value expansion in each component separately which gives this more complicated expression.

Proof. (Proof of Theorem 2.2.2) We first notice that

$$\widehat{\Omega}_n^{-1/2}\sqrt{n}\big(\hat{\theta}_n - \theta_n^*\big) = \Omega_n^{-1/2}\sqrt{n}\big(\hat{\theta}_n - \theta_n^*\big) + \big(\widehat{\Omega}_n^{-1/2} - \Omega_n^{-1/2}\big)\sqrt{n}\big(\hat{\theta}_n - \theta_n^*\big). \tag{2.A.10}$$

From Proposition 2.2.2, we obtain that $\Omega_n^{-1/2}\sqrt{n}\big(\hat{\theta}_n - \theta_n^*\big) \xrightarrow{d} \mathcal{N}\big(0, I_k\big)$. Furthermore, as $\big(\widehat{\Omega}_n^{-1/2} - \Omega_n^{-1/2}\big) = o_P(1)$ by assumption, we apply Slutzky's theorem in order to get that $\big(\widehat{\Omega}_n^{-1/2} - \Omega_n^{-1/2}\big)\sqrt{n}\big(\hat{\theta}_n - \theta_n^*\big) = o_P(1)$. Thus, $\widehat{\Omega}_n^{-1/2}\sqrt{n}\big(\hat{\theta}_n - \theta_n^*\big) \xrightarrow{d} \mathcal{N}\left(0, I_k\right)$ and the result based on the selection matrices R follows by applying the continuous mapping theorem, which concludes the proof of this theorem. □

Appendix 2.B Additional data generating processes

In the following, we describe two additional data generating processes (DGPs) used in the extensions of the simulation study in Section 2.3.3.

The GAS DGP

We introduce two specifications of the GAS models (Creal et al., 2013), where the second candidate potentially generates model misspecification in the strict ES encompassing test. For this, we generate $\tilde{Y}_{1,t+1}$, $\hat{q}_{1,t}$ and $\hat{e}_{1,t}$ from a GAS model with Gaussian innovations, which corresponds to the standard GARCH specification given in (2.3.2). We obtain the second sequence of forecasts from a GAS model with Student-t residuals with time-varying variance and degrees of freedom, given by

$$(\hat{\mu}_2, \hat{\sigma}_{2,t}^2, \hat{\nu}_{2,t})^\top = \kappa + B \cdot (\hat{\mu}_2, \hat{\sigma}_{2,t-1}^2, \hat{\nu}_{2,t-1})^\top + AH_t\nabla_t, \tag{2.B.1}$$

where $H_t\nabla_t$ is the forcing variable of the model, the scaling matrix H_t is the Hessian and ∇_t the derivative of the log-likelihood function. We calibrate both models to daily IBM returns resulting in the parameter values $\kappa = (0.0659, 0.00599, -1.737)$, $A = \text{diag}(0, 0.146, 7.563)$ and $B = \text{diag}(0, 0.994, 7.381)$. This model implies that $\tilde{Y}_{2,t+1} \sim t_{\hat{\nu}_{2,t}}\big(\hat{\mu}_2, \hat{\sigma}_{2,t}^2\big)$ and we obtain the VaR and ES forecasts from this t-distribution. In order to simulate returns which follow a convex combination of these two distributions, we simulate Bernoulli draws $\pi_{t+1} \sim \text{Bern}(\pi)$ and let $Y_{t+1} = (1 - \pi_{t+1})\tilde{Y}_{1,t+1} + \pi_{t+1}\tilde{Y}_{2,t+1}$, as for the VaR/ES-GAS models in Section 2.3.1.

The VaR/ES CAViaR DGP

This simulation setup follows the dynamic ES models of Taylor (2019), which we denote by ES-CAViaR as they augment the CAViaR models of Engle and Manganelli (2004) with a dynamic ES specification. The asymmetric slope AS-ES-CAViaR model is given by

$$\hat{q}_{1,t} = -0.0003 - 0.05|\tilde{Y}_{1,t}|\mathbb{1}_{\{\tilde{Y}_{1,t}\geq 0\}} - 0.15|\tilde{Y}_{1,t}|\mathbb{1}_{\{\tilde{Y}_{1,t}<0\}} + 0.8\hat{q}_{1,t-1}, \qquad \text{and} \tag{2.B.2}$$

$$\hat{e}_{1,t} = \hat{q}_{1,t} - x_t, \qquad \text{where} \tag{2.B.3}$$

$$x_t = \begin{cases} 0.00017 + 0.125(\hat{q}_{1,t-1} - \tilde{Y}_{1,t}) + 0.84\hat{q}_{1,t-1} & \text{if } \hat{q}_{1,t-1} \leq \tilde{Y}_{1,t}, \\ x_{t-1} & \text{if } \hat{q}_{1,t-1} > \tilde{Y}_{1,t}. \end{cases} \tag{2.B.4}$$

The second model variant we consider is the symmetric absolute value SAV-ES-CAViaR model, where the quantile equation is given by

$$\hat{q}_{2,t} = -0.0003 - 0.1|\tilde{Y}_{2,t}| + 0.8\hat{q}_{2,t-1}, \tag{2.B.5}$$

and $\hat{e}_{2,t}$ and x_t follow the dynamic specifications in (2.B.3) and (2.B.4). These parameter choices are slightly modified values of the ones obtained by Taylor (2019). In this setup, we simulate data according to the additive model $Y_{t+1} = \big((1-\pi)\hat{e}_{1,t} + \pi\hat{e}_{2,t}\big) + \varepsilon_{t+1}$, where $\varepsilon_{t+1} \sim \mathcal{N}\big(-\sigma\xi_\alpha, \sigma^2\big)$, for $\sigma = 0.1$. This implies that for $\pi = 0$, $ES_\alpha(Y_{t+1}|\mathcal{F}_t) = \hat{e}_{1,t}$ almost surely, and the same holds inversely for $\pi = 1$. This setup generalizes the CAViaR DGP used in the simulations for the VaR encompassing test of Giacomini and Komunjer (2005) to the ES.

Appendix 2.C Technical proofs

Lemma 2.C.1. Given the conditions from Assumption 2.2.1, the function $\rho\big(Y_{t+1}, g_t^q(\beta), g_t^e(\eta)\big)$ is Lipschitz-L_1 on Θ with $\mathcal{F}_t$-measurable and integrable Lipschitz-constant.

Proof. We split the ρ-function $\rho\big(Y_{t+1}, g_t^q(\beta), g_t^e(\eta)\big) = \rho_1\big(Y_{t+1}, g_t^q(\beta), g_t^e(\eta)\big) + \rho_2\big(Y_{t+1}, g_t^q(\beta), g_t^e(\eta)\big)$, where

$$\rho_1\big(Y_{t+1}, g_t^q(\beta), g_t^e(\eta)\big) = -\mathbb{1}_{\{Y_{t+1}\leq g_t^q(\beta)\}}\frac{1}{\alpha g_t^e(\eta)}(g_t^q(\beta) - Y_{t+1}),$$

$$\rho_2\big(Y_{t+1}, g_t^q(\beta), g_t^e(\eta)\big) = \frac{g_t^q(\beta) - g_t^e(\eta)}{g_t^e(\eta)} - \log(-g_t^e(\eta)).$$

Local Lipschitz continuity of ρ_2 follows since it is a continuously differentiable function in θ (such that $g_t^e(\eta) \neq 0$) and thus (locally) Lipschitz-L_1. We consequently get that for all

$\theta^o \in \Theta$, there exists a $\delta^o > 0$ such that for all $\theta \in U_{\delta^o}(\theta^o) := \left\{\theta \in \Theta \middle| ||\theta - \theta^o|| \leq \delta^o\right\}$, it holds that

$$\begin{aligned}&\left|\rho_2\left(Y_{t+1}, g_t^q(\beta^o), g_t^e(\eta^o)\right) - \rho_2\left(Y_{t+1}, g_t^q(\beta), g_t^e(\eta)\right)\right| \\ &\leq \left\|\theta - \theta^o\right\| \cdot \sup_{\theta \in U_{\delta^o}(\theta^o)} \left(\left\| \frac{\nabla_\beta g_t^q(\beta) + \nabla_\eta g_t^e(\eta)}{g_t^e(\eta)} \right\| + \left\| \frac{g_t^q(\beta) \nabla_\eta g_t^e(\eta)}{(g_t^e(\eta))^2} \right\| \right),\end{aligned} \tag{2.C.1}$$

where the sequences $\frac{1}{n}\sum_{t=m}^{T-1} \mathbb{E}\left[\left\|\frac{\nabla_\beta g_t^q(\beta)+\nabla_\eta g_t^e(\eta)}{g_t^e(\eta)}\right\|\right]$ and $\frac{1}{n}\sum_{t=m}^{T-1} \mathbb{E}\left[\left\|\frac{g_t^q(\beta)\nabla_\eta g_t^e(\eta)}{(g_t^e(\eta))^2}\right\|\right]$ are bounded for all $\theta^o \in \Theta$ by the conditions (h) in Assumption 2.2.1.

For the function ρ_1, we consider the following four cases. First, let $\Gamma_1 = \left\{\omega \in \Omega, \theta \in U_{\delta^o}(\theta^o) \middle| g_t^q(\beta^o)(\omega) < Y_{t+1}(\omega) \text{ and } g_t^q(\beta)(\omega) < Y_{t+1}(\omega)\right\}$. Then, on Γ_1, it holds that,

$$\rho_1\left(Y_{t+1}, g_t^q(\beta), g_t^e(\eta)\right) = \rho_1\left(Y_{t+1}, g_t^q(\beta^o), g_t^e(\eta^o)\right) = 0, \tag{2.C.2}$$

which is Lipschitz-L_1.

Second, let $\Gamma_2 = \left\{\omega \in \Omega, \theta \in U_{\delta^o}(\theta^o) \middle| g_t^q(\beta^o)(\omega) \geq Y_{t+1}(\omega) \text{ and } g_t^q(\beta)(\omega) \geq Y_{t+1}(\omega)\right\}$. On Γ_2, for both $\tilde{\theta} \in \{\theta, \theta^o\}$, it holds that

$$\rho_1\left(Y_{t+1}, g_t^q(\tilde{\beta}), g_t^e(\tilde{\eta})\right) = -\frac{1}{\alpha g_t^e(\tilde{\eta})}\left(g_t^q(\tilde{\beta}) - Y_{t+1}\right), \tag{2.C.3}$$

which is a continuously differentiable function. Thus,

$$\begin{aligned}&\left|\rho_1\left(Y_{t+1}, g_t^q(\beta^o), g_t^e(\eta^o)\right) - \rho_1\left(Y_{t+1}, g_t^q(\beta), g_t^e(\eta)\right)\right| \\ &\leq \left\|\theta^o - \theta\right\| \cdot \left(\sup_{\theta \in U_{\delta^o}(\theta^o)} \left\| \frac{\nabla_\beta g_t^q(\beta)}{\alpha g_t^e(\eta)} \right\| + \sup_{\theta \in U_{\delta^o}(\theta^o)} \left\| \frac{\nabla_\eta g_t^e(\eta)}{\alpha (g_t^e(\eta))^2} (g_t^q(\beta) - Y_{t+1}) \right\| \right),\end{aligned} \tag{2.C.4}$$

where the average of the expectations of the suprema sequences in the last two lines are bounded by the conditions (h) in Assumption 2.2.1.

Finally, let $\Gamma_3 = \left\{\omega \in \Omega, \theta \in U_{\delta^o}(\theta^o) \middle| g_t^q(\beta)(\omega) < Y_{t+1}(\omega) \leq g_t^q(\beta^o)(\omega)\right\}$. As on Γ_3, $|g_t^q(\beta^o) - Y_{t+1}| \leq |g_t^q(\beta^o) - g_t^q(\beta)|$ almost surely, it holds that

$$\begin{aligned}&\left|\rho_1\left(Y_{t+1}, g_t^q(\beta^o), g_t^e(\eta^o)\right) - \rho_1\left(Y_{t+1}, g_t^q(\beta), g_t^e(\eta)\right)\right| = \left| \frac{1}{\alpha g_t^e(\eta^o)} (g_t^q(\beta^o) - Y_{t+1}) \right| \\ &\leq \left| \frac{1}{\alpha g_t^e(\eta^o)} (g_t^q(\beta^o) - g_t^q(\beta)) \right| \leq \left\|\theta - \theta^o\right\| \cdot \sup_{\theta \in U_{\delta^o}(\theta^o)} \left\| \frac{\nabla_\beta g_t^q(\beta)}{\alpha g_t^e(\eta)} \right\|.\end{aligned}$$

Equivalently as above, the average of the expectations of the suprema sequences in the last two lines are bounded by the conditions (h) in Assumption 2.2.1. An equivalent argument

holds for $\Gamma_4 = \left\{\omega \in \Omega, \theta \in U_{\delta^o}(\theta^o) \,\middle|\, g_t^q(\beta^o)(\omega) < Y_{t+1}(\omega) \leq g_t^q(\beta)(\omega)\right\}$. As $\Omega = \bigcup_{i=1}^4 \Gamma_i$, we can conclude that the function $\rho_1\left(Y_{t+1}, g_t^q(\beta^o), g_t^e(\eta^o)\right)$ is Lipschitz-L_1 on Θ. □

Lemma 2.C.2. Given the conditions from Assumption 2.2.1, there exist constants $a, d_0 > 0$ such that

$$\left\|\Psi_n^0(\theta)\right\| \geq a||\theta - \theta_n^*|| \qquad \text{for any } \theta \in \Theta \text{ such that } ||\theta - \theta_n^*|| \leq d_0, \tag{2.C.5}$$

and for all $n \geq n_0$, where $n_0 \in \mathbb{N}$ is large enough.

Proof. Let $\theta \in \Theta$ such that $||\theta - \theta_n^*|| \leq d_0$ for some (small) constant $d_0 > 0$ and define

$$\Psi_{n,q}^0(\theta) = \frac{1}{n}\sum_{t=m}^{T-1} \mathbb{E}\left[-\frac{\nabla_\beta g_t^q(\beta)}{\alpha g_t^e(\eta)}\left(F_t(g_t^q(\beta)) - \alpha\right)\right] \qquad \text{and} \tag{2.C.6}$$

$$\Psi_{n,e}^0(\theta) = \frac{1}{n}\sum_{t=m}^{T-1} \mathbb{E}\left[\frac{\nabla_\eta g_t^e(\eta)}{(g_t^e(\eta))^2}\left(g_t^e(\eta) - g_t^q(\beta) + \frac{1}{\alpha}(g_t^q(\beta) - Y_{t+1})\mathbb{1}_{\{Y_{t+1} \leq g_t^q(\beta)\}}\right)\right], \tag{2.C.7}$$

such that $\Psi_n^0(\theta)^\top = \left(\Psi_{n,q}^0(\theta)^\top, \Psi_{n,e}^0(\theta)^\top\right)$. Henceforth, we use the following short notations

$$G_t^q(\beta) = \nabla_\beta g_t^q(\beta)\nabla_\eta g_t^q(\beta)^\top \tag{2.C.8}$$

$$G_t^{qe}(\beta, \eta) = \nabla_\beta g_t^q(\beta)\nabla_\eta g_t^e(\eta)^\top \tag{2.C.9}$$

$$G_t^{eq}(\beta, \eta) = \nabla_\eta g_t^e(\eta)\nabla_\beta g_t^q(\beta)^\top \tag{2.C.10}$$

$$G_t^e(\eta) = \nabla_\eta g_t^e(\eta)\nabla_\eta g_t^e(\eta)^\top, \tag{2.C.11}$$

$H_t^q(\beta)$ is the $k_\beta \times k_\beta$ Hessian matrix of $g_t^q(\beta)$ and equivalently, $H_t^e(\eta)$ is the $k_\eta \times k_\eta$ Hessian matrix of $g_t^e(\eta)$.

In the following, we apply the mean-value theorem to the individual rows of $\Psi_n^0(\theta)$ instead of to the complete vector, as the mean-value theorem cannot be generalized directly to vector-valued functions. Then, by applying the mean-value theorem to the j-th row of of $\Psi_n^0(\theta)$ for all $j = 1, \ldots, k$, we get that

$$\Psi_n^0(\theta) - \Psi_n^0(\theta_n^*) = \Delta_n(\tilde{\theta}_1, \ldots, \tilde{\theta}_k) \cdot \left(\theta - \theta_n^*\right), \tag{2.C.12}$$

where

$$\Delta_n(\tilde{\theta}_1, \ldots, \tilde{\theta}_k) = \begin{pmatrix} \Delta_{n,qq} & \Delta_{n,qe} \\ \Delta_{n,eq} & \Delta_{n,ee} \end{pmatrix}. \tag{2.C.13}$$

For all $j = 1, \dots, k_\beta$, the j-th row of $\Delta_{n,qq}$ is given by

$$\Delta_{n,qq,j}(\tilde{\beta}_j) = \frac{1}{n}\sum_{t=m}^{T-1}\mathbb{E}\left[\frac{H_{t,j}^q(\tilde{\beta}_j)}{\alpha g_t^e(\tilde{\eta}_j)}\big(F_t(g_t^q(\tilde{\beta}_j)) - \alpha\big) + \frac{G_t^q(\tilde{\beta}_j)}{\alpha g_t^e(\tilde{\eta}_j)} h_t(g_t^q(\tilde{\beta}_j))\right], \tag{2.C.14}$$

where $H_{t,j}^q(\tilde{\beta}_j)$ denotes the j-th row of $H_t^q(\tilde{\beta}_j)$, and the j-th row of $\Delta_{n,qe}$ is given by

$$\Delta_{n,qe,j}(\tilde{\theta}_j) = \frac{1}{n}\sum_{t=m}^{T-1}\mathbb{E}\left[-\frac{G_{t,j}^{qe}(\tilde{\beta}_j, \tilde{\eta}_j)}{\alpha g_t^e(\tilde{\eta}_j)^2}\big(F_t(g_t^q(\tilde{\beta}_j)) - \alpha\big)\right]. \tag{2.C.15}$$

For all $j = k_\beta + 1, \dots, k_\beta + k_\eta$, the j-th row of $\Delta_{n,eq}$ is given by

$$\Delta_{n,eq,j}(\tilde{\theta}_j) = \frac{1}{n}\sum_{t=m}^{T-1}\mathbb{E}\left[\frac{G_{t,j}^{eq}(\tilde{\beta}_j, \tilde{\eta}_j)}{\alpha g_t^e(\tilde{\eta}_j)^2}\big(F_t(g_t^q(\tilde{\beta}_j)) - \alpha\big)\right] \tag{2.C.16}$$

and the j-th row of $\Delta_{n,ee}$ is given by

$$\begin{aligned}\Delta_{n,ee,j}(\tilde{\theta}_j) = \frac{1}{n}\sum_{t=m}^{T-1}\mathbb{E}\Bigg[&\frac{H_{t,j}^e(\tilde{\eta}_j)}{g_t^e(\tilde{\eta}_j)^2}\left(g_t^e(\tilde{\eta}_j) - g_t^q(\tilde{\beta}_j) + \frac{1}{\alpha}(g_t^q(\tilde{\beta}_j) - Y_{t+1})\mathbb{1}_{\{Y_{t+1}\le g_t^q(\tilde{\beta}_j)\}}\right)\\ &+\frac{G_{t,j}^{ee}(\tilde{\eta}_j)}{g_t^e(\tilde{\eta}_j)^2} - 2\frac{G_{t,j}^{ee}(\tilde{\eta}_j)}{g_t^e(\tilde{\eta}_j)^3}\left(g_t^e(\tilde{\eta}_j) - g_t^q(\tilde{\beta}_j) + \frac{1}{\alpha}(g_t^q(\tilde{\beta}_j) - Y_{t+1})\mathbb{1}_{\{Y_{t+1}\le g_t^q(\tilde{\beta}_j)\}}\right)\Bigg].\end{aligned}$$

In the following, we show that $\left\|\Delta_n\big(\tilde{\theta}_1, \dots, \tilde{\theta}_k\big) - \Lambda_n(\theta_n^*)\right\| \le c_1||\theta - \theta_n^*||$ by considering the individual components again. For each $j, i = 1, \dots, n_\beta$, (corresponding to the upper-left quantile-specific part of the Hessian matrix)

$$\begin{aligned}&||\Delta_{n,ji}\big(\tilde{\theta}_j\big) - \Lambda_{n,ji}(\theta_n^*)||\\ =&\left|\frac{1}{n}\sum_{t=m}^{T-1}\mathbb{E}\left[\frac{H_{t,ji}^q(\tilde{\beta}_j)}{\alpha g_t^e(\tilde{\eta}_j)}\big(F_t(g_t^q(\tilde{\beta}_j)) - \alpha\big) + \frac{G_{t,ji}^q(\tilde{\beta}_j)}{\alpha g_t^e(\tilde{\eta}_j)} h_t(g_t^q(\tilde{\beta}_j))\right]\right.\\ &\left. - \frac{1}{n}\sum_{t=m}^{T-1}\mathbb{E}\left[\frac{H_{t,ji}^q(\beta_n^*)}{\alpha g_t^e(\eta_n^*)}\big(F_t(g_t^q(\beta_n^*)) - \alpha\big) + \frac{G_{t,ji}^q(\beta_n^*)}{\alpha g_t^e(\eta_n^*)} h_t(g_t^q(\beta_n^*))\right]\right|\\ =&\left\|\frac{1}{n}\sum_{t=m}^{T-1}\mathbb{E}\left[\frac{\nabla H_{t,ji}^q(\bar{\beta}_j)}{\alpha g_t^e(\bar{\eta}_j)}\big(F_t(g_t^q(\bar{\beta}_j)) - \alpha\big) - \nabla g_t^e(\bar{\eta}_j)\frac{1}{\alpha g_t^e(\bar{\eta}_j)^2} H_{t,ji}^q(\bar{\beta}_j)\big(F_t(g_t^q(\bar{\beta}_j)) - \alpha\big)\right.\right.\\ &\left.\left. + \nabla_\beta g_t^q(\bar{\beta}_j)\frac{H_{t,ji}^q(\bar{\beta}_j)}{\alpha g_t^e(\bar{\eta}_j)} h_t(g_t^q(\bar{\beta}_j))\right]\right\| \cdot \left\|\tilde{\theta}_j - \theta_n^*\right\|,\end{aligned}$$

for some $\bar{\theta}_j = (\bar{\beta}_j, \bar{\eta}_j)$ on the line between $\tilde{\theta}_j$ and θ_n^*. Furthermore, for all $j = 1, \ldots, n_\beta$ and $i = n_\beta + 1, \ldots, n_\beta + n_\eta$ (corresponding to the upper-right quantile/ES-specific part of the Hessian matrix), it holds that

$$
\begin{aligned}
&||\Delta_{n,ji}(\tilde{\theta}_j) - \Lambda_{n,ji}(\theta_n^*)|| \\
&= \left| \frac{1}{n} \sum_{t=m}^{T-1} \mathbb{E} \left[\frac{G_{t,ji}^{qe}(\tilde{\beta}_j, \tilde{\eta}_j)}{\alpha g_t^e(\tilde{\eta}_j)^2} \left(F_t(g_t^q(\tilde{\beta}_j)) - \alpha \right) \right] - \frac{1}{n} \sum_{t=m}^{T-1} \mathbb{E} \left[\frac{G_{t,ji}^{qe}(\beta_n^*, \eta_n^*)}{\alpha g_t^e(\eta_n^*)^2} \left(F_t(g_t^q(\beta_n^*)) - \alpha \right) \right] \right| \\
&= \left\| \frac{1}{n} \sum_{t=m}^{T-1} \mathbb{E} \left[\frac{\nabla G_{t,ji}^{qe}(\bar{\beta}_j, \bar{\eta}_j)}{\alpha g_t^e(\bar{\eta}_j)^2} \left(F_t(g_t^q(\bar{\beta}_j)) - \alpha \right) + \nabla g_t^q(\bar{\beta}_j) \frac{G_{t,ji}^{qe}(\bar{\beta}_j, \bar{\eta}_j)}{\alpha g_t^e(\bar{\eta}_j)^2} h_t(g_t^q(\bar{\beta}_j)) \right. \right. \\
&\qquad \left. \left. - 2 \ \nabla g_t^e(\bar{\eta}_j) \frac{G_{t,ji}^{qe}(\bar{\beta}_j, \bar{\eta}_j)}{\alpha g_t^e(\bar{\eta}_j)^3} \left(F_t(g_t^q(\bar{\beta}_j)) - \alpha \right) \right] \right\| \cdot \left\| \tilde{\theta}_j - \theta_n^* \right\|,
\end{aligned}
$$

for some $\bar{\theta}_j = (\bar{\beta}_j, \bar{\eta}_j)$ on the line between $\tilde{\theta}_j$ and θ_n^*. This holds equivalently for the lower-left block of Δ_n and Λ_n. Eventually for the lower-right block, i.e. for each $j, i = n_\beta + 1, \ldots, n_\beta + n_\eta$, we get that

$$
\begin{aligned}
&\left| \Delta_{n,ji}(\tilde{\theta}_j) - \Lambda_{n,ji}(\theta_n^*) \right| \\
&= \left| \frac{1}{n} \sum_{t=m}^{T-1} \mathbb{E} \left[\frac{H_{t,ji}^e(\tilde{\eta}_j)}{(g_t^e(\tilde{\eta}_j))^2} \left(g_t^e(\tilde{\eta}_j) - g_t^q(\tilde{\beta}_j) + \frac{1}{\alpha} (g_t^q(\tilde{\beta}_j) - Y_{t+1}) \mathbb{1}_{\{Y_{t+1} \leq g_t^q(\tilde{\beta}_j)\}} \right) \right. \right. \\
&\qquad \left. + \frac{G_{t,ji}^{ee}(\tilde{\eta}_j)}{(g_t^e(\tilde{\eta}_j))^2} - 2 \frac{G_{t,ji}^{ee}(\tilde{\eta}_j)}{(g_t^e(\tilde{\eta}_j))^3} \left(g_t^e(\tilde{\eta}_j) - g_t^q(\tilde{\beta}_j) + \frac{1}{\alpha} (g_t^q(\tilde{\beta}_j) - Y_{t+1}) \mathbb{1}_{\{Y_{t+1} \leq g_t^q(\tilde{\beta}_j)\}} \right) \right] \\
&\quad - \frac{1}{n} \sum_{t=m}^{T-1} \mathbb{E} \left[\frac{H_{t,ji}^e(\eta_n^*)}{(g_t^e(\eta_n^*))^2} \left(g_t^e(\eta_n^*) - g_t^q(\beta_n^*) + \frac{1}{\alpha} (g_t^q(\beta_n^*) - Y_{t+1}) \mathbb{1}_{\{Y_{t+1} \leq g_t^q(\beta_n^*)\}} \right) \right. \\
&\qquad \left. \left. + \frac{G_{t,ji}^{ee}(\eta_n^*)}{(g_t^e(\eta_n^*))^2} - 2 \frac{G_{t,ji}^{ee}(\eta_n^*)}{(g_t^e(\eta_n^*))^3} \left(g_t^e(\eta_n^*) - g_t^q(\beta_n^*) + \frac{1}{\alpha} (g_t^q(\beta_n^*) - Y_{t+1}) \mathbb{1}_{\{Y_{t+1} \leq g_t^q(\beta_n^*)\}} \right) \right] \right| \\
&= \left\| \frac{1}{n} \sum_{t=m}^{T-1} \mathbb{E} \left[\left\{ \frac{\nabla H_{t,ji}^e(\bar{\eta}_j)}{(g_t^e(\bar{\eta}_j))^2} - 2 \nabla g_t^e(\bar{\eta}_j) \frac{H_{t,ji}^e(\bar{\eta}_j)}{(g_t^e(\bar{\eta}_j))^3} - 2 \frac{\nabla G_{t,ji}^{ee}(\bar{\eta}_j)}{(g_t^e(\bar{\eta}_j))^3} + 6 \nabla g_t^e(\bar{\eta}_j) \frac{G_{t,ji}^{ee}(\bar{\eta}_j)}{(g_t^e(\bar{\eta}_j))^4} \right\} \times \right. \right. \\
&\qquad \left\{ g_t^e(\bar{\eta}_j) - g_t^q(\bar{\beta}_j) + \frac{1}{\alpha} (g_t^q(\bar{\beta}_j) - Y_{t+1}) \mathbb{1}_{\{Y_{t+1} \leq g_t^q(\bar{\beta}_j)\}} \right\} \\
&\qquad + \frac{\nabla G_{t,ji}^{ee}(\bar{\eta}_j)}{(g_t^e(\bar{\eta}_j))^2} - 2 \nabla g_t^e(\bar{\eta}_j) \frac{G_{t,ji}^{ee}(\bar{\eta}_j)}{(g_t^e(\bar{\eta}_j))^3} \\
&\qquad \left. \left. + \left\{ \frac{H_{t,ji}^e(\bar{\eta}_j)}{(g_t^e(\bar{\eta}_j))^2} - 2 \frac{G_{t,ji}^{ee}(\bar{\eta}_j)}{(g_t^e(\bar{\eta}_j))^2} \right\} \cdot \left\{ \nabla g_t^e(\bar{\eta}_j) - \nabla g_t^q(\bar{\beta}_j) + \frac{1}{\alpha} \nabla g_t^q(\bar{\beta}_j) F_t(\nabla g_t^q(\bar{\beta}_j)) \right\} \right] \right\| \\
&\qquad \cdot \left\| \tilde{\theta}_j - \theta_n^* \right\|.
\end{aligned}
$$

for some $\bar{\theta}_j = (\bar{\beta}_j, \bar{\eta}_j)$ on the line between $\tilde{\theta}_j$ and θ_n^*. As the respective moments are finite given the moment conditions in (h) in Assumption 2.2.1 and since $||\tilde{\theta}_j - \theta_n^*|| \leq ||\theta - \theta_n^*||$ for all j, we have shown that for all n sufficiently large enough, there exists a constant $c_1 > 0$ such that

$$\left\|\Delta_n(\tilde{\theta}_1, \ldots, \tilde{\theta}_k) - \Lambda_n(\theta_n^*)\right\| \leq c_1 ||\theta - \theta_n^*||. \tag{2.C.17}$$

Furthermore, as the matrix $\Lambda_n(\theta_n^*)$ has Eigenvalues bounded from below (for n large enough) by assumption, there exists a constant $c_2 > 0$, such that

$$\left\|\Lambda_n(\theta_n^*) \cdot (\theta - \theta_n^*)\right\| \geq c_2 ||\theta - \theta_n^*||. \tag{2.C.18}$$

Thus, we choose $d_0 > 0$ small enough such that $d_0 < \frac{c_2}{2c_1}$. Then $||\theta - \theta_n^*|| \leq d_0 < \frac{c_2}{2c_1}$ and thus, $2c_1||\theta - \theta_n^*||^2 \leq c_2||\theta - \theta_n^*||$. Consequently, $\left\|\left(\Delta_n(\tilde{\theta}_1, \ldots, \tilde{\theta}_k) - \Lambda_n(\theta_n^*)\right) \cdot (\theta - \theta_n^*)\right\| \leq c_1||\theta - \theta_n^*||^2 \leq c_2/2||\theta - \theta_n^*||$ and thus

$$\begin{aligned}
\left\|\Psi_n^0(\theta)\right\| &= \left\|\Delta_n(\tilde{\theta}_1, \ldots, \tilde{\theta}_k) \cdot (\theta - \theta_n^*)\right\| \\
&= \left\|\Lambda_n(\theta_n^*) \cdot (\theta - \theta_n^*) + \left(\Delta_n(\tilde{\theta}_1, \ldots, \tilde{\theta}_k) - \Lambda_n(\theta_n^*)\right) \cdot (\theta - \theta_n^*)\right\| \\
&\geq \left| \left\|\Lambda_n(\theta_n^*) \cdot (\theta - \theta_n^*)\right\| - \left\|\left(\Delta_n(\tilde{\theta}_1, \ldots, \tilde{\theta}_k) - \Lambda_n(\theta_n^*)\right) \cdot (\theta - \theta_n^*)\right\| \right| \\
&\geq \frac{c_2}{2}||\theta - \theta_n^*||,
\end{aligned} \tag{2.C.19}$$

by applying the mean value expansion and the inverse triangular inequality. □

Lemma 2.C.3. Given the conditions in Assumption 2.2.1 it holds that

$$\Sigma_n^{-1/2}(\theta_n^*)\sqrt{n}\,\Psi_n(\theta_n^*) \xrightarrow{d} \mathcal{N}(0, I_k). \tag{2.C.20}$$

Proof. We show this multivariate result by applying the Cramér–Wold theorem, i.e. by showing that the conditions for the univariate CLT for strong mixing sequences given in Theorem 5.20 in White (2001), p. 130 hold for all linear combinations $u^\top \psi\left(Y_{t+1}, g_t^q(\beta_n^*), g_t^e(\eta_n^*)\right)$ for all $u \in \mathbb{R}^k$ such that $||u|| = 1$. By Theorem 3.49 in White (2001) p. 50, we get that the sequences $\psi\left(Y_{t+1}, g_t^q(\beta_n^*), g_t^e(\eta_n^*)\right)$ and $u^\top \psi\left(Y_{t+1}, g_t^q(\beta_n^*), g_t^e(\eta_n^*)\right)$ are strong mixing of size $-r/(r-2)$ for some $r > 2$. Furthermore, for all $t \in \mathbb{N}$, it holds that

$$\begin{aligned}
&\mathbb{E}\left[\left|u^\top \psi\left(Y_{t+1}, g_t^q(\beta_n^*), g_t^e(\eta_n^*)\right)\right|^r\right] \leq \mathbb{E}\left[\left\|\psi\left(Y_{t+1}, g_t^q(\beta_n^*), g_t^e(\eta_n^*)\right)\right\|^r\right] \\
&\leq 4^{r-1}\left\{\max\left(\frac{1-\alpha}{\alpha}, 1\right)^r \mathbb{E}\left[\left\|\frac{\nabla_\beta g_t^q(\beta_n^*)}{g_t^e(\eta_n^*)}\right\|^r\right] + \mathbb{E}\left[\left\|\frac{\nabla_\eta g_t^e(\eta_n^*) g_t^e(\eta_n^*)}{(g_t^e(\eta_n^*))^2}\right\|^r\right]\right.
\end{aligned}$$

$$
\begin{aligned}
&+ \left(1+\frac{1}{\alpha}\right)^r \mathbb{E}\left[\left\|\frac{\nabla_\eta g_t^e(\eta_n^*) g_t^q(\beta_n^*)}{(g_t^e(\eta_n^*))^2}\right\|^r\right] + \mathbb{E}\left[\left\|\frac{\nabla_\eta g_t^e(\eta_n^*) Y_{t+1}}{\alpha (g_t^e(\eta_n^*))^2}\right\|^r\right]\Bigg\} \\
\leq 4^{r-1} &\left\{\max\left(\frac{1-\alpha}{\alpha}, 1\right)^r \frac{1}{K^r}\mathbb{E}\left[\left\|\nabla_\beta g_t^q(\beta_n^*)\right\|^r\right] + \frac{1}{K^r}\mathbb{E}\left[\left\|\nabla_\eta g_t^e(\eta_n^*)\right\|^r\right]\right. \\
&\left. + \frac{1}{K^{2r}}\left(1+\frac{1}{\alpha}\right)^r \mathbb{E}\left[\left\|\nabla_\eta g_t^e(\eta_n^*) g_t^q(\beta_n^*)\right\|^r\right] + \frac{1}{\alpha K^{2r}}\mathbb{E}\left[\left\|\nabla_\eta g_t^e(\eta_n^*) Y_{t+1}\right\|^r\right]\right\} < \infty,
\end{aligned}
$$

by applying Jensen's inequality and by the moment conditions (h) in Assumption 2.2.1, where $r > 2$ (from condition (a)). As the sequence $\psi\left(Y_{t+1}, g_t^q(\beta_n^*), g_t^e(\eta_n^*)\right)$ is uncorrelated by condition (c) in Assumption 2.2.1, we get that for all $n \geq 1$,

$$
\begin{aligned}
&\operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=m}^{T-1}\psi\left(Y_{t+1}, g_t^q(\beta_n^*), g_t^e(\eta_n^*)\right)\right) \\
&= \frac{1}{n}\sum_{t=m}^{T-1}\mathbb{E}\left[\psi\left(Y_{t+1}, g_t^q(\beta_n^*), g_t^e(\eta_n^*)\right)\cdot\psi\left(Y_{t+1}, g_t^q(\beta_n^*), g_t^e(\eta_n^*)\right)^\top\right] = \Sigma_n(\theta_n^*).
\end{aligned} \tag{2.C.21}
$$

As $\Sigma_n(\theta_n^*)$ is real and symmetric and positive definite, it can be diagonalized with a real orthogonal matrix S, i.e. $S^\top\Sigma_n(\theta_n^*)S = D_n$, where D_n is a diagonal matrix containing the Eigenvalues of $\Sigma_n(\theta_n^*)$, denoted by $\{\lambda_{1,n}, \ldots, \lambda_{k,n}\}$. Consequently, for any $u \in \mathbb{R}^k$,

$$
\begin{aligned}
\operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=m}^{T-1}u^\top\psi\left(Y_{t+1}, g_t^q(\beta_n^*), g_t^e(\eta_n^*)\right)\right) &= u^\top\Sigma_n(\theta_n^*)u = u^\top S^\top D_n S u = v^\top D_n v \\
&> \min_{i=1,\ldots,k}\lambda_{i,n},
\end{aligned} \tag{2.C.22}
$$

where $v = Su$, i.e. $||v|| = 1$ as S is orthogonal and where the Eigenvalues $\{\lambda_{1,n}, \ldots, \lambda_{k,n}\}$ are bounded away from zero for n sufficiently large. Thus, we can apply Theorem 5.20 in White (2001) p. 130 for asymptotic normality of the sequences $u^\top\psi\left(Y_{t+1}, g_t^q(\beta_n^*), g_t^e(\eta_n^*)\right)$ for all $u \in \mathbb{R}^k$ such that $||u|| = 1$. Applying the Cramér-Wold theorem concludes the proof. □

Appendix 2.D Additional tables

Table 2.D.1: Empirical sizes of the forecast encompassing tests

	$\mathbb{H}_0^{(1)}$				$\mathbb{H}_0^{(2)}$			
	Str ES	Aux ES	VaR ES	VaR	Str ES	Aux ES	VaR ES	VaR
n	GARCH							
500	3.05	3.00	8.40	10.10	2.80	2.80	8.20	10.10
1000	1.45	1.75	5.90	7.80	2.20	1.95	7.70	9.30
2500	1.85	1.80	5.85	7.30	1.85	1.75	5.45	6.45
5000	1.25	1.25	3.90	5.05	0.80	0.80	4.10	5.15
n	VaR/ES GAS							
500	5.65	5.30	9.40	11.20	4.40	4.20	9.20	11.50
1000	4.15	4.05	6.65	7.75	3.55	3.25	6.55	8.40
2500	2.70	2.65	4.80	5.80	1.35	1.45	4.70	5.95
5000	1.80	1.90	3.10	4.10	1.40	1.20	4.50	5.55
n	GAS-*t*							
500	5.45	5.50	7.70	8.00	4.90	5.30	7.75	9.15
1000	4.15	4.45	6.05	6.75	2.00	2.25	4.75	6.00
2500	2.00	1.90	3.10	3.40	1.25	1.35	3.10	3.90
5000	1.70	1.80	3.95	4.05	1.00	1.05	2.15	2.70
n	ES-CAViaR							
500	2.05	1.30	4.30	6.00	2.35	1.45	5.05	6.50
1000	1.85	1.25	3.55	5.55	1.65	1.25	3.10	4.85
2500	1.00	1.15	2.30	3.10	1.00	0.90	2.05	3.00
5000	1.15	0.85	1.55	2.15	1.10	1.15	1.15	1.85

Notes: This table presents the empirical sizes (in %) of our three forecast encompassing tests for the ES together with the VaR encompassing test of Giacomini and Komunjer (2005) for a nominal size of 1%. The results are shown for the four DGPs described in Section 2.3.1 and Appendix 2.B in the horizontal panels, for both tested hypotheses in the vertical panels and for different sample sizes.

Table 2.D.2: Empirical sizes of the forecast encompassing tests

	$\mathbb{H}_0^{(1)}$				$\mathbb{H}_0^{(2)}$			
	Str ES	Aux ES	VaR ES	VaR	Str ES	Aux ES	VaR ES	VaR
n	GARCH							
500	9.20	9.30	13.90	17.50	8.75	9.30	13.45	17.55
1000	6.90	6.45	11.40	14.40	6.90	6.35	12.65	17.10
2500	6.35	6.40	11.10	13.55	5.90	5.75	9.85	12.05
5000	5.65	5.25	8.65	9.75	5.00	5.05	9.00	10.65
n	VaR/ES GAS							
500	17.75	18.65	16.00	20.40	13.35	13.20	17.35	20.60
1000	13.75	13.30	13.10	16.50	11.00	11.05	12.65	16.55
2500	9.65	9.70	10.05	12.20	6.85	7.10	9.90	12.95
5000	7.80	7.10	8.25	9.80	5.45	5.70	8.65	12.05
n	GAS-*t*							
500	14.50	14.30	14.40	14.50	11.80	11.50	13.90	16.20
1000	11.80	11.75	12.30	13.75	7.90	8.10	9.60	11.00
2500	7.00	6.85	9.85	9.75	6.05	6.05	7.25	9.40
5000	7.05	7.05	9.50	8.85	5.30	5.35	6.65	7.15
n	ES-CAViaR							
500	6.75	5.95	9.15	12.90	7.20	5.75	9.70	13.80
1000	7.00	6.15	8.40	11.05	6.35	5.30	7.85	10.65
2500	5.10	4.45	5.60	8.05	5.05	4.40	6.20	8.70
5000	5.40	4.80	5.20	7.15	5.25	5.10	5.00	6.85

Notes: This table presents the empirical sizes (in %) of our three forecast encompassing tests for the ES together with the VaR encompassing test of Giacomini and Komunjer (2005) for a nominal size of 5%. The results are shown for the four DGPs described in Section 2.3.1 and Appendix 2.B in the horizontal panels, for both tested hypotheses in the vertical panels and for different sample sizes.

Table 2.D.3: Empirical sizes for two additional DGPs

	$\mathbb{H}_0^{(1)}$				$\mathbb{H}_0^{(2)}$			
	Str ES	Aux ES	VaR ES	VaR	Str ES	Aux ES	VaR ES	VaR
n	GAS-*t*							
500	21.95	21.75	20.25	19.50	18.50	18.10	18.20	21.75
1000	18.75	18.35	18.75	19.35	14.25	13.95	13.80	17.45
2500	12.40	12.05	15.45	14.90	11.65	11.75	12.55	15.85
5000	12.50	12.35	15.25	14.50	9.60	9.20	11.50	12.90
n	ES-CAViaR							
500	12.95	11.80	13.55	19.00	13.05	11.55	15.05	19.40
1000	12.30	11.70	12.70	17.25	11.40	10.60	11.95	16.50
2500	10.35	9.45	9.65	13.75	10.85	9.55	10.20	13.10
5000	9.65	9.65	10.65	12.65	10.35	9.75	10.20	11.70

Notes: This table presents the empirical sizes (in %) of our three forecast encompassing tests for the ES together with the VaR encompassing test of Giacomini and Komunjer (2005) for a nominal size of 10%. The results are shown for the two additional DGPs described in Appendix 2.B in the horizontal panels, for both tested hypotheses in the vertical panels and for different sample sizes.

Table 2.D.4: Empirical test sizes for different loss functions

	$\mathbb{H}_0^{(1)}$			$\mathbb{H}_0^{(2)}$		
	Str ES	Aux ES	VaR ES	Str ES	Aux ES	VaR ES
n	GARCH DGP and $\phi(z) = 1/\sqrt{-z}$					
500	13.80	14.30	19.40	14.60	14.90	19.10
1000	13.40	13.00	17.90	13.00	12.70	17.50
2500	10.90	10.60	15.70	11.20	11.00	13.20
5000	10.30	10.40	14.50	9.30	9.70	14.00
n	GARCH DGP and $\phi(z) = -1/z$					
500	14.80	14.60	19.30	14.50	14.50	18.30
1000	12.40	11.60	16.00	14.10	14.10	18.40
2500	12.20	11.60	16.10	10.70	10.30	14.50
5000	11.00	10.80	15.30	9.90	10.30	13.60
n	VaR/ES GAS DGP and $\phi(z) = 1/\sqrt{-z}$					
500	27.60	27.30	24.30	20.70	20.20	23.20
1000	22.10	21.60	19.10	17.60	17.10	18.90
2500	15.00	16.40	14.90	14.60	14.80	17.90
5000	13.30	13.00	14.70	12.00	12.30	14.20
n	VaR/ES GAS DGP and $\phi(z) = -1/z$					
500	30.90	31.00	26.70	21.60	21.90	21.90
1000	22.90	22.70	22.80	17.70	17.60	18.10
2500	15.00	14.80	16.30	13.90	13.80	15.90
5000	11.00	11.80	12.50	11.70	11.30	13.70

Notes: This table presents the empirical sizes (in %) of our three forecast encompassing tests for the ES for a nominal size of 10% for two additional strictly consistent loss functions and for the two DGPs described in Section 2.3.1. The first and third horizontal panels consider the choices $\mathfrak{g}(z) = 0$ and $\phi(z) = 1/\sqrt{-z}$ in the loss function in (2.3.7) while the second and fourth panel consider the choices $\mathfrak{g}(z) = 0$ and $\phi(z) = -1/z$.

Table 2.D.5: Empirical sizes for different link functions

	$\mathbb{H}_0^{(1)}$			$\mathbb{H}_0^{(2)}$		
	Str ES	Aux ES	VaR ES	Str ES	Aux ES	VaR ES
n	Affine Link: GARCH DGP					
500	12.93	11.52	13.83	14.21	12.31	14.01
1000	14.20	12.70	14.00	11.80	9.70	13.70
2500	13.20	11.40	13.40	11.70	9.90	12.80
5000	11.60	9.90	11.80	12.20	9.60	15.60
n	Affine Link: VaR/ES GAS DGP					
500	23.52	22.72	15.82	14.70	13.70	13.50
1000	21.30	18.20	14.60	10.90	10.40	13.00
2500	16.20	14.30	12.80	10.50	9.80	12.80
5000	13.80	11.00	10.70	9.90	9.90	12.00
n	Affine Link: GAS-*t* DGP					
500	18.63	16.72	15.21	12.80	10.50	11.40
1000	12.90	10.30	12.00	11.20	9.00	8.80
2500	12.80	11.20	12.20	10.50	8.90	9.20
5000	15.10	12.90	12.70	12.80	9.70	10.60
n	Affine Link: ES-CAViaR DGP					
500	11.88	10.39	11.56	11.33	10.40	11.54
1000	12.45	12.75	14.26	12.42	10.71	11.01
2500	9.80	9.40	10.30	12.10	11.90	12.00
5000	8.50	8.60	9.80	10.40	10.30	10.80
n	Nonlinear Link: Nonlinear DGP					
500	8.12	8.82	13.53	8.82	8.72	16.03
1000	8.70	9.10	14.80	8.90	9.10	13.60
2500	10.40	10.10	11.90	9.50	10.10	14.30
5000	11.30	11.30	17.20	11.10	10.80	12.50

Notes: This table presents the empirical sizes (in %) of our three forecast encompassing tests for the ES for a nominal size of 10% for the affine and the nonlinear link functions described in Section 2.3.3. The upper four horizontal panels consider the affine link functions for the four DGPs described in Sections 2.3.1 and in Appendix 2.B. The lowest panel presents results for the nonlinear link functions based on the nonlinear GARCH DGP described in (2.3.10).

Table 2.D.6: Correlation matrices of the VaR and ES forecasts in the empirical application

Panel A: IBM Stock

	Quantile Forecasts								ES Forecasts							
	HS	RM	GJR	GAS	G1F	G2F	ASES	SAVES	HS	RM	GJR	GAS	G1F	G2F	ASES	SAVES
HS	1.00	0.58	0.54	0.65	0.31	0.41	0.40	0.46	1.00	0.58	0.54	0.65	0.31	0.41	0.40	0.46
RM		1.00	0.96	0.95	0.78	0.86	0.84	0.89		1.00	0.96	0.95	0.78	0.86	0.84	0.89
GJR			1.00	0.92	0.83	0.91	0.85	0.85			1.00	0.92	0.83	0.91	0.85	0.85
GAS				1.00	0.70	0.82	0.81	0.87				1.00	0.70	0.82	0.81	0.87
G1F					1.00	0.88	0.88	0.82					1.00	0.88	0.88	0.82
G2F						1.00	0.84	0.82						1.00	0.84	0.82
ASES							1.00	0.96							1.00	0.96
SAVES								1.00								1.00

Panel B: S&P 500 Index

	Quantile Forecasts								ES Forecasts							
HS	1.00	0.66	0.58	0.67	0.55	0.55	0.48	0.57	1.00	0.66	0.58	0.67	0.55	0.55	0.48	0.57
RM		1.00	0.98	1.00	0.92	0.91	0.90	0.97		1.00	0.98	1.00	0.92	0.91	0.90	0.97
GJR			1.00	0.97	0.95	0.94	0.96	0.97			1.00	0.97	0.95	0.94	0.96	0.97
GAS				1.00	0.90	0.90	0.90	0.97				1.00	0.90	0.90	0.90	0.97
G1F					1.00	0.96	0.89	0.90					1.00	0.96	0.89	0.90
G2F						1.00	0.89	0.90						1.00	0.89	0.90
ASES							1.00	0.94							1.00	0.94
SAVES								1.00								1.00

Panel C: DAX 30 Index

	Quantile Forecasts								ES Forecasts							
HS	1.00	0.68	0.53	0.62	0.67	0.60	0.57	0.62	1.00	0.68	0.53	0.62	0.67	0.60	0.57	0.62
RM		1.00	0.92	0.98	0.93	0.94	0.95	0.98		1.00	0.92	0.98	0.93	0.94	0.95	0.98
GJR			1.00	0.96	0.91	0.95	0.99	0.95			1.00	0.96	0.91	0.95	0.99	0.95
GAS				1.00	0.93	0.95	0.97	0.99				1.00	0.93	0.95	0.97	0.99
G1F					1.00	0.96	0.90	0.91					1.00	0.96	0.90	0.91
G2F						1.00	0.94	0.93						1.00	0.94	0.93
ASES							1.00	0.98							1.00	0.98
SAVES								1.00								1.00

Notes: This table shows the correlations of the respective quantile and ES forecasts obtained from the eight forecasting models described in Section 2.4 for the IBM stock, the S&P 500 and the DAX 30 indices estimated on an in-sample window of $m = 2000$ days. The models are abbreviated as follows: Historical simulation (HS), RiskMetrics (RM), GAS-t model (GAS), GAS one factor (G1F) and GAS two factor (G2F) model, dynamic AS-ES-CAViaR (ASES) and dynamic SAV-ES-CAViaR (SAVES).

Table 2.D.7: Forecast encompassing test p-values for the IBM stock

	Joint VaR ES encompassing test							
	HS	RM	GJR	GAS	G1F	G2F	ASES	SAVES
HS		0.000*	0.000*	0.000*	0.000*	0.000	0.000	0.000
RM	0.035*		0.057*	0.000*	0.002*	0.001	0.000	0.000
GJR	0.006*	0.002*		0.000*	0.000*	0.000	0.000	0.000
GAS	0.000*	0.000*	0.000*		0.000*	0.000	0.000*	0.000
G1F	0.000*	0.000*	0.000*	0.000*		0.000*	0.000*	0.000
G2F	0.574	0.397	0.200	0.475	0.091*		0.000	0.000
ASES	0.577	0.648	0.621	0.024*	0.006*	0.419		0.341
SAVES	0.393	0.352	0.265	0.750	0.111	0.173	0.730	

	VaR encompassing test							
	HS	RM	GJR	GAS	G1F	G2F	ASES	SAVES
HS		0.000*	0.000*	0.000*	0.000*	0.000	0.000	0.000
RM	0.079*		0.111	0.000*	0.001*	0.001	0.000	0.000
GJR	0.002*	0.002		0.004*	0.000*	0.000*	0.000	0.000
GAS	0.000*	0.000*	0.000*		0.000*	0.000	0.000*	0.000
G1F	0.000*	0.000*	0.000*	0.000*		0.000*	0.000*	0.000
G2F	0.827	0.175	0.087*	0.338	0.017*		0.000	0.000
ASES	0.809	0.368	0.737	0.009*	0.004*	0.302		0.115
SAVES	0.335	0.920	0.930	0.766	0.172	0.522	0.564	

	Auxiliary ES encompassing test							
	HS	RM	GJR	GAS	G1F	G2F	ASES	SAVES
HS		0.000*	0.000*	0.000*	0.001	0.000	0.000	0.000
RM	0.016*		0.040*	0.002*	0.055	0.033	0.000	0.000
GJR	0.023*	0.011*		0.000*	0.052	0.019	0.000	0.000
GAS	0.002*	0.020*	0.007*		0.029*	0.018	0.001	0.001
G1F	0.298	0.260	0.388	0.076*		0.242	0.000	0.001
G2F	0.265	0.710	0.788	0.276	0.825		0.002	0.002
ASES	0.497	0.713	0.528	0.890	0.446	0.472		0.824
SAVES	0.185	0.183	0.136	0.570	0.254	0.212	0.473	

	Strict ES encompassing test							
	HS	RM	GJR	GAS	G1F	G2F	ASES	SAVES
HS		0.000*	0.000*	0.000*	0.000	0.000	0.000	0.000
RM	0.013*		0.043*	0.003*	0.053	0.057	0.000	0.000
GJR	0.023*	0.011*		0.000*	0.048	0.014	0.000	0.000
GAS	0.001*	0.022*	0.007*		0.034*	0.034	0.000	0.001
G1F	0.307	0.244	0.391	0.073*		0.192	0.000	0.001
G2F	0.250	0.729	0.756	0.282	0.793		0.006	0.005
ASES	0.485	0.589	0.408	0.866	0.338	0.580		0.729
SAVES	0.255	0.170	0.125	0.563	0.222	0.310	0.454	

Notes: This table reports the p-values of the three different ES encompassing tests introduced in Section 2.2 and of the VaR encompassing test of Giacomini and Komunjer (2005) applied to returns from the IBM stock and an in-sample period of $m = 2000$ days. The p-value in the i-th row and j-th column of the respective matrices corresponds to testing $\mathbb{H}_0$: the forecasts from model i encompass the forecasts from model j. The symbol * denotes model pairs where both encompassing tests (test whether model i encompasses model j and vice versa) are significant at the 10% level. The models are abbreviated as follows: Historical simulation (HS), RiskMetrics (RM), GAS-t model (GAS), GAS one factor (G1F) and GAS two factor (G2F) model, dynamic AS-ES-CAViaR (ASES) and dynamic SAV-ES-CAViaR (SAVES).

Table 2.D.8: Forecast encompassing test p-values for the S&P 500 index

	Joint VaR ES encompassing test								VaR encompassing test							
	HS	RM	GJR	GAS	G1F	G2F	ASES	SAVES	HS	RM	GJR	GAS	G1F	G2F	ASES	SAVES
HS		0.000	0.000	0.000	0.000*	0.000	0.000*	0.000		0.000	0.000	0.000	0.000	0.000	0.000*	0.000
RM	0.389		0.000	0.067	0.000*	0.001	0.000	0.099	0.809		0.000	0.206	0.000*	0.001	0.000	0.070*
GJR	0.255	0.281		0.682	0.193	0.035*	0.000*	0.860	0.505	0.412		0.540	0.177	0.005*	0.000*	0.736
GAS	0.128	0.226	0.000		0.000*	0.010	0.000	0.047	0.106	0.134	0.001		0.000	0.003*	0.000*	0.013
G1F	0.057*	0.092*	0.009	0.056*		0.039*	0.000*	0.027*	0.170	0.082*	0.003	0.121		0.107	0.001*	0.125
G2F	0.309	0.406	0.000*	0.156	0.096*		0.005*	0.119	0.127	0.166	0.000*	0.056*	0.051		0.007*	0.094*
ASES	0.011*	0.479	0.002*	0.421	0.006*	0.048*		0.662	0.008*	0.144	0.000*	0.055*	0.003*	0.007*		0.395
SAVES	0.556	0.131	0.007	0.467	0.000*	0.002	0.000		0.473	0.055*	0.002	0.154	0.000	0.000*	0.000	

	Auxiliary ES encompassing test								Strict ES encompassing test							
	HS	RM	GJR	GAS	G1F	G2F	ASES	SAVES	HS	RM	GJR	GAS	G1F	G2F	ASES	SAVES
HS		0.000	0.000	0.000	0.000*	0.000	0.000	0.000		0.000	0.000	0.000	0.000*	0.000	0.000	0.000
RM	0.131		0.000	0.033	0.003*	0.020	0.002	0.072	0.119		0.000	0.015	0.002*	0.019	0.002	0.059
GJR	0.544	0.110		0.476	0.252	0.575	0.376	0.853	0.580	0.108		0.432	0.239	0.621	0.364	0.900
GAS	0.222	0.181	0.003		0.015*	0.100	0.020	0.794	0.220	0.158	0.003		0.017*	0.133	0.019	0.811
G1F	0.039*	0.028*	0.008	0.015*		0.015	0.000*	0.007*	0.031*	0.035*	0.009	0.019*		0.016	0.001*	0.008*
G2F	0.632	0.516	0.123	0.242	0.131		0.043	0.293	0.501	0.429	0.101	0.172	0.157		0.029	0.258
ASES	0.779	0.674	0.454	0.500	0.056*	0.410		0.817	0.778	0.651	0.464	0.497	0.053*	0.515		0.802
SAVES	0.819	0.936	0.034	0.486	0.009*	0.089	0.024		0.821	0.945	0.039	0.465	0.013*	0.088	0.021	

Notes: This table reports the p-values of the three different ES encompassing tests introduced in Section 2.2 and of the VaR encompassing test of Giacomini and Komunjer (2005) applied to returns from the S&P 500 index and an in-sample period of $m = 2000$ days. The p-value in the i-th row and j-th column of the respective matrices corresponds to testing $\mathbb{H}_0$: the forecasts from model i encompass the forecasts from model j. The symbol * denotes model pairs where both encompassing tests (test whether model i encompasses model j and vice versa) are significant at the 10% level. The models are abbreviated as follows: Historical simulation (HS), RiskMetrics (RM), GAS-t model (GAS), GAS one factor (G1F) and GAS two factor (G2F) model, dynamic AS-ES-CAViaR (ASES) and dynamic SAV-ES-CAViaR (SAVES).

Table 2.D.9: Forecast encompassing test p-values for the DAX 30 index

	Joint VaR ES encompassing test							
	HS	RM	GJR	GAS	G1F	G2F	ASES	SAVES
HS		0.000	0.000	0.000	0.000	0.000*	0.000	0.000
RM	0.587		0.000	0.000*	0.160	0.125	0.000	0.004
GJR	0.375	0.122		0.175	0.455	0.775	0.000	0.085*
GAS	0.936	0.096*	0.049		0.832	0.647	0.005	0.252
G1F	0.732	0.011	0.000	0.000		0.037*	0.000	0.000
G2F	0.012*	0.001	0.000	0.000	0.020*		0.000	0.000*
ASES	0.869	0.905	0.537	0.867	0.965	0.997		0.079*
SAVES	0.673	0.476	0.047*	0.239	0.273	0.068*	0.000*	

	VaR encompassing test							
	HS	RM	GJR	GAS	G1F	G2F	ASES	SAVES
HS		0.000	0.000	0.000	0.000	0.000*	0.000	0.000
RM	0.252		0.000*	0.000*	0.020*	0.030*	0.000	0.025
GJR	0.279	0.046*		0.073*	0.456	0.917	0.000	0.144
GAS	0.889	0.042*	0.080*		0.692	0.304	0.001	0.854
G1F	0.590	0.036*	0.000	0.000		0.013*	0.000	0.000*
G2F	0.022*	0.004*	0.000	0.001	0.097*		0.000	0.000*
ASES	0.848	0.964	0.793	0.986	0.927	0.964		0.543
SAVES	0.742	0.760	0.022	0.072	0.028*	0.029*	0.000	

	Auxiliary ES encompassing test							
	HS	RM	GJR	GAS	G1F	G2F	ASES	SAVES
HS		0.000	0.000	0.000	0.000	0.000	0.000	0.000
RM	0.518		0.152	0.043	0.271	0.229	0.034	0.001
GJR	0.188	0.093		0.052	0.170	0.421	0.011	0.025
GAS	0.994	0.796	0.838		0.892	0.951	0.333	0.101
G1F	0.929	0.048	0.047	0.001		0.290	0.000	0.000
G2F	0.131	0.014	0.026	0.001	0.035		0.000	0.000
ASES	0.625	0.746	0.445	0.736	0.837	0.982		0.268
SAVES	0.809	0.662	0.898	0.800	0.520	0.929	0.830	

	Strict ES encompassing test							
	HS	RM	GJR	GAS	G1F	G2F	ASES	SAVES
HS		0.000	0.000	0.000	0.000	0.000*	0.000	0.000
RM	0.459		0.161	0.036	0.247	0.278	0.027	0.001
GJR	0.148	0.070		0.057	0.156	0.465	0.005	0.007
GAS	0.967	0.808	0.813		0.880	0.945	0.339	0.041
G1F	0.923	0.059	0.051	0.001		0.359	0.000	0.000
G2F	0.087*	0.012	0.033	0.001	0.021		0.000	0.000
ASES	0.600	0.743	0.303	0.695	0.820	0.960		0.208
SAVES	0.766	0.680	0.991	0.817	0.612	0.846	0.694	

Notes: This table reports the p-values of the three different ES encompassing tests introduced in Section 2.2 and of the VaR encompassing test of Giacomini and Komunjer (2005) applied to returns from the DAX 30 index and an in-sample period of $m = 2000$ days. The p-value in the i-th row and j-th column of the respective matrices corresponds to testing $\mathbb{H}_0$: the forecasts from model i encompass the forecasts from model j. The symbol * denotes model pairs where both encompassing tests (test whether model i encompasses model j and vice versa) are significant at the 10% level. The models are abbreviated as follows: Historical simulation (HS), RiskMetrics (RM), GAS-t model (GAS), GAS one factor (G1F) and GAS two factor (G2F) model, dynamic AS-ES-CAViaR (ASES) and dynamic SAV-ES-CAViaR (SAVES).

Table 2.D.10: Average losses for the VaR and ES forecasts for the IBM stock

Panel A: Forecast Combination Losses & Joint VaR ES Encompassing Weights

Model	Loss	HS	RM	GJR	GAS	G1F	G2F	ASES	SAVES
HS	1.544		1.369	1.381	1.358	1.380	1.375	1.326	1.324
RM	1.504			1.370	1.365	1.368	1.370	1.325	1.321
GJR	1.476				1.363	1.375	1.376	1.324	1.320
GAS	1.420					1.360	1.363	1.325	1.324
G1F	1.403						1.375	1.323	1.318
G2F	1.382							1.322	1.316
ASES	1.329								1.322
SAVES	1.325								

Panel B: Forecast Combination Losses & Strict ES Encompassing Weights

Model	Loss	HS	RM	GJR	GAS	G1F	G2F	ASES	SAVES
HS	1.544		1.369	1.378	1.355	1.380	1.367	1.334	1.323
RM	1.504			1.370	1.364	1.369	1.372	1.324	1.320
GJR	1.476				1.361	1.375	1.376	1.323	1.320
GAS	1.420					1.359	1.364	1.326	1.323
G1F	1.403						1.370	1.325	1.317
G2F	1.382							1.323	1.317
ASES	1.329								1.320
SAVES	1.325								

Panel C: Forecast Combination Losses & VaR Encompassing Weights

Model	Loss	HS	RM	GJR	GAS	G1F	G2F	ASES	SAVES
HS	1.120		0.993	1.002	0.988	0.998	0.997	0.971	0.969
RM	1.008			0.998	0.994	0.988	0.989	0.970	0.970
GJR	1.010				0.996	0.992	0.994	0.971	0.970
GAS	1.008					0.985	0.988	0.969	0.970
G1F	1.022						0.993	0.971	0.969
G2F	0.998							0.971	0.969
ASES	0.976								0.968
SAVES	0.970								

Notes: This table shows average out-of-sample forecast losses for the eight stand-alone models described in Section 2.4 and the respective linear forecast combinations with estimated weights for the IBM stock. The column labeled "Loss" reports the average losses for the individual forecasting models and the remaining eight columns report the average losses of the forecast combinations with the respective estimated combination parameters. In Panel A, the estimated weights are obtained from the underlying regression of the "joint VaR and ES encompassing test" (or equivalently from the "auxiliary ES encompassing test"), in Panel B from the "strict ES encompassing test" and in Panel C from the "VaR encompassing test". The values in Panel C are multiplied by 10. The losses given in Panel A and B refer to the joint VaR and ES loss function given in (2.2.8) whereas in Panel C, the values are obtained by using the quantile-specific piecewise linear loss function.

Table 2.D.11: Average losses for the VaR and ES forecasts for the S&P 500 index

Panel A: Forecast Combination Losses & Joint VaR ES Encompassing Weights

Model	Loss	HS	RM	GJR	GAS	G1F	G2F	ASES	SAVES
HS	1.267		1.031	0.985	1.020	0.978	0.995	0.948	1.009
RM	1.157			0.976	1.014	0.982	0.996	0.960	1.011
GJR	1.070				0.980	0.976	0.981	0.962	0.986
GAS	1.090					0.981	0.993	0.961	1.011
G1F	1.001						0.981	0.955	0.977
G2F	1.009							0.961	0.989
ASES	0.960								0.960
SAVES	1.013								

Panel B: Forecast Combination Losses & Strict ES Encompassing Weights

Model	Loss	HS	RM	GJR	GAS	G1F	G2F	ASES	SAVES
HS	1.267		1.031	0.986	1.015	0.979	0.998	0.952	1.011
RM	1.157			0.976	1.003	0.982	0.996	0.960	1.011
GJR	1.070				0.980	0.976	0.981	0.962	0.985
GAS	1.090					0.981	0.992	0.961	1.009
G1F	1.001						0.981	0.955	0.977
G2F	1.009							0.960	0.990
ASES	0.960								0.960
SAVES	1.013								

Panel C: Forecast Combination Losses & VaR Encompassing Weights

Model	Loss	HS	RM	GJR	GAS	G1F	G2F	ASES	SAVES
HS	9.426		7.644	7.407	7.592	7.420	7.464	7.307	7.581
RM	7.876			7.382	7.591	7.414	7.442	7.368	7.566
GJR	7.630				7.399	7.381	7.369	7.357	7.411
GAS	7.876					7.400	7.427	7.364	7.557
G1F	7.463						7.404	7.305	7.391
G2F	7.539							7.319	7.436
ASES	7.414								7.389
SAVES	7.608								

Notes: This table shows average out-of-sample forecast losses for the eight stand-alone models described in Section 2.4 and the respective linear forecast combinations with estimated weights for the S&P 500 index. The column labeled "Loss" reports the average losses for the individual forecasting models and the remaining eight columns report the average losses of the forecast combinations with the respective estimated combination parameters. In Panel A, the estimated weights are obtained from the underlying regression of the "joint VaR and ES encompassing test" (or equivalently from the "auxiliary ES encompassing test"), in Panel B from the "strict ES encompassing test" and in Panel C from the "VaR encompassing test". The values in Panel C are multiplied by 100. The losses given in Panel A and B refer to the joint VaR and ES loss function given in (2.2.8) whereas in Panel C, the values are obtained by using the quantile-specific piecewise linear loss function.

Table 2.D.12: Average losses for the VaR and ES forecasts for the DAX 30 index

Panel A: Forecast Combination Losses & Joint VaR ES Encompassing Weights

Model	Loss	HS	RM	GJR	GAS	G1F	G2F	ASES	SAVES
HS	1.331		1.140	1.117	1.125	1.150	1.150	1.107	1.121
RM	1.202			1.116	1.124	1.138	1.136	1.108	1.121
GJR	1.159				1.116	1.121	1.124	1.106	1.111
GAS	1.166					1.124	1.124	1.108	1.120
G1F	1.152						1.151	1.108	1.120
G2F	1.173							1.108	1.118
ASES	1.109								1.105
SAVES	1.121								

Panel B: Forecast Combination Losses & Strict ES Encompassing Weights

Model	Loss	HS	RM	GJR	GAS	G1F	G2F	ASES	SAVES
HS	1.331		1.138	1.118	1.125	1.151	1.148	1.110	1.112
RM	1.202			1.116	1.124	1.138	1.131	1.110	1.111
GJR	1.159				1.116	1.121	1.124	1.108	1.104
GAS	1.166					1.124	1.123	1.110	1.110
G1F	1.152						1.151	1.110	1.111
G2F	1.173							1.109	1.106
ASES	1.109								1.106
SAVES	1.121								

Panel C: Forecast Combination Losses & VaR Encompassing Weights

Model	Loss	HS	RM	GJR	GAS	G1F	G2F	ASES	SAVES
HS	9.853		8.362	8.142	8.237	8.416	8.370	8.097	8.276
RM	8.622			8.134	8.215	8.330	8.291	8.097	8.277
GJR	8.290				8.144	8.167	8.174	8.098	8.140
GAS	8.447					8.235	8.225	8.097	8.235
G1F	8.431						8.392	8.097	8.246
G2F	8.561							8.097	8.217
ASES	8.105								8.084
SAVES	8.283								

Notes: This table shows average out-of-sample forecast losses for the eight stand-alone models described in Section 2.4 and the respective linear forecast combinations with estimated weights for the DAX 30 index. The column labeled "Loss" reports the average losses for the individual forecasting models and the remaining eight columns report the average losses of the forecast combinations with the respective estimated combination parameters. In Panel A, the estimated weights are obtained from the underlying regression of the "joint VaR and ES encompassing test" (or equivalently from the "auxiliary ES encompassing test"), in Panel B from the "strict ES encompassing test" and in Panel C from the "VaR encompassing test". The values in Panel C are multiplied by 100. The losses given in Panel A and B refer to the joint VaR and ES loss function given in (2.2.8) whereas in Panel C, the values are obtained by using the quantile-specific piecewise linear loss function.

Appendix 2.E Additional figures

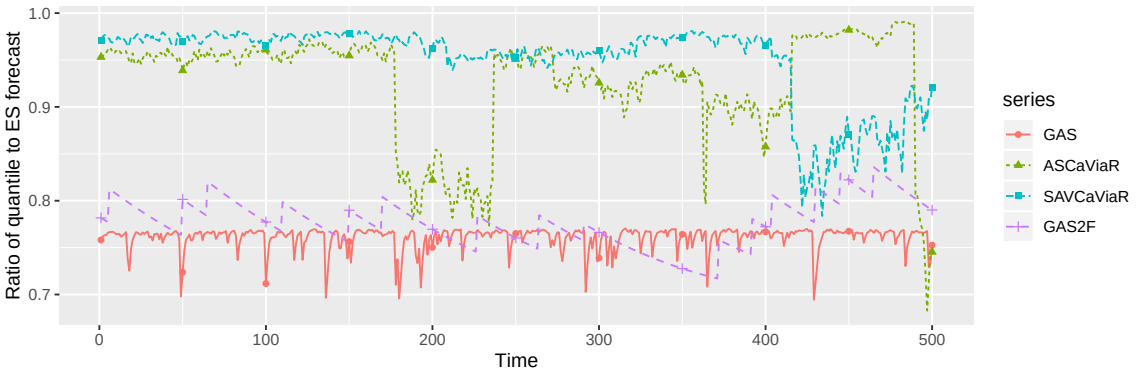


Figure 2.E.1: This figure shows the ratio of the true VaR and ES forecasts $\hat{q}_t/\hat{e}_t$ for simulated paths of the GAS-t model of Creal et al. (2013), the two factor GAS model for the VaR and ES introduced in Patton et al. (2019), the AS-ES-CAViaR model and the SAV-ES-CAViaR model proposed by Taylor (2019).

Chapter 3

Empirical Asset Pricing in a DSGE Framework: Reconciling Calibration and Econometrics using Partial Indirect Inference

This project is joint work with Joachim Grammig and Dalia Elshiaty. For helpful comments, we thank Tim Landvoigt, Jantje Sönksen and the seminar participants at the University of Konstanz, the 2020 World Conference of the Econometric Society (virtual), the 2020 EFA conference (virtual) and the 2020 EEA conference (virtual).

3.1. Introduction

Embedding preference-based asset pricing within a dynamic stochastic general equilibrium (DSGE) framework holds the promise to resolve the prominent puzzles of financial economics by using a pricing kernel that is consistent with macroeconomic processes. However, the apparent empirical success of such approaches is predominantly based on calibration methods, which Hansen and Heckman (1996) have criticized for lacking the rigor of econometric analysis.

This study aims at delivering a critical assessment of the DSGE asset pricing approach, and to return the empirical analysis of these models to the econometrics fold. For that purpose, we focus on the most recent exponent proposed by Chen (2017). Applying the partial indirect inference method (PII) introduced by Dridi et al. (2007), we acknowledge that parts of the model (the production technology) are misspecified, while other parts (the pricing kernel) have the claim of capturing financial economic reality. Adapting the PII philosophy to the present problem, we use binding functions that facilitate the consistent estimation of some structural model parameters of interest, while treating others as nuisance parameters. The latter do not truly capture economic reality, but are necessary to generate model-implied data within the simulation-based estimation procedure.

The class of DSGE models considered here connects asset prices to an exogenous technology process and an endogenously generated consumption process. In this model economy, households try to smooth their consumption stream in response to technology shocks, and the fluctuations in consumption determine asset prices. Capital adjustment costs induce that the supply of new capital via household investment and consumption decisions is not perfectly flexible. Jermann (1998) pioneers this literature by considering a real business cycle (RBC) model that explains both business cycle moments and asset pricing facts. He introduces linear habit persistence, which creates volatility in the pricing kernel without the need for an unreasonably high relative risk aversion. While the model-implied time-varying risk premium resolves the equity premium puzzle, the linear habit specification also generates an implausibly high volatility of the risk-free rate. Chen (2017) accounts for this drawback by accounting for nonlinear habit preferences as in Campbell and Cochrane (1999), and convex capital adjustment costs. The empirical validation of this most recent exponent of DSGE asset pricing models is performed by means of a calibration study. The results are interpreted such that the calibrated model resolves many empirical asset pricing puzzles.

Quite different to such a calibration study, econometric analysis challenges a model's specification, instead of seeking confirmatory evidence. The model parameters are estimated using empirical data, which entails a careful discussion of identifying restrictions. The key questions are whether these restrictions facilitate consistent parameter estimation, and how

informative the available data are to accomplish that task. The analysis is deeply concerned with parameter estimation uncertainty, which translates into confidence intervals for the parameters and model-implied indicators of interest. These are all non-issues for calibration studies.

However, the econometric analysis of DSGE asset pricing models is a challenging task. Proponents of calibration may even argue that such highly stylized models are not suitable for econometric analysis in the first place. Notwithstanding, there exists a large body of literature that deals with the estimation of DSGE models (not necessarily concerned with asset pricing). Early studies rely on likelihood methods[1], although the complex model structures often render the likelihood function intractable. A more fundamental caveat concerns the notion of "true" model parameters that one seeks to estimate by maximum likelihood. Parts of a DSGE asset pricing model are openly misspecified; as put by Dridi et al. (2007), they represent a "caricature of reality". As such, it is unclear what an asymptotically efficient estimation achieves in the first place. Bayesian methods are regularly used for the estimation of DSGE models[2], but they lack the intuitive appeal of frequentist approaches, which are capable of directly linking the structural parameters to the DSGE model's equilibrium conditions, as noted by Fernández-Villaverde (2010). Moment matching strategies (generalized or simulated method of moments) are an alternative to likelihood-based methods[3], but the aforementioned caveat applies, if one doubts that data generated from the DSGE model endowed with the "true" parameters corresponds to the real world data generating process.

The present study pursues a matching approach that acknowledges that parts of the DSGE asset pricing model are evidently misspecified, while others claim to reflect economic reality. Specifically, we implement, for the first time in the context of DSGE asset pricing, the partial indirect inference philosophy proposed by Dridi et al. (2007) and focus our estimation efforts on those parameters and parts of the DSGE asset pricing model that are assumed to reflect economic reality: its asset pricing implications. We thereby build a bridge between the data generated by the DSGE asset pricing model and the data that we observe in the real world. To the best of our knowledge, this is the first study that provides a comprehensive

[1] See, e.g. Altug (1989), Leeper and Sims (1994), McGrattan et al. (1997) and Ireland (2004).

[2] An and Schorfheide (2007) and Fernández-Villaverde et al. (2016) provide reviews of the Bayesian DSGE model estimation approach.

[3] Christiano and Eichenbaum (1992) and Burnside et al. (1993) pioneer the literature that uses the generalized method of moments to estimate the full vector of structural parameters of DSGE models with analytically tractable unconditional moments. Regarding more complex models, for which unconditional moments cannot be provided in closed form, the literature resorts to matching the impulse response function parameters from VARs generated from both simulated data and empirical data, see for example Rotemberg and Woodford (1997), Christiano et al. (2005) and Altig et al. (2011). Ruge-Murcia (2007) and Ruge-Murcia (2012) match simulated moments a la Duffie and Singleton (1993) and McFadden (1989), while Ruge-Murcia (2014) uses indirect inference estimation introduced by Smith (1993) and Gourieroux et al. (1993) to match the parameters of a non-linear VAR.

application of partial indirect inference to econometrically analyze a DSGE asset pricing model. We make the general assumptions of Dridi et al. (2007) explicit, and derive the limiting distribution implied by the instrumental model that we employ for our study, which facilitates statistical inference about the estimated parameters. Applying the PII method to estimate the structural parameters of interest, we provide confidence intervals for these parameters as well as model-implied key economic indicators, which rely on the derived limiting distribution.

Our econometric analysis shows that the empirical performance of the model proposed by Chen (2017) is not unfavorable, but the claim that the DSGE approach can resolve the notorious asset pricing puzzles should be viewed with caution. On the upside, the model-implied equity premium and Sharpe ratio as well as the business cycle and consumption growth moments are economically plausible, also when taking estimation uncertainty into consideration. Moreover, the macroeconomic moments are largely unaffected by the choice of preference parameters and are mainly impacted by a sensible calibration choice. However, there are two notable caveats. First, the economically plausible equity premium and Sharpe ratio are not achieved by a level of risk aversion conventionally thought to be plausible.[4] Provided that one is willing to accept a steady-state relative risk aversion coefficient of about 20, it can be claimed that the DSGE asset pricing model with nonlinear habit persistence may help to resolve the equity premium puzzle. Second, while the estimated model does not imply an excessively volatile risk-free rate, which is often encountered in production-based models (Jermann, 1998), it fails to match the empirical level of the risk-free rate. The elevated level of the model-implied risk-free rate is associated with a time preference point estimate above unity. Accordingly, the risk-free rate puzzle remains unresolved.

The remainder of the paper is organized as follows: In Section 3.2.1, we present the anatomy of the DSGE asset pricing model to be analyzed, and we highlight its key features in Section 3.2.2. Calibration results are presented in Section 3.2.3. Section 3.3.1 outlines the econometric methodology. We use a brief exposition in the main text and explain methodological details in the appendix. We elaborate on the specific assumptions used for our instrumental model choice in Appendix 3.A and we derive the asymptotic distribution of the PII estimator in Appendix 3.B. Section 3.3.2 motivates the parameter separation into structural parameters of interest and nuisance parameters. Section 3.4 presents the estimation results and a critical assessment of the empirical performance of the estimated model. Section 3.5 concludes.

[4]This is is a well-documented feature of partial equilibrium habit-based asset pricing models (see, e.g. Campbell and Cochrane (1999), Cochrane (2016), Cochrane (2017)).

3.2. Asset pricing in a DSGE framework

In this section, we elaborate on the DSGE asset pricing model proposed by Chen (2017). This model is built on a standard RBC model that is augmented by nonlinear habit preferences in the spirit of Campbell and Cochrane (1999) and by convex capital adjustment costs. Section 3.2.1 gives details about the model specification. We focus on the model's asset pricing features in Section 3.2.2. More specifically, we investigate the model-implied risk-free rate and Sharpe ratio. In Section 3.2.3, we present calibration results of the model similar to those in Chen (2017).

3.2.1. Anatomy of the DSGE asset pricing model

The representative household chooses consumption and hours of labor such that the following expected lifetime utility function is maximized.

$$\mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \frac{(C_t - H_t)^{1-\gamma} - 1}{1-\gamma} \right], \tag{3.2.1}$$

where C_t and H_t are consumption and habit in time t, respectively, while β and γ represent the time preference and the utility curvature parameters. Chen (2017) assumes that there is no leisure, and that, in equilibrium, $N_t = 1$, i.e. the household is endowed with a single unit of labor. Following Campbell and Cochrane (1999), surplus consumption, S_t, determines the evolution of habit, such that $S_t \equiv \frac{C_t - H_t}{C_t}$. Surplus consumption follows an autoregressive process

$$s_{t+1} = (1 - \rho_s)\bar{s} + \rho_s s_t + \lambda(\Delta c_{t+1} - \mu), \tag{3.2.2}$$

where $s_t = \ln S_t$ indicates the log of the variable, ρ_s is an autoregressive parameter, μ is the steady-state growth rate of technology and λ is a constant fixed to $\lambda = 1/\bar{S} - 1$ as in Campbell and Cochrane (1999).

The utility function in Equation (3.2.1) is a power utility augmented with nonlinear external habit preferences. This specification ensures that the model has a time-varying risk premium given by $rra_t = \gamma / S_t$.

Markets are complete and the stochastic discount factor (SDF) of the households is given by

$$M_{t,t+1} = \beta \left(\frac{C_{t+1}}{C_t} \frac{S_{t+1}}{S_t} \right)^{-\gamma}. \tag{3.2.3}$$

On the production side, the representative firm produces output, Y_t, according to a constant returns to scale production function $Y_t = Z_t K_t^{\alpha} (X_t N_t)^{1-\alpha}$, where Z_t, K_t and N_t refer to the level of production technology, capital and labor available at time t, respectively. X_t is the

deterministic long-run growth component of productivity which evolves as $x_{t+1} = x_t + \mu$. As in a standard RBC model, the log of the production technology evolves in an autoregressive manner

$$z_{t+1} = \rho_z z_t + \sigma_z \varepsilon_{z,t+1}, \tag{3.2.4}$$

where $\varepsilon_{z,t+1} \sim \mathcal{N}(0,1)$ i.i.d. The technology shock is a homoskedastic shock, and it constitutes the only source of uncertainty in the model. Capital accumulation in this economy evolves according to

$$K_{t+1} = I_t + (1-\delta)K_t, \tag{3.2.5}$$

where I_t refers to investment at time t, and δ refers to the capital depreciation rate. Moreover, the representative firm faces convex capital adjustment costs $\Phi(I_t, K_t) = \frac{\phi}{2}\left(\frac{I_t}{K_t} - (e^{\mu} - 1 + \delta)\right)^2 K_t$.

These adjustment costs become zero in the steady state. The firm seeks to maximize the expected discounted value of future dividends, i.e.

$$\max_{\{I_t, K_{t+1}, N_t\}} \mathbb{E}_0\Big[\sum_{t=0}^{\infty} M_{0,t}\left\{Z_t K_t^{\alpha}(X_t N_t)^{1-\alpha} - W_t N_t - \Phi(I_t, K_t) - I_t\right\}\Big], \tag{3.2.6}$$

subject to capital accumulation in Equation (3.2.5).

In equilibrium, both the labor market and the goods market clear which gives rise to the equilibrium wage and the budget constraint, respectively

$$W_t = (1-\alpha) Z_t X_t K_t^{\alpha}(X_t N_t)^{-\alpha}, \tag{3.2.7}$$

$$C_t + I_t = Z_t K_t^{\alpha}(X_t N_t)^{1-\alpha} - \Phi(I_t, K_t), \tag{3.2.8}$$

as well as the asset pricing equation for investment return, $R^I_{t,t+1}$.

$$\mathbb{E}_t\left[M_{t,t+1} R^I_{t,t+1}\right] = 1, \tag{3.2.9}$$

with

$$R^I_{t,t+1} = \frac{\alpha \frac{Y_{t+1}}{K_{t+1}} + \left(1 + \phi\left(\frac{I_{t+1}}{K_{t+1}}\right)\right)(1-\delta) + \frac{\phi}{2}\left(\frac{I_{t+1}}{K_{t+1}}\right)^2}{1 + \phi\left(\frac{I_t}{K_t}\right)}, \tag{3.2.10}$$

where $\alpha \frac{Y_{t+1}}{K_{t+1}}$ is the marginal product of capital, $\frac{I_{t+1}}{K_{t+1}}$ is the investment rate, and ϕ is the quadratic adjustment cost parameter.

3.2.2. Features of the model

In the following, we discuss the main features of this model that resolve the asset pricing puzzles. We are primarily concerned with the risk-free rate puzzle and the equity premium puzzle.

Empirical data show that the firm or equity return volatility, $\sigma(R^I_{t,t+1})$, is 15% while the volatility of the investment return rate, $\frac{I_{t+1}}{K_{t+1}}$, is only 2% annually. Thus, the volatility of the investment return will not reconcile, see Equation (3.2.10), unless the adjustment cost parameter, ϕ, is calibrated at a very high value. In fact, Chen sets $\phi = 100$. This implies that productivity shocks are absorbed by asset prices rather than investment. This fact is also demonstrated in Jermann (1998), Kogan (2004), Jermann (2010), and Kogan and Papanikolaou (2012). Thus, high adjustment costs discourage volatile investments. Consequently, market clearing indicates that productivity shocks will alternatively imply volatile consumption which defies consumption smoothing preferences by households as indicated in empirical data. This dilemma, in turn, is solved by incurring low elasticity of intertemporal substitution EIS $= \frac{\partial \mathbb{E}_t[\Delta c_{t+1}]}{\partial r^f_{t+1}} = \frac{\bar{s}}{\gamma} \approx 0.035$.

However, low EIS propagates the problem to a different part in the model, as it implies a more volatile risk-free rate relative to empirical data (see the caveat in Jermann (1998)), which is precisely the risk-free rate puzzle. Persistent external habit preferences in this model smooth risk-free rate volatility via a precautionary savings effects. The log-normal approximation of the risk-free rate is given by

$$r^f_{t+1} \approx -\log\beta + \frac{\gamma}{\bar{S}}\mathbb{E}_t[\Delta c_{t+1}] - \gamma(1-\rho_s)(s_t - \bar{s}) - \frac{1}{2}\frac{\gamma^2}{\bar{S}^2}\mathrm{Var}_t[\Delta c_{t+1}], \tag{3.2.11}$$

where $\overline{rra} = \gamma/\bar{S}$ gives the steady-state relative risk aversion of the representative investor. Here, intertemporal substitution, the second term in the equation above, implies that in bad times, investors want to borrow from the future by selling the risk-free asset which, in turn, pushes r^f_{t+1} up. On the other hand, precautionary savings, the third term above, work in the opposite direction as they entail that, in a volatile economy, investors are more reluctant to sell the risk-free asset. As a result, given high persistence of habit, i.e. $\rho_s = 0.98$, a relatively smooth risk-free rate can coexist together with low EIS.

In this model, consumption volatility risk is endogenously generated. As the model is driven by an homoskedastic productivity shock, the heteroskedasticity arises from the nonlinearity in the law of motions of the model. This nonlinearity leads to counter-cyclical consumption volatility, as consumption is more sensitive to shocks in bad times. The precautionary savings channel is summarized as follows: a negative shock to wealth strengthens the motive to save which leads to an increase in savings and a decrease in

consumption. The opposite effect occurs for a positive wealth shock. Consequently, this uncertainty in the need for precautionary savings causes consumption volatility. Chen (2017) further shows that external habit amplifies the effect of this channel.

Consumption volatility risk in this model plays a crucial role in valuing assets as the asset pricing dynamics are derived from the time-varying risk premium which, in turn, is driven by the time-varying consumption volatility. From the model, the conditional maximum Sharpe ratio is derived via a log-normal approximation of the SDF as follows

$$\max_{\{\text{all assets}\}} \left[\frac{\mathbb{E}_t(R_{t+1} - R^f_{t+1})}{\sigma_t(R_{t+1})} \right] \approx \frac{\gamma}{\bar{S}} \sigma_t(\Delta c_{t+1}). \tag{3.2.12}$$

Equation (3.2.12) shows that the habit sensitivity is capable of amplifying the risk aversion. Thus, a small increase in consumption volatility can result in a large increase in risk premium.

3.2.3. Calibration

In this section, we replicate the results of Chen (2017). Towards this end, we calibrate the model and calculate the asset pricing implications and the business cycle moments. The moments that are targeted and their resulting parameter values are comparable to those reported in Campbell and Cochrane (1999) and Kaltenbrunner and Lochstoer (2010). The parameter values used to simulate the model are provided in Table 3.2.1. The technology parameters are calibrated to values common in the RBC literature: The capital share of GDP, α, is matched to the empirical target of this ratio and set to 0.35. The depreciation rate is guided by the mean investment rate and chosen to be $\delta = 0.016$. The total factor productivity (TFP) parameters are chosen as follows: long-run productivity growth $\mu = 0.0045$ is given by the mean output growth, the volatility of TFP is matched to the volatility of HP-filtered GDP, i.e. $\sigma_z = 0.012$, whereas the persistence of TFP is set to the persistence of the output-to-capital ratio, i.e. $\rho_z = 0.98$. The quadratic capital adjustment cost is chosen to be $\phi = 100$ comparable to Guvenen (2009) and Kaltenbrunner and Lochstoer (2010). This choice implies that productivity shocks are absorbed by asset prices and leads to less than 1% mean adjustment costs in relation to output.

The subjective time-preference rate of the representative agent, β, is chosen to match the mean return on the 90-day Treasury bill and is given by 0.995. Relative risk aversion is captured by the ratio of the utility curvature parameter, γ, to the surplus consumption S_t, i.e. $rra_t = \gamma/S_t$, and is time-varying. In the steady-state, risk aversion is given by $\overline{rra} = \gamma/\bar{S}$, and we cannot uniquely identify the two associated parameters. Here, the utility curvature parameter, γ, is set to 2 without targeting a specific moment, and the steady-state surplus consumption is chosen to be $\bar{S} = 0.07$ as in Campbell and Cochrane

Table 3.2.1: Calibrated values

Parameter		Value	Target Moment	Data
Preferences				
β	Time Preference	0.995	90-Day T-bill Return (%)	0.17
ρ_s	Persistence of Surplus Consumption	0.980		
$\bar{S}$	Steady-State Surplus Consumption	0.070		
γ	Utility Curvature	2.00	for comparison with Campbell-Cochrane	
$\overline{rra} = \gamma/\bar{S}$	Steady-State Relative Risk Aversion	28.57		
Long-run Technology				
α	Capital Share	0.35	Mean Output/Capital	0.143
δ	Depreciation Rate	0.016	Mean Investment Rate	0.025
μ	TFP Growth	0.0045	Mean Output Growth (%)	0.48
Cyclical Technology				
σ_z	Volatility of TFP	0.012	Vol HP-filtered GDP (%)	1.72
ρ_z	Persistence of TFP	0.980	Persistence of Output/Capital	0.997
ϕ	Adjustment Cost	100.00		
Financial Leverage				
ν	Debt to Firm Value	0.30		
J	Maturity New Debt	40		

Notes: This table replicates Table 1 in Chen (2017). It reports the parameter values obtained from matching empirical target moments. The time series used span a period from 1948 to 2012 on quarterly frequency from US data.

(1999).[5] Hence, the steady state relative risk aversion is calibrated to $\overline{rra} = 28.57$. Lastly, the persistence of surplus consumption, ρ_s, is guided by the persistence of Tobin's Q and the first autocorrelation of the price-dividend ratio and calibrated to 0.98.

We simulate the model 500 times over a time series length of $T = 260$, i.e. the length is equivalent to the number of quarters in our observed sample. The calibration results, as well as their counterparts from Chen (2017), are provided in Table 3.2.2. For the business cycle moments, the results implied by the simulations and the observed data, are quite similar, for both pairs of moments, with the exception of the mean output to capital ratio.[6] As for

[5] In Campbell and Cochrane (1999) a closed form solution for the steady-state surplus consumption exists and risk aversion is only captured in γ. Here, no closed form solution exist as the system is impacted by endogenous consumption volatility risk.

[6] This value is reported to be perfectly matched in Table 1 in Chen (2017), as he is reporting the empirical moment twice. We provide the corrected values in our Table 3.2.2.

Table 3.2.2: Calibration results

Moment		**Calibration**	**Chen (2017)**
	Data	**Simulations**	**Simulations**
Asset Prices			
90-Day T-bill Return (%)	0.17	0.65	0.35
Vol of R^f (%)	0.90	0.74	0.52
Persistence of Tobin's Q		0.96	0.96
Mean Sharpe Ratio of CRSP Index	0.22	0.23	0.16
Equity Premium R^e (%)	1.82	2.46	1.72
Vol of R^e (%)	8.24	10.70	11.04
Business Cycle			
Mean Output/Capital	0.143	0.089	0.142
Mean Investment Rate	0.025	0.021	0.021
Mean Output Growth (%)	0.48	0.45	0.45
Vol HP-filtered GDP (%)	1.72	1.52	1.53
Persistence of Output/Capital	0.997	0.986	0.983
Relative Volatility of Consumption Growth	0.52	0.51	0.46
Mean Adj Cost / Output (%)		0.11	0.11
Consumption Growth			
Mean of ln of Consumption Growth (%)	0.47	0.45	0.46
Std of ln of Consumption Growth (%)	0.52	0.61	0.56

Notes: This table shows simulated moments implied by the calibrated values given in Table 3.2.1. The calibrated structural parameters of interest are given by $\beta = 0.995$ and $\overline{rra} = 28.57$. The column labeled Chen (2017) contains the published values in his Tables 1, 2 and Table 3. The data used is taken from the online appendix of Chen (2017). It is described in more detail in Appendix 4.E.

the consumption growth moments, we find that the mean and the standard deviation of log consumption growth, implied by the model, match the empirical moments well.

For the asset pricing moments, the mean quarterly Sharpe ratio of the CRSP index is given by 0.22 which is equivalent to the sample average over the ensembles. The observed quarterly equity premium is equal to 1.82%, whereas the ensemble mean is 2.46%. Furthermore, the volatility of the excess return is sensible in the simulations. We use the 90-day T-bill return as a proxy of the risk-free rate. The average quarterly risk-free rate is given by 0.17% in the data, whereas it is considerably higher in our simulations and given by 0.65%. The volatility

of the model-implied risk-free rate at 0.74% is not excessively volatile and is in line with the observed counterpart at 0.90%. We can, therefore, conclude that the trade-off between the intertemporal substitution and the precautionary savings channel balances the volatility of risk-free rate well. Yet, the level of the risk-free rate is not compatible with our empirical data. The model is, thus, affected by the risk-free rate puzzle described in Weil (1989). Chen (2017) reports a lower level of the risk-free rate. However, this value cannot be replicated in neither his nor our code. Overall, we conclude from our simulations and that of Chen's (2017) that the model fares fairly well over the business cycle dimension, the consumption growth moments and for the equity premium, though it is not capable of matching the level of the empirical risk-free rate.

The previous calibration exercise seeks confirmatory evidence for the model by matching model simulated moments to those observed empirically using economically plausible parameter values. Yet, what if economically implausible parameter values are able to achieve such a match? From a pure calibration point of view, those will be deemed acceptable, though economically inconceivable. This is where econometric analysis goes beyond the scope of calibration. With the aid of estimated confidence intervals, we can assess the validity of the entire range of possible parameter values that are compatible with empirical data. In the next section, we lay out the partial indirect inference framework which we use to estimate the structural parameters of interest.

3.3. Returning to the econometrics fold

The indirect inference estimation philosophy is introduced and refined by Smith (1993), Gourieroux et al. (1993) and Gallant and Tauchen (1996). This approach is used in the presence of a nonlinear analytically intractable structural model, which implies that the parameters of such a model (known as the structural parameters), ξ, cannot be directly estimated. The underlying idea is to identify a binding function, $\theta(\xi)$, which relates the structural parameters ξ to instrumental parameters θ which, in turn, are estimated from an instrumental model that is easily estimated. Indirect inference implies that two sets of instrumental parameters are estimated from the instrumental model; one using the simulated data from the structural model, and the other from empirical data. The structural parameters are, then, estimated such that the minimum of a quadratic loss function calculated between the two sets of instrumental parameters is achieved.

3.3.1. The partial indirect inference approach

Intuition and motivation

The DSGE based asset pricing model does not represent a realistic model of the economic process. In spite of this, its stochastic discount factor may be able to price assets and convey economic reality. As a result of this misspecification, the empirical data generating process and the simulated data from the structural model do not coincide. This renders any indirect estimation strategy in which all structural parameters are estimated based on a correctly specified structural model, infeasible.

Alternatively, Dridi et al. (2007) proposes an extension to the indirect inference theory to accommodate misspecified structural models. This approach is called partial indirect inference, and it focuses on the consistent estimation of only a subset of the structural parameters; those that are believed not to be misspecified, while the rest of the parameters remain calibrated. In this manner, calibration can be viewed as econometrics for misspecified models. In their terminology, we split the parameters $\xi = (\xi_1, \xi_2)^\top$, such that the only parameters of interest for the estimation process are ξ_1, while ξ_2 refers to nuisance parameters that will remain calibrated for the entire exercise. The aim is to obtain a consistent estimator of the true unknown value ξ_1^0 when solving the sample or simulation-based counterpart of the binding function

$$\theta^0 = \tilde{\theta}^0 \left(\xi_1^0, \xi_2^* \right) \tag{3.3.1}$$

with respect to ξ_1. In this context, ξ_2^* are the pseudo-true values of the nuisance parameters ξ_2. Equation (3.3.1) constitutes a necessary condition which implies that the structural model fully encompasses the instrumental one in the presence of misspecification. Thus, if we choose a convenient instrumental model, Equation (3.3.1) characterizes the true unknown parameters ξ_1^0 given the misspecification in ξ_2.

If, however, only a subset of the encompassing conditions in (3.3.1) are fulfilled, we are in a partial encompassing case. Hence, the identification of the structural parameters ξ is incomplete, but we achieve consistent estimation of the structural parameters of interest ξ_1, if the under-identification is only about a subset of the pseudo-parameters ξ_2 that do not appear in the binding function. In this case, the necessary condition becomes

$$\theta_1^0 = \tilde{\theta}_1^0 \left(\xi_1^0, \bar{\xi}_{21} \right). \tag{3.3.2}$$

This means that the structural model's nuisance parameters are now $\xi_2 = (\xi_{21}, \xi_{22})^\top$, where ξ_{22} do not interfere with the binding function, and thus are allowed to escape the encompassing condition (3.3.2). These parameters, in turn, remain calibrated during the

estimation process, while their counterparts, ξ_{21}, become part of the estimation process, i.e. they are identified from the instrumental parameters, θ_1, and are thus necessary for the consistent estimation of ξ_1.

PII estimator

In this subsection, we detail the construction of an indirect estimation strategy for the structural parameters of interest ξ_1. We denote the empirical time series as $\{y_t\}_{t=1}^T$, and the model-implied series obtained from simulations by $\{\tilde{y}_s(\xi_1, z_0)\}_{s=1}^S$, where z_0 refers to the initial value of the state variable.

The crucial step is to carefully choose an instrumental model which is capable of capturing the model-implied financial time series, and has a tractable estimation function for its parameters, i.e. the instrumental parameters. Similar to the recommendations of Dridi et al. (2007), we advocate the use of well-chosen moment restrictions that reflect the asset pricing implications of the DSGE asset pricing model rather than using a full parametric model as an instrumental model. We, first, describe the estimation of the instrumental parameters based on empirical data.[7] The set of k moment restrictions is given by

$$\mathbb{E}\left[u_{1,t}\left(\{y_t\}_{t-l}^t, \theta_1^0\right)\right] = 0. \tag{3.3.3}$$

Moreover, the number of moment restrictions is at least as large as the number of structural parameters, i.e. $k = \dim \theta_1 \geq p_1 + p_{21} = \dim \xi_1 + \dim \xi_{21}$. To estimate the instrumental parameters θ_1, we maximize a GMM-type criterion function given by

$$\min_{\{\theta_1 \in \Theta_1\}} Q_{1,T}\left(\{y_t\}_{t=1}^T, \theta_1\right), \tag{3.3.4}$$

$$\text{where} \quad Q_{1,T}\left(\{y_t\}_{t=1}^T, \theta_1\right) = \frac{1}{2}\, g_{1,T}\left(\{y_t\}_{t=1}^T, \theta_1\right)^\top \cdot \widehat{W}_{1,T} \cdot g_{1,T}\left(\{y_t\}_{t=1}^T, \theta_1\right), \tag{3.3.5}$$

$$g_{1,T}\left(\{y_t\}_{t=1}^T, \theta_1\right) = \frac{1}{T}\sum_{t=1}^T u_{1,t}\left(\{y_t\}_{t-l}^t, \theta_1\right), \tag{3.3.6}$$

with $\widehat{W}_{1,T} \xrightarrow{a.s.} W_1$, a positive semi-definite weighting matrix.

$$\hat{\theta}_{1,T} = \underset{\{\theta_1 \in \Theta_1\}}{\operatorname{argmin}}\, Q_{1,T}\left(\{y_t\}_{t=1}^T, \theta_1\right)$$

is a consistent estimator of θ_1^0.

[7] We use the notation of the partial encompassing case in Dridi et al. (2007). All quantities are indexed by a 1.

As the simulated data stems from a misspecified simulator, its probability limit is not the same as that of the observed data. For consistency of the instrumental parameter estimates associated with the simulated time series, we impose an identification assumption for the simulated time series given by

$$\mathbb{E}^* \left[u_{1,t} \left(\{\tilde{y}_t^s(\xi)\}_{t-l}^t, \tilde{\theta}_1^0(\xi_1^0, \bar{\xi}_{21}) \right) \right] = 0.$$

Under appropriate GMM assumptions given in Appendix 3.A,

$$\tilde{\theta}_{1,T}^s(\xi_1, \xi_{21}, \bar{\xi}_{22}) = \underset{\{\theta_1 \in \Theta_1\}}{\operatorname{argmin}} \, Q_{1,T} \left(\{\tilde{y}_T^s(\xi)\}_{t=1}^T, \theta_1^s \right)$$

and $\tilde{\theta}_{1,T}(\xi_1, \xi_{21}, \bar{\xi}_{22}) = \frac{1}{S} \tilde{\theta}_{1,T}^s(\xi_1, \xi_{21}, \bar{\xi}_{22})$ is a consistent estimator of $\tilde{\theta}_1^0 \left(\xi_1^0, \bar{\xi}_{21} \right)$.

For the indirect estimator of the structural parameters, we require partial encompassing given by

$$\theta_1^0 = \tilde{\theta}_1^0(\xi_1^0, \bar{\xi}_{21}), \tag{3.3.7}$$

where θ_1^0 is probability limit of the instrumental parameters associated with the observed data and $\tilde{\theta}_1^0(\xi_1^0, \bar{\xi}_{21})$ is the probability limit for the simulated time series. Equation (3.3.7) is the partial encompassing condition in Dridi et al. (2007) and the necessary condition in Equation (3.3.2) needed in the partial encompassing case. We refer to it as the partial encompassing null hypothesis $H_0^{(1)} \left(\bar{\xi}_{22} \right)$.

The partial indirect inference (PII) estimator is defined as

$$\begin{aligned} \begin{pmatrix} \hat{\xi}_{1,TS}(\bar{\xi}_{22}) \\ \hat{\xi}_{21,TS}(\bar{\xi}_{22}) \end{pmatrix} = \underset{\{(\xi_1, \xi_{21}) \in \Xi_1 \times \Xi_{21}\}}{\operatorname{argmin}} & \left[\hat{\theta}_{1,T} - \frac{1}{S} \sum_{s=1}^{S} \tilde{\theta}_{1,T}^s(\xi_1, \xi_{21}, \bar{\xi}_{22}) \right]^\top \cdot \widehat{\Omega}_{1,T} \\ & \cdot \left[\hat{\theta}_{1,T} - \frac{1}{S} \sum_{s=1}^{S} \tilde{\theta}_{1,T}^s(\xi_1, \xi_{21}, \bar{\xi}_{22}) \right], \end{aligned} \tag{3.3.8}$$

$$\hat{\theta}_{1,T} = \underset{\{\theta_1 \in \Theta_1\}}{\operatorname{argmin}} \, Q_{1,T} \left(\{y_t\}_{t=1}^T, \theta_1 \right), \tag{3.3.9}$$

$$\tilde{\theta}_{1,T}^s \left(\xi_1, \xi_{21}, \bar{\xi}_{22} \right) = \underset{\{\theta_1 \in \Theta_1\}}{\operatorname{argmin}} \, Q_{1,T} \left(\{\tilde{y}_T^s(\xi)\}_{t=1}^T, \theta_1^s \right), \tag{3.3.10}$$

and $\widehat{\Omega}_{1,T}$ is positive definite weighting matrix that converges almost surely to the non-stochastic positive definite matrix Ω_1.

Given that $\hat{\xi}_{1,TS}\left(\bar{\xi}_{22}\right)$ is a consistent estimator of ξ_1^0, the structural parameters estimates of ξ_1 and ξ_{21} follow an asymptotic normal distribution

$$\sqrt{T}\begin{pmatrix}\hat{\xi}_{1,TS}\left(\bar{\xi}_{22}\right)-\xi_1^0\\ \hat{\xi}_{21,TS}\left(\bar{\xi}_{22}\right)-\bar{\xi}_{21}\end{pmatrix}\xrightarrow{d}\mathcal{N}\left(0,\Sigma\left(S,\Omega_1\right)\right), \tag{3.3.11}$$

where the asymptotic covariance $\Sigma\left(S,\Omega_1\right)$ is given in Equation (3.B.1) in Appendix 3.B. This asymptotic covariance matrix can be used to construct confidence intervals of our parameters of interest. Using this technique, we can return calibration to the econometrics fold.

The underlying assumptions that ensures the consistency of the PII estimator under the partial encompassing case are discussed in Appendix 3.A and 3.B. We refer the reader to Dridi et al. (2007), as well as Smith (1993), Gourieroux et al. (1993) and Gallant and Tauchen (1996) for the assumptions governing the consistency of indirect inference estimation in general.

3.3.2. Specification of the instrumental model

The PII estimation methodology is adapted to the structural model at hand, the DSGE asset pricing model proposed by Chen (2017). Here, we conceive the asset pricing implications of the model, the implied SDF, to be the economic reality that we want to capture. Hence, the investor's preference parameters are of relevance, i.e. $\xi_1 = (\beta, \overline{rra})^\top$ where $\overline{rra} = \gamma/\bar{S}$, while the remaining parameters are understood as caricatural aspects of the model perceived as the nuisance parameters, ξ_2. We investigate which informational content the empirical data has on the implied timing and risk preference. As the size of the relative risk aversion and the time preference rate carry economic meaning, this is crucial for the empirical assessment of the model. These parameters play a key role in consumption-preference based asset pricing which determine the model-implied equilibrium risk-free rate and the market equity premium. If the parameter estimates are outside the range of economically plausible values, this is empirical evidence against the structural model. Moreover, if the estimated confidence intervals of the parameters include a wide range of plausible and implausible values, then the observed data are not informative for assessing the empirical performance of the model.

The structural nuisance parameters, ξ_2, are needed to specify a complete model, but they do not convey economic reality. The bridge between the theoretical model and the economic reality is fragile, hence, we rely on calibration for the macroeconomic part of the model. We argue that the risk-free rate and the market portfolio are present in the structural model and empirical equivalents exist in the real world. Proxies of both quantities are traded on asset markets. Moreover, we can justify an econometric analysis by matching empirical and

simulated moments. In contrast, the Cobb-Douglas production function with adjustment costs in our structural model does not have an empirical equivalent. We resort to calibration and choose plausible values which may not be identifiable from the observed data. Thus, the nuisance parameters constitute the macroeconomic parameters as well as the persistence of surplus consumption. They are given by $\bar{\xi}_2 = (\rho_s, \alpha, \delta, \phi, \mu, \rho_z, \sigma_z)^\top$ and are calibrated to the values specified in Section 3.2.3, i.e. $\bar{\xi}_2 = (0.98, 0.35, 0.016, 100, 0.0045, 0.98, 0.012)^\top$.

The choice of moment conditions for the instrumental parameters is guided by the model-implied equations for the risk-free rate and the conditional Sharpe ratio which are given in Equations (3.2.11) and (3.2.12). The dynamics of the risk-free rate is affected by the time preference parameter, β, as well as the intertemporal substitution and the precautionary savings motive. These, in turn, are driven by the steady state relative risk aversion $\overline{rra} = \gamma/\bar{S}$. In addition, the conditional Sharpe ratio is influenced by the steady state relative risk aversion. Thus, we make use of the unconditional time-series averages of the risk-free rate $\mathbb{E}(R_t^f) = \mu_{r_f}$. Moreover, we would like to use the Sharpe ratio given by $\frac{\mathbb{E}(R_t - R_t^f)}{\sigma(R_t - R_t^f)}$ to identify the risk aversion parameter. However, as no Law of Iterated Expectations exists for ratios of moments, the conditional Sharpe ratio cannot be transformed into an unconditional Sharpe ratio readily. Moreover, it is also unclear whether the ratio of the empirical equivalents of the expectation and standard deviation converges in probability to the population equivalents. Hence instead of the Sharpe ratio, we propose to use the mean of the equity premium $\mathbb{E}(R_t - R_t^f) = \mathbb{E}(R_t^e) = \mu_{r_e}$ and its second moment $\text{Var}(R_t - R_t^f) = \sigma^2(R_t - R_t^f) = \sigma^2(R_t^e) = \sigma_{r_e}^2$ separately. As a fourth moment, we use the variance of the risk-free rate $\text{Var}(R_t^f) = \sigma^2(R_t^f) = \sigma_{r_f}^2$.

Hence, the moment restrictions used to estimate the instrumental parameters $\theta_1 = \left(\mu_{r_f}, \mu_{r_e}, \sigma_{r_e}^2, \sigma_{r_f}^2\right)^\top$ are given by

$$\mathbb{E}\left[u_{1,t}\left(\{y_t\}_{t=1}^T, \theta_1\right)\right] = \mathbb{E}\begin{bmatrix} R_t^f - \mu_{r_f} \\ R_{m,t}^e - \mu_{r_m^e} \\ (R_{m,t}^e)^2 - (\mu_{r_e})^2 - (\sigma_{r_m^e})^2 \\ (R_t^f)^2 - (\mu_{r_f})^2 - (\sigma_{r_f})^2 \end{bmatrix} = 0. \tag{3.3.12}$$

Note the moment conditions yield an exactly identified problem. Hence there is no need to numerically optimize the instrumental objective function given in Equation (3.3.5). Instead, we use the sample equivalents of the moments specified above.

The classification of the structural parameters at the beginning of this section raises the question of whether the partial encompassing condition holds for our parameters of interest $\xi_1 = (\beta, \overline{rra})$ and the chosen moment conditions. Equation (3.2.11) includes the

persistence of habit, ρ_s, as a component of the intertemporal substitution, yet ρ_s is considered a nuisance parameter in our classification. A more refined classification is also plausible, where ρ_s qualifies as a ξ_{21} parameter. This choice would imply that the moment conditions in our instrumental model should identify ρ_s, e.g. by the first order autocorrelation of the price-dividend ratio. We leave this parameter classification and choice of instrumental model to future research.

3.4. Results

The econometric assessment of Chen's (2017) model is conducted in various steps. We start by estimating the structural parameters of interest, ξ_1. Then, we compute approximate confidence intervals around these estimates to evaluate which preference parameter values are compatible with the observed data. Finally, we implement a parametric bootstrap to assess the impact of the estimation uncertainty on the DSGE asset pricing-implied moments that are matched to the empirical data.

To find the estimates of ξ_1, we initially perform a grid search for the ξ_1 parameter combination that returns the minimum value of the PII objective function described in Equation (3.3.8) together with our choice of moment conditions given by Equation (3.3.12). Details of the grid are described in Appendix 3.C.[8] From the minimum value of the grid search, we start a numerical optimization of the objective function with respect to ξ_1. The optimization is based on the non-gradient based Neldor-Mead algorithm.[9] It yields an estimate of the time preference rate, $\hat{\beta}$, of 1.0008, hence, a time preference rate larger than one, and a relative risk aversion estimate, $\widehat{\overline{rra}}$, of 22.6621.

We, then, compute approximate confidence intervals based on the asymptotic distribution described in Appendix 3.A. The estimates, the standard errors and the 95% confidence interval are given in Table 3.4.1. Both confidence intervals are relatively tight around the estimated values, as the standard errors are small in magnitude. Hence, only a small range of preference parameters values are capable of matching the simulated moments to their empirical counterparts. This implies that these parameters are indeed identifiable from the data.

The steady-state relative risk aversion estimate is given by $\widehat{\overline{rra}} = 22.66$, and its confidence interval ranges from around 18.3 to 27. The calibrated value of 28.57, the rra value also used by Campbell and Cochrane (1999) and Chen (2017), is not included in the confidence

[8]The grid is minimized for values of $\beta = 1.001$ and $\overline{rra} = 22.5$. These parameter values are inner solutions of the specified grid, and the objective function is not flat around the minimum.

[9]As a robustness check, we also start the numerical optimization from different starting values and obtain similar results. This is unsurprising as the objective function has a clear minimum.

Table 3.4.1: Partial indirect inference estimation results

			Confidence Interval	
Parameter	**Estimate**	**Standard Error**	**2.5%**	**97.5%**
β	1.0008	0.0034	0.9941	1.0075
$\overline{rra}$	22.66	2.20	18.34	26.98

Notes: This table reports the estimation results for the parameters of interest ξ_1, i.e. β and $\overline{rra}$. It also provides standard errors and 95% confidence intervals.

interval and, hence, it is not compatible with the empirical data. Another concern is that the model implied steady-state relative risk aversion is not within the microeconomic evidence based relative risk aversion range of 1 to 5 proposed by Mehra and Prescott (1985). Hence, risk aversion in our model is in fact not low, and the estimated parameter value, along with its confidence interval, is economically not plausible.[10] It is, however, smaller than in a pure consumption-based asset pricing model with power utility preferences, where astronomical risk aversion values, in a magnitude of 40 up to 200, are needed to resolve the equity premium puzzle (c.f. Weil (1989)). In the partial-equilibrium habit persistence model proposed by Campbell and Cochrane (1999) and the general equilibrium model evaluated here, the steady state value is given by $\overline{rra} = 28.57$. Hence, it is even larger than our $\overline{rra}$ estimate. The critique of a relatively high risk aversion coefficient value applies to all habit driven asset pricing models. Cochrane (2016) and (2017) refers to this problem as the equity-premium risk-free rate puzzle. In a nutshell, habit persistence models yield plausible equity premia with a relatively low risk-free rate and sensible consumption growth moments, but the model-implied risk aversion is not low.[11] The results on the risk aversion coefficient are, thus, in line with the habit formation asset pricing literature.

The implications of the time preference rate, β, results are even more troubling. The parameter estimate is given by $\hat{\beta} = 1.0008$, and its confidence interval ranges from 0.994 to 1.0075. The calibrated time preference rate of $\beta = 0.995$ is included in the approximate confidence interval and it is compatible with the observed data. In the asset pricing literature, e.g. in Mehra and Prescott (1985), Weil (1989), Campbell and Cochrane (1999) and Cochrane (2009), we typically find values of the time preference rate, β, smaller than

[10]Note that parameter of the utility function, γ, is set to 2 but risk aversion in our model is given by $rra_t = \frac{\gamma}{S_t}$.

[11]The puzzle is, hence, distilled in sensible asset pricing moments obtained without having low utility curvature and risk aversion. He concludes that no model so far has achieved a full solution of the equity premium puzzle.

unity. Any value of the time preference rate above one implies that the representative agent prefers consumption tomorrow over consumption today; a fact that is not supported by microeconomic experiments. The confidence interval for the time preference rate includes many values that are in excess of unity and, hence, these values are economically implausible. This finding is linked to the risk-free rate puzzle described by Weil (1989). The large value of β can be explained by a model-implied risk-free rate that is too large in comparison to its empirical equivalent. The high model-implied risk-free rate is slightly mitigated by using a value of the time preference rate that is larger than one when we match empirical and simulated moments in the PII estimation. In this case, the first term in Equation (3.2.11) in the risk-free rate, $-\log(\beta)$, is negative and decreases the level of the risk-free rate.

In addition to the sensibility of the preference parameter estimates, we assess the economic plausibility of the model-implied moments using a parametric bootstrap. We achieve this by drawing randomly from the joint normal distribution of the estimates of the structural parameters of interest based on the asymptotic results, i.e. from

$$\xi_1^* \sim \mathcal{N}\left(\begin{pmatrix} 1.0008 \\ 22.6621 \end{pmatrix}, \begin{pmatrix} 0.0000116 & -0.0064 \\ -0.0064 & 4.8617 \end{pmatrix}\right). \tag{3.4.1}$$

We use $M = 1000$ draws of β and $\overline{rra}$ to solve and simulate the model at the respective values. Then, we compute the implied asset pricing and macroeconomic moments and calculate the mean over all M samples, as well as the 2.5% and 97.5% bootstrap quantiles. The bootstrap results are given in Table 3.4.2.

The table shows that the macroeconomic moments governing the business cycle and consumption growth do not change qualitatively with the change of β and $\overline{rra}$ as compared to Table 3.2.2. Moreover, the business cycle moments are not sensitive to our bootstrapping exercise. Most of those moments are not affected by different draws of the preference parameters; the bootstrap intervals are either just one value or their width is very small. Hence, the impact of estimation uncertainty on these model-implied moments is, thus, negligible.

The resulting asset pricing moments are impacted by the different choice of preference parameters in comparison to the calibration results in Table 3.2.2. The model-implied Sharpe ratio is slightly lower than that implied by US data. This result is due to the fact that the volatility of the excess return is slightly too high, and the level of expected excess return is slightly too low. Both of these quantities are, however, better matched to US data in comparison to the calibrated values. The simulated mean risk-free rate is higher than the one implied by US data, while its volatility is slightly lower. Furthermore, the asset pricing moments are impacted by the different bootstrap draws of the preference parameters

Table 3.4.2: Estimation-implied economic plausibility check

Target Moment			**Confidence Interval**	
	Data	**Implied Estimate**	**2.5%**	**97.5%**
Asset Prices				
90-Day T-bill Return (%)	0.17	0.61	0.36	0.82
Vol of R^f (%)	0.90	0.67	0.59	0.76
Persistence of Tobin's Q		0.96	0.96	0.96
Mean Sharpe Ratio of CRSP Index	0.22	0.17	0.11	0.23
Equity Premium R^e (%)	1.82	1.64	1.04	2.35
Vol of R^e (%)	8.24	9.85	9.14	10.78
Business Cycle				
Mean Output/Capital	0.143	0.076	0.064	0.091
Mean Investment Rate	0.025	0.021	0.021	0.021
Mean Output Growth (%)	0.48	0.45	0.45	0.45
Vol HP-filtered GDP (%)	1.72	1.53	1.52	1.53
Persistence of Output/Capital	0.997	0.984	0.981	0.985
Relative Volatility of Consumption Growth	0.52	0.47	0.40	0.53
Mean Adj Cost / Output (%)		0.11	0.10	0.11
Consumption Growth				
Mean of ln of Consumption Growth (%)	0.47	0.45	0.45	0.45
Std of ln of Consumption Growth (%)	0.52	0.56	0.48	0.63

Notes: This table shows the results of a parametric bootstrap using $M = 1000$ draws from the joint normal distribution of β and $\overline{rra}$ around the PII estimates $\hat{\beta} = 1.0008$ and $\widehat{\overline{rra}} = 22.6621$. It shows the bootstrapped moments of asset prices, business cycle and consumption growth along with the estimate-implied and empirical means.

and are, thus, affected by estimation uncertainty. The equity premium moments lie within the bootstrap interval bounds, and the empirical Sharpe ratio is equivalent to the upper bound. The widths of the bootstrapped intervals for all three quantities are small and contain economically plausible values. Hence, the equity premium implied by our model is economically sensible. This finding is in line with the habit formation literature summarized by Cochrane (2017).

Additionally, the bootstrapping results confirm that the DSGE asset pricing model is not able to match the risk-free rate moments. The mean as well as the 2.5% and the 97.5% bootstrapping interval bounds for the expected risk-free rate are larger than the empirical

counterpart. Hence, the risk-free rate puzzle is also apparent here. The volatility of the risk-free rate implied by the bootstrap draws is slightly lower than the empirical one. Hence, the model-implied risk-free rate is not excessively volatile, but its level is too high.

Over all, the empirical performance of the DGSE extension to the asset pricing model should be taken with a grain of salt as it does not solve the asset pricing puzzles as well as implied in the calibration results of Chen (2017). We find that the equity premium puzzle is only resolved if we accept a value for the relative risk aversion that is around 20. This value is, however, above the range of risk aversion values conventionally found in the microeconomic evidence based literature (c.f. Mehra and Prescott (1985)). More problematic, however, is the unresolved risk-free rate puzzle, as we find that the model cannot match the risk-free rate level in the U.S. economy. This, in turn, is reflected in a time preference rate in excess of unity which is economically inconceivable.

3.5. Conclusion

Incorporating preference-based asset pricing in a DSGE framework holds the promise to resolve prominent puzzles of financial economics with a stochastic discount factor that is consistent with macroeconomic processes. However, the apparent empirical success of the DSGE asset pricing approaches, the most recent exponent of which is the model by Chen (2017), is predominantly based on calibration studies, which have been criticized by Hansen and Heckman (1996) and Dridi et al. (2007) for lacking the discipline of econometric analysis. On the other hand, Kim and Pagan (1999; p. 328) argue that when dealing with such highly stylized models, "the specification errors being committed are of sufficient magnitude to make conventional estimation and testing of dubious value".

In the debate between proponents of calibration and econometricians, we concur with Dridi et al. (2007) that even if a model is misspecified, economic reality is captured by certain parameters of interest that one should aim to estimate consistently. Pursuing such a goal entails focusing on economically meaningful identifying restrictions that are associated with the parts of a DSGE asset pricing model that claim to capture economic reality. Moreover, empirical tests of the model should reach beyond the confirmatory nature of calibration practices. To implement these ideas, we employ the partial indirect inference (PII) framework proposed by Dridi et al. (2007). In line with the PII philosophy, we use binding functions that aim at the consistent estimation of some structural model parameters of interest (reflecting investor preferences), while treating others (associated with macroeconomic dynamics) as nuisance parameters.

Our results indicate that the very positive calibration results regarding the empirical performance of DSGE asset pricing models should not be overstated. The estimated model

is able to match the equity premium and the Sharpe ratio, as well as the business cycle and the consumption growth moments. These moments remain economically plausible even when we take estimation uncertainty into consideration. The macroeconomic moments are mainly impacted by a sensible calibration choice and only slightly affected when varying the estimated preference parameters. However, although the equity premium and the Sharpe ratio implied by the estimated model are plausible, they are associated with a point estimate of relative risk aversion that is generally considered as too high (as are the bounds of the 95% confidence interval). The estimated model only provides a solution to the equity premium puzzle, if we one is willing to accept a risk aversion coefficient in the magnitude of around 20. This is a familiar caveat, which applies to partial equilibrium habit-based asset pricing models (e.g. Campbell and Cochrane (1999), Cochrane (2016), Cochrane (2017)). Our results suggest that it applies to DSGE asset pricing models with nonlinear habit preferences, too. The estimation also does not yield economically sensible model-implied values of the risk-free rate, neither in terms of the point estimate of the mean risk-free rate, nor in terms of the bounds of the 95% confidence interval. Whereas the model does not suffer from excessively volatile risk-free rates, it fails to match the empirical level of the risk-free rate. The implausibly high value for the model-implied mean risk-free rate is associated with a point estimate of the time preference parameter greater than unity. Hence, the DSGE asset pricing model is unable to resolve the risk-free rate puzzle described in Weil (1989). Our econometric analysis, thus, suggests that empirical asset pricing within a DSGE framework, although it shows some potential, is not yet a panacea to resolve the prominent asset pricing puzzles.

References

Altig, D., L. J. Christiano, M. Eichenbaum, and J. Linde (2011). "Firm-specific capital, nominal rigidities and the business cycle". *Review of Economic Dynamics* 14. (2), 225–247.

Altug, S. (1989). "Time-to-build and aggregate fluctuations: some new evidence". *International Economic Review* 30. (4), 889–920.

An, S. and F. Schorfheide (2007). "Bayesian analysis of DSGE models". *Econometric Reviews* 26. (2-4), 113–172.

Burnside, C., M. Eichenbaum, and S. Rebelo (1993). "Labor hoarding and the business cycle". *The Journal of Political Economy* 101. (2), 245–273.

Campbell, J. Y. and J. H. Cochrane (1999). "By force of habit: A consumption-based explanation of aggregate stock market behavior". *The Journal of Political Economy* 107, 205–251.

Chen, A. Y. (2017). "External habit in a production economy: A model of asset prices and consumption volatility risk". *The Review of Financial Studies* 30. (8), 2890–2932.

Christiano, L. J. and M. Eichenbaum (1992). "Current real-business-cycle theories and aggregate labor-market fluctuations". *The American Economic Review* 82. (3), 430–450.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (2005). "Nominal rigidities and the dynamic effects of a shock to monetary policy". *The Journal of Political Economy* 113. (1), 1–45.

Cochrane, J. H. (2009). *Asset Pricing*. Princeton University Press.

— (2016). *The Habit Habit*. Economics Working Papers 16105. Hoover Institution, Stanford University.

— (2017). "Macro-finance". *Review of Finance* 21. (3), 945–985.

Dridi, R., A. Guay, and E. Renault (2007). "Indirect inference and calibration of dynamic stochastic general equilibrium models". *Journal of Econometrics* 136. (2), 397–430.

Dridi, R. and E. Renault (2000). "Semi-parametric indirect inference". *LSE STICERD Research Paper No. EM392*.

Duffie, D. and K. J. Singleton (1993). "Simulated moments estimation of Markov models of asset prices". *Econometrica* 61. (4), 929–952.

Fernald, J. (2014). "A quarterly, utilization-adjusted series on total factor productivity". In: Federal Reserve Bank of San Francisco.

Fernández-Villaverde, J. (2010). "The econometrics of DSGE models". *SERIEs* 1. (1-2), 3–49.

Fernández-Villaverde, J., J. F. Rubio-Ramírez, and F. Schorfheide (2016). "Solution and estimation methods for DSGE models". *Handbook of Macroeconomics* 2, 527–724.

Gallant, A. R. and G. Tauchen (1996). "Which moments to match?" *Econometric Theory* 12, 657–681.

Gourieroux, C., A. Monfort, and E. Renault (1993). "Indirect inference". *Journal of Applied Econometrics* 8, 85–118.

Guvenen, F. (2009). "A parsimonious macroeconomic model for asset pricing". *Econometrica* 77. (6), 1711–1750.

Hansen, L. P. (1982). "Large sample properties of generalized method of moments estimators". *Econometrica* 50. (4), 1029–1054.

Hansen, L. P. and J. J. Heckman (1996). "The empirical foundations of calibration". *The Journal of Economic Perspectives* 10. (1), 87–104.

Ireland, P. N. (2004). "Technology shocks in the new Keynesian model". *Review of Economics and Statistics* 86. (4), 923–936.

Jermann, U. J. (1998). "Asset pricing in production economies". *Journal of Monetary Economics* 41. (2), 257–275.

Jermann, U. J. (2010). "The equity premium implied by production". *Journal of Financial Economics* 98. (2), 279–296.

Kaltenbrunner, G. and L. A. Lochstoer (2010). "Long-run risk through consumption smoothing". *The Review of Financial Studies* 23. (8), 3190–3224.

Kim, K. and A. R. Pagan (1999). "The econometric analysis of calibrated macroeconomic models". In: *Handbook of Applied Econometrics*. Ed. by Pesaran, H. and Wickens, M. Vol. 1: Macroeconomics. Oxford: Blackwell Publishing Ltd. Chap. 7, 309–338.

Kogan, L. (2004). "Asset prices and real investment". *Journal of Financial Economics* 73. (3), 411–431.

Kogan, L. and D. Papanikolaou (2012). "Economic activity of firms and asset prices". *Annual Review Financial Economics* 4. (1), 361–384.

Leeper, E. M. and C. A. Sims (1994). "Toward a modern macroeconomic model usable for policy analysis". *NBER Macroeconomics Annual* 9, 81–118.

McFadden, D. (1989). "A method of simulated moments for estimation of discrete response models without numerical integration". *Econometrica* 57. (5), 995–1026.

McGrattan, E. R., R. Rogerson, and R. Wright (1997). "An equilibrium model of the business cycle with household production and fiscal policy". *International Economic Review* 38. (2), 267–290.

Mehra, R. and E. C. Prescott (1985). "The equity premium: A puzzle". *Journal of Monetary Economics* 15, 145–161.

Newey, W. K. and K. D. West (1987). "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix". *Econometrica* 55. (3), 703–708.

Rotemberg, J. J. and M. Woodford (1997). "An optimization-based econometric framework for the evaluation of monetary policy". *NBER Macroeconomics Annual* 12, 297–346.

Rouwenhorst, G. K. (1995). *Asset returns and business cycles in frontiers of business cycle research*. Ed. by Cooley, T. F. Princeton University Press.

Ruge-Murcia, F. (2012). "Estimating nonlinear DSGE models by the simulated method of moments: With an application to business cycles". *Journal of Economic Dynamics and Control* 36. (6), 914–938.

— (2014). *Indirect inference estimation of nonlinear dynamic general equilibrium models: with an application to asset pricing under skewness risk*. Tech. rep. Centre interuniversitaire de recherche en économie quantitative, CIREQ.

Ruge-Murcia, F. J. (2007). "Methods to estimate dynamic stochastic general equilibrium models". *Journal of Economic Dynamics and Control* 31. (8), 2599–2636.

Singleton, K. J. (2009). *Empirical dynamic asset pricing: Model specification and econometric assessment*. Princeton University Press.

Smith, A. A. (1993). "Estimating nonlinear time-series models using simulated vector autoregressions". *Journal of Applied Econometrics* 8, 63–84.

Weil, P. (1989). "The equity premium puzzle and the risk-free rate puzzle". *Journal of Monetary Economics* 24, 401–421.

Appendix 3.A The partial indirect inference estimator with instrumental parameters estimated by moment conditions

We construct an indirect estimation strategy for the structural parameters of interest ξ_1. In the following, we denote empirical time series as $\{y_t\}_{t=1}^T$ and we obtain the model-implied series through simulations from $\{\tilde{y}_T^s(\xi, z_0^s)\}_{s=1}^S$ where z_0^s contains the initial value of the state variables. As in Dridi et al. (2007), we assume that the observed data is a realization of a stochastic process with true unknown p.d.f. $P_0 : \{y_t, t \in \mathbb{Z}\}$

Assumption 3.A.1. (i) P_0 belongs to a family $\mathcal{P}$ of p.d.f. on $\mathcal{Y}^{\mathbb{Z}}$,

(ii) $\tilde{\xi}_1$ is an application of $\mathcal{P}$ onto a part $\Xi_1 = \tilde{\xi}_1(\mathcal{P})$ of $\mathbb{R}^{p_1}$,

(iii) $\tilde{\xi}_1(\mathcal{P}_0) = \xi_1^0$, the true parameters of interest belongs to the interior Ξ_1^0 of Ξ_1.

These structural parameters of interest, ξ_1, are defined through a set of identifying moment conditions given by a partial parametric model

$$\mathbb{E}_{\mathcal{P}}\left[h(y_t, \ldots, y_{t-p}; \xi_1)\right] = 0 \quad \rightarrow \ \xi_1 = \tilde{\xi}_1(\mathcal{P}).$$

To obtain a fully parametric model, we have to make additional nominal assumptions. This model is likely inconsistent with the true data-generating process. The partial parametric model defined by Assumption 3.A.1 is plugged into the structural model which is fully parametric but misspecified.

Assumption 3.A.2 (Nominal assumption). $\{y_t, t \in \mathbb{Z}\}$ follows a stationary and ergodic process conformable to the nonlinear simultaneous model

(i) $r(y_t, y_{t-1}, v_t, \xi) = 0, \qquad \varphi(v_t, v_{t-1}, \tilde{\varepsilon}_t, \xi) = 0,$

$\xi = (\xi_1, \xi_2) \in (\Xi_1 \times \Xi_2) = \Xi$ which is a compact subset of $\mathbb{R}^{p_1+p_2}$,

(ii) $\{\tilde{\varepsilon}_t, t \in \mathbb{Z}\}$ is white noise with known distribution G_*,

where $r(\cdot)$ describes the evolution of the observable variables given the state variables, and $\varphi(\cdot)$ that of the state variables given the exogenous variables. This assumption is comparable to the fully parametric model described in Gourieroux et al. (1993) which they perceive to describe the true data-generating process. In case of our misspecified simulator, this only constitutes a nominal assumption. For given values of $\xi = (\xi_1, \xi_2)^\top$, we simulate values $\{\tilde{y}_1^s(\xi, z_0^s), \ldots, \tilde{y}_T^s(\xi, z_0^s)\}$ given the initial condition $z_0^s = (y_0^s, v_0^s)^\top$ from the simulated values $\{\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_T\}$ from G_*. P_* denotes the probability limit of the simulated processes.[12]

[12] For brevity, we omit the dependence of the simulation on the initial values of the state variables in the following.

Indirect inference about the true value of the structural parameters of interest ξ_1^0 of ξ_1 builds on the misspecified structural model introduced in Assumption 3.A.2 and the instrumental model $\mathcal{N}_\theta$. The pseudo-true values of instrumental parameters are defined as probability limits of an extremum estimator associated with the criterion function $Q_{1,T}(\{y_t\}_{t=1}^T, \theta_1)$.[13] Hence, the idea of PII is to use calibration as econometrics of a misspecified structural model and to use this misspecified model as a simulator.

The instrumental model has to be chosen carefully as the estimation of the instrumental variables must be tractable and it should be capable of capturing the properties of the model-implied financial time series. Similar to the recommendations of Dridi et al. (2007), we make use of moment restrictions that reflect the asset pricing implications of the DSGE asset pricing model rather than using a full parametric model as an instrumental model. The set of k moment restrictions is given by

$$\mathbb{E}\left[u_{1,t}\left(\{y_t\}_{t-l}^t, \theta_1^0\right)\right] = 0.$$

The number of moment restrictions has to be at least as large as the number of structural parameters, i.e. $k = \dim\ \theta_1 \ \geq\ p_1 + p_{21} = \dim\ \xi_1 + \ \dim\ \xi_{21}$ and $\theta_1 \in \Theta_1 \subset \mathbb{R}^k$. We assume that the true parameter θ_1^0 is the only value within the parameter space that gives a solution to the moment condition. Hence, the identification assumption reads

Assumption 3.A.3.

$$\mathbb{E}\left[u_{1,t}\left(\{y_t\}_{t-l}^t, \theta_1\right)\right] \neq 0 \quad \text{for all} \ \ \theta_1 \neq \theta_1^0 \in \Theta_1.$$

We minimize a GMM-type criterion function given below to estimate the instrumental parameters θ_1,

$$\min_{\{\theta_1 \in \Theta_1\}} Q_{1,T}\left(\{y_t\}_{t=1}^T, \theta_1\right),$$

$$\text{where} \quad Q_{1,T}\left(\{y_t\}_{t=1}^T, \theta_1\right) = \frac{1}{2}\, g_{1,T}\left(\{y_t\}_{t=1}^T, \theta_1\right)^\top \cdot \widehat{W}_{1,T} \cdot g_{1,T}\left(\{y_t\}_{t=1}^T, \theta_1\right),$$

$$g_{1,T}\left(\{y_t\}_{t=1}^T, \theta_1\right) = \frac{1}{T}\sum_{t=1}^T u_{1,t}\left(\{y_t\}_{t-l}^t, \theta_1\right),$$

with $\widehat{W}_{1,T} \xrightarrow{a.s.} W_1$, a positive semi-definite weighting matrix.

Following Hansen (1982), we impose the following regularity conditions

Assumption 3.A.4. (i) Θ_1 is a compact subset of $\mathbb{R}^k$,

[13] Here, $\theta_1 \in \Theta_1$ which is a compact subset of $\mathbb{R}^k$ and $\{y_t\}_{t=1}^T$ denote lagged values of y_t for a fixed number of M lags. Note that the instrumental parameters are denoted by θ_1 in accordance with the partial encompassing case in Dridi et al. (2007). In our setting, there is no second set of instrumental parameters θ_{21}.

(ii) $u_{1,t}(\cdot, \theta_1)$ is Borel measurable for each $\theta_1 \in \Theta_1$,

(iii) $\mathbb{E}\left[u_{1,t}\left(\{y_t\}_{t-l}^{t}, \theta_1\right)\right]$ exists and is finite for all $\theta_1 \in \Theta_1$,

(iv) $u_{1,t}\left(\{y_t\}_{t-l}^{t}, \theta_1\right)$ is first-moment continuous at all $\theta_1 \in \Theta_1$.

Using Assumption 3.A.4 and the stationarity and ergodictiy assumption of the time series results in sufficient conditions that the GMM-type criterion in Equation (3.3.5) converges almost surely uniformly to a non-stochastic limit criterion

$$Q_{1,\infty}\left(G_*, \xi_1^0, \bar{\xi}_{21}, \theta_1\right) = \frac{1}{2}\mathbb{E}\left[u_{1,t}\left(\{y_t\}_{t-l}^{t}, \theta_1\right)\right]^{\top} \cdot W_1 \cdot \mathbb{E}\left[u_{1,t}\left(\{y_t\}_{t-l}^{t}, \theta_1\right)\right].$$

Moreover, Assumption 3.A.3 implies that the limit criterion is uniquely minimized by

$$\theta_1^0 = \underset{\{\theta_1 \in \Theta_1\}}{\operatorname{argmin}} \, Q_{1,\infty}\left(G_*, \xi_1, \xi_{21}, \theta_1\right). \tag{3.A.1}$$

Hence, under the nominal assumption 3.A.2, the identification assumption 3.A.3, and the equicontinuity assumption 3.A.4,

$$\hat{\theta}_{1,T} = \underset{\{\theta_1 \in \Theta_1\}}{\operatorname{argmin}} \, Q_{1,T}\left(\{y_t\}_{t=1}^{T}, \theta_1\right)$$

is a consistent estimator of θ_1^0 (Singleton, 2009).

In case of a misspecified simulator, the probability limits of the simulated and observed time series are not the same. In addition to the assumptions for the data-generating process of the observed time series, we need assumptions for the misspecified simulator. As above, we impose an identification assumption for the simulated time series

Assumption 3.A.5.

$$\mathbb{E}^*\left[u_{1,t}\left(\{\tilde{y}_t^s(\xi)\}_{t-l}^{t}, \theta_1^s\right)\right] \neq 0 \quad \text{for all} \quad \theta_1^s \neq \tilde{\theta}_1^0(\xi_1^0, \bar{\xi}_{21}) \in \Theta_1^s.$$

We also impose an equivalent GMM regularity conditions for the moment conditions evaluated at the simulated series, i.e.

Assumption 3.A.6. (i) Θ_1^s is a compact subset of $\mathbb{R}^k$,

(ii) $u_{1,t}(\cdot, \theta_1^s)$ is Borel measurable for each $\theta_1^s \in \Theta_1^s$,

(iii) $\mathbb{E}^*\left[u_{1,t}\left(\{\tilde{y}_t^s(\xi)\}_{t-l}^{t}, \theta_1^s\right)\right]$ exists and is finite for all $\theta_1^s \in \Theta_1^s$,

(iv) $u_{1,t}\left(\{\tilde{y}_t^s(\xi)\}_{t-l}^{t}, \theta_1^s\right)$ is first-moment continuous at all $\theta_1^s \in \Theta_1^s$.

Comparable to the observed time series case and under Assumption 3.A.6 for stationary and ergodic simulated series, the GMM-type criterion in Equation (3.3.5) evaluated at the simulated data, i.e. $Q_{1,T}\left(\{\tilde{y}_T^s(\xi)\}_{t=1}^T, \theta_1^s\right)$, converges almost surely uniformly to a non-stochastic limit criterion.[14] Given Assumption 3.A.5, the limit criterion is uniquely minimized by $\tilde{\theta}_1^0\left(\xi_1^0, \bar{\xi}_{21}\right)$. Hence, under the nominal assumption 3.A.2, the identification assumption 3.A.5, and the equicontinuity assumption 3.A.6 for the simulated time series,

$$\tilde{\theta}_{1,T}^s(\xi_1, \xi_{21}, \bar{\xi}_{22}) = \underset{\{\theta_1 \in \Theta_1\}}{\operatorname{argmin}} Q_{1,T}\left(\{\tilde{y}_T^s(\xi)\}_{t=1}^T, \theta_1^s\right)$$

is a consistent estimator of $\tilde{\theta}_1^0\left(\xi_1^0, \bar{\xi}_{21}\right)$.

For the indirect estimator of the structural parameters, we impose the following assumptions on the binding function

Assumption 3.A.7. (i) $\theta_1(\cdot,\cdot)$ is one-to-one,

(ii) $$P_* \lim_{T\to\infty} \frac{\partial \tilde{\theta}_{1,T}^s}{\partial \begin{pmatrix} \xi_1 \\ \xi_{21} \end{pmatrix}} \left(\xi_1^0, \bar{\xi}_{21}, \bar{\xi}_{22}\right)^\top = \frac{\partial \tilde{\theta}_1}{\partial \begin{pmatrix} \xi_1 \\ \xi_{21} \end{pmatrix}} \left(\xi_1^0, \bar{\xi}_{21}\right)^\top \quad \text{has rank } p_1 + p_{21}.$$

The one-to-one assumption in 3.A.7 (i) is comparable to assumption (A4) in Gourieroux et al. (1993) and the rank assumption in 3.A.7 (ii) is equivalent to assumption (A14) in Dridi et al. (2007). We further require partial encompassing given by

Assumption 3.A.8. $\theta_1^0 = \tilde{\theta}_1^0(\xi_1^0, \bar{\xi}_{21})$,

where θ_1^0 is probability limit of the instrumental parameters associated with the observed data defined in Equation (3.A.1) and $\tilde{\theta}_1^0(\xi_1^0, \bar{\xi}_{21})$ is the probability limit for the simulated time series. Assumption 3.A.8 is the partial encompassing condition in Dridi et al. (2007). We refer to it as the partial encompassing null hypothesis $H_0^{(1)}\left(\bar{\xi}_{22}\right)$.

It follows that under Assumptions 3.A.3 to 3.A.8, the instrumental parameter estimators converge uniformly in (ξ_1, ξ_{21}) to

$$P_0 \lim_{T\to\infty} \hat{\theta}_{1,T} = \theta_1^0 \qquad \text{and} \qquad P_* \lim_{T\to\infty} \tilde{\theta}_{1,T}^s(\xi_1, \xi_{21}) = \tilde{\theta}_1^0(\xi_1^0, \bar{\xi}_{21}),$$

where $P_0 \lim_{T\to\infty}$ and $P_* \lim_{T\to\infty}$ denote the limit with respect to the probability distributions P_0 and P_* for the observed and simulated series.

[14] This limit criterion is given by

$$Q_{1,\infty}^*\left(G_*, \xi_1^0, \bar{\xi}_{21}, \theta_1^s\right) = \frac{1}{2}\mathbb{E}^*\left[u_{1,t}\left(\{\tilde{y}_t^s(\xi)\}_{t-l}^t, \theta_1^s\right)\right]^\top \cdot W_1 \cdot \mathbb{E}^*\left[u_{1,t}\left(\{\tilde{y}_t^s(\xi)\}_{t-l}^t, \theta_1^s\right)\right].$$

The partial indirect inference (PII) estimator is defined as

$$\begin{aligned}\begin{pmatrix}\hat{\xi}_{1,TS}(\bar{\xi}_{22})\\ \hat{\xi}_{21,TS}(\bar{\xi}_{22})\end{pmatrix} = \underset{\{(\xi_1,\xi_{21})\in\Xi_1\times\Xi_{21}\}}{\operatorname{argmin}} & \left[\hat{\theta}_{1,T}-\frac{1}{S}\sum_{s=1}^{S}\tilde{\theta}^s_{1,T}(\xi_1,\xi_{21},\bar{\xi}_{22})\right]^{\top}\cdot\widehat{\Omega}_{1,T}\\ & \cdot\left[\hat{\theta}_{1,T}-\frac{1}{S}\sum_{s=1}^{S}\tilde{\theta}^s_{1,T}(\xi_1,\xi_{21},\bar{\xi}_{22})\right]\\ \hat{\theta}_{1,T} = \underset{\{\theta_1\in\Theta_1\}}{\operatorname{argmin}} & Q_{1,T}\left(\{y_t\}_{t=1}^{T},\theta_1\right)\\ \tilde{\theta}^s_{1,T}\left(\xi_1,\xi_{21},\bar{\xi}_{22}\right) = \underset{\{\theta_1\in\Theta_1\}}{\operatorname{argmin}} & Q_{1,T}\left(\{\tilde{y}^s_T(\xi)\}_{t=1}^{T},\theta^s_1\right)\end{aligned}$$

and $\widehat{\Omega}_{1,T}$ is a positive definite weighting matrix that converges almost surely to the non-stochastic positive definite matrix Ω_1.

Proposition 3.A.1 (Consistency). Under assumptions 3.A.1, 3.A.3 - 3.A.8, $\hat{\xi}_{1,TS}\left(\bar{\xi}_{22}\right)$ is a consistent estimator of ξ_1^0.

The misspecified model defined in Assumption 3.A.2 endowed with the pseudo-true value $\begin{pmatrix}\xi_1^0\\ \bar{\xi}_{21}\end{pmatrix}$ partially encompasses $\mathcal{N}_\theta$ under assumptions 3.A.7 and 3.A.8. This partial encompassing provides a sufficient condition for the consistency of the partial indirect inference estimator of the parameter of interest $\hat{\xi}_{1,TS}\left(\bar{\xi}_{22}\right)$ as described in Proposition 3.3. in Dridi et al. (2007). The proof of the proposition is an adapted version of those found in Dridi and Renault (2000) and Dridi et al. (2007). In our case, we use a GMM-type instrumental model rather than a general extremum estimator. This difference features in the identification assumptions 3.A.3 and 3.A.5 and the regularity conditions 3.A.4 and 3.A.6 which are specific to the GMM-type instrumental model. Assumptions 3.A.1, 3.A.2 and 3.A.7 are comparable to those in Dridi et al. (2007) for the partial encompassing case in the absence of exogenous variables.

Asymptotic distribution of the partial indirect inference estimator

To derive the asymptotic distribution of the structural parameters, we follow a sequential approach. In a first step, we derive the asymptotic distribution of the instrumental parameters θ_1 for the GMM-type instrumental model. We impose additional assumptions on the moment conditions. More specifically, we assume that a central limit theorem applies to appropriately scaled moment conditions evaluated at the observed data $\{y_t\}_{t=1}^{T}$ and the simulated data $\left\{\tilde{y}^s_t(\xi)\right\}_{t=1}^{T}$.

Assumption 3.A.9. A multivariate central limit theorem applies, such that under Assumption 3.A.3

$$\sqrt{T}\left[\frac{1}{T}\sum_{t=1}^{T} u_{1,t}\left(\{y_t\}_{t=1}^{T}, \theta_1^0\right)\right] \underset{d}{\to} \mathcal{N}(0, V)$$

with

$$V = \Gamma_0 + \sum_{j=1}^{\infty}\left(\Gamma_j + \Gamma_j^\top\right)$$
$$\Gamma_j = \mathbb{E}\left[u_{1,t}\left(\{y_t\}_{t=1}^{T}, \theta_1^0\right) u_{1,t-j}\left(\{y_t\}_{t=1}^{T}, \theta_1^0\right)^\top\right]$$

and

Assumption 3.A.10. A multivariate central limit theorem applies, such that under Assumption 3.A.5

$$\sqrt{T}\left[\frac{1}{T}\sum_{t=1}^{T} u_{1,t}\left(\{\tilde{y}_T^s(\xi)\}_{t=1}^{T}, \tilde{\theta}_1^0\left(\bar{\xi}_{22}\right)\right)\right] \underset{d}{\to} \mathcal{N}(0, V^*)$$

with

$$V^* = \Gamma_0^* + \sum_{j=1}^{\infty}\left(\Gamma_j^* + \Gamma_j^{*\top}\right)$$
$$\Gamma_j^* = \mathbb{E}^*\left[u_{1,t}\left(\{\tilde{y}_T^s(\xi)\}_{t=1}^{T}, \tilde{\theta}_1^0\left(\bar{\xi}_{22}\right)\right) u_{1,t-j}\left(\{\tilde{y}_T^s(\xi)\}_{t=1}^{T}, \tilde{\theta}_1^0\left(\bar{\xi}_{22}\right)\right)^\top\right].$$

Under the null hypothesis of partial encompassing $H_0^{(1)}\left(\bar{\xi}_{22}\right) : \theta_1^0 = \tilde{\theta}_1^0\left(\bar{\xi}_{22}\right)$ given in Assumption 3.A.8, the following result holds.

Proposition 3.A.2 (Asymptotic distribution of the instrumental parameters). Under Assumptions 3.A.3, 3.A.5, 3.A.9, 3.A.10 and partial encompassing given in 3.A.8, the instrumental parameters are asymptotically distributed as

$$\sqrt{T}\left(\hat{\theta}_{1,T} - \frac{1}{S}\sum_{s=1}^{S}\tilde{\theta}_{1,T}^s\left(\xi_1^0, \bar{\xi}_{21}\right)\right) \xrightarrow{d} \mathcal{N}\left(0, \Phi_1^0(S, W_1)\right) \tag{3.A.2}$$

where the asymptotic covariance matrix is given by

$$\Phi_1^0(S, W_1) = C_1^0 V C_1^{0\top} + \frac{1}{S} C_1^{*0} V^* C_1^{*0\top} \tag{3.A.3}$$

and

$$C_1^0 = \left\{ \mathbb{E}\left[\frac{\partial u_{1,t}^\top}{\partial \theta_1}\left(y, \theta_1^0\right)\right] W_1 \mathbb{E}\left[\frac{\partial u_{1,t}}{\partial \theta_1^\top}\left(y, \theta_1^0\right)\right]\right\}^{-1} \mathbb{E}\left[\frac{\partial u_{1,t}^\top}{\partial \theta_1}\left(y, \theta_1^0\right)\right] W_1, \tag{3.A.4}$$

$$C_1^{*0} = \left\{ \mathbb{E}^*\left[\frac{\partial u_{1,t}^\top}{\partial \theta_1}\left(\tilde{y}, \theta_1^0\right)\right] W_1 \mathbb{E}^*\left[\frac{\partial u_{1,t}}{\partial \theta_1^\top}\left(\tilde{y}, \theta_1^0\right)\right]\right\}^{-1} \mathbb{E}^*\left[\frac{\partial u_{1,t}^\top}{\partial \theta_1}\left(\tilde{y}, \theta_1^0\right)\right] W_1, \tag{3.A.5}$$

$$\Gamma_j^* = \mathbb{E}^*\left[u_{1,t}\left(\tilde{y}, \theta_1^0\right) u_{1,t-j}\left(\tilde{y}, \theta_1^0\right)^\top\right]. \tag{3.A.6}$$

where y and $\tilde{y}$ are short-hand notations for $\{y_t\}_{t=1}^T$ and $\{\tilde{y}_T^s(\xi)\}_{t=1}^T$, respectively.

The asymptotic distribution of the estimated structural parameters is described in the following proposition.

Proposition 3.A.3 (Asymptotic distribution of the structural parameters). Under assumptions 3.A.1 - 3.A.7, 3.A.8, 3.A.9 and 3.A.10, the structural parameters estimates of ξ_1 and ξ_{21} follow an asymptotic normal distribution

$$\sqrt{T}\begin{pmatrix} \hat{\xi}_{1,TS}\left(\bar{\xi}_{22}\right) - \xi_1^0 \\ \hat{\xi}_{21,TS}\left(\bar{\xi}_{22}\right) - \bar{\xi}_{21} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \Sigma\left(S, \Omega\right)\right) \tag{3.A.7}$$

where the asymptotic covariance $\Sigma(S, \Omega)$ is given in Equations (3.B.1) in Appendix 3.B, Equations (3.A.3), and (3.A.4) - (3.A.6).

The proof of this proposition and the assumptions needed for the asymptotic distribution are provided in Appendix 3.B.

Appendix 3.B Derivation of the asymptotic distribution

In this section, we derive the asymptotic distribution of the parameters of interest ξ_1 (and of ξ_{21}). The objective function of a partial encompassing indirect inference estimator is given by the following quadratic form

$$\min_{\{(\xi_1, \xi_{21}) \in (\Xi_1 \times \Xi_{21})\}} \left[\hat{\theta}_{1,T} - \frac{1}{S}\sum_{s=1}^{S} \tilde{\theta}_{1,T}^s\left(\xi_1, \xi_{21}, \bar{\xi}_{22}\right)\right]^\top \widehat{\Omega}_{1,T} \left[\hat{\theta}_{1,T} - \frac{1}{S}\sum_{s=1}^{S} \tilde{\theta}_{1,T}^s\left(\xi_1, \xi_{21}, \bar{\xi}_{22}\right)\right].$$

The first order condition (FOC) to this minimization problem with respect to the estimated structural parameters reads

$$\frac{1}{S}\sum_{s=1}^{S}\frac{\partial\tilde{\theta}_{1,T}^{s}}{\partial\begin{pmatrix}\xi_1\\ \xi_{21}\end{pmatrix}}\left(\hat{\xi}_{1,TS}\left(\bar{\xi}_{22}\right),\hat{\xi}_{21,TS}\left(\bar{\xi}_{22}\right),\bar{\xi}_{22}\right)^{\top}\widehat{\Omega}_{1,T}$$
$$\cdot\left[\hat{\theta}_{1,T}-\frac{1}{S}\sum_{s=1}^{S}\tilde{\theta}_{1,T}^{s}\left(\hat{\xi}_{1,TS}\left(\bar{\xi}_{22}\right),\hat{\xi}_{21,TS}\left(\bar{\xi}_{22}\right),\bar{\xi}_{22}\right)\right]=0.$$

We evaluate the FOC around the limit values ξ_1^0 and $\bar{\xi}_{21}$. We need the following two results. The derivative of the instrumental parameters with respect to the structural parameters of interest follows a uniform convergence in the neighborhood of θ_0, i.e.

$$\frac{\partial\tilde{\theta}_{1,T}^{s}}{\partial\begin{pmatrix}\xi_1\\ \xi_{21}\end{pmatrix}}\left(\hat{\xi}_{1,TS}\left(\bar{\xi}_{22}\right),\hat{\xi}_{21,TS}\left(\bar{\xi}_{22}\right),\bar{\xi}_{22}\right)^{\top}\underset{a.s.}{\rightarrow}\frac{\partial\tilde{\theta}_{1}}{\partial\begin{pmatrix}\xi_1\\ \xi_{21}\end{pmatrix}}\left(\xi_1^0,\bar{\xi}_{21}\right)^{\top},$$

the second is the partial encompassing assumption 3.A.7 *(ii)* given by

$$P_*\lim_{T\rightarrow\infty}\frac{\partial\tilde{\theta}_{1,T}^{s}}{\partial\begin{pmatrix}\xi_1\\ \xi_{21}\end{pmatrix}}\left(\xi_1^0,\bar{\xi}_{21},\bar{\xi}_{22}\right)^{\top}=\frac{\partial\tilde{\theta}_{1}}{\partial\begin{pmatrix}\xi_1\\ \xi_{21}\end{pmatrix}}\left(\xi_1^0,\bar{\xi}_{21}\right)^{\top}\quad\text{has rank } p_1+p_{21}.$$

The expansion is given by

$$\frac{1}{S}\sum_{s=1}^{S}\frac{\partial\tilde{\theta}_{1,T}^{s}}{\partial\begin{pmatrix}\xi_1\\ \xi_{21}\end{pmatrix}}\left(\hat{\xi}_{1,TS}\left(\bar{\xi}_{22}\right),\hat{\xi}_{21,TS}\left(\bar{\xi}_{22}\right),\bar{\xi}_{22}\right)^{\top}\widehat{\Omega}_{1,T}\sqrt{T}\left[\hat{\theta}_{1,T}-\left(\frac{1}{S}\sum_{s=1}^{S}\tilde{\theta}_{1,T}^{s}\left(\xi_1^0,\bar{\xi}_{21}\right)\right.\right.$$
$$\left.\left.+\frac{1}{S}\sum_{s=1}^{S}\frac{\partial\tilde{\theta}_{1,T}^{s}}{\partial\begin{pmatrix}\xi_1\\ \xi_{21}\end{pmatrix}^{\top}}\left(\xi_1^0,\bar{\xi}_{21}\right)\begin{pmatrix}\hat{\xi}_{1,TS}\left(\bar{\xi}_{22}\right)-\xi_1^0\\ \hat{\xi}_{21,TS}\left(\bar{\xi}_{22}\right)-\bar{\xi}_{21}\end{pmatrix}\right)\right].$$

We rearrange the equation for the structural parameters. Asymptotically, they can be represented as

$$\sqrt{T}\begin{pmatrix}\hat{\xi}_{1,TS}\left(\bar{\xi}_{22}\right)-\xi_1^0\\ \hat{\xi}_{21,TS}\left(\bar{\xi}_{22}\right)-\bar{\xi}_{21}\end{pmatrix} \simeq \left\{\frac{\partial\tilde{\theta}_1}{\partial\begin{pmatrix}\xi_1\\ \xi_{21}\end{pmatrix}}\left(\xi_1^0,\bar{\xi}_{21}\right)^\top \Omega_1 \frac{\partial\tilde{\theta}_1}{\partial\begin{pmatrix}\xi_1\\ \xi_{21}\end{pmatrix}^\top}\left(\xi_1^0,\bar{\xi}_{21}\right)\right\}^{-1} \cdot \frac{\partial\tilde{\theta}_1}{\partial\begin{pmatrix}\xi_1\\ \xi_{21}\end{pmatrix}}\left(\xi_1^0,\bar{\xi}_{21}\right)^\top \Omega_1 \cdot \sqrt{T}\left[\hat{\theta}_{1,T}-\frac{1}{S}\sum_{s=1}^{S}\tilde{\theta}_{1,T}^s\left(\xi_1^0,\bar{\xi}_{21}\right)\right]$$

3.B.1 Asymptotic distribution of the instrumental parameters

As moment conditions in a GMM-type manner are used to estimate the instrumental parameters θ_1, the objective function evaluated at the observed time series $\{y_t\}_{t=1}^T$ is given by

$$Q_{1,T}\left(\{y_t\}_{t=1}^T,\theta_1\right)=\frac{1}{2}\,g_{1,T}\left(\{y_t\}_{t=1}^T,\theta_1\right)^\top\cdot\widehat{W}_{1,T}\cdot g_{1,T}\left(\{y_t\}_{t=1}^T,\theta_1\right),$$

where $g_{1,T}\left(\{y_t\}_{t=1}^T,\theta_1\right)=\frac{1}{T}\sum_{t=1}^T u_{1,t}\left(\{y_t\}_{t=1}^T,\theta_1\right)$ is the sample average of the moment condition whose limit is given as $\mathbb{E}\left[u_{1,t}\left(\{y_t\}_{t=1}^T,\theta_1\right)\right]\neq 0$ for all $\theta_1\neq\theta_1^0$.

First, we show how $\sqrt{T}\left(\hat{\theta}_{1,T}-\theta_1^0\right)$ is distributed. We apply a similar strategy as for the quadratic form above. We take the FOC of the objective function

$$\frac{\partial g_{1,T}^\top}{\partial\theta_1}\left(\{y_t\}_{t=1}^T,\hat{\theta}_{1,T}\right)\cdot\widehat{W}_{1,T}\cdot g_{1,T}\left(\{y_t\}_{t=1}^T,\hat{\theta}_{1,T}\right)$$

and use a mean value expansions of $g_{1,T}(\cdot)$ around the limit value θ_1^0

$$g_{1,T}\left(\{y_t\}_{t=1}^T,\hat{\theta}_{1,T}\right)\simeq g_{1,T}\left(\{y_t\}_{t=1}^T,\theta_1^0\right)+\frac{\partial g_{1,T}}{\partial\theta_1^\top}\left(\{y_t\}_{t=1}^T,\theta_1^0\right)\left(\hat{\theta}_{1,T}-\theta_1^0\right).$$

Then, we plug the MVT into the FOC and rearrange the equation for the instrumental parameters

$$\left(\hat{\theta}_{1,T}-\theta_1^0\right)=-\left\{\frac{\partial g_{1,T}^\top}{\partial\theta_1}\left(\{y_t\}_{t=1}^T,\hat{\theta}_{1,T}\right)\widehat{W}_{1,T}\frac{\partial g_{1,T}}{\partial\theta_1^\top}\left(\{y_t\}_{t=1}^T,\theta_1^0\right)\right\}^{-1}\frac{\partial g_{1,T}^\top}{\partial\theta_1}\left(\{y_t\}_{t=1}^T,\hat{\theta}_{1,T}\right)\widehat{W}_{1,T}\cdot g_{1,T}\left(\{y_t\}_{t=1}^T,\theta_1^0\right).$$

Asymptotically the equation above can be written as

$$\begin{aligned}
&\sqrt{T}\left(\hat{\theta}_{1,T}-\theta_1^0\right)\\
&\simeq -\left\{\mathbb{E}\left[\frac{\partial u_{1,t}^{\top}}{\partial\theta_1}\left(\theta_1^0\right)\right]W_1\mathbb{E}\left[\frac{\partial u_{1,t}}{\partial\theta_1^{\top}}\left(\theta_1^0\right)\right]\right\}^{-1}\mathbb{E}\left[\frac{\partial u_{1,t}^{\top}}{\partial\theta_1}\left(\theta_1^0\right)\right]W_1\cdot\sqrt{T}g_{1,T}\left(\{y_t\}_{t=1}^T,\theta_1^0\right)\\
&\simeq -C_1^0\cdot\sqrt{T}g_{1,T}\left(\{y_t\}_{t=1}^T,\theta_1^0\right),
\end{aligned}$$

where $g_{1,T}\left(\{y_t\}_{t=1}^T,\theta_1^0\right)=\frac{1}{T}\sum_{t=1}^T u_{1,t}\left(\{y_t\}_{t=1}^T,\theta_1^0\right)$.
Under the CLT in Assumption 3.A.9, the instrumental parameters evaluated at the observed data are distributed as

$$\sqrt{T}\left(\hat{\theta}_{1,T}-\theta_1^0\right)\underset{d}{\to}\mathcal{N}\left(0,C_1^0VC_1^{0^{\top}}\right).$$

Second, we elaborate on how the instrumental parameters for the simulated time series $\sqrt{T}\left(\tilde{\theta}_{1,T}^s\left(\xi_1^0,\bar{\xi}_{21}\right)-\tilde{\theta}_1^0\left(\xi_1^0,\bar{\xi}_{21}\right)\right)$ are distributed. For simplicity, define $\tilde{\theta}_{1,T}^s\left(\xi_1^0,\bar{\xi}_{21}\right)\equiv\tilde{\theta}_{1,T}^s\left(\bar{\xi}_{22}\right)$ and $\tilde{\theta}_1^0\left(\bar{\xi}_{22}\right)\equiv\tilde{\theta}_1^0\left(\xi_1^0,\bar{\xi}_{21}\right)$. Following the same steps for the simulated series as for the objective function evaluated at the observed data, we calculate the FOC

$$\frac{\partial g_{1,T}^{\top}}{\partial\theta_1}\left(\{\tilde{y}_T^s(\xi)\}_{t=1}^T,\tilde{\theta}_{1,T}^s\left(\bar{\xi}_{22}\right)\right)\cdot\widehat{W}_{1,T}\cdot g_{1,T}\left(\{\tilde{y}_T^s(\xi)\}_{t=1}^T,\tilde{\theta}_{1,T}^s\left(\bar{\xi}_{22}\right)\right)$$

and use a mean value expansions of $g_{1,T}(\cdot)$ around the limit value $\tilde{\theta}_1^0\left(\bar{\xi}_{22}\right)$

$$g_{1,T}\left(\tilde{\theta}_{1,T}^s\left(\bar{\xi}_{22}\right)\right)\simeq g_{1,T}\left(\{\tilde{y}_T^s(\xi)\}_{t=1}^T,\tilde{\theta}_1^0\left(\bar{\xi}_{22}\right)\right)+\frac{\partial g_{1,T}}{\partial\theta_1^{\top}}\left(\tilde{\theta}_1^0\left(\bar{\xi}_{22}\right)\right)\left(\tilde{\theta}_{1,T}^s\left(\bar{\xi}_{22}\right)-\tilde{\theta}_1^0\left(\bar{\xi}_{22}\right)\right).$$

Then, we plug the MVT into the FOC and rearrange the equation for the instrumental parameters

$$\begin{aligned}
&\left(\tilde{\theta}_{1,T}^s\left(\bar{\xi}_{22}\right)-\tilde{\theta}_1^0\left(\bar{\xi}_{22}\right)\right)\\
&=-\left\{\frac{\partial g_{1,T}^{\top}}{\partial\theta_1}\left(\tilde{\theta}_{1,T}^s\left(\bar{\xi}_{22}\right)\right)\widehat{W}_{1,T}\frac{\partial g_{1,T}}{\partial\theta_1^{\top}}\left(\tilde{\theta}_1^0\left(\bar{\xi}_{22}\right)\right)\right\}^{-1}\frac{\partial g_{1,T}^{\top}}{\partial\theta_1}\left(\tilde{\theta}_{1,T}^s\left(\bar{\xi}_{22}\right)\right)\widehat{W}_{1,T}\\
&\cdot g_{1,T}\left(\{\tilde{y}_T^s(\xi)\}_{t=1}^T,\tilde{\theta}_1^0\left(\bar{\xi}_{22}\right)\right).
\end{aligned}$$

Based on the simulated time series $\{\tilde{y}_T^s(\xi)\}_{t=1}^T$, the instrumental parameters can asymptotically by written as

$$\begin{aligned}
&\sqrt{T}\left(\tilde{\theta}_{1,T}^s\left(\bar{\xi}_{22}\right)-\tilde{\theta}_1^0\left(\bar{\xi}_{22}\right)\right)\\
\simeq&-\left\{\mathbb{E}^*\left[\frac{\partial u_{1,t}^\top}{\partial\theta_1}\left(\{\tilde{y}_T^s(\xi)\}_{t=1}^T,\tilde{\theta}_1^0\left(\bar{\xi}_{22}\right)\right)\right]W_1\mathbb{E}^*\left[\frac{\partial u_{1,t}}{\partial\theta_1^\top}\left(\{\tilde{y}_T^s(\xi)\}_{t=1}^T,\tilde{\theta}_1^0\left(\bar{\xi}_{22}\right)\right)\right]\right\}^{-1}\\
&\cdot\mathbb{E}^*\left[\frac{\partial u_{1,t}^\top}{\partial\theta_1}\left(\{\tilde{y}_T^s(\xi)\}_{t=1}^T,\tilde{\theta}_1^0\left(\bar{\xi}_{22}\right)\right)\right]W_1\cdot\sqrt{T}g_{1,T}\left(\{\tilde{y}_T^s(\xi)\}_{t=1}^T,\tilde{\theta}_1^0\left(\bar{\xi}_{22}\right)\right),
\end{aligned}$$

where $g_{1,T}\left(\{\tilde{y}_T^s(\xi)\}_{t=1}^T,\tilde{\theta}_1^0\left(\bar{\xi}_{22}\right)\right)=\frac{1}{T}\sum_{t=1}^T u_{1,t}\left(\{\tilde{y}_T^s(\xi)\}_{t=1}^T,\tilde{\theta}_1^0\left(\bar{\xi}_{22}\right)\right)$.
Under the null hypothesis of partial encompassing $H_0^{(1)}\left(\bar{\xi}_{22}\right):\ \theta_1^0=\tilde{\theta}_1^0\left(\bar{\xi}_{22}\right)$ given in Assumption 3.A.8 and the CLT in Assumption 3.A.10, it follows that

$$\begin{aligned}
\sqrt{T}\left(\tilde{\theta}_{1,T}^s\left(\bar{\xi}_{22}\right)-\tilde{\theta}_1^0\right)\simeq&-\left\{\mathbb{E}^*\left[\frac{\partial u_{1,t}^\top}{\partial\theta_1}\left(\{\tilde{y}_T^s(\xi)\}_{t=1}^T,\theta_1^0\right)\right]W_1\mathbb{E}^*\left[\frac{\partial u_{1,t}}{\partial\theta_1^\top}\left(\{\tilde{y}_T^s(\xi)\}_{t=1}^T,\theta_1^0\right)\right]\right\}^{-1}\\
&\cdot\mathbb{E}^*\left[\frac{\partial u_{1,t}^\top}{\partial\theta_1}\left(\{\tilde{y}_T^s(\xi)\}_{t=1}^T,\theta_1^0\right)\right]W_1\cdot\sqrt{T}g_{1,T}\left(\{\tilde{y}_T^s(\xi)\}_{t=1}^T,\theta_1^0\right)\\
\simeq&-C_1^{*0}\cdot\sqrt{T}g_{1,T}\left(\{\tilde{y}_T^s(\xi)\}_{t=1}^T,\theta_1^0\right),
\end{aligned}$$

where $\sqrt{T}g_{1,T}\left(\{\tilde{y}_T^s(\xi)\}_{t=1}^T,\theta_1^0\right)=\sqrt{T}\frac{1}{T}\sum_{t=1}^T u_{1,t}\left(\{\tilde{y}_T^s(\xi)\}_{t=1}^T,\theta_1^0\right)\underset{d}{\to}\mathcal{N}(0,V^*)$.
Under the null hypothesis of partial encompassing, the instrumental parameters evaluated at the simulated series are distributed as

$$\sqrt{T}\left(\tilde{\theta}_{1,T}^s\left(\bar{\xi}_{22}\right)-\theta_1^0\right)\underset{d}{\to}\mathcal{N}\left(0,C_1^{*0}V^*C_1^{*0^\top}\right).$$

Third, we investigate the distribution of $\sqrt{T}\left(\frac{1}{S}\sum_{s=1}^S\tilde{\theta}_{1,T}^s\left(\xi_1^0,\bar{\xi}_{21}\right)-\theta_1^0\right)$. Under the null hypothesis of partial encompassing, the instrumental parameters evaluated at the simulated series are distributed as

$$\sqrt{T}\left(\tilde{\theta}_{1,T}^s\left(\bar{\xi}_{22}\right)-\theta_1^0\right)\underset{d}{\to}\mathcal{N}\left(0,\frac{1}{S}C_1^{*0}V^*C_1^{*0^\top}\right).$$

Finally, the previous results are combined to show the distribution of $\sqrt{T}\left(\hat{\theta}_{1,T} - \frac{1}{S}\sum_{s=1}^{S}\tilde{\theta}_{1,T}^{s}\left(\xi_1^0, \bar{\xi}_{21}\right)\right)$. Based on the results above, the expression can asymptotically be written as

$$\sqrt{T}\left(\hat{\theta}_{1,T} - \frac{1}{S}\sum_{s=1}^{S}\tilde{\theta}_{1,T}^{s}\left(\xi_1^0, \bar{\xi}_{21}\right)\right) \simeq -C_1^0 \cdot \sqrt{T} g_{1,T}\left(\{y_t\}_{t=1}^{T}, \theta_1^0\right) + C_1^{*0}\sqrt{T} g_{1,T}\left(\{\tilde{y}_T^s(\xi)\}_{t=1}^{T}, \theta_1^0\right)$$

and is distributed asymptotically normal with asymptotic covariance matrix $\Phi_1^0(S, W_1)$

$$\Phi_1^0(S, W_1) = C_1^0 V C_1^{0\top} - C_1^0 K C_1^{*0\top} - C_1^{*0} K' C_1^{0\top} + \frac{1}{S} C_1^{*0} V^* C_1^{*0\top} + \left(1 - \frac{1}{S}\right) C_1^{*0} K^* C_1^{*0\top}$$

where K and K^* are asymptotic covariance matrices of the observed and simulated data which are non-zero if exogenous variables exist in the model. If there are no exogenous variables in the model, the asymptotic covariance is given by

$$\Phi_1^0(S, W_1) = C_1^0 V C_1^{0\top} + \frac{1}{S} C_1^{*0} V^* C_1^{*0\top}.$$

Given the results above, we can show the asymptotic distribution of the instrumental parameters. Under Assumptions 3.A.3, 3.A.5, 3.A.9, 3.A.10 and 3.A.8, the instrumental parameters are asymptotically distributed as

$$\sqrt{T}\left(\hat{\theta}_{1,T} - \frac{1}{S}\sum_{s=1}^{S}\tilde{\theta}_{1,T}^{s}\left(\xi_1^0, \bar{\xi}_{21}\right)\right) \xrightarrow{d} \mathcal{N}\left(0, \Phi_1^0(S, W_1)\right)$$

where the asymptotic covariance matrix is given by

$$\Phi_1^0(S, W_1) = C_1^0 V C_1^{0\top} + \frac{1}{S} C_1^{*0} V^* C_1^{*0\top}$$

and

$$C_1^0 = \left\{\mathbb{E}\left[\frac{\partial u_{1,t}^\top}{\partial \theta_1}\left(y, \theta_1^0\right)\right] W_1 \mathbb{E}\left[\frac{\partial u_{1,t}}{\partial \theta_1^\top}\left(y, \theta_1^0\right)\right]\right\}^{-1} \mathbb{E}\left[\frac{\partial u_{1,t}^\top}{\partial \theta_1}\left(y, \theta_1^0\right)\right] W_1,$$

$$C_1^{*0} = \left\{\mathbb{E}^*\left[\frac{\partial u_{1,t}^\top}{\partial \theta_1}\left(\tilde{y}, \theta_1^0\right)\right] W_1 \mathbb{E}^*\left[\frac{\partial u_{1,t}}{\partial \theta_1^\top}\left(\tilde{y}, \theta_1^0\right)\right]\right\}^{-1} \mathbb{E}^*\left[\frac{\partial u_{1,t}^\top}{\partial \theta_1}\left(\tilde{y}, \theta_1^0\right)\right] W_1,$$

$$\Gamma_j^* = \mathbb{E}^*\left[u_{1,t}\left(\tilde{y}, \theta_1^0\right) u_{1,t-j}\left(\tilde{y}, \theta_1^0\right)^\top\right].$$

where y and $\tilde{y}$ are short-hand notations for $\{y_t\}_{t=1}^{T}$ and $\{\tilde{y}_T^s(\xi)\}_{t=1}^{T}$, respectively.

3.B.2 Asymptotic distribution of the structural parameters

Overall the estimated structural parameters are distributed as follows: Under assumptions 3.A.1 - 3.A.8, 3.A.9 and 3.A.10, the structural parameters estimates of ξ_1 and ξ_{21} follow an asymptotic normal distribution

$$\sqrt{T}\begin{pmatrix}\hat{\xi}_{1,TS}\left(\bar{\xi}_{22}\right)-\xi_1^0\\ \hat{\xi}_{21,TS}\left(\bar{\xi}_{22}\right)-\bar{\xi}_{21}\end{pmatrix}\xrightarrow{d}\mathcal{N}\left(0,\Sigma\left(S,\Omega\right)\right)$$

where the asymptotic covariance matrix is given by

$$\begin{aligned}\Sigma\left(S,\Omega\right)=&\left\{\frac{\partial\tilde{\theta}_1}{\partial\begin{pmatrix}\xi_1\\ \xi_{21}\end{pmatrix}}\left(\xi_1^0,\bar{\xi}_{21}\right)^\top\Omega_1\frac{\partial\tilde{\theta}_1}{\partial\begin{pmatrix}\xi_1\\ \xi_{21}\end{pmatrix}^\top}\left(\xi_1^0,\bar{\xi}_{21}\right)\right\}^{-1}\frac{\partial\tilde{\theta}_1}{\partial\begin{pmatrix}\xi_1\\ \xi_{21}\end{pmatrix}}\left(\xi_1^0,\bar{\xi}_{21}\right)^\top\Omega_1\\ &\cdot\Phi_1^0\left(S,W_1\right)\Omega_1\frac{\partial\tilde{\theta}_1}{\partial\begin{pmatrix}\xi_1\\ \xi_{21}\end{pmatrix}^\top}\left(\xi_1^0,\bar{\xi}_{21}\right)\left\{\frac{\partial\tilde{\theta}_1}{\partial\begin{pmatrix}\xi_1\\ \xi_{21}\end{pmatrix}}\left(\xi_1^0,\bar{\xi}_{21}\right)^\top\Omega_1\frac{\partial\tilde{\theta}_1}{\partial\begin{pmatrix}\xi_1\\ \xi_{21}\end{pmatrix}^\top}\left(\xi_1^0,\bar{\xi}_{21}\right)\right\}^{-1}.\end{aligned}\tag{3.B.1}$$

The asymptotic covariance matrix of the auxiliary parameters is given by

$$\Phi_1^0\left(S,W_1\right)=C_1^0VC_1^{0\top}+\frac{1}{S}C_1^{*0}V^*C_1^{*0\top},$$

where the components read

$$\begin{aligned}C_1^0&=\left\{\mathbb{E}\left[\frac{\partial u_{1,t}^\top}{\partial\theta_1}\left(y,\theta_1^0\right)\right]W_1\mathbb{E}\left[\frac{\partial u_{1,t}}{\partial\theta_1^\top}\left(y,\theta_1^0\right)\right]\right\}^{-1}\mathbb{E}\left[\frac{\partial u_{1,t}^\top}{\partial\theta_1}\left(y,\theta_1^0\right)\right]W_1,\\ C_1^{*0}&=\left\{\mathbb{E}^*\left[\frac{\partial u_{1,t}^\top}{\partial\theta_1}\left(\tilde{y},\theta_1^0\right)\right]W_1\mathbb{E}^*\left[\frac{\partial u_{1,t}}{\partial\theta_1^\top}\left(\tilde{y},\theta_1^0\right)\right]\right\}^{-1}\mathbb{E}^*\left[\frac{\partial u_{1,t}^\top}{\partial\theta_1}\left(\tilde{y},\theta_1^0\right)\right]W_1,\\ \Gamma_j^*&=\mathbb{E}^*\left[u_{1,t}\left(\tilde{y},\theta_1^0\right)u_{1,t-j}\left(\tilde{y},\theta_1^0\right)^\top\right].\end{aligned}$$

where y and $\tilde{y}$ are short-hand notations for $\{y_t\}_{t=1}^T$ and $\{\tilde{y}_T^s(\xi)\}_{t=1}^T$, respectively.

3.B.3 Efficient weighting matrix

Proposition 3.B.1 (Asymptotic distribution of the optimal structural parameters). Under assumptions 3.A.1 - 3.A.7, 3.A.8, 3.A.9 and 3.A.10, and using the optimal weighting

matrix $\Omega_1^*\left(\bar{\xi}_{22}\right) = \Phi_1^0(S,W_1)^{-1}$, the optimal structural parameters estimates of ξ_1^* and ξ_{21}^* follow an asymptotic normal distribution

$$\sqrt{T}\begin{pmatrix}\hat{\xi}_{1,TS}^*\left(\bar{\xi}_{22}\right)-\xi_1^0\\ \hat{\xi}_{21,TS}^*\left(\bar{\xi}_{22}\right)-\bar{\xi}_{21}\end{pmatrix}\xrightarrow{d}\mathcal{N}\left(0,\Sigma\left(S,\Omega_1^*\left(\bar{\xi}_{22}\right)\right)\right)$$

where the asymptotic covariance matrix is given by

$$\Sigma\left(S,\Omega_1^*\left(\bar{\xi}_{22}\right)\right)=\left\{\frac{\partial\tilde{\theta}_1}{\partial\begin{pmatrix}\xi_1\\ \xi_{21}\end{pmatrix}}\left(\xi_1^0,\bar{\xi}_{21}\right)^\top\Phi_1^0\left(S,W_1\right)^{-1}\frac{\partial\tilde{\theta}_1}{\partial\begin{pmatrix}\xi_1\\ \xi_{21}\end{pmatrix}^\top}\left(\xi_1^0,\bar{\xi}_{21}\right)\right\}^{-1}.$$

3.B.4 Estimation of the asymptotic covariance matrix

For the calculation of $\Sigma\left(S,\Omega_1\right)$, we rely on Equations (3.B.1), (3.A.3), and (3.A.4) - (3.A.6) in Appendix 3.A.

The first building block is calculating V and V^*, the variance-covariance matrices of the moment conditions in Equation (3.3.12) evaluated using the empirical data and the model simulated data, respectively. These are computed using the Newey and West (1987) method to account for serial correlation, up to the 10th order, between the moment conditions.

The next step is calculating C_1^0 and C_1^{*0} shown in Equations (3.A.4) and (3.A.5) for the empirical and simulated data, respectively. In our implementation, W_1 is chosen to be an identity matrix. Given all the necessary inputs, $\Phi_1^0\left(S,W_1\right)$ can now be calculated.

As 500 ensembles are simulated from the DSGE model for each parameter value, the second part of Equation (3.A.3) is evaluated 500 times for each of the ensembles, and the average is used to compute Φ_1^0.

Finally, $\Sigma\left(S,\Omega_1\right)$ is calculated using Equation (3.B.1). Again, Ω_1 is chosen to be an identity matrix to assign identical weights to the sensitivity of each instrumental parameter to changes in each of the structural parameters.

Appendix 3.C Grid Search

To find estimates of ξ_1, we initially perform a grid search for the ξ_1 parameter combination that returns the minimum value of the PII objective function described in Equation (3.3.8) and our choice of moment conditions given by Equations (3.3.12). The β grid ranges from 0.98 to 1.01 in increments of 0.001, and the $\overline{rra}$ grid is spanned between 15.5 and 39.5 in unit increments. As values of β close to 1 are reasonable, we allow for time preference rates larger than one to find a minimum within the grid. As for the $\overline{rra}$, we base our specification

of the grid on the steady-state values reported in Campbell and Cochrane (1999) and Chen (2017). We use their steady-state relative risk aversion value, $\overline{rra} = 28.6$, in the middle of the grid and vary around it. Note that we fix the utility parameter $\gamma = 2$ and vary the steady state surplus consumption from 0.068 to 0.129.

The grid is minimized for values of $\beta = 1.001$ and $\overline{rra} = 22.5$ and the objective function is not flat around the minimum.

Appendix 3.D Data

The empirical data used for the calibration of the parameter values and shown in Tables 3.2.1, 3.2.2 and 3.4.2 are real U.S. post-war quarterly data from 1948 to 2012 and are chosen similar to Chen (2017). The data, together with their sources, are summarized in Table 3.D.1. The macroeconomic data are publicly available and can be easily retrieved. Output, consumption, investment, wages and capital are calculated from quantities available in the National Income and Product Accounts (NIPA) with the exception of private fixed assets which is retrieved from Fixed Asset Tables (FAT). All the quantities are deflated using their price indices and divided by the population to get their real per capita values. Productivity, however, is mimicked by total factor productivity (TFP) which is retrieved from Fernald (2014).

Unlike the macroeconomic data, the financial time series are not publicly available. The risk-free rate is calculated as the forecast of the ex-post real return of the 90-day Treasury bill. The data needed for this forecast is the current treasury bill yield and the inflation rate for the past 12 months. The earlier is retrieved from the Center for Research in Security Prices (CRSP) while the latter is retrieved from the Bureau of Labor Statistics (BLS). Firm returns are calculated as the weighted average of the CRSP index and the Barclays U.S. corporate investment grade index. The latter is retrieved from Lehman Bond indices and is weighted by the firm value. Firm value, in turn, is calculated as debt plus market equity, both of which are retrieved from Compustat. Finally, Tobin's Q time series can be retrieved from FRED.

Table 3.D.1: Data sources

Variable	**Data**	**Source**	**URL**
Output	GDP	NIPA table 1.1.4	https://apps.bea.gov/iTable/iTable.cfm?reqid=19&step=4&isuri=1&categories=flatfiles
Consumption	Non durable goods	NIPA table 1.1.4	same as before
	Services	NIPA table 1.1.4	same as before
Investment	Private fixed investments	NIPA table 1.1.4	same as before
	Durable goods	NIPA table 1.1.4	same as before
Wages	Compensation of Employees	NIPA table 3.9.4	same as before
Capital	Durable goods	NIPA table 1.1.4	same as before
	Government fixed assets	NIPA table 1.1.4	same as before
	Private fixed assets	FAT table 1.1	https://apps.bea.gov/iTable/iTable.cfm?reqid=10&step=2&isuri=1
Productivity	Total Factor Productivity (TFP)	Fernald (2014)	http://www.johnfernald.net/TFP
Risk-free rate	Forecast of the 90-day Treasury bill	CRSP	Not available
Firm returns	Weighted average of CRSP	CRSP	Not available
	Barclays U.S. corporate investment grade index	Lehman Bond Indices from Datastream	Not available
CPI	Seasonally adjusted inflation	BLS	https://fred.stlouisfed.org/series/CPIAUCSL
Firm value	Debt and Market Equity	Compustat	Not available
Tobin's Q	Tobin's Q	FRED	Not available

Notes: This table summarizes the U.S. post war empirical data used in the paper.

Appendix 3.E Numerical solution of the model

To numerically solve the DSGE model, we rely on the projection method developed by Chen (2017)[15]. The projection method entails a homotopy method, whereby the surplus consumption growth is added to the SDF in increments. This is done by including a parameter in the SDF that is an exponent, ν, to the surplus consumption growth, and thus controls its effect. This parameter ranges from zero to 1 in 11 incremental steps. At $\nu = 0$, the SDF is merely that of a power utility model, while at $\nu = 1$, the full extent of the external habit persistence model is included in the SDF. At each of these 11 steps, the model is solved, and the new solution is the starting point for the next step. The homotopy method helps stabilize the nonlinear solver.

At each of these incremental steps, the model is solved by discretizing the productivity process into 13 states using a 13-point Markov chain via the Rouwenhorst (1995) method. The laws of motion for capital and surplus consumption are approximated by two-dimensional cubic splines of 6th order and 8th order, respectively. Broyden's method is used to calculate the cubic spline coefficients that satisfy the firm's Euler equation given in Equation (3.2.9). Given the productivity states today and tomorrow as well as tomorrow's values for capital and surplus consumption that follow from their laws of motion, the remaining endogenous variables in the model can be simulated.

Appendix 3.F Details on the model solution

The firm's Euler equation

Chen (2017) claims that the habit in consumption leads to a consumption distortion. Due to that, the welfare theorems are no longer valid, and we cannot solve for an equilibrium using a social planner problem. A valid alternative in this setting is a recursive equilibrium which amounts to transforming the infinite-time problem into a recursive problem. Thus, we solve the problem via value functions which only depend on the value at time t and $t+1$.

$$\begin{aligned} V(K';K,S,Z) = \max_{\{I,N,K'\}} \; & [\Pi(K,Z,N) - W(K,S,Z)N - \Phi(I,K) - I \\ & + \mathbb{E}_Z\left(M(K,S,Z;Z')V(K';K',S',Z')\right)] \end{aligned}$$

$$\text{s.t.} \quad K' = (1-\delta)K + I$$

[15] Some of the MATLAB code is provided on his website https://sites.google.com/site/chenandrewy/. Our model solution is based on the code in the RFS folder.

where $V(\cdot)$ is the value function, K, S and Z are state variables that we can observe at time t which describe the state of the economy, and K' [16], I and N are called control variables which have to be chosen optimally at time t. The intuition is that given the optimal decision of these variables in $t+1$, what is the optimal choice at time t.

In order to make the optimization problem computationally less demanding, Chen (2017) discretizes the productivity process to a 13 point Markov-chain with a transition matrix $\pi_Z(Z_i, Z_j)$ which gives the probability of moving from state i in period t to state j in period $t+1$. Thus, the expected value and the optimization problem can be written in terms of a discrete random variable Z. When the economy is at state i in period t, its Bellman equation can be written as

$$\begin{aligned} V(K'; K, S, Z_i) = \max_{\{I,N,K'\}} \Big[& \Pi(K, Z_i, N) - W(K, S, Z_i)N - \Phi(I, K) - I \\ & + \sum_{Z_j} \pi_Z(Z_i, Z_j) M(K, S, Z_i; Z_j) V(K'; K', S', Z_j) \Big] \end{aligned}$$

$$\text{s.t.} \quad K' = (1-\delta)K + I.$$

A solution to this optimization problem consists of the F.O.C. w.r.t. labor which gives the equilibrium wage, the F.O.C. w.r.t. investment and the Envelope and firm's Euler equation. Equation (3.F.1) shows the equilibrium condition that sets the marginal productivity of labor equivalent to the wage:

$$\frac{\partial V}{\partial N} = \Pi_N(K, Z_i, N) - W(K, S, Z_i) \stackrel{!}{=} 0 \tag{3.F.1}$$

where Π_N denotes the derivative of the production function w.r.t. labor N.

Equation (3.F.2), in turn, denote the F.O.C. w.r.t investment.

$$\begin{aligned} & \frac{\partial V}{\partial I} = -1 - \Phi_I(I, K) + \sum_{Z_j} \pi_Z(Z_i, Z_j) M(\cdot) \frac{\partial V}{\partial K'} \frac{\partial K'}{\partial I} \stackrel{!}{=} 0 \\ & \text{or} \quad 1 + \Phi_I(I, K) = \sum_{Z_j} \pi_Z(Z_i, Z_j) M(\cdot) V_{K'}(\cdot) \end{aligned} \tag{3.F.2}$$

where $V_{K'}(\cdot)$ is

$$\frac{\partial V}{\partial K'} = -\frac{\partial \Phi}{\partial I}\frac{\partial I}{\partial K'} - \frac{\partial I}{\partial K'} + \sum_{Z_j} \pi_Z(Z_i, Z_j) M(\cdot) \frac{\partial V}{\partial K'} \stackrel{!}{=} 0 \tag{3.F.3}$$

[16] variables with ' denote values at time $t+1$

In order to solve the F.O.C. seen in Equation (3.F.3), we have to find the partial derivative of the value function w.r.t. K'. This can be done using the following trick

1. plug the optimal K' into the value function
2. take the partial derivative of the value function w.r.t. K

$$\frac{\partial V}{\partial K} = \Pi_K(K, Z_i, N) - \Phi_K(I, K) - \frac{\partial \Phi}{\partial I}\frac{\partial I}{\partial K} - \frac{\partial I}{\partial K}$$
$$V_K(\cdot) = \Pi_K(\cdot) + (1-\delta)(1 + \Phi_I(I, K)) - \Phi_K(I, K)$$

3. Envelope theorem: as in equilibrium K and K' are almost the same, we omit the partial derivative of K' w.r.t. K and the partial derivative of the value function w.r.t. K' can be obtained by iterating the derivative above one period forward:

$$V_{K'}(\cdot; Z_j) = \Pi_{K'}(K', Z_j, N') + (1-\delta)(1 + \Phi_{I'}(I_j, K')) - \Phi_{K'}(I_j, K')$$

4. Plug this partial derivative into the F.O.C. w.r.t. investment above and obtain the firm's Euler equation.

$$\begin{aligned}
\Phi_I(I, K) + 1 &= \sum_{Z_j} \pi_Z(Z_i, Z_j) M(\cdot) \left[\Pi_{K'}(K', Z_j, N') + (1-\delta)(1 + \Phi_{I'}(I_j, K')) - \Phi_{K'}(I_j, K')\right] \\
1 &= \sum_{Z_j} \pi_Z(Z_i, Z_j) M(\cdot) \frac{\Pi_{K'}(K', Z_j, N') + (1-\delta)(1 + \Phi_{I'}(I_j, K')) - \Phi_{K'}(I_j, K')}{\Phi_I(I, K) + 1} \\
1 &= \mathbb{E}_Z \left[M(\cdot) \frac{\Pi_{K'}(K', Z_j, N') + (1-\delta)(1 + \Phi_{I'}(I_j, K')) - \Phi_{K'}(I_j, K')}{\Phi_I(I, K) + 1} \right]
\end{aligned}$$

Household's Euler equation

Based on the intertemporal consumption and savings decision of the household, we obtain the usual Euler equation which holds for returns. The risky asset in this setting is investment and the Euler equation in terms of the return to investment R_I reads

$$\mathbb{E}_Z \left[M(K, S, Z_i; Z_j) R_I(K, S, Z_i; Z_j) \right] = 1$$

or with the discretized productivity process we have

$$\sum_{Z_j} \pi_Z(Z_i, Z_j) M(K, S, Z_i; Z_j) R_I(K, S, Z_i; Z_j) = 1.$$

Equilibrium

The household does not value labor, i.e. $N = 1$ and we can obtain the return of investment by equating the two Euler equations which yields a state dependent return on investment

$$R_I(K, S, Z_i; Z_j) = \frac{\Pi_{K'}(K', Z_j, 1) + (1-\delta)(1 + \Phi_{I'}(I_j, K')) - \Phi_{K'}(I_j, K')}{\Phi_I(I, K) + 1}$$

as the capital, labor and investment chosen by the firms and the households are the same in equilibrium.

We obtain the following equilibrium conditions:

$$\begin{aligned}
1 &= \mathbb{E}_Z \left[M(K, S, Z_i; Z_j) R_I(K, S, Z_i; Z_j) \right] \\
\text{with} \quad R_I(K, S, Z_i; Z_j) &= \frac{\Pi_{K'}(K', Z_j, 1) + (1-\delta)(1 + \Phi_{I'}(I_j, K')) - \Phi_{K'}(I_j, K')}{\Phi_I(I, K) + 1} \\
\text{where} \quad M &= M(K, S, Z_i; Z_j) = \beta \left(\frac{C_j}{C} \frac{S_j}{S} \right)^{-\gamma} \qquad \text{from household optimization} \\
C_j &= \Pi(K', Z_j, 1) - \Phi(I_j, K') - I_j \qquad \text{from accounting identity} \\
C &= \Pi(K, Z_i, 1) - \Phi(I, K) - I \\
K' &= G(K, S, Z_i) \qquad \text{time-invariant policy function, solution to Bellman eq.} \\
I &= G(K, S, Z_i) - (1-\delta)K \\
I_j &= G(K', S', Z_j) - (1-\delta)K' \\
&= G(G(K, S, Z_i), S', Z_j) - (1-\delta)G(K, S, Z_i)
\end{aligned}$$

and the evolution of surplus consumption satisfies

$$s_j = (1 - \rho_s)\bar{s} + \rho_s s + \lambda(c_j - c)$$

Chapter 4

Econometric Analysis of Partial Equilibrium Asset Pricing Models: A Two-Step Partial Indirect Inference Approach

I thank Ralf Brüggemann, Joachim Grammig, Dalia Elshiaty and the seminar participants of the joint Econometrics Webinar for helpful suggestions.

4.1. Introduction

In this paper, we propose a new estimation strategy for partial equilibrium asset pricing models. The approach combines a two-step estimation approach based on the model structure with the idea underlying partial indirect inference.

In partial equilibrium asset pricing models, the technology processes are modeled to provide consumption and dividend dynamics in an exogenous manner. This part of the model is independent of the specification of the representative agent's intertemporal decision. The consumption dynamics are then used in the asset pricing part of the model whose focus lies on specifying the representative agent's preferences given by the utility function specification. The two-step approach is motivated as in Grammig and Küchlin (2018) and makes use of this structure; the technology parameters are estimated in a first step using macroeconomic time series. In a second step, the first step macroeconomic parameters are taken as given and the asset pricing part of the model is estimated based on financial data.

We augment the two-step approach by the notion of partial indirect inference (PII) proposed by Dridi et al. (2007). In PII, the model parameters are divided based on whether they are part of the model that reflect economic reality or a part that is merely seen as a caricature of economic truth. Using the two-step PII strategy, we focus on the consistent estimation of model parameters that are economically of interest. The other (nuisance) parameters that stem from parts of the model that are potentially misspecified are secondary and under certain circumstances we may choose to calibrate them. The application of the PII approach specifies encompassing conditions under which the structural parameters of interest can be consistently estimated in the presence of potential model misspecification.

In the case of partial equilibrium models, we believe that the asset pricing part of the model yields testable restrictions. Hence, we are interested in the consistent estimation of the preference parameters that pertain to this part of the model. This parameter separation of structural parameters of interest and nuisance parameters affects the limiting distribution of our two-step partial indirect inference estimator.

Our estimation approach is applied to a partial equilibrium asset pricing model with habit formation as in Campbell and Cochrane (1999) and a model that accounts for long-run risk as in Bansal and Yaron (2004). The two-step partial indirect inference approach enables an evaluation of the asset pricing models that reflects both the model structure and the modeling focus. The model structure is captured by the two-step estimation strategy and the modeling focus is reflected in the parameter classification of structural parameters of interest and nuisance parameters.

The general structure of partial equilibrium asset pricing models is based on the consumption and savings decision of a representative agent whose primary aim is her

intertemporal decision to smooth consumption over time.[1] Hence, the representative agent has to decide whether she uses her current wealth to consume today or to save and postpone consumption into a future period. Towards this purpose, the representative agent optimizes her expected lifetime utility subject to a budget constraint. The price of a risky asset is determined by this intertemporal choice problem. The first order condition can be reformulated to the basic asset pricing equation (Euler equation) which stipulates that the price of the risky asset is equal to the expected value of the discounted payoff that is weighted by a stochastic discount factor (SDF). The SDF, in turn, is equivalent to the representative agent's intertemporal marginal rate of substitution. The excess return of the risky asset, or its equity premium, is determined by the quantity of risk times the price of risk. In historical data, the equity premium is higher than reasonable risk aversion and consumption volatility estimates imply (Cochrane, 2009), an anomaly called equity premium puzzle. In the standard consumption based asset pricing literature, the utility function is often specified as a power utility function defined over aggregate consumption (e.g. Rubinstein (1976), Lucas (1978) and Mehra and Prescott (1985)) for which the quantity of risk is equal to the covariance of the excess stock returns with consumption growth. As the observed consumption growth is smooth, the covariance is rather small and cannot explain much of the equity risk premium. As the price of risk is equal to the coefficient of relative risk aversion which is constant, the price of risk is time-invariant. Thus, to explain the equity risk premium, the risk aversion coefficient has to take on implausibly large values which is not compatible with the range of plausible risk aversion coefficients between one and ten suggested by Mehra and Prescott (1985). Furthermore, in order to enforce an economically plausible real risk-free rate, the time discount factor often takes on values greater than one implying a negative time preference rate. This phenomenon is known as the risk-free rate puzzle and is first described in Weil (1989, 1990).

These findings are not compatible with the stylized facts observed in U.S. data described by Campbell (2003) among others. Campbell and Cochrane (1999) develop a habit persistence model with a time-varying price of risk that aims to alleviate inadequacies of the classical consumption based asset pricing model. This is achieved by complementing the representative agent's utility function with habit formation preferences. Bansal and Yaron (2004) take a different approach towards generating a time-varying risk premium. In their long-run risk (LRR) model, they model a small predictable long run growth component to consumption growth and introduce stochastic economic uncertainty. Moreover, they use recursive Epstein-Zin-Weil preferences proposed by Epstein and Zin (1989) and Weil

[1] Early contribution to consumption-based asset pricing are treated in Rubinstein (1976) and Lucas (1978).

(1989) rather than power utility. In their model, the economy is affected by long-run growth prospects and economic uncertainty which gives rise to the time-varying risk premium.

Both partial equilibrium asset pricing models have in common that the consumption and dividend growth dynamics are exogenous. Whereas the habit persistence model specifies i.i.d. growth dynamics, the LRR model allows for a small persistent long-run growth component. They are partial equilibrium models as the optimization problem is set up as a representative agent's optimal intertemporal decision. The macro economy is independently modeled from the asset pricing part, however, the latter depends on the specification of the technology process in the macro economy.

There are various approaches to estimating the habit persistence and long-run risk model in the literature. Bansal et al. (2007) propose an efficient method of moments based strategy, and Gallant and Tauchen (1996) and Aldrich and Gallant (2011) avocate a Bayesian indirect inference estimation approach for both models. Gallant and McCulloch (2009) estimate the habit persistence model using a GARCH-type instrumental model in the Bayesian setting and Schorfheide et al. (2018) propose a Bayesian mixed frequency approach for the long-run risk model. Moreover, there exists extensive literature on frequentist estimation approaches for the long-run risk model; Constantinides and Ghosh (2011), Hasseltoft (2012), Bansal et al. (2012) and Calvet and Czellar (2015) estimate the model parameters in a single step, whereas Grammig and Küchlin (2018) propose a two-step estimation strategy that builds on the intricate structure of the long-run risk model. We favor moment matching based frequentist estimation approaches over Bayesian ones for these models. These models build on equilibrium conditions expressed by first order conditions that we can use to derive moment conditions. Method of moments and indirect inference are intuitive procedures with a transparent link between these first order conditions/ moments and the equilibrium equations (Fernández-Villaverde, 2010).

All attempts to estimate the two partial equilibrium asset pricing models are based on the implicit assumption that the respective models are correctly specified. Our partial indirect inference based approach allows to distinguish parts of the model that reflect economic reality and others that are openly misspecified. It gives a rationale to calibrate (some) nuisance parameters which index the misspecified part of the model without compromising the consistent estimation of the structural parameters of interest.

We use the habit persistence and the long-run risk models to apply the proposed two-step partial indirect inference approach. In case of the habit persistence model, we show the effect that calibration of a subset of model parameters has on the estimation uncertainty of the remaining parameters. We find that estimating all model parameters in a two-step approach does not allow to identify all the structural parameters. Calibration of the model

parameters for which we do not find an appropriate mapping to the empirical data and which do not describe an economic mechanism appropriately, stabilizes the estimation. Moreover, calibrating all nuisance parameters of the model does not change the estimation results qualitatively. The advantage of estimating all parameters of the model, for which we find suitable moment conditions, is that it allows us to assess the impact that the estimation uncertainty of the technology parameters has on the estimates of the preference parameters.

Using the LRR model, we compare two strategies for estimating the parameters of the model; the PII strategy and the indirect inference approach of Grammig and Küchlin (2018). This sheds light on the difference between the two approaches which entails whether we conceive the structural model to be a correctly specified model that is it describes economic reality appropriately or whether we believe that the model is partially misspecified. This misspecification of the model results in different probability limits for the simulated and empirical data in the asymptotic covariance. Comparing the two strategies enables a discussion of the degree of misspecification encountered in the model. The misspecification appears to be negligible for the LRR model.

The remainder of the paper is organized as follows: In Section 4.2, we present the habit persistence model introduced by Campbell and Cochrane (1999) and the long-run risk model proposed by Bansal and Yaron (2004). The two models introduce a time-varying risk premium based on the most prominent utility function specifications in the asset pricing literature. The two-step partial indirect inference estimator and its asymptotic distribution is developed and discussed in Section 4.3. Section 4.4 elaborates on the specifications of the instrumental models for the partial equilibrium asset pricing models. The estimation results are shown in Section 4.5 and we discuss the implication of allowing for misspecification of the structural model. Section 4.6 concludes. The appendix is composed of a detailed discussion of the model solution details of both asset pricing models in Appendix 4.A, the regularity conditions of the two-step partial indirect inference estimator in Appendix 4.B, the derivation of the asymptotic distribution of the two-step partial indirect inference estimator in Appendix 4.C, detailed information on the instrumental model for the LRR model in Appendix 4.D and a description of the data in Appendix 4.E.

4.2. Partial equilibrium asset pricing

We introduce the two models in Section 4.2.1 and 4.2.2. We provide the relevant model equations and elaborate on the technology parameters describing the macroeconomy as well as the preference parameters that index the asset pricing part in both models.

4.2.1. Habit persistence model

We present the habit persistence model as it is described in Campbell and Cochrane (1999). The habit persistence model builds on the stylized fact that risk premia are not time-invariant but are higher in business cycle troughs. The standard representative agent consumption based asset pricing model is augmented by a slow moving habit level added to the power utility. Hence, the utility depends on the consumption level C_t and a habit level of consumption X_t. The basic mechanism of the habit persistence model is as follows: In a business cycle trough, consumption declines towards habit which causes the curvature of the utility function to rise. Thus, the coefficient of relative risk aversion increases and the risk premium rises. Through these dynamics, risky asset prices fall and expected returns rise. The core of the model specification is a slowly time-varying counter-cyclical risk premium. Stock market volatility in the habit persistence model is explained by consumption risk, which is amplified by the high risk aversion in bad states. The habit persistence model explains the equity premium puzzle by the high stock market volatility and a high level of risk aversion.

The model of the technology process in the macroeconomy is based on two assumptions. First, consumption growth is log normally distributed with the same mean and standard deviation as observed consumption growth. Second, a constant risk-free interest rate is modeled in an endowment economy which is consistent with a linear production technology. Campbell and Cochrane (1999) specify an endowment economy where consumption growth follows an i.i.d. log normal process, i.e.

$$\Delta c_{t+1} = c_{t+1} - c_t = g + v_{t+1} \qquad \text{where} \quad v_{t+1} \sim \mathcal{N}\left(0, \sigma_v^2\right) \quad \text{i.i.d.}, \tag{4.2.1}$$

where g is the mean of log consumption growth, σ_v^2 is the variance of log consumption growth and $c_t = \log C_t$. The vector of technology parameters is given by the consumption growth moments $\xi^M = (g, \sigma_v)^\top$.

Concerning the specification of the representive agent's preferences, Campbell and Cochrane (1999) model the dynamics of the surplus consumption ratio $S_t = \frac{C_t - X_t}{C_t}$ rather than habit X_t itself. This surplus consumption ratio is increasing in consumption. Hence, a low value of S_t implies a bad state.

They assume that a representative agent maximizes her expected lifetime utility

$$\mathbb{E}\left[\sum_{j=0}^{\infty} \delta^j \frac{(C_{t+j} - X_{t+j})^{1-\gamma} - 1}{1-\gamma}\right], \tag{4.2.2}$$

where δ is the subjective time discount factor or time preference rate and γ is the risk-aversion coefficient, subject to her budget constraint. To generate a nonlinear and slow response to

consumption, Campbell and Cochrane (1999) specify the log surplus consumption ratio $s_t = \log S_t$ to be a heteroskedastic AR(1) process, i.e.

$$s_{t+1} = (1-\phi)\bar{s} + \phi s_t + \lambda(s_t)\,(c_{t+1} - c_t - g)\,, \tag{4.2.3}$$

where ϕ is a habit persistence parameter which describes the adjustment of habit to consumption and $\lambda(s_t)$ is a sensitivity function which measures the sensitivity of surplus consumption to changes in consumption. The sensitivity function is defined as

$$\lambda(s_t) = \begin{cases} 1/\bar{S}\sqrt{1-2(s_t-\bar{s})} - 1 & \text{for} \quad s_t \leq s_{\max} \\ 0 & \text{for} \quad s_t > s_{\max}, \end{cases}$$

where $\bar{S} = \sigma_v\sqrt{\frac{\gamma}{1-\phi}}$ links the steady state surplus consumption ratio to the model parameters. Moreover, $\bar{s} = \log \bar{S}$ and $s_{max} = \bar{s} + \frac{1}{2}\left(1-\bar{S}^2\right)$. This specification of preferences and technology leads to an intertemporal marginal rate of substitution or stochastic discount factor M_{t+1} given by

$$M_{t+1} = \delta\left(\frac{S_{t+1}}{S_t}\frac{C_{t+1}}{C_t}\right)^{-\gamma}. \tag{4.2.4}$$

The preference parameters are collected in a vector $\xi^P = (\delta, \gamma, \phi)^\top$ and contain the time preference rate and risk aversion along with the persistence of habit. To complete the model of the macroeconomy and the agent's preferences, Campbell and Cochrane (1999) describe asset prices by modeling a stock as a claim to a consumption stream. The price-consumption ratio for a consumption claim can be expressed as a function of the surplus consumption ratio

$$\frac{P_t}{C_t}(s_t) = \mathbb{E}_t\left[M_{t+1}\left(1 + \frac{P_{t+1}}{C_{t+1}}(s_{t+1})\right)\frac{C_{t+1}}{C_t}\right] \tag{4.2.5}$$

and is solved numerically over the state variable s_t. Geometric returns on the consumption asset are calculated as

$$r_{c,t+1} = \log\left[\frac{1 + P_{t+1}/C_{t+1}}{P_t/C_t}\frac{C_{t+1}}{C_t}\right]. \tag{4.2.6}$$

A thorough description of the habit persistence model, the evolution of the constant log risk-free rate, the maximum Sharpe ratio and details on the model solution can be found in Appendix 4.A.1.

4.2.2. Long run risk model

In this section, we briefly describe the long-run risk (LRR) model introduced by Bansal and Yaron (2004). The model uses recursive Epstein-Zin-Weil preferences (Epstein and

Zin (1989) and Weil (1989)) and is built on a small predictable component in the growth dynamics.

The macroeconomy is composed of consumption and dividend growth, denoted by g_t and $g_{d,t}$ which are guided by two latent processes. These are given by a small predictable long-run component, x_t, and stochastic variance, σ_t^2, accounting for fluctuations in economic uncertainty, as shown in the following equations

$$g_{t+1} = \mu_c + x_t + \sigma_t \eta_{t+1}, \tag{4.2.7}$$

$$x_{t+1} = \rho x_t + \varphi_e e_{t+1}, \tag{4.2.8}$$

$$g_{d,t+1} = \mu_d + \phi x_t + \varphi_d \sigma_t u_{t+1}, \tag{4.2.9}$$

$$\sigma_{t+1}^2 = \bar{\sigma}^2 + \nu \left(\sigma_t^2 - \bar{\sigma}^2 \right) + \sigma_\omega \omega_{t+1}, \tag{4.2.10}$$

where η, e, u and ω are i.i.d. $\mathcal{N}(0,1)$ and are contemporaneously uncorrelated. Moreover, μ_c and μ_d denote expected consumption and dividend growth, ρ gives the persistence of the long-run risk component, $\bar{\sigma}^2$ is the unconditional variance, ϕ is the leverage ratio on expected consumption growth, and φ_e and φ_d are variance scaling parameters. Moreover, ν is the mean-reversion parameter of economic uncertainty and σ_ω^2 is the variance of an economic uncertainty shock.

The vector of technology parameters is given by $\xi^M = (\mu_c, \mu_d, \rho, \varphi_e, \bar{\sigma}, \phi, \varphi_d, \nu, \sigma_\omega)^\top$. In this paper, we follow the special case outlined in Bansal and Yaron (2004) and Grammig and Küchlin (2018) that does not allow for economic uncertainty as the calibrated values of both parameters are close to zero. This implies that $\nu = \sigma_\omega = 0$ and the variance is given by $\sigma_{t+1}^2 = \bar{\sigma}^2$. Hence, the underlying macroeconomic parameters are given by $\xi^M = (\mu_c, \mu_d, \rho, \varphi_e, \bar{\sigma}, \phi, \varphi_d)^\top$.

The representative agent in this model has recursive Epstein-Zin-Weil preferences (Epstein and Zin (1989) and Weil (1989)) with the corresponding utility function given by

$$U_t = \left[(1-\delta) C_t^{\frac{1-\gamma}{\theta}} + \delta \left(\mathbb{E}\left(U_{t+1}^{1-\gamma} \right) \right)^{\frac{1}{\theta}} \right]^{\frac{\theta}{1-\gamma}}, \tag{4.2.11}$$

where C_t is aggregate consumption and the constant is defined as $\theta = \frac{1-\gamma}{1-\frac{1}{\psi}}$. As in the habit persistence model, δ is the subjective discount factor or time preference rate, γ is the coefficient of relative risk aversion and ψ gives the intertemporal elasticity of substitution. The representative agent optimizes her expected life-time utility subject to the following budget constraint

$$W_{t+1} = (W_t - C_t) R_{a,t+1}, \tag{4.2.12}$$

where W denotes aggregate wealth and R_a is the gross return of the latent aggregate wealth portfolio which is given by the claim to aggregate consumption. The resulting Euler equation derived by Epstein and Zin (1989) reads

$$\mathbb{E}_t \left[\delta^\theta G_{t+1}^{-\frac{\theta}{\psi}} R_{a,t+1}^{-(1-\theta)} R_{i,t+1} \right] = 1, \tag{4.2.13}$$

where G is gross consumption growth and R_i is the gross return of a risky asset. The vector of preference parameters is hence given by $\xi^P = (\delta, \gamma, \psi)^\top$.

The model is solved as outlined in Grammig and Küchlin (2018). The solution is based on linear approximations introduced in Campbell and Shiller (1988) and the linear approach described in Bansal and Yaron (2004). A detailed discussion of the long-run risk model details, the simulation approach and the time aggregation can be found in the Appendix of Grammig and Küchlin (2018). A brief summary of the model solution and a description of the log risk-free rate and the equity premium can be found in Appendix 4.A.2.

4.3. A two-step partial indirect inference approach

The essence of the partial indirect inference (PII) approach introduced by Dridi et al. (2007) (DGR) is that it acknowledges that parts of the (partial equilibrium) asset pricing model are misspecified, while other parts claim to capture economic reality. This is achieved by the using a binding function that enables the consistent estimation of the structural parameters of interest which describe the part of the model that conveys economic truth. The remaining parameters are treated as nuisance parameters that are needed to generate model-implied data for the simulation-based approach but do not capture economic reality.

PII is an extension of the classical indirect inference estimation approach, proposed by Gourieroux et al. (1993) and Smith (1993) and refined by Gallant and Tauchen (1996), that puts emphasis on the consistent estimation of the structural parameters of interest which describe a model that is believed to be correctly specified. The theory thus allows for misspecified structural models. Both approaches are indirect in that they are used to estimate analytically intractable models whose underlying structural parameters, ξ, are not directly estimable. Indirect inference estimation relies on the identification of a binding function $\theta(\xi)$ that connects structural parameters ξ to instrumental parameters θ. The latter are obtained from the estimation of an instrumental model, $\mathcal{N}_\theta$, that is chosen to convey information about the structural parameters. These instrumental parameters are estimated based on empirical data and on simulated data from the structural model. We obtain estimates of the structural parameters by minimizing the (quadratic) distance between the two sets of instrumental parameters. Implicitly, indirect inference is based on the assumption that the

structural model is correctly specified and that the simulations from the structural model stem from the same data generating process (DGP) as the empirical data. In our setting, we believe that the underlying partial equilibrium model is misspecified in some dimension. As a consequence of this misspecification, we can no longer assume that simulations from the structural model are realizations from the same DGP as the empirical data.

In the notation of DGR, the structural parameters of interest are denoted by ξ_1, whereas the nuisance parameters are labeled as ξ_2, i.e. $\xi = (\xi_1, \xi_2)^\top$. At the core of PII lies the consistent estimation of the true unknown value ξ_1^0 which is obtained by solving the simulation and sample-based counterpart of the binding function $\theta(\xi)$. The necessary condition for consistency is provided by encompassing conditions. The structural parameters fully encompass the instrumental parameters if it holds that

$$\theta^0 = \tilde{\theta}^0 \left(\xi_1^0, \bar{\xi}_2 \right), \tag{4.3.1}$$

where $\bar{\xi}_2$ denotes the pseudo-true value of the nuisance parameters. This amounts to finding a suitable instrumental model that characterizes the true unknown parameters ξ_1^0 given the misspecification reflected in the nuisance parameters $\bar{\xi}_2$. The encompassing condition implies that even though the data generating process of the empirical data and the simulated time series are not the same, the instrumental model is chosen appropriately to identify ξ_1 such that the probability limits of the instrumental parameter are the same. Hence, the probability limit of the binding function is the same as the true instrumental parameter obtained for the empirical data. This instrumental model has to be able to capture the dynamics of both, the empirical data and the simulator. When encompassing holds, the impact of the misspecified part of the model described by the pseudo-true parameters $\bar{\xi}_2$ does not impair the consistent estimation of the structural parameters of interest. Full encompassing comes into play when all structural parameters are estimated. In case full encompassing does not hold, i.e. not all encompassing conditions hold or only a subset of the structural parameters are estimated, consistent estimation of the structural parameters of interest, ξ_1, can be achieved when partial encompassing holds. This is given by

$$\theta^0 = \tilde{\theta}^0 \left(\xi_1^0, \bar{\xi}_{21} \right), \tag{4.3.2}$$

where $\xi_2 = (\xi_{21}, \xi_{22})^\top$ and $\bar{\xi}_{21}$ are the pseudo-true values of the nuisance parameters that are part of the binding function. In this case, a subset of the nuisance parameters given by ξ_{22} remain fixed or calibrated. The condition implies that the under-identification in the binding function in Equation (4.3.1) is with respect to these parameters and not with respect to ξ_1 or

ξ_{21}. In case partial encompassing holds, the probability limit of the binding function does not depend on the nuisance parameters ξ_{22}.[2]

Next, we provide details on how the partial indirect inference strategy can be combined with a two-step estimation. We use moment-condition based instrumental models and derive the estimator and discuss the associated inference. We describe the two-step partial indirect inference methodology in Section 4.3.1 and elaborate on the asymptotic distribution of the estimators in Section 4.3.2.

4.3.1. The two-step partial indirect inference estimator

In addition to the partial indirect inference framework, we follow Grammig and Küchlin (2018) and implement a two-step estimation procedure that is based on the intricate structure of partial equilibrium asset pricing models. Partial equilibrium asset pricing models imply stochastic processes for consumption and dividend growth that only depend on the technology parameters, ξ^M, and stochastic processes for the risk-free rate, the return of the market portfolio and the price-dividend ratio indexed by the technology and preference parameters, ξ^M and ξ^P. Our estimation methodology builds on a two-step strategy that separates the estimation of the technology and preference parameters. In case of the Campbell and Cochrane (1999) habit persistence model, we believe that the preference parameters of the utility function, the time preference rate δ and the risk-aversion coefficient γ, contain economic truth reflected in the implied asset pricing fluctuations, i.e. $\xi_1^P = (\delta, \gamma)^\top$. The remaining parameters, the habit persistence $\xi_2^P = \phi$ and the technology parameters $\xi^M = (g, \sigma_v)^\top$ are conceived as nuisance parameters that are used to close the model. A similar decomposition of the parameters can be conducted for the long-run risk model of Bansal and Yaron (2004). In this model, the preference parameters that govern the asset pricing dynamics are given by $\xi^P = (\delta, \gamma, \psi)^\top$, and the long-run risk component is modeled by $\xi_1^M = (\rho, \varphi_e)^\top$, whereas the nuisance parameters that describe the remaining macroeconomy are given by $\xi_2^M = (\mu_c, \mu_d, \bar{\sigma}, \phi, \varphi_d)^\top$. For both models, the simulations of the technology or macroeconomic variables and the financial variables can be carried out sequentially. For the LRR model, the macroeconomic time series are the consumption and dividend growth series $y^M = \left(g_t, g_{d,t}\right)^\top$ and the financial time series consist of the risk-free rate, the return of the market portfolio and the price-dividend ratio $y^P = \left(r_t^f, r_t^m, z_{m,t}\right)^\top$. In case of the habit persistence model, the macroeconomy is described by the consumption growth dynamics, i.e. $y^M = c_t$ and the financial time series are given as in the LRR model.

In the following exposition, we treat a subset of the preference parameters ξ_1^P as the structural parameters of interest (ξ_1 in the DGR notation), whereas the technology parameters

[2] In finite sample, the binding function may depend on the nuisance parameters ξ_{22}.

ξ^M and the remaining preference parameters ξ_2^P are treated as nuisance parameters (ξ_2 in the DGR notation). Note that this notation is consistent with the habit persistence model. In case of the LRR model, the structural parameters of interest are given by $\xi_1 = \left(\xi^P, \xi_1^M\right)^\top$, whereas the nuisance parameters are those of $\xi_2 = \xi_2^M$. Moreover, we limit the derivations of the two-step PII estimator to the full encompassing case as given in Equation (4.3.1). Extensions to the partial compassing case as given in Equation (4.3.2) are straightforward and are treated in more detail in footnotes whenever helpful.

We denote the empirical time series as $\{y_t^M\}_{t=1}^T$ and $\{y_t^P\}_{t=1}^T$ and the model-implied series obtained through simulations of the technology part by $\left\{\tilde{y}_t^{M,s}(\xi^M)\right\}_{t=1}^T$ and the financial part by $\left\{\tilde{y}_t^{P,s}(\xi^M, \xi^P)\right\}_{t=1}^T$. Comparable to Dridi et al. (2007), we perceive the observed data to be a realization of a stochastic process with the true unknown p.d.f. $P_0 : \{y_t, t \in \mathbb{Z}\}$

Assumption 4.3.1. (i) P_0 belongs to a family $\mathcal{P}$ of p.d.f. on $\mathcal{Y}^{\mathbb{Z}}$,

(ii) $\tilde{\xi}_1^P$ is an application of $\mathcal{P}$ onto $\Xi_1^P = \tilde{\xi}_1^P(\mathcal{P})$,

(iii) $\tilde{\xi}_1^P(\mathcal{P}_0) = \xi_{1,0}^P$ belongs to the interior of Ξ_1^P.

This assumption is concerned with the structural parameters of interest, ξ_1^P. These parameters are defined through a set of identifying moment conditions given by a partial parametric model

$$\mathbb{E}_{\mathcal{P}}\left[h(y_t, \ldots, y_{t-p}; \xi_1^P)\right] = 0 \quad \rightarrow \ \xi_1^P = \tilde{\xi}_1^P(\mathcal{P}).$$

In case of a misspecified simulator, the partial parametric model defined by Assumption 4.3.1 is plugged into the structural model which is fully parametric but misspecified. Additional nominal assumptions are needed to specify this fully parametric structural model which is likely inconsistent with the true data-generating process.

Assumption 4.3.2 (Nominal assumption). $\{y_t, t \in \mathbb{Z}\}$ follows a stationary and ergodic process that follows the recursive structure of the partial equilibrium model based on the macroeconomic time series indexed by ξ^M and financial variables modeled by ξ^M and ξ^P. The nonlinear simultaneous model of $y_t = \left(y_t^M, y_t^P\right)^\top$ is conformable to

(i) $r\left(y_t, y_{t-1}, v_t, \xi\right) = 0, \qquad \varphi\left(v_t, v_{t-1}, \tilde{\varepsilon}_t, \xi\right) = 0,$

$\xi = \left(\xi_1^P, \xi_2^P, \xi^M\right) \in \left(\Xi_1^P \times \Xi_2^P \times \Xi^M\right) = \Xi$ which is a compact subset of $\mathbb{R}^{p_1+p_2+p_M}$,

(ii) $\{\tilde{\varepsilon}_t, t \in \mathbb{Z}\}$ is white noise with known distribution G_*,

where $r(\cdot)$ and $\varphi(\cdot)$ describe the evolution of the observed time series given the state variables and that of the state variables given the exogenous time series, respectively. This assumption is comparable to Gourieroux et al. (1993). Whereas they assume that this process is the true data-generating process, here this is only a nominal assumption that specifies a fully parametric model from which we can simulate but which is misspecified. For given values of $\xi = (\xi_1^P, \xi_2^P, \xi^M)$, we simulate the macroeconomic variables $\left\{\tilde{y}_t^{M,s}(\xi^M)\right\}_{t=1}^T$ and the financial variables by $\left\{\tilde{y}_t^{P,s}(\xi^M, \xi^P)\right\}_{t=1}^T$ from the simulated values $\{\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_T\}$ from G_*. P_* denotes the probability limit of the simulated processes which is different from that of the observed data, P_0.

Indirect inference about the true value $\xi_{1,0}^P$ of ξ_1^P is conducted by means of the misspecified structural model defined in Assumption 4.3.2 and separate instrumental models for the macroeconomic and the financial variables, $\mathcal{N}_\theta^M$ and $\mathcal{N}_\theta^P$, which are indexed by two sets of pseudo-true values of instrumental parameters θ^M and θ^P. These pseudo-true values are defined as the probability limits of extremum estimators associated with the criterion functions $Q_T^M\left(y_t^M, \theta^M\right)$ and $Q_T^P\left(y_t^P, \theta^P\right)$ for each of the models.[3] Thus, the idea of the partial indirect inference (PII) estimator can be seen as econometrics of a misspecified structural model.

First estimation step

This first step of the estimation strategy focuses on the macroeconomic nuisance parameters which describe parts of the partial equilibrium models that are conceived as a caricature of economic reality. We base the macroeconomic instrumental model on a set of k^M GMM-type moment restrictions. This set is given by

$$\mathbb{E}\left[u_t^M\left(y_t^M, \theta_0^M\right)\right] = 0, \tag{4.3.3}$$

where y_t^M denotes the observed macroeconomic data. The number of moment restrictions is at least as large as the number of structural parameters, i.e. $k^M = \dim \theta^M \geq \dim \xi^M$ and $\theta^M \in \Theta^M \subset \mathbb{R}^{k^M}$.

To estimate the instrumental parameters θ^M, we minimize a GMM-type criterion function given by

$$\min_{\{\theta^M \in \Theta^M\}} Q_T^M\left(y_t^M, \theta^M\right), \tag{4.3.4}$$

[3] Here, $\theta^M \in \Theta^M$ which is a compact subset of $\mathbb{R}^{k^M}$ and y_t^M denote lagged values of y_t^M for a fixed number of K lags and $\theta^P \in \Theta^P$ which is a compact subset of $\mathbb{R}^{k^P}$ and y_t^P denote lagged values of y_t^P for a fixed number of K lags.

$$\text{with } Q_T^M\left(y_t^M,\theta^M\right) = \frac{1}{2}\, g_T^M\left(y_t^M,\theta^M\right)^\top \widehat{W}_T^M\, g_T^M\left(y_t^M,\theta^M\right), \tag{4.3.5}$$

$$g_T^M\left(y_t^M,\theta^M\right) = \frac{1}{T}\sum_{t=1}^{T} u_t^M\left(y_t^M,\theta^M\right), \tag{4.3.6}$$

and $\widehat{W}_T^M$ is a positive semi-definite weighting matrix that converges almost surely to a non-stochastic positive semi-definite matrix W^M.

Under appropriate GMM regularity conditions given in Appendix 4.B,

$$\hat{\theta}_T^M = \underset{\{\theta^M\in\Theta^M\}}{\operatorname{argmin}}\; Q_T^M\left(y_t^M,\theta^M\right) \tag{4.3.7}$$

is a consistent estimator of θ_0^M.

Note that the probability limit of the simulated macroeconomic data is not the same as that of the observed data. Hence, we need additional assumptions for the misspecified simulator. We write the moment conditions evaluated using the simulated macroeconomic time series as

$$\mathbb{E}^*\left[u_t^M\left(\tilde{y}_t^{M,s}(\xi^M),\tilde{\theta}^M(\bar{\xi}^M)\right)\right] = 0, \tag{4.3.8}$$

where the expectation is taken with respect to G_*, the distribution of the white noise innovations given in the nominal assumption 4.3.2. Under the assumptions given in Appendix 4.B,

$$\tilde{\theta}_T^{M,s}(\xi^M) = \underset{\{\theta^M\in\Theta^M\}}{\operatorname{argmin}}\; Q_T^M\left(\tilde{y}_T^{M,s}(\xi^M),\theta^M\right) \tag{4.3.9}$$

and $\tilde{\theta}_T^M(\xi^M) = \frac{1}{S}\sum_{s=1}^{S}\tilde{\theta}_T^{M,s}(\xi^M)$ is a consistent estimator of $\tilde{\theta}^M(\bar{\xi}^M)$. Note that the instrumental parameter estimator and its probability limit are dependent on the macroeconomic parameters ξ^M used to simulate the macroeconomic time series.

For the indirect estimator of the macroeconomic nuisance parameters ξ^M, we further require assumptions on the binding function given by

Assumption 4.3.3. (i) $\tilde{\theta}^M(\cdot,\cdot)$ is one-to-one,

(ii) $P_*\lim_{T\to\infty}\frac{\partial\tilde{\theta}_T^{M,s}}{\partial\xi^{M\top}}(\bar{\xi}^M) = \frac{\partial\tilde{\theta}^M}{\partial\xi^{M\top}}(\bar{\xi}^M)$ has full column rank dim ξ^M,

and

Assumption 4.3.4. $\theta_0^M = \tilde{\theta}^M(\bar{\xi}^M)$.

The one-to-one assumption in 4.3.3 *(i)* is similar to assumption (A4) used by Gourieroux et al. (1993). It implies that the mapping between the structural and instrumental parameters

is one-to-one. Moreover, the rank assumption in 4.3.3 *(ii)* is equivalent to assumption (A14) in Dridi et al. (2007). This rank condition implies that the macroeconomic structural parameters $\bar{\xi}^M$ are identified from the first step instrumental model. Assumptions 4.3.4 is the full encompassing condition in Dridi et al. (2007). We refer to Assumption 4.3.4 as the full encompassing null hypothesis $H_0^{(M)}$ for the first step estimates. Full encompassing requires that even though the model is a misspecified simulator, the (probability limits of the) instrumental parameters, associated with the empirical and simulated data, are identical. This is the most critical assumption which depends on the choice of the instrumental model.

The first step partial indirect inference estimator of the macroeconomic nuisance parameters ξ^M is defined as

$$\hat{\xi}_T^M = \underset{\{\xi^M \in \Xi^M\}}{\operatorname{argmin}} \left[\hat{\theta}_T^M - \frac{1}{S}\sum_{s=1}^{S} \tilde{\theta}_T^{M,s}(\xi^M)\right]^\top \widehat{\Omega}_T^M \left[\hat{\theta}_T^M - \frac{1}{S}\sum_{s=1}^{S} \tilde{\theta}_T^{M,s}(\xi^M)\right], \tag{4.3.10}$$

$$\text{with} \quad \hat{\theta}_T^M = \underset{\{\theta^M \in \Theta^M\}}{\operatorname{argmin}} \, Q_T^M\left(y_t^M, \theta^M\right),$$
$$\tilde{\theta}_T^{M,s}(\xi^M) = \underset{\{\theta^M \in \Theta^M\}}{\operatorname{argmin}} \, Q_T^M\left(\tilde{y}_t^{M,s}(\xi^M), \theta^M\right),$$

and $P_* \lim_{T\to\infty} \widehat{\Omega}_T^M = \Omega^M$ is positive definite.

Proposition 4.3.1 (Consistency of the macroeconomic nuisance parameters). Under Assumptions 4.3.1 - 4.3.4 and the regularity conditions given in Appendix 4.B, $\hat{\xi}_T^M$ is a consistent estimator of the pseudo-true value $\bar{\xi}^M$.

Assumption 4.3.4 implies that the misspecified model defined in Assumption 4.3.2 endowed with the pseudo-true value $\bar{\xi}^M$ fully encompasses $\mathcal{N}_\theta^M$. Following Proposition 3.3. in Dridi et al. (2007), this full encompassing condition provides a sufficient condition for the consistency of the partial indirect inference estimator of the macroeconomic nuisance parameters $\hat{\xi}_T^M$.[4] The proof of the proposition is an adapted version of those found in Dridi and Renault (2000) and Dridi et al. (2007). In our case, we use a GMM-type instrumental model rather than a general extremum estimator. This difference features in the identification assumptions 4.B.1 and 4.B.3 and regularity conditions 4.B.2 and 4.B.4 which are specific to the GMM-type instrumental model. Assumptions 4.3.1, 4.3.2, 4.3.3 and 4.3.4 are comparable

[4] If full encompassing does not hold, we could also specify the assumption as partial encompassing, i.e. $\theta_0^M = \tilde{\theta}(\bar{\xi}_1^M)$, where ξ_1^M is a subset of the full vector of first step parameters. If we assume partial encompassing instead, then under the assumptions it implies that the misspecified model endowed with the pseudo-true value $\bar{\xi}_1^M$ partially encompasses $\mathcal{N}_\theta^M$. It can be shown that the partial encompassing condition is a sufficient condition for the consistency of the partial indirect inference estimator of the macroeconomic nuisance parameters $\hat{\xi}_{1,T}^M$

to those in Dridi et al. (2007) for the full encompassing case in the absence of exogenous variables. The extension is straightforward and we omit the proof in this paper.

Second estimation step

In the second estimation step, we take the first step macroeconomic nuisance parameter estimates $\hat{\xi}_T^M$ as given and we estimate the structural parameters of interest, the preference parameter ξ^P. The second step instrumental model captures the key asset pricing implications of the partial equilibrium models. These implications are collected in k^P moment conditions indexed by instrumental parameters $\theta^P \subset \Theta^P \in \mathbb{R}^{k^P}$ whose dimensions must be at least as large as the number of preference parameters $\dim \xi_1^P + \dim \xi_2^P$ in the full encompassing case and at least $\dim \xi_1^P$ in the partial encompassing case. These moment conditions are given by

$$\mathbb{E}\left[u_t^P\left(y_t^P, \theta_0^P\right)\right] = 0, \tag{4.3.11}$$

where y_t^P denotes the observed financial time series.
The instrumental parameter estimates $\hat{\theta}_T^P$ can be obtained as the solution of the GMM-type criterion function given by

$$\min_{\{\theta^P \in \Theta^P\}} Q_T^P\left(y_t^P, \theta^P\right), \tag{4.3.12}$$

$$\text{with } Q_T^P\left(y_t^P, \theta^P\right) = \frac{1}{2}\, g_T^P\left(y_t^P, \theta^P\right)^\top \widehat{W}_T^P g_T^P\left(y_t^P, \theta^P\right), \tag{4.3.13}$$

$$g_T^P\left(y_t^P, \theta^P\right) = \frac{1}{T}\sum_{t=1}^{T} u_t^P\left(y_t^P, \theta^P\right), \tag{4.3.14}$$

and $P_0 \lim_{T\to\infty} \widehat{W}_T^P = W^P$ is positive semi-definite.

$$\hat{\theta}_T^P = \underset{\{\theta^P \in \Theta^P\}}{\operatorname{argmin}}\, Q_T^P\left(y_t^P, \theta^P\right) \tag{4.3.15}$$

is a consistent estimator of θ_0^P under the GMM regularity conditions given in Appendix 4.B.

Due to the misspecified structural model, the probability limit for the instrumental parameter θ^P based on the simulated data $\{\tilde{y}_t^{M,s}(\xi^M)\}_{t=1}^T$ and $\{\tilde{y}_t^{P,s}(\xi^M, \xi^P)\}_{t=1}^T$ is denoted by P_*. As in the first estimation step, we need additional assumptions for the misspecified simulator of the financial variables. We impose an identification assumption for the simulated time series, i.e.

$$\mathbb{E}^*\left[u_t^P\left(\tilde{y}_t^{P,s}(\xi^M, \xi^P), \tilde{\theta}^P(\bar{\xi}^M, \xi_{1,0}^P, \bar{\xi}_2^P)\right)\right] = 0. \tag{4.3.16}$$

Under the assumptions for the GMM-type instrumental parameters given in Appendix 4.B,

$$\tilde{\theta}_T^{P,s}(\xi^M,\xi^P) = \underset{\{\theta^P\in\Theta^P\}}{\operatorname{argmin}} Q_T^P\left(\tilde{y}_T^{P,s}(\xi^M,\xi^P),\theta^P\right) \tag{4.3.17}$$

and $\tilde{\theta}_T^P(\xi^M,\xi^P) = \frac{1}{S}\sum_{s=1}^S \tilde{\theta}_T^{P,s}(\xi^M,\xi^P)$ is a consistent estimator of $\tilde{\theta}^P(\bar{\xi}^M,\xi_{1,0}^P,\bar{\xi}_2^P)$.[5]

For the indirect estimator of the structural preference parameters ξ^P, we make the following assumptions on the binding function

Assumption 4.3.5. (i) $\tilde{\theta}^P(\cdot,\cdot)$ is one-to-one,

(ii) $P_*\lim_{T\to\infty}\frac{\partial\tilde{\theta}_T^{P,s}}{\partial\xi_1^{P\top}}(\bar{\xi}^M,\xi_{1,0}^P,\bar{\xi}_2^P) = \frac{\partial\tilde{\theta}^P}{\partial\xi_1^{P\top}}(\bar{\xi}^M,\xi_{1,0}^P,\bar{\xi}_2^P)$ has full column rank dim ξ^P.

We differentiate different cases of the full and partial encompassing condition. The following is stated in terms of the full encompassing condition.

Assumption 4.3.6. $\theta_0^P = \tilde{\theta}^P(\bar{\xi}^M,\xi_{1,0}^P,\bar{\xi}_2^P)$

Assumption 4.3.5 is similar to Assumption 4.3.3. We refer to Assumption 4.3.6 as the full encompassing null hypothesis $H_0^{(P)}$ for the second step estimates. The partial encompassing condition can either be written as $\theta_0^P = \tilde{\theta}^P(\bar{\xi}^M,\xi_{1,0}^P)$, where $\xi_{1,0}^P$ is a subset of ξ_0^P which we estimate. In this case, there are nuisance parameters in the specification of the preferences, $\bar{\xi}_2^P$, that do not impair the consistent estimation of the preference parameters of interest. These nuisance parameters may affect the binding function in finite samples but do not affect it asymptotically. Alternatively, in case that only the structural parameters of interest, ξ_1^P, are estimated and all nuisance parameters, ξ_2^P and ξ^M, are calibrated, partial encompassing is expressed as $\theta_0^P = \tilde{\theta}^P(\xi_{1,0}^P)$. In this case, the asymptotic binding function only depends on the preference parameters of interested and is not impacted by the preference or macroeconomic nuisance parameters. The proofs given in Appendix 4.C can be adapted to the partial encompassing case.

The second step partial indirect inference estimator of the structural preference parameters ξ^P is defined as

$$\hat{\xi}_T^P = \underset{\{\xi^P\in\Xi^P\}}{\operatorname{argmin}}\left[\hat{\theta}_T^P - \frac{1}{S}\sum_{s=1}^S\tilde{\theta}_T^{P,s}(\xi^M,\xi^P)\right]^\top\widehat{\Omega}_T^P\left[\hat{\theta}_T^P - \frac{1}{S}\sum_{s=1}^S\tilde{\theta}_T^{P,s}(\xi^M,\xi^P)\right], \tag{4.3.18}$$

[5]The dependence of the estimator and probability limit on the underlying structural parameters associated with the simulated data is made explicit here.

$$\text{with } \hat{\theta}_T^P = \underset{\{\theta^P \in \Theta^P\}}{\operatorname{argmin}} \, Q_T^P\left(y_t^P, \theta^P\right),$$
$$\tilde{\theta}_T^{P,s}(\xi^M, \xi^P) = \underset{\{\theta^P \in \Theta^P\}}{\operatorname{argmin}} \, Q_T^P\left(\tilde{y}_t^{P,s}(\xi^M, \xi^P), \theta^P\right),$$

and $\widehat{\Omega}_T^P$ is a positive definite weighting matrix that converges almost surely to a non-stochastic positive definite matrix Ω^P.

Proposition 4.3.2 (Consistency of the structural preference parameters). Under Assumptions 4.3.1, 4.3.2, 4.3.5, 4.3.6 and the regularity conditions given in Appendix 4.B, $\hat{\xi}_T^P$ is a consistent estimator of ξ_0^P.

Given assumption 4.3.6, the misspecified model defined in the nominal assumption 4.3.2 endowed with the pseudo-true value $\bar{\xi}^M$ and the underlying true parameters ξ_0^P fully (or partially) encompasses $\mathcal{N}_\theta^P$. Hence, Proposition 3.3. in Dridi et al. (2007) applies and this full (partial) encompassing condition provides a sufficient condition for the consistency of the partial indirect inference estimator of the structural preference parameters $\hat{\xi}_T^P$. As for the first step estimates, the proof of the proposition is adapted from Dridi and Renault (2000) and Dridi et al. (2007). The GMM-type instrumental model features in the identification assumptions 4.B.5 and 4.B.7 and the regularity conditions 4.B.6 and 4.B.8. Assumptions 4.3.1, 4.3.2 and 4.3.5 are similar to the assumptions for the full encompassing case in absence of exogenous variables in Dridi et al. (2007).

4.3.2. Asymptotic distribution of the two-step partial indirect inference estimator

To derive the asymptotic distribution of our two-step partial indirect inference estimator, we make use of the identification and equicontinuity assumptions in Appendix 4.B to invoke a central limit theorem (CLT) for the moment conditions evaluated at the observed and simulated data. These CLTs are given in Appendix 4.C. Moreover, by the full encompassing null hypotheses, $H_0^{(M)} : \theta_0^M = \tilde{\theta}^M(\bar{\xi}^M)$ in Assumption 4.3.4 and $H_0^{(P)} : \theta_0^P = \tilde{\theta}^P(\bar{\xi}^M, \xi_{1,0}^P, \bar{\xi}_2^P)$ in Assumption 4.3.6, the following result is derived for the instrumental parameters for both steps.[6]

Proposition 4.3.3 (Asymptotic distribution of the instrumental parameters). Under the regularity conditions for the instrumental parameters given in Appendix 4.B, the central

[6]Using the partial encompassing null hypotheses, the line of reasoning is comparable to the full encompassing case. In this paper, we limit the exposition to the derivations in the full encompassing case.

limit theorems given in Assumptions 4.C.1 and 4.C.2 in Appendix 4.C and the encompassing conditions 4.3.4 and 4.3.6, the instrumental parameters are asymptotically distributed as

$$\sqrt{T}\begin{bmatrix} \hat{\theta}_T^M - \frac{1}{S}\sum_{s=1}^{S}\tilde{\theta}_T^{M,s}(\bar{\xi}^M) \\ \hat{\theta}_T^P - \frac{1}{S}\sum_{s=1}^{S}\tilde{\theta}_T^{P,s}(\bar{\xi}^M, \xi_{1,0}^P, \bar{\xi}_2^P) \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \Phi_0\left(S, W^M, W^P\right)\right) \tag{4.3.19}$$

where the asymptotic covariance matrix is given by

$$\Phi_0\left(S, W^M, W^P\right) = \begin{bmatrix} C_0^M & 0 \\ 0 & C_0^P \end{bmatrix} V \begin{bmatrix} C_0^{M\top} & 0 \\ 0 & C_0^{P\top} \end{bmatrix} + \frac{1}{S}\begin{bmatrix} C_0^{M*} & 0 \\ 0 & C_0^{P*} \end{bmatrix} V^* \begin{bmatrix} C_0^{M*\top} & 0 \\ 0 & C_0^{P*\top} \end{bmatrix}, \tag{4.3.20}$$

where the matrices C_0^M, C_0^{M*}, C_0^P, C_0^{P*}, V and V^* are defined in Appendix 4.C.

The asymptotic distribution of the structural parameters is described in the following proposition.

Proposition 4.3.4 (Asymptotic distribution of the structural parameters). Under Assumptions 4.3.1 to 4.3.6, the regularity conditions given in Appendix 4.B and the central limit theorems 4.C.1 and 4.C.2 given in Appendix 4.C, the structural parameter estimates of the macroeconomic nuisance parameters, ξ^M, and structural preference parameters, ξ^P,[7] follow an asymptotic normal distribution

$$\sqrt{T}\begin{bmatrix} \hat{\xi}_T^M - \bar{\xi}^M \\ \hat{\xi}_T^P - \xi_0^P \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \Sigma\left(S, \Omega^M, \Omega^P\right)\right) \tag{4.3.21}$$

where

$$\begin{aligned} &\Sigma\left(S, \Omega^M, \Omega^P\right) \\ &= \begin{bmatrix} A\left(\Omega^M\right) C_0^M & 0 \\ B\left(\Omega^P\right) C\left(\Omega^M\right) C_0^M & B\left(\Omega^P\right) C_0^P \end{bmatrix} V \begin{bmatrix} C_0^{M\top} A\left(\Omega^M\right)^\top & C_0^{M\top} C\left(\Omega^M\right)^\top B\left(\Omega^P\right)^\top \\ 0 & C_0^{P\top} B\left(\Omega^P\right)^\top \end{bmatrix} \\ &\quad + \frac{1}{S}\begin{bmatrix} A\left(\Omega^M\right) C_0^{M*} & 0 \\ B\left(\Omega^P\right) C\left(\Omega^M\right) C_0^{M*} & B\left(\Omega^P\right) C_0^{P*} \end{bmatrix} V^* \begin{bmatrix} C_0^{M*\top} A\left(\Omega^M\right)^\top & C_0^{M*\top} C\left(\Omega^M\right)^\top B\left(\Omega^P\right)^\top \\ 0 & C_0^{P*\top} B\left(\Omega^P\right)^\top \end{bmatrix} \end{aligned} \tag{4.3.22}$$

[7] Note that ξ_0^P is composed of $\xi_{1,0}^P$ and $\bar{\xi}_2^P$, the true and pseudo-true values of the preference parameters, respectively.

with

$$A\left(\Omega^M\right) = \left\{\frac{\partial\tilde{\theta}^M}{\partial\xi^M}\left(\bar{\xi}^M\right)^\top \Omega^M \frac{\partial\tilde{\theta}^M}{\partial\xi^{M\top}}\left(\bar{\xi}^M\right)\right\}^{-1} \frac{\partial\tilde{\theta}^M}{\partial\xi^M}\left(\hat{\xi}_T^M\right)^\top \Omega^M, \tag{4.3.23}$$

$$B\left(\Omega^P\right) = \left\{\frac{\partial\tilde{\theta}^P}{\partial\xi^P}\left(\bar{\xi}^M, \xi_{1,0}^P, \bar{\xi}_2^P\right)^\top \Omega^P \frac{\partial\tilde{\theta}^P}{\partial\xi^{P\top}}\left(\bar{\xi}^M, \xi_{1,0}^P, \bar{\xi}_2^P\right)\right\}^{-1} \frac{\partial\tilde{\theta}^P}{\partial\xi^P}\left(\hat{\xi}_T^M, \hat{\xi}_T^P\right)^\top \Omega^P, \tag{4.3.24}$$

$$C\left(\Omega^M\right) = -\frac{\partial\tilde{\theta}^P}{\partial\xi^{M\top}}\left(\bar{\xi}^M, \xi_{1,0}^P, \bar{\xi}_2^P\right) A\left(\Omega^M\right). \tag{4.3.25}$$

The proofs of propositions 4.3.3 and 4.3.4 are provided in Appendix 4.C. The block matrix structure, $A(\Omega^M)$, $B(\Omega^P)$ and $C(\Omega^M)$ and the associated zero blocks stem from the two-step nature of the estimator. The two summands with separate estimated quantities for the observed, C_0^M, C_0^P and V, and the simulated time series, $C_0^{M^*}$, $C_0^{P^*}$ and V^*, are a result of the PII approach.

To illustrate the effect of the partial indirect inference approach in comparison to indirect inference, we provide the limiting distribution of the Grammig and Küchlin (2018) two-step indirect inference estimator in Appendix 4.C.3. Moreover, the estimation of the components of the asymptotic covariance matrix is detailed in Appendix 4.C.4.

4.4. Specification of the instrumental models

In this section, we describe the parameter classification and model specification for both models in more details. We elaborate on the parts of the model that we believe to reflect economic reality and on other parts that are potentially misspecified. Moreover, we discuss the choice of moment conditions for the instrumental models for both partial equilibrium asset pricing models.

4.4.1. Model specification for habit persistence model

In the habit persistence model proposed by Campbell and Cochrane (1999), we conceive the implied stochastic discount factor that describes the investor's preferences to be the economic reality that is captured. Hence, the time preference rate, δ, and the risk aversion, γ, form the structural parameters of interest, i.e. $\xi_1^P = (\delta, \gamma)^\top$, whereas the persistence of surplus consumption, $\xi_2^P = \phi$, and the dynamics of consumption growth, $\xi^M = (g, \sigma_v)^\top$, index parts of the model that are misspecified.

In the following, we detail a two-step estimation approach that makes use of separate instrumental models for the consumption growth dynamics and the asset pricing implications. We estimate the technology part given by the dynamics of consumption growth in a first

step using an AR(1) model as the instrumental model. We further augment the instrumental model by the mean and variance of consumption growth. The resulting moment conditions that are used to estimate the instrumental parameters $\theta^M = \left(\kappa, \alpha, \sigma_\zeta, \mu_c, \sigma_c\right)^\top$ are given by

$$\mathbb{E}\left[u_t^M\left(y_t^M, \theta^M\right)\right] = \mathbb{E}\begin{bmatrix} \zeta_t \\ \zeta_t \cdot c_{t-1} \\ (\zeta_t)^2 - (\sigma_\zeta)^2 \\ c_t - \mu_c \\ (c_t)^2 - (\mu_c)^2 - (\sigma_c)^2 \end{bmatrix}, \tag{4.4.1}$$

where $\zeta_t = c_t - \kappa - \alpha c_{t-1}$.

The choice of moment conditions for the second step instrumental parameters is guided by the model-implied equations for the risk-free rate and the conditional Sharpe ratio given in Equations (4.A.1) and (4.A.2). The risk-free rate is affected by the time preference parameter, δ, as well as the intertemporal substitution and the precautionary savings motive. These, in turn, are driven by the risk aversion parameter γ. We use the expected value of the risk-free rate as a moment condition, i.e. $\mathbb{E}(R_t^f) = \mu_{r_f}$. In addition, the conditional Sharpe ratio is influenced by the risk aversion parameter. To capture the evolution of the risky asset, we make use of the expected value of the equity premium, $\mathbb{E}(R_t - R_t^f) = \mathbb{E}(R_t^e) = \mu_{r_m^e}$, and its second moment, $\text{Var}(R_t - R_t^f) = \sigma^2(R_t - R_t^f) = \sigma^2(R_t^e) = \sigma^2_{r_m^e}$, separately. Finally, we pin down the persistence of habit $\xi_2^P = \phi$ by using the first order autocovariance of the price-dividend ratio $PD_t = z_{m,t}$, i.e. $\mathbb{E}(z_{m,t} z_{m,t-1}) = \rho_{z_m}$ as an instrumental moment.

Hence, the moment restrictions used to estimate the instrumental parameters $\theta^P = \left(\mu_{r_m^e}, \mu_{r_f}, \sigma_{r_m^e}, \rho_{z_m}\right)^\top$ are given by

$$\mathbb{E}\left[u_t^P\left(y_t^P, \theta^P\right)\right] = \mathbb{E}\begin{bmatrix} R_{m,t}^e - \mu_{r_m^e} \\ R_t^f - \mu_{r_f} \\ (R_{m,t}^e)^2 - (\mu_{r_m^e})^2 - (\sigma_{r_m^e})^2 \\ z_{m,t} z_{m,t-1} - \rho_{z_m} \end{bmatrix}. \tag{4.4.2}$$

Note that the moment conditions yield an exactly identified problem. Hence, there is no need to numerically optimize the instrumental objective function given in Equation (4.3.13). Instead, we use sample equivalents of the moments specified above.

The model is based on plausible microeconomic foundations but it can be critized as an attempt to reverse engineer the statistical inadequacies of the classical consumption based asset pricing model (c.f. Gallant and McCulloch (2009)). This can be seen most clearly when keeping in mind that in all formulations of the model, the habit persistence parameter ϕ is close to but smaller than unity to avoid that the surplus consumption ratio process

becomes nonstationary. There is no plausible economic interpretation for this necessity. However, if we fail to implement this condition, the risk-free rate varies heavily which is not compatible with the observed U.S. data. As the persistence of surplus consumption is not identified properly in the data or by economic theory, we choose to calibrate it in a second setting, i.e. we set $\bar{\xi}_2^P = \phi$. Hence, we drop the fourth moment condition and are left with the moment restrictions

$$\mathbb{E}\left[u_t^P\left(y_t^P, \theta^P\right)\right] = \mathbb{E}\begin{bmatrix} R_{m,t}^e - \mu_{r_m^e} \\ R_t^f - \mu_{r_f} \\ (R_{m,t}^e)^2 - (\mu_{r_m^e})^2 - (\sigma_{r_m^e})^2 \end{bmatrix}, \tag{4.4.3}$$

which are used to estimate the instrumental parameters $\theta^P = \left(\mu_{r_m^e}, \mu_{r_f}, \sigma_{r_m^e}\right)^\top$.

4.4.2. Model specification for LRR model

In case of the LRR model, the asset pricing dynamics are driven by the utility specification that advocates an early resolution of uncertainty risk and a small persistent and predictable long-run risk component. Hence, the structural parameters of interest are given by the recursive utility parameters, $\xi^P = (\delta, \gamma, \psi)^\top$ and the dynamics of the long run-risk component, $\xi_1^M = (\rho, \varphi_e)^\top$. The remaining parameters index parts of the model understood as caricatural aspects. These nuisance parameters are given by $\xi_2^M = (\mu_c, \mu_d, \bar{\sigma}, \phi, \varphi_d)^\top$.

As for the habit persistence model, we specify a sequential estimation approach based on two instrumental models that is based on the intricate two-step model structure. The main challenge in the specification of the instrumental model for the technology parameters in the LRR model is to account for the predictable growth component, x_t. It introduces very persistent but small serial correlations in consumption and dividend growth. We use the heterogeneous autoregressive (HAR) model introduced by Corsi (2009) to estimate the dynamics of the consumption and dividend growth. The model includes different sampling frequencies in the autoregressive model and thereby induces long memory. This allows to capture deviations from i.i.d. growth. The HAR model is given by

$$\begin{bmatrix} g_t \\ g_{d,t} \end{bmatrix} = \begin{bmatrix} \kappa_1 \\ \kappa_2 \end{bmatrix} + \sum_{i=1}^{\tau} \Phi_i L^i \begin{bmatrix} g_t \\ g_{d,t} \end{bmatrix} + \Phi_{\tau+1} \begin{bmatrix} g_t^{f(h_1)} \\ g_{d,t}^{f(h_1)} \end{bmatrix} + \Phi_{\tau+2} \begin{bmatrix} g_t^{f(h_2)} \\ g_{d,t}^{f(h_2)} \end{bmatrix} + \begin{bmatrix} \zeta_{1,t} \\ \zeta_{2,t} \end{bmatrix}, \tag{4.4.4}$$

where $f(h_i)$ denotes the time aggregation of the base frequency over h_i periods and Φ_i, $\Phi_{\tau+1}$ and $\Phi_{\tau+2}$ are (2×2) autoregressive matrices over the respective aggregation frequency. We follow Grammig and Küchlin (2018) and use six lags on the base frequency (quartely), i.e. $\tau = 6$. Moreover, the time aggregation of the consumption and dividend growth series is

conducted as in Calvet and Czellar (2015) and we use an annual, i.e. $h_1 = 4$, and tri-annual aggregation, i.e. $h_2 = 12$. The first step instrumental parameter vector is augmented by the mean and standard deviations of consumption and dividend growth based on the different frequencies. The first step instrumental parameters consist of the parameters of the HAR model, the expected values of the consumption and dividend growth and the standard deviations of both series and are given by

$$\begin{aligned}
\theta^M &= \left(\theta^{HAR\top}, \mu_g^\top, \sigma_g^\top\right)^\top \\
\theta^{HAR} &= \left(\kappa_1, \kappa_2, \text{vec}(\Phi_1)^\top, \ldots, \text{vec}(\Phi_{\tau+1})^\top, \text{vec}(\Phi_{\tau+2})^\top, \sigma_{\zeta_1}, \sigma_{\zeta_2}, \sigma_{\zeta_1\zeta_2}\right)^\top, \\
\mu_g &= \left(\mu_c, \mu_d, \mu_c^{f(h_1)}, \mu_d^{f(h_1)}, \mu_c^{f(h_2)}, \mu_d^{f(h_2)}\right)^\top, \\
\sigma_g &= \left(\sigma_c, \sigma_d, \sigma_c^{f(h_1)}, \sigma_d^{f(h_1)}, \sigma_c^{f(h_2)}, \sigma_d^{f(h_2)}\right)^\top.
\end{aligned}$$

The associated moment conditions are given in Equation (4.D.1) in Appendix 4.D. The connections between the structural technology parameters, ξ^M, and the instrumental parameters, θ^M, are described in Grammig and Küchlin (2018). The persistence of the long-run predictable component, ρ, and the leverage ratio on the expected consumption growth, ϕ, are linked to the autoregressive parameters of the HAR model, Φ_i, $\Phi_{\tau+1}$ and $\Phi_{\tau+2}$. The unconditional expected values of log consumption and dividend growth, μ_c and μ_d, are captured by the constants, κ_1 and κ_2, and the means of the growth series. Lastly, the unconditional variance, $\bar{\sigma}^2$, and the variance scaling parameters, φ_e and φ_d are identified from the (co)variance of the HAR innovation distributions, given by σ_{ζ_1}, σ_{ζ_2} and $\sigma_{\zeta_1\zeta_2}$. As the number of instrumental parameters is equal to the number of moment conditions, the optimization is independent of the weighting matrix choice $\widehat{W}_T^M$.

Next, we turn to the specification for the second estimation step. The second step instrumental parameters, θ^P, are defined by a set of moment conditions which capture the basic asset pricing implications of the LRR model. The choice is guided by the model-implied risk-free rate and the market equity premium. Closed form solutions of both quantites can be found in Grammig and Küchlin (2018), Bansal and Yaron (2004) and in Appendix 4.A.2.

In our specification of the instrumental model, we follow the selection of moment conditions of Grammig and Küchlin (2018) to identify the structural parameters of interest

$\xi^P = (\delta, \gamma, \psi)^\top$. The moment restrictions used to estimate the instrumental parameters $\theta^P = \left(\alpha, \beta, \mu_{r_m^e}, \mu_{r_f}, \mu_{z_m}, \sigma_{r_m^e}, \sigma_{r_f}, \sigma_{z_m}\right)^\top$ are given by

$$\mathbb{E}\left[u_t^P\left(y_t^P, \theta^P\right)\right] = \mathbb{E}\begin{bmatrix} \zeta_{3,t} \\ \zeta_{3,t} \cdot r_{f,t} \\ r_{m,t}^e - \mu_{r_m^e} \\ r_t^f - \mu_{r_f} \\ z_{m,t} - \mu_{z_m} \\ (r_{m,t}^e)^2 - (\mu_{r_m^e})^2 - (\sigma_{r_m^e})^2 \\ (r_t^f)^2 - (\mu_{r_f})^2 - (\sigma_{r_f})^2 \\ (z_{m,t})^2 - (\mu_{z_m})^2 - (\sigma_{z_m})^2 \end{bmatrix}, \tag{4.4.5}$$

where $\zeta_{3,t} = z_{m,t} - \alpha - \beta r_t^f$ and $r_{m,t}^e = r_{m,t} - r_t^f$. The expected value of the risk-free rate, $\mathbb{E}(r_t^f) = \mu_{r_f}$, conveys information about the intertemporal elasticity of substitution, ψ, the subjective time preference rate, δ, and risk aversion, γ, that is contained in the precautionary savings term. The expected value of the market excess return, $\mathbb{E}(r_{m,t} - r_t^f) = \mu_{r_m^e}$, is mainly linked to risk aversion. We exploit the contemporaneous relationship between the log price-dividend ratio and the log risk-free rate to disentangle intertemporal substitution and risk aversion. This relation is predominantly affected by the IES. It is captured by a linear regression of $z_{m,t}$ on r_t^f and the associated regression coefficients, α and β. Whereas the expected value of the price-dividend ratio, $\mathbb{E}(z_{m,t}) = \mu_{z_m}$, depends on all preference parameters, its standard deviation given by $\sigma(z_{m,t}) = \sigma_{z_m}$ only depends on the risk aversion and time preference. It hence provides separate information about risk aversion and time preference rate. Lastly, the standard deviations of the risk-free rate and the market excess returns, σ_{r_f} and $\sigma_{r_m^e}$, are used as instrumental parameters. As for the technology parameters, the number of instrumental parameters is equal to the number of moment conditions, the optimization is independent of the choice of the weighting matrix, $\widehat{W}_T^P$.

4.5. Results

Using the model specifications described in the previous sections, we estimate the structural parameters of interest for the habit persistence and the long-run risk model. First, we estimate the parameters of the habit persistence model. Based on this example, we illustrate the effect calibrating a subset of model parameters has on the estimation uncertainty. Detailed estimation results are described in Section 4.5.1.

Second, we use the estimation results of the long-run risk model to show which impact a partial indirect inference strategy has in comparison to the indirect inference approach of

Grammig and Küchlin (2018). The key difference between the two approaches is whether we conceive the structural model to be a correctly specified model that describes economic reality appropriately or whether we believe that the model is misspecified. The latter materializes in different probability limits and thus estimated quantities for the simulated and empirical data in the asymptotic covariance of the instrumental models. Section 4.5.2 presents estimation results based on both estimation strategies and discusses the degree of misspecification encountered in the LRR model.

The empirical application of the habit persistence model is based on annual U.S. data from 1930 to 2008 and the long run risk model is estimated on quarterly U.S. data from 1947Q2 to 2014Q4. The former data set is based on data taken from Beeler and Campbell (2012) and the latter is obtained using similar calculations and is comparable to the data set in Grammig and Küchlin (2018). A detailed description of the data and summary statistics can be found in Appendix 4.E.

4.5.1. Habit persistence model

We estimate the structural parameters of interest, δ and γ, using the two-step partial indirect inference approach in three different settings. Column 1 in Tables 4.5.1 to 4.5.3 report the calibrated values found in Campbell and Cochrane (1999) and Wachter (2005). The preference parameters are calibrated as follows; the risk aversion coefficient in the utility function is set to $\gamma = 2$ and the implied steady state relative risk aversion coefficient is given by $\overline{rra} = \frac{\gamma}{\bar{S}} = 28.57$. The monthly time preference parameter is given by $\delta = 0.9913$. The persistence of surplus consumption has to be chosen close to unity but it should imply a stationary heteroskedastic AR(1) process. It is calibrated to $\phi = 0.9885$. Furthermore, the consumption growth moments are specified with a mean of $g = 0.0016$ and a standard deviation of $\sigma_v = 0.0077$. Campbell and Cochrane (1999) and Wachter (2005) show that the model matches the level of the risk-free rate well. It slightly overestimates the equity premium and underestimates its volatility. Moreover, the moments of consumption growth are well matched.

The first estimation setting is based on the estimation of all model parameters, the technology and preference parameters. This is an instance of full encompassing as described in Equation (4.3.1). The estimation is split in a two-step procedure based on the macro and asset pricing part of the model and makes use of the instrumental model for the technology parameters laid out in Equation (4.4.1) and the asset pricing-implied moment conditions given in Equation (4.4.2).

The structural parameter estimates obtained in the full encompassing case displayed in Table 4.5.1 are economically plausible. The point estimate of the risk aversion coefficient

Table 4.5.1: Estimation results for the habit persistence model in the full encompassing case

	Calibration	Estimate	Standard Error	Confidence Interval 2.5%	Confidence Interval 97.5%
			preference parameters		
γ	2	2.08	13609.99	-26673.00	26677.16
δ	0.9913	0.9998	12.6560	-23.8056	25.8052
$\overline{rra}$	28.57	9.02			
			technology parameters		
ϕ	0.9885	0.9977	12.1511	-22.8180	24.8134
g	0.0016	0.0016	0.0005	0.0006	0.0026
σ	0.0077	0.0077	0.0019	0.0039	0.0115

Notes: This table reports the estimation results obtained when estimating all model parameters (full encompassing case). The empirical application is based on annual observations from 1930 to 2008. The standard errors and confidence intervals are based on a consistent estimate of the asymptotic covariance matrix in Proposition 4.3.4.

is given by $\hat{\gamma} = 2.08$ and the time preference rate is estimated at $\hat{\delta} = 0.9998$. Moreover, the model-implied steady-state relative risk aversion amounts to $\widehat{\overline{rra}} = \frac{\hat{\gamma}}{\bar{S}} = 9.02$. This value lies within the range of plausible risk aversion values proposed by Mehra and Prescott (1985). Moreover, the estimated values for the habit persistence parameter and the moments of consumption growth are economically reasonable. Although the point estimates are plausible, the estimation uncertainty reflected in the standard errors and the confidence intervals is large for the preference parameters. Thus, the second step moment conditions cannot identify the parameters of interest. The problem arises due to the estimation of the persistence of habit parameter which cannot be pinned down well by empirical data. There does not exist a straightforward bridge between economic theory on habit evolution and empirical data.

To achieve identification of the parameters of interest, we introduce a second setting in which we calibrate the persistence of habit parameter to $\phi = 0.9885$. In opposition to the first approach, this strategy is embedded in the partial encompassing setting introduced in Equation (4.3.2). This two-step estimation approach is also based on the instrumental model for the macroeconomic parameters and the second set of moment conditions given in Equation (4.4.3).

Table 4.5.2: Estimation results for the habit persistence model in the partial encompassing case

	Calibration	Estimate	Standard Error	Confidence Interval 2.5%	97.5%
			preference parameters		
γ	2	0.74	2.39	-3.94	5.42
δ	0.9913	0.9974	0.0044	0.9887	1.0061
$\overline{rra}$	28.57	11.99			
			technology parameters		
g	0.0016	0.0016	0.0005	0.0006	0.0026
σ	0.0077	0.0077	0.0019	0.0039	0.0115

Notes: This table reports the estimation results obtained when the persistence of habit parameter ϕ is calibrated to 0.9885 and the other parameters are estimated using the two-step partial indirect inference method. The empirical application is based on annual observations from 1930 to 2008. The standard errors and confidence intervals are based on a consistent estimate of the asymptotic covariance matrix in Proposition 4.3.4.

Table 4.5.2 presents the estimation results of the second setup. The point estimate of the time preference rate is smaller than unity and given by $\hat{\delta} = 0.9974$, whereas the estimate of relative risk aversion is given by $\hat{\gamma} = 0.74$. The steady-state risk aversion value implied by the estimation setup is $\widehat{\overline{rra}} = 11.99$ and, thus, smaller than the value implied by the calibration exercise in Campbell and Cochrane (1999). Moreover, the estimation uncertainty of the structural parameters of interest is a lot smaller which hints at the identification of the parameters by the chosen moment conditions. The 95% confidence interval around γ has length 9.22 and the one around δ is tight, but overlaps unity.

As a consistency check, we compare this estimation strategy to the setting where only the structural parameters of interest are estimated, whereas all nuisance parameters are calibrated. This approach is based on the moment conditions given in Equation (4.4.3).

The estimation results are displayed in Table 4.5.3. They are quantitatively comparable those of Table 4.5.2 both in terms of the parameter estimates and the standard errors. This implies that the estimation of the technology parameters in the first step does not deteriorate the second-step estimation results. It is crucial for the identification of the structural parameters of interest that we fix the persistence of habit as we do not find a moment condition that identifies this parameter.

Table 4.5.3: Estimation results for the habit persistence model for preference parameters

	Calibration	Estimate	Standard Error	Confidence Interval 2.5%	97.5%
			preference parameters		
γ	2	0.74	2.75	-4.64	6.12
δ	0.9913	0.9974	0.0038	0.9900	1.0049
$\overline{rra}$	28.57	11.87			

Notes: This table reports the estimation results obtained when we restrict the estimation to the structural parameters of interest. Hence, the persistence of habit is calibrated to $\phi = 0.9885$ and the mean and standard deviation of consumption are given by $g = 0.0016$ and $\sigma_v = 0.0135$. The empirical application is based on annual observations from 1930 to 2008.

4.5.2. LRR model

We also implement our two-step partial indirect inference estimation approach for the long-run risk model. The aim is to estimate the structural parameters of interest, evaluate the associated estimation uncertainty and compare the two-step indirect inference strategy advocated by Grammig and Küchlin (2018) with our two-step partial indirect inference strategy.[8] The main difference between the two approaches is that we allow the model to be misspecified, whereas Grammig and Küchlin (2018) implicitly assume that the model is correctly specified.

To assess the economic plausibility of the structural parameters, we present the calibrated values of the LRR model in the first column of Tables 4.5.4 and 4.5.5, respectively. Bansal and Yaron (2004) calibrate the parameters in the LRR model without economic uncertainty as follows: the persistence of the long-run predictable growth component is calibrated to $\rho = 0.979$. To match the unconditional variance and the first order autocorrelation of annual consumption growth, they calibrate $\bar{\sigma} = 0.0078$ and $\varphi_e = 0.044$. Following Abel (1999), they set $\phi = 3$ to capture the leveraged nature of dividends. Moreover, $\varphi_d \bar{\sigma} = 0.0351$ is chosen to match the unconditional variance of dividend growth and its annual correlation with consumption growth. It implies a calibrated value of $\varphi_d = 4.5$. The expected values of consumption and dividend growth are set to $\mu_c = \mu_d = 0.0015$.

[8]To replicate the two-step indirect inference results, we use the code provided by Grammig and Küchlin (2018) on their webpage http://tinyurl.com/GS2017-Web-Appendix. In the implementation of the two-step partial indirect inference strategy, we adapt their code to allow for S ensembles rather than one long simulation and implement the inference for our estimator.

Table 4.5.4: Estimation results LRR model using two-step indirect inference

	Calibration	Estimate	Standard Error	Confidence Interval 2.5%	97.5%
			preference parameters		
γ	10	11.78	15.78	-19.15	42.70
δ	0.998	1.0000	0.0260	0.9489	1.0510
ψ	1.5	0.29	0.18	-0.06	0.64
			technology parameters		
μ_c	0.0015	0.0017	0.0002	0.0013	0.0020
μ_d	0.0015	0.0019	0.0009	0.0001	0.0036
ρ	0.979	0.9910	0.0096	0.9723	1.0098
φ_e	0.044	0.0643	0.0292	0.0070	0.1215
$\bar{\sigma}$	0.0078	0.0024	0.0005	0.0014	0.0034
ϕ	3	5.14	1.24	2.71	7.56
φ_d	4.5	3.06	0.80	1.50	4.63

Notes: This table replicates the estimation results obtained by Grammig and Küchlin (2018) when using their two-step indirect inference approach. The results are based on the assumption that the model is correctly specified. The results are based on U.S. data on quarterly frequency from 1947Q2 to 2014Q4.

The economic uncertainty implied in the full LRR model is knocked out which implies that the associated parameters are calibrated to be zero, i.e. $\nu = \sigma_\omega = 0$. This has implications for the estimation approach which is set in the partial encompassing world. The two economic uncertainty parameters are assumed to be nuisance parameters of type ξ_{22} in the notation of Dridi et al. (2007).

The two-step estimation strategy as detailed in Section 4.4.2 is the same as in Grammig and Küchlin (2018). The instrumental models for the technology parameters and the preference parameters given in Equations (4.4.4) and (4.4.5) are in line with the choices of Grammig and Küchlin (2018). Table 4.5.4 gives the estimation results of Grammig and Küchlin (2018), whereas Table 4.5.5 presents the results for our setting where the number of simulated ensembles is chosen to be $S = 1000$.

The point estimates of the structural parameters are qualitatively comparable for the two estimation settings. The most critical estimate is that of the intertemporal elasticity of substitution, ψ, which is smaller than unity. This is at odds with the calibrated value and the early resolution of risk that is advocated by Bansal and Yaron (2004), Bansal et al. (2007)

Table 4.5.5: Estimation results LRR model using two-step partial indirect inference

	Calibration	Estimate	Standard Error	Confidence Interval 2.5%	97.5%
			preference parameters		
γ	10	11.79	13.56	-14.79	38.38
δ	0.998	0.99997	0.0269	0.9472	1.0527
ψ	1.5	0.29	0.16	-0.03	0.61
			technology parameters		
μ_c	0.0015	0.0017	0.0002	0.0013	0.0020
μ_d	0.0015	0.0019	0.0002	0.0015	0.0022
ρ	0.979	0.9910	0.0185	0.9547	1.0273
φ_e	0.044	0.0643	0.0296	0.0062	0.1223
$\bar{\sigma}$	0.0078	0.0024	0.0003	0.0018	0.0031
ϕ	3	5.13	1.07	3.04	7.23
φ_d	4.5	3.06	0.57	1.95	4.18

Notes: This table reports the estimation results obtained for the two-step partial indirect inference method. All parameters of the model without economic uncertainty are estimated using this approach. The standard errors and confidence intervals are based on a consistent estimate of the asymptotic covariance matrix in Proposition 4.3.4. The empirical analysis is based on $S = 1000$ simulated ensembles and on U.S. data on quarterly frequency from 1947Q2 to 2014Q4.

and Bansal et al. (2012). A thorough discussion of the estimation results and the plausibility of the estimates can be found in Grammig and Küchlin (2018).

The main difference between the two-step indirect inference and the two-step partial indirect inference strategy is that the latter allows for misspecification of the structural model, whereas the two-step indirect inference approach assumes that the LRR model is correctly specified. This translates to different asymptotic covariance matrices. The main difference in Grammig and Küchlin (2018) compared to the quantities described in Section 4.3.1 is that Assumption 4.3.2 is understood as an accurate description of the simulator of the structural model that is believed to be correctly specified. Hence, the probability limits of all quantities associated with the empirical and simulated data is assumed to be the same, i.e. $P_* = P_0$. This implies that the asymptotic distribution of the instrumental models given in Proposition 4.3.3 simplifies remarkably as it holds that $C_0^M = C_0^{M*}$ and $C_0^P = C_0^{P*}$. The limiting distribution of the instrumental parameters in the correctly specified case derived in Grammig and Küchlin (2018) can be found in Appendix 4.C.3.

Note that, the two-step approach uses one long simulation of length $T = 100000$, whereas our strategy is based on S simulated ensembles of length $T = 271$. Our strategy depends on the number of simulated ensembles. By increasing the number of simulated ensembles to a sufficiently large number, the point estimates and standard errors stabilize to the results found in Table 4.5.5 which are independent of the realizations of the random variables in the simulations.

Concerning the estimation uncertainty, we find similar results for both estimation strategies. Whereas the standard errors for the risk aversion coefficient, γ, the elasticity of substitution, ψ, and the persistence of the predictable long-run risk component, ρ, are slightly smaller when accounting for potential model misspecification, those associated with the time preference rate, δ, and the variance scaling parameter of the long-run risk component, φ_e, increase marginally. Hence, accounting for potential model-misspecification has a negligible effect on the magnitude of estimation uncertainty. The estimated quantities needed for the construction of the asymptotic covariance of the instrumental model are almost identical for the empirical and simulated data. Hence, the degree of model misspecification is rather small for the LRR model.

4.6. Conclusion

In this paper, we introduce a two-step partial indirect inference strategy to estimate the parameters of partial equilibrium asset pricing models. Partial indirect inference introduced by Dridi et al. (2007) allows for the estimation of misspecified models. It enables the consistent estimation of the structural parameters of interest that describe parts of the model that capture economic reality, while acknowledging that other parts of the model are misspecified. It gives a rationale that some model parameters that index the misspecified part of the model may be fixed or calibrated without compromising the consistent estimation of the parameters of interest.

The two-step nature of the estimation approach builds on the separability of the technology description of the model and the asset pricing part. We follow the arguments in Grammig and Küchlin (2018) and apply the two-step approach to other partial equilibrium asset pricing models. Our estimation approach combines the two notions.

We estimate the parameters of the habit persistence model introduced by Campbell and Cochrane (1999). We find that estimating all model parameters using our two-step strategy based on separate instrumental models for the macro and preference parameters leads to unidentified parameter estimates. For one of the parameters, the one that captures the persistence of habit, we cannot find appropriate moment conditions that identify the parameter. There is no evident bridge between the evolution of latent habit and the imposed

structure in the habit persistence model, and thus, the persistence parameter only represents a caricature of reality. When we calibrate this parameter, we find point estimates of the parameters of interest that are economically plausible. Additionally, the estimation uncertainty implied by the estimated asymptotic covariance is moderately small. Calibrating all nuisance parameters of the model yields comparable point estimates and standard errors.

Using the parameter estimation of the long-run risk model proposed by Bansal and Yaron (2004), we illustrate the impact of allowing for model misspecification on the estimation uncertainty and contrast our estimation results to those of Grammig and Küchlin (2018). This comparison between the estimated parameters obtained by either indirect inference or partial indirect inference and the associated estimation uncertainty allows for an assessment of the degree of misspecification of the model. We find standard errors of similar magnitude which implies that the simulated time series are comparable to the empirical data.

The proposed two-step partial indirect inference approach allows to estimate and evaluate the partial equilibrium asset pricing models based on the sequential model structure and the modeling focus, i.e. the part of the model that we believe to reflect economic truth. It can be used to estimate any model with a sequential structure. Moreover, it sheds light on the relationship between misspecified models, calibration and the consistent estimation of the structural parameters of interest.

References

Abel, A. B. (1999). "Risk premia and term premia in general equilibrium". *Journal of Monetary Economics* 43. (1), 3–33.

Aldrich, E. M. and A. R. Gallant (2011). "Habit, long-run risks, prospect? A statistical inquiry". *Journal of Financial Econometrics* 9, 589–618.

Bansal, R., A. R. Gallant, and G. Tauchen (2007). "Rational pessimism, rational exuberance, and asset pricing models". *The Review of Economic Studies* 74, 1005–1033.

Bansal, R., D. Kiku, and A. Yaron (2012). "An empirical evaluation of the long-run risks model for asset prices". *Critical Finance Review* 1, 183–221.

Bansal, R. and A. Yaron (2004). "Risks for the long run: A potential resolution of asset pricing puzzles". *The Journal of Finance* 59, 1481–1509.

Beeler, J. and J. Y. Campbell (2012). "The long-run risks model and aggregate asset prices: An empirical assessment". *Critical Finance Review* 1, 141–182.

Calvet, L. E. and V. Czellar (2015). "Through the looking glass: Indirect inference via simple equilibria". *Journal of Econometrics* 185. (2), 343–358.

Campbell, J. Y. (2003). "Consumption-based asset pricing". *Handbook of the Economics of Finance* 1, 803–887.

Campbell, J. Y. and J. H. Cochrane (1999). "By force of habit: A consumption-based explanation of aggregate stock market behavior". *The Journal of Political Economy* 107, 205–251.

Campbell, J. Y. and R. J. Shiller (1988). "The dividend-price ratio and expectations of future dividends and discount factors". *The Review of Financial Studies* 1. (3), 195–228.

Cochrane, J. H. (2009). *Asset Pricing*. Princeton University Press.

Constantinides, G. M. and A. Ghosh (2011). "Asset pricing tests with long-run risks in consumption growth". *The Review of Asset Pricing Studies* 1. (1), 96–136.

Corsi, F. (2009). "A simple approximate long-memory model of realized volatility". *Journal of Financial Econometrics* 7. (2), 174–196.

Dridi, R., A. Guay, and E. Renault (2007). "Indirect inference and calibration of dynamic stochastic general equilibrium models". *Journal of Econometrics* 136. (2), 397–430.

Dridi, R. and E. Renault (2000). "Semi-parametric indirect inference". *LSE STICERD Research Paper No. EM392*.

Epstein, L. G. and S. E. Zin (1989). "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework". *Econometrica* 57. (4), 937–969.

Fernández-Villaverde, J. (2010). "The econometrics of DSGE models". *SERIEs* 1. (1-2), 3–49.

Gallant, A. R. and R. E. McCulloch (2009). "On the determination of general scientific models with application to asset pricing". *Journal of the American Statistical Association* 104. (485), 117–131.

Gallant, A. R. and G. Tauchen (1996). "Which moments to match?" *Econometric Theory* 12, 657–681.

Gourieroux, C., A. Monfort, and E. Renault (1993). "Indirect inference". *Journal of Applied Econometrics* 8, 85–118.

Grammig, J. and E.-M. Küchlin (2018). "A two-step indirect inference approach to estimate the long-run risk asset pricing model". *Journal of Econometrics* 205. (1), 6–33.

Hansen, L. P. (1982). "Large sample properties of generalized method of moments estimators". *Econometrica* 50. (4), 1029–1054.

Hasseltoft, H. (2012). "Stocks, bonds, and long-run consumption risks". *Journal of Financial and Quantitative Analysis* 47. (2), 309–332.

Lucas, R. E. (1978). "Asset prices in an exchange economy". *Econometrica* 46. (6), 1429–1445.

Mehra, R. and E. C. Prescott (1985). "The equity premium: A puzzle". *Journal of Monetary Economics* 15, 145–161.

Newey, W. K. and K. D. West (1987). "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix". *Econometrica* 55. (3), 703–708.

Rubinstein, M. (1976). "The valuation of uncertain income streams and the pricing of options". *The Bell Journal of Economics* 7. (2), 407–425.

Schorfheide, F., D. Song, and A. Yaron (2018). "Identifying long-run risks: A Bayesian mixed-frequency approach". *Econometrica* 86. (2), 617–654.

Singleton, K. J. (2009). *Empirical dynamic asset pricing: Model specification and econometric assessment*. Princeton University Press.

Smith, A. A. (1993). "Estimating nonlinear time-series models using simulated vector autoregressions". *Journal of Applied Econometrics* 8, 63–84.

Wachter, J. A. (2005). "Solving models with external habit". *Finance Research Letters* 2, 210–226.

Weil, P. (1989). "The equity premium puzzle and the risk-free rate puzzle". *Journal of Monetary Economics* 24, 401–421.
— (1990). "Nonexpected utility in macroeconomics". *The Quarterly Journal of Economics* 105. (1), 29–42.

Appendix 4.A Model solution details of the partial equilibrium asset pricing models

4.A.1 Details on the habit persistence model

In this section, we present the asset pricing part of the habit persistence model as described in Campbell and Cochrane (1999) in more detail.

The specification of preferences and technology in Section 4.2.1 leads to an intertemporal marginal rate of substitution or stochastic discount factor M_{t+1}

$$M_{t+1} = \delta \left(\frac{S_{t+1}}{S_t} \frac{C_{t+1}}{C_t} \right)^{-\gamma}$$

and a constant real risk-free rate $r^f_{t+1} = \log R^f_{t+1}$

$$r^f_{t+1} = -\log \delta + \gamma g + \gamma(1-\phi)(\bar{s} - s_t) - \frac{\gamma^2 \sigma_v^2}{2} \left[1 + \lambda(s_t)\right]^2 .$$

We refer to $\gamma(1-\phi)(\bar{s} - s_t)$ as intertemporal substitution. The basic intuition of this term is that when the surplus consumption ratio is low, marginal utility of consumption is high. Thus, consumers borrow to finance their consumption, driving up the risk-free rate. The term $\frac{\gamma^2 \sigma_v^2}{2} \left[1 + \lambda(s_t)\right]^2$ is called precautionary savings. The logic is that with increasing uncertainty, consumers are more willing to save, hence, the risk-free rate decreases. To assure that the risk-free interest rate is constant, the serial correlation parameter in the habit process has to be close to unity but never unity, i.e. $\phi \approx 1$ and $\phi < 1$, which implies only a small degree of intertemporal substitution. Furthermore, the sensitivity function $\lambda(s_t)$ is modeled as a function that is declining in s_t. The consequence of these two adjustments of the model necessary to assure a constant risk-free rate is that when uncertainty is high or when S_t is low, precautionary savings offset intertemporal substitution.

Next, we turn to the specification of the functional form of the sensitivity function $\lambda(s_t)$. The requirements that this functional form should fulfill are that the sensitivity function is declining in s_t, that it implies a constant real risk-free rate, that habit moves non-negatively with consumption, and that habit is predetermined at the steady state $s_t = \bar{s}$. The last

condition can be implemented by linking the steady state surplus consumption ratio to the model parameters, i.e. $\bar{S} = \sigma_v \sqrt{\frac{\gamma}{1-\phi}}$. Furthermore, the sensitivity function is defined as

$$\lambda(s_t) = \begin{cases} \frac{1}{\bar{S}}\sqrt{1 - 2(s_t - \bar{s})} - 1 & \text{for} \quad s_t \leq s_{max} \\ 0 & \text{for} \quad s_t > s_{max}, \end{cases}$$

where $\bar{s} = \log \bar{S}$ and $s_{max} = \bar{s} + \frac{1}{2}\left(1 - \bar{S}^2\right)$. This specification of the sensitivity function and the dynamics of the surplus consumption ratio, in combination with lognormal consumption growth lead to a constant real risk-free rate

$$r^f_{t+1} = -\log \delta + \gamma g - \left(\frac{\gamma}{\bar{S}}\right)^2 \frac{\sigma_v^2}{2} = -\log \delta + \gamma g - \frac{\lambda}{2}(1 - \phi). \tag{4.A.1}$$

The maximum Sharpe ratio of the habit persistence model is given by

$$\max_{\{j\}} \frac{\mathbb{E}_t\left[R^e_{t+1}\right]}{\sigma_t\left(R^e_{t+1}\right)} \approx \gamma\sigma_v[1 + \lambda(s_t)]. \tag{4.A.2}$$

This formualtion of the habit persistence model is proposed by Wachter (2005). She omits to specify a dividend growth process but models a stock price solely as the price of the claim to a consumption stream. This formulation of the model is useful because it does not incorporate dividend dynamics which are more noisy than consumption dynamics. Furthermore, it restricts the number of parameters to two preference parameters, δ, the subjective discount factor and, γ, the risk aversion parameter of a representative agent, two technology parameters associated with the consumption growth dynamics, g, the mean and, σ_v, the standard deviation of log consumption growth, and, ϕ, the habit persistence parameter.

In addition to the problem that the dividend growth process poses, the correlation between consumption and dividend growth is not observable. Thus, omitting dividend growth altogether facilitates to determine the structural parameters without altering the general structure of the habit persistence model.

Gallant and McCulloch (2009) and Aldrich and Gallant (2011) use yet another specification of the habit persistence model. They stick with the framework described in Campbell and Cochrane (1999), however, they model a stock price only as the price of a claim to a dividend stream. Hence, they have seven structural parameters to determine: the five parameters as in Wachter (2005), the standard deviation of log dividend growth σ_w and the correlation between consumption and dividend growth ρ.

Stochastic discount factor, risk-free rate and mean-variance frontier

The habit persistence model developed by Campbell and Cochrane (1999) is a consumption based asset pricing model in which the standard power utility is augmented by the habit level of consumption. Thus, the stochastic discount factor used in the Euler equation is altered. In this section, we exploit the properties of the stochastic discount factor implied by the habit persistence model in detail.[9]

As described in Section 4.2.1, the intertemporal marginal rate of substitution or stochastic discount factor M_{t+1} can be expressed as

$$M_{t+1} = \delta \left(\frac{C_{t+1}}{C_t} \frac{S_{t+1}}{S_t} \right)^{-\gamma},$$

where δ is the subjective discount factor, γ is a risk aversion parameter of a representative agent. Furthermore, $S_t = \frac{C_t - X_t}{C_t}$ is the surplus consumption ratio, where X_t is the unobserved external habit level.

Given the dynamics of (log) consumption and (log) surplus consumption ratio growth, Δc_{t+1} and Δs_{t+1}, specified in Section 4.2.1

$$\begin{aligned} \Delta c_{t+1} &= g + v_{t+1} \qquad \text{with} \quad v_{t+1} \sim \mathcal{N}(0, \sigma_v^2) \quad \text{i.i.d.}, \\ \Delta s_{t+1} &= (1 - \phi)(\bar{s} - s_t) + \lambda(s_t) v_{t+1}, \end{aligned}$$

we can write the stochastic discount factor as

$$M_{t+1} = \delta e^{-\gamma[g + (1-\phi)(\bar{s} - s_t)] - \gamma[1 + \lambda(s_t)]v_{t+1}}.$$

Real risk-free interest rate

We show that the risk-free rate $R_{t+1}^f = (\mathbb{E}_t[M_{t+1}])^{-1}$ with log normal consumption growth $\frac{C_{t+1}}{C_t}$ can be written as

$$r_t^f = -\log \delta + \gamma g + \gamma(1 - \phi)(\bar{s} - s_t) - \frac{\gamma^2 \sigma_v^2}{2}[1 + \lambda(s_t)]^2.$$

From the specification of log consumption and log surplus consumption ratio growth which are both normally distributed, we can infer that consumption growth and surplus consumption ratio growth, $\frac{C_{t+1}}{C_t}$ and $\frac{S_{t+1}}{S_t}$, are lognormally distributed. For the expected value

[9] Part of the derivations found here are mentioned in Campbell and Cochrane (1999) and their Appendix. However, our presentation of the results is more detailed.

of a lognormally distributed random variable $Y = e^Z$ it holds that when $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Y \sim \text{LN}(\mu, \sigma^2)$ and its expected value is equal to $\mathbb{E}(Y) = e^{\mu + \frac{1}{2}\sigma^2}$.

In order to assess the risk-free rate, we are interested in the expected value of the stochastic discount factor which incorporates both growth rates

$$R_{t+1}^f = \frac{1}{\mathbb{E}_t[M_{t+1}]} = \frac{1}{\delta \mathbb{E}_t \left[\left(\frac{C_{t+1}}{C_t} \frac{S_{t+1}}{S_t} \right)^{-\gamma} \right]} = \frac{1}{\delta \mathbb{E}_t \left[e^{-\gamma(\Delta c_{t+1} + \Delta s_{t+1})} \right]}.$$

As both growth rates are lognormally distributed, and depend on the consumption innovation v_{t+1}, the expected value can be written as

$$\mathbb{E}_t \left[e^{-\gamma(\Delta c_{t+1} + \Delta s_{t+1})} \right] = e^{-\gamma \mathbb{E}_t[\Delta c_{t+1} + \Delta s_{t+1}] + \frac{1}{2}\gamma^2 \sigma_t^2 (\Delta c_{t+1} + \Delta s_{t+1})},$$

with $\mathbb{E}_t[\Delta c_{t+1} + \Delta s_{t+1}] = g + (1-\phi)(\bar{s} - s_t)$ and $\sigma_t^2(\Delta c_{t+1} + \Delta s_{t+1}) = [1 + \lambda(s_t)]^2 \sigma_v^2$. Plugging this into the equation above and rearranging the terms, we obtain

$$R_{t+1}^f = \frac{1}{\delta e^{-\gamma g - \gamma(1-\phi)(\bar{s} - s_t) + \frac{1}{2}\gamma^2 \sigma_v^2 [\lambda(s_t)^2 + 1]}}.$$

Finally, by taking logs, we can show that the real risk-free rate is constant and equal to the term above

$$r_t^f = -\log \delta + \gamma g + \gamma(1-\phi)(\bar{s} - s_t) - \frac{\gamma^2 \sigma_v^2}{2}[1 + \lambda(s_t)]^2.$$

Slope of the mean-standard deviation frontier or the maximum Sharpe ratio

In addition to the expected value of the stochastic discount factor which we implicitly explore when calculating the risk-free interest rate, we can link the stochastic discount factor M_{t+1} to excess returns.

The basic asset pricing or Euler equation for excess returns reads

$$\begin{aligned} \mathbb{E}_t \left[M_{t+1} R_{t+1}^e \right] &= \text{Cov}_t \left(M_{t+1}, R_{t+1}^e \right) + \mathbb{E}_t \left[M_{t+1} \right] \mathbb{E}_t \left[R_{t+1}^e \right] \\ &= \rho_t \left(M_{t+1}, R_{t+1}^e \right) \sigma_t \left(M_{t+1} \right) \sigma_t \left(R_{t+1}^e \right) + \mathbb{E}_t \left[M_{t+1} \right] \mathbb{E}_t \left[R_{t+1}^e \right] = 0. \end{aligned}$$

The Euler equation can be rearranged to form the Sharpe ratio of excess returns, i.e. the ratio of expected excess return and the standard deviation of excess return given by

$$\frac{\mathbb{E}_t \left[R_{t+1}^e \right]}{\sigma_t \left(R_{t+1}^e \right)} = -\rho_t \left(M_{t+1}, R_{t+1}^e \right) \frac{\sigma_t \left(M_{t+1} \right)}{\mathbb{E}_t \left[M_{t+1} \right]} \leq \frac{\sigma_t \left(M_{t+1} \right)}{\mathbb{E}_t \left[M_{t+1} \right]},$$

which is a function of the moments of the stochastic discount factor. As the correlation coefficient $\rho_t(\cdot)$ can only take on values between -1 and 1, the inequality holds and the maximum Sharpe ratio is equal to the last term.

In the habit persistence model, the stochastic discount factor is conditionally lognormally distributed. We can show (in sequential steps) that the maximum Sharpe ratio can be approximated by a function of model parameters, i.e.

$$\max_{\{j\}} \frac{\mathbb{E}_t\left[R^e_{t+1}\right]}{\sigma_t\left(R^e_{t+1}\right)} = \sqrt{e^{\gamma^2\sigma_v^2[1+\lambda(s_t)]^2} - 1} \approx \gamma\sigma_v[1+\lambda(s_t)].$$

We start by showing that the first equality holds. For this purpose, we calculate the first and second moment of the stochastic discount factor. The expected value is equal to

$$\mathbb{E}_t\left[M_{t+1}\right] = \delta e^{-\gamma(g+(1-\phi)(\bar{s}-s_t))+\frac{1}{2}\gamma^2\sigma^2[1+\lambda(s_t)]^2}.$$

The variance of the stochastic discount factor can be written as

$$\begin{aligned}
\mathrm{Var}_t(M_{t+1}) &= \mathbb{E}_t\left[M^2_{t+1}\right] - (\mathbb{E}_t[M_{t+1}])^2 \\
&= \delta^2 e^{-2\gamma[g+(1-\phi)(\bar{s}-s_t)]+2\gamma^2\sigma_v^2[1+\lambda(s_t)]^2} - \delta^2 e^{-2\gamma[g+(1-\phi)(\bar{s}-s_t)]+\gamma^2\sigma_v^2[1+\lambda(s_t)]^2} \\
&= \delta^2 e^{-2\gamma[g+(1-\phi)(\bar{s}-s_t)]+\gamma^2\sigma_v^2[1+\lambda(s_t)]^2}\left(e^{\gamma^2\sigma_v^2[1+\lambda(s_t)]^2} - 1\right) \\
&= (\mathbb{E}_t[M_{t+1}])^2\left(e^{\gamma^2\sigma_v^2[1+\lambda(s_t)]^2} - 1\right).
\end{aligned}$$

Rearranging this term, we get $\frac{\mathrm{Var}_t(M_{t+1})}{(\mathbb{E}_t[M_{t+1}])^2} = \left(e^{\gamma^2\sigma_v^2[1+\lambda(s_t)]^2} - 1\right)$. When taking the square root, we obtain the maximum Sharpe ratio implied by the habit persistence model, i.e.

$$\max_{\{j\}} \frac{\mathbb{E}_t\left[R^e_{t+1}\right]}{\sigma_t\left(R^e_{t+1}\right)} = \sqrt{e^{\gamma^2\sigma_v^2[1+\lambda(s_t)]^2} - 1}.$$

Finally, we use a first order Taylor approximation[10] of $e^{\gamma^2\sigma_v^2[1+\lambda(s_t)]^2}$ at $\sigma^2_{v,0} = 0$ and find that $f(\sigma_v^2) = e^{\gamma^2\sigma_v^2[1+\lambda(s_t)]^2} \approx 1 + \gamma^2\sigma_v^2[1+\lambda(s_t)]^2$. Plugging this into the equation above, we

[10] $f(\sigma_v^2) \approx f(\sigma^2_{v,0}) + f'(\sigma^2_{v,0})(\sigma_v^2 - \sigma^2_{v,0}) + \ldots$ where $\sigma^2_{v,0} = 0$. We obtain $f(\sigma^2_{v,0}) = 1$, $f'(\sigma_v^2) = e^{\gamma^2\sigma_v^2[1+\lambda(s_t)]^2}\gamma^2[1+\lambda(s_t)]^2$ and $f'(\sigma^2_{v,0}) = \gamma^2[1+\lambda(s_t)]^2$.

can show that the maximum Sharpe ratio implied by the habit persistence model is indeed given by the following equation

$$\max_{\{j\}} \frac{\mathbb{E}_t\left[R^e_{t+1}\right]}{\sigma_t\left(R^e_{t+1}\right)} = \frac{\sigma_t(M_{t+1}}{\mathbb{E}_t[M_{t+1}]} = \sqrt{e^{\gamma^2\sigma_v^2[1+\lambda(s_t)]^2} - 1} \approx \gamma\sigma_v[1+\lambda(s_t)].$$

Solution mechanisms of the Euler equation

In the previous sections, we present the habit persistence model and derive Euler equations for different assets. This section elaborates on the series method described in Wachter (2005) to solve the nonlinear Euler equations numerically. The Euler equation for the price-dividend ratio is solved over a grid of the state variable s_t using a Gauss-Legendre quadrature approach based on 40 points to numerically integrate the expected value.

The basic asset pricing equation for the price-dividend ratio is set up in the formulations provided by Campbell and Cochrane (1999) and Wachter (2005) as follows

$$\frac{P_t}{C_t}(s_t) = \mathbb{E}_t\left[M_{t+1}\left(\frac{P_{t+1}}{C_{t+1}}(s_{t+1}) + 1\right)\frac{C_{t+1}}{C_t}\right]. \tag{4.A.3}$$

The aim of this analysis is to solve for the price-dividend ratio of the consumption claim for each state s_t.

Series method

The series method described in Wachter (2005) does not solve the Euler equation recursively. In contrast to the fixed point method, the basic asset pricing equation (4.A.3) is iterated N times forward

$$G^N(s_t) = \sum_{n=1}^{N}\mathbb{E}_t\left[\left(\prod_{j=1}^{n}M_{t+j}\right)\frac{C_{t+n}}{C_t}\right] + \mathbb{E}_t\left[\left(\prod_{j=1}^{N}M_{t+j}\right)\frac{C_{t+N}}{C_t}G^0(s_{t+N})\right],$$

where we choose the initial price-dividend ratio $G^0(s_{t+N})$ such that $\lim_{N\to\infty}\mathbb{E}_t\left[\left(\prod_{j=1}^{N}M_{t+j}\right)\frac{C_{t+N}}{C_t}G^0(s_{t+N})\right] = 0$. Thus, the price-dividend ratio of a consumption claim is an infinite sum of expectations

$$G(s_t) = \lim_{N\to\infty}G^N(s_t) = \sum_{n=1}^{\infty}\mathbb{E}_t\left[\left(\prod_{j=1}^{n}M_{t+j}\right)\frac{C_{t+n}}{C_t}\right] = \sum_{n=1}^{\infty}F_n(s_t).$$

Hence, we conceive the series method as an alternative solution for the price dividend ratio that consists of the approximation of an infinite sum.

In order to do so, we solve the Euler equation, $\mathbb{E}_t\left[M_{t+1}R_{t+1}\right] = 1$, recursively. We can rewrite $F_n(s_t)$, the price of a zero coupon equity maturing in n periods using its one period return, i.e. $R_{n,t+1} = \frac{F_{n-1}(s_{t+1})C_{t+1}}{F_n(s_t)C_t}$, and the Euler equation as

$$F_n(s_t) = \mathbb{E}_t\left[M_{t+1}\frac{C_{t+1}}{C_t}F_{n-1}(s_{t+1})\right]. \tag{4.A.4}$$

The series method consists of iterating on Equation (4.A.4) and summing up the terms to solve for the price dividend ratio of a consumption claim. The price-dividend ratio has converged when the price of the zero coupon maturing n periods in the future is smaller then 0.0001 in any state for large n. Hence, the approximation to the infinite sum is adequate enough.

Details on the Euler equation for consumption claims and the risk-free rate

We elaborate in more details on the integrals resulting from the Euler equation that need to be solved numerically. We start by analyzing the stochastic discount factor M_{t+1} and its expected value given the dynamics of (log) consumption growth and the (log) surplus consumption ratio.

Log consumption growth follows a Gaussian white noise process

$$\Delta c_{t+1} = c_{t+1} - c_t = g + v_{t+1} \qquad \text{or} \qquad \frac{C_{t+1}}{C_t} = e^{\Delta c_{t+1}} = e^{g+v_{t+1}},$$

where $v_{t+1} \sim \mathcal{N}(0,\sigma_v)$ i.i.d. which follows the probability distribution function of a normal distribution, i.e. $p(v) = \frac{1}{\sqrt{2\pi}\sigma_v}\exp(-\frac{1}{2}\frac{v^2}{\sigma_v^2})$. And the (log) surplus consumption ratio $s_t = \log\frac{C_t - X_t}{C_t}$ is a heteroskedastic AR(1)

$$s_{t+1} = (1-\phi)\bar{s} + \phi s_t + \lambda(s_t)v_{t+1} \quad \text{or} \quad \Delta s_{t+1} = s_{t+1} - s_t = (1-\phi)(\bar{s} - s_t) + \lambda(s_t)v_{t+1}.$$

Thus, the intertemporal marginal rate of substitution or the stochastic discount factor M_{t+1} can be written as

$$M_{t+1} = \delta\left(\frac{C_{t+1}}{C_t}\frac{S_{t+1}}{S_t}\right)^{-\gamma} = \delta e^{-\gamma g - \gamma[[1-\phi)(\bar{s}-s_t)+[\lambda(s_t)+1]v_{t+1}]}.$$

The real risk-free rate can also be obtained from the Euler equation as $R^f_{t+1} = [\mathbb{E}_t[M_{t+1}]]^{-1}$ or $e^{-r^f_{t+1}} = \mathbb{E}_t[M_{t+1}]$. Hence, we are interested in the expected value of the stochastic discount factor, which can be obtained by numerical integration on

$$\begin{aligned}\mathbb{E}_t[M_{t+1}] &= \delta e^{-\gamma[g+(1+\phi)(\bar{s}-s_t)]}\mathbb{E}_t\left[e^{-\gamma[\lambda(s_t)+1]v_{t+1}}\right] \\ &= \delta e^{-\gamma[g+(1+\phi)(\bar{s}-s_t)]}\int_{-\infty}^{\infty} p(v)e^{-\gamma[\lambda(s_t)+1]v}\,dv.\end{aligned}$$

When using the series method, we iterate on

$$F_n(s_t) = \mathbb{E}_t\left[M_{t+1}\frac{C_{t+1}}{C_t}F_{n-1}(s_{t+1})\right].$$

and sum up the resulting terms to obtain the price-dividend ratio $G(s_t) = \sum_{n=1}^{\infty} F_n(s_t)$ where $F_0(s_t) = 1$. For the price-dividend ratio of a consumption claim Equation 4.A.4 reads

$$\begin{aligned}F_n(s_t) = \delta e^{(1-\gamma)g-\gamma(1-\phi)(\bar{s}-s_t)} \\ \times\int_{-\infty}^{\infty} p(v)e^{(1-\gamma)v-\gamma\lambda(s_t)v}F_{n-1}\left((1-\phi)\bar{s}+\phi s_t+\lambda(s_t)v\right)\,dv.\end{aligned}$$

4.A.2 Model solution of the LRR model and model implications

In this section, we elaborate on the LRR model solution results. This section is a summary of the exposition in Grammig and Küchlin (2018), as well as that in Bansal and Yaron (2004) and Bansal et al. (2012). It is provided for completeness rather than because it offers new insights. The Euler equation given in Equation (4.2.13) is solved using linear approximations as suggested by Campbell and Shiller (1988). Bansal and Yaron (2004) adapt this to the log return of the aggregate wealth portfolio, r_a, and the log return of the market portfolio which is given by a claim to the dividend stream, r_m given by

$$r_{a,t+1} = c_0 + c_1 z_{t+1} - z_t + g_{t+1}, \tag{4.A.5}$$

$$r_{m,t+1} = c_{0,m} + c_{1,m} z_{m,t+1} - z_{m,t} + g_{d,t+1}, \tag{4.A.6}$$

where z is the log price-consumption ratio and z_m denotes the log price-dividend ratio. The parameters are given by $c_1 = \frac{\exp(\bar{z})}{1+\exp(\bar{z})}$, $c_{1,m} = \frac{\exp(\bar{z}_m)}{1+\exp(\bar{z}_m)}$, $c_0 = \ln(1+\exp(\bar{z})) - c_1\bar{z}$, and $c_{0,m} = \ln(1+\exp(\bar{z}_m)) - c_{1,m}\bar{z}_m$, where $\bar{z}$ and $\bar{z}_m$ are time series averages.

For the solution of the LRR model, the price-consumption and price-dividend ratios are expressed as linear functions of the latent state variables, i.e.

$$\begin{aligned} z_t &= A_0 + A_1 x_t + A_2 \sigma_t^2, \\ z_{m,t} &= A_{0,m} + A_{1,m} x_t + A_{2,m} \sigma_t^2. \end{aligned}$$

The A-coefficients[11] depend on the structural model parameters $\xi^M = (\mu_c, \mu_d, \rho, \varphi_e, \bar{\sigma}, \phi, \varphi_d, \nu, \sigma_\omega)^\top$ and $\xi^P = (\delta, \gamma, \psi)^\top$, given by

$$\begin{aligned} A_0 &= \frac{1}{1-c_1}\left[\ln\delta + \left(1-\frac{1}{\psi}\right)\mu_c + c_0 + c_1 A_2 \bar{\sigma}^2(1-\nu) + \frac{\theta}{2}(c_1 A_2 \sigma_\omega)^2\right], \\ A_1 &= \frac{1-\frac{1}{\psi}}{1-c_1\rho}, \qquad A_2 = \frac{1}{2}\frac{\left(\theta-\frac{\theta}{\psi}\right)^2 + (\theta A_1 c_1 \varphi_e)^2}{\theta(1-c_1\nu)}, \\ A_{0,m} &= \frac{1}{1-c_{1,m}}\left[\theta\ln\delta - \frac{\theta}{\psi}\mu_c + (\theta-1)\left[c_0 + c_1 A_0 + c_1 A_2(1-\nu)\bar{\sigma}^2 - A_0 + \mu_c\right]\right. \\ &\quad \left. + c_{0,m} + c_{1,m} A_{2,m}\bar{\sigma}^2(1-\nu) + \mu_d + \frac{1}{2}[(\theta-1)c_1 A_2 + c_{1,m}A_{2,m}]^2\sigma_\omega^2\right], \\ A_{1,m} &= \frac{\phi-\frac{1}{\psi}}{1-c_{1,m}\rho}, \qquad A_{2,m} = \frac{(1-\theta)(1-c_1\nu)A_2}{(1-c_{1,m}\nu)} \\ &\quad + \frac{0.5\left[(-\frac{\theta}{\psi}+\theta-1)^2 + ((c_{1,m}A_{1,m}\varphi_e) - ((1-\theta)c_1 A_1 \varphi_e))^2 + \varphi_d^2\right]}{1-c_{1,m}\nu}. \end{aligned}$$

The model-implied log risk-free rate is given by

$$r_{f,t} = -\theta\ln\delta + \frac{\theta}{\psi}(\mu_c + x_t) + (1-\theta)\mathbb{E}_t(r_{a,t+1}) - \frac{1}{2}\mathrm{Var}_t(m_{t+1}), \tag{4.A.7}$$

where m_{t+1} is the logarithm of the stochastic discount factor M_{t+1} and given by

$$\begin{aligned} m_{t+1} &= \theta\ln\delta - \frac{\theta}{\psi}g_{t+1} + (\theta-1)r_{a,t+1}, \\ \mathbb{E}_t(r_{a,t+1}) &= c_0 + c_1[A_0 + A_1\rho x_t + A_2(\bar{\sigma}^2 + \nu(\sigma_t^2 - \bar{\sigma}^2))] \\ &\quad + A_0 - A_1 x_t - A_2\sigma_t^2 + \mu_c + x_t, \\ \mathrm{Var}_t(m_{t+1}) &= \left(\frac{\theta}{\psi} + 1 - \theta\right)^2\sigma_t^2 + [(1-\theta)c_1 A_1 \varphi_e]^2\sigma_t^2 + [(1-\theta)c_1 A_2]\sigma_\omega^2. \end{aligned}$$

[11] Detailed derivations can be found in Grammig and Küchlin (2018).

Moreover, the equity premium can be expressed as

$$\mathbb{E}_t(r_{m,t+1} - r_{f,t}) = \beta_{m,e}\lambda_{m,e}\sigma_t^2 + \beta_{m,\omega}\lambda_{m,\omega}\sigma_\omega^2 - \frac{1}{2}\mathrm{Var}_t(r_{m,t+1}), \tag{4.A.8}$$

where $\lambda_{m,e} = (1-\theta)c_1 A_1 \varphi_e$, $\beta_{m,e} = c_{1,m}\left(\phi - \frac{1}{\phi}\right)\frac{\varphi_e}{1-c_{1,m}\rho}$, $\lambda_{m,\omega} = (1-\theta)A_2 c_1$, $\beta_{m,\omega} = c_{1,m}A_{2,m}$, and $\mathrm{Var}_t(r_{m,t+1}) = \beta_{m,e}^2\sigma_t^2 + \varphi_d^2\sigma_t^2 + \beta_{m,\omega}^2\sigma_\omega^2$. $\lambda_{m,e}$ gives the market compensation for the long-run predictable component and $\lambda_{m,\omega}$ is the market compensation for the stochastic consumption volatility risk, whereas $\beta_{m,e}$ is the exposure of the market return to the expected growth rate news and $\beta_{m,\omega}$ is the exposure to the economic uncertainty shocks.

Appendix 4.B Regularity conditions for the two-step partial indirect inference estimator

This section elaborates on the GMM regularity conditions imposed for the consistency of the instrumental parameter estimates. We impose a set of assumptions for the first and second step and for the observed and simulated data, respectively. For the four cases, there is an identification assumption and an equicontinuity assumption for the moment conditions.

4.B.1 First estimation step

We base the macroeconomic instrumental model on a set of k^M GMM-type moment restrictions. We assume that the (pseudo-true) parameter θ_0^M is the only value within the parameter space that gives a solution to the moment condition. Hence, the identification assumption reads

Assumption 4.B.1.

$$\mathbb{E}\left[u_t^M\left(y_t^M, \theta^M\right)\right] \neq 0 \quad \text{for all} \quad \theta^M \neq \theta_0^M \in \Theta^M.$$

Following Hansen (1982), we assume

Assumption 4.B.2. (i) Θ^M is a compact subset of $\mathbb{R}^{k^M}$,

(ii) $u_t^M(\cdot, \theta^M)$ is Borel measurable for each $\theta^M \in \Theta^M$,

(iii) $\mathbb{E}\left[u_t^M\left(y_t^M, \theta^M\right)\right]$ exists and is finite for all $\theta^M \in \Theta^M$,

(iv) $u_t^M\left(y_t^M, \theta^M\right)$ is first-moment continuous at all $\theta^M \in \Theta^M$.

Assumption 4.B.2 in combination with the stationarity and ergodictiy assumption in the nominal assumption 4.3.2 of the time series form sufficient conditions ensuring that the GMM-type criterion in Equation (4.3.5) converges almost surely uniformly to a non-stochastic limit criterion

$$Q_{\infty}^{M}\left(G_{*},\bar{\xi}^{M},\theta^{M}\right)=\frac{1}{2}\mathbb{E}\left[u_{t}^{M}\left(y_{t}^{M},\theta^{M}\right)\right]^{\top}W^{M}\mathbb{E}\left[u_{t}^{M}\left(y_{t}^{M},\theta^{M}\right)\right].$$

Assumption 4.B.1 implies that the limit criterion is uniquely minimized by θ_0^M, i.e.

$$\theta_{0}^{M}=\underset{\{\theta^{M}\in\Theta^{M}\}}{\operatorname{argmin}}\ Q_{\infty}^{M}\left(G_{*},\xi^{M},\theta^{M}\right).$$

Hence, under Assumptions 4.3.2, 4.B.1, and 4.B.2,

$$\hat{\theta}_{T}^{M}=\underset{\{\theta^{M}\in\Theta^{M}\}}{\operatorname{argmin}}\ Q_{T}^{M}\left(y_{t}^{M},\theta^{M}\right)$$

is a consistent estimator of θ_0^M. A proof of this claim can be found in Singleton (2009) Ch.3.

The probability limit of the simulated macroeconomic data is not the same as that of the observed data. Hence, we require additional assumptions for the misspecified simulator. We impose an identification assumption for the simulated macroeconomic time series given by

Assumption 4.B.3.

$$\mathbb{E}^{*}\left[u_{t}^{M}\left(\tilde{y}_{t}^{M,s}(\xi^{M}),\theta^{M}\right)\right]\neq 0\quad\text{for all}\ \ \theta^{M}\neq\tilde{\theta}^{M}(\bar{\xi}^{M})\in\Theta^{M},$$

and equivalent GMM regularity conditions on the moment conditions evaluated at the simulated macroeconomic time series, i.e.

Assumption 4.B.4. (i) Θ^M is a compact subset of $\mathbb{R}^{k^M}$,

(ii) $u_t^M(\cdot,\theta^M)$ is Borel measurable for each $\theta^M\in\Theta^M$,

(iii) $\mathbb{E}^{*}\left[u_{t}^{M}\left(\tilde{y}_{t}^{M,s}(\xi^{M}),\theta^{M}\right)\right]$ exists and is finite for all $\theta^M\in\Theta^M$,

(iv) $u_{t}^{M}\left(\tilde{y}_{t}^{M,s}(\xi^{M}),\theta^{M}\right)$ is first-moment continuous at all $\theta^M\in\Theta^M$.

In analogy to the observed series case, the GMM-type criterion in Equation (4.3.5) evaluated at the simulated macroeconomic data, $Q_{T}^{M}\left(\tilde{y}_{t}^{M,s}(\xi^{M}),\theta^{M}\right)$, converges almost surely uniformly to a non-stochastic limit criterion[12] under Assumption 4.B.3 for stationary and ergodic

[12]This limit criterion is given by

$$Q_{\infty}^{*,M}\left(G_{*},\bar{\xi}^{M},\theta^{M}\right)=\frac{1}{2}\mathbb{E}^{*}\left[u_{t}^{M}\left(\tilde{y}_{t}^{M,s}(\xi^{M}),\theta^{M}\right)\right]^{\top}W^{M}\mathbb{E}^{*}\left[u_{t}^{M}\left(\tilde{y}_{t}^{M,s}(\xi^{M}),\theta^{M}\right)\right].$$

simulated series. Assumption 4.B.4 implies that the limit criterion is uniquely minimized by $\tilde{\theta}^M(\bar{\xi}^M)$. Hence, under Assumptions 4.3.2, 4.B.3, and 4.B.4,

$$\tilde{\theta}_T^{M,s}(\bar{\xi}^M) = \underset{\{\theta^M \in \Theta^M\}}{\operatorname{argmin}} \; Q_T^M \left(\tilde{y}_T^{M,s}(\xi^M), \theta^M \right)$$

and $\tilde{\theta}_T^M(\bar{\xi}^M) = \frac{1}{S} \sum_{s=1}^S \tilde{\theta}_T^{M,s}(\bar{\xi}^M)$ is a consistent estimator of $\tilde{\theta}^M(\bar{\xi}^M)$.

Under Assumptions 4.B.1 to 4.3.4, the instrumental parameter estimator converges uniformly in ξ^M to

$$P_0 \lim_{T\to\infty} \hat{\theta}_T^M = \theta_0^M \qquad \text{and} \qquad P_* \lim_{T\to\infty} \tilde{\theta}_T^{M,s}(\xi^M) = \tilde{\theta}^M(\bar{\xi}^M).$$

4.B.2 Second estimation step

In the second estimation step, we take the first step macroeconomic nuisance parameter estimates $\hat{\xi}_T^M$ as given and we estimate the structural parameters of interest, the preference parameter ξ^P. We make the identification assumption

Assumption 4.B.5.

$$\mathbb{E}\left[u_t^P \left(y_t^P, \theta^P \right) \right] \neq 0 \quad \text{for all} \quad \theta^P \neq \theta_0^P \in \Theta^P.$$

For the second step, we proceed in a similar fashion as for the first step instrumental parameters. We assume that

Assumption 4.B.6. (i) Θ^P is a compact subset of $\mathbb{R}^{k^P}$,

(ii) $u_t^P(\cdot, \theta^P)$ is Borel measurable for each $\theta^P \in \Theta^P$,

(iii) $\mathbb{E}\left[u_t^P \left(y_t^P, \theta^P \right) \right]$ exists and is finite for all $\theta^P \in \Theta^P$,

(iv) $u_t^P \left(y_t^P, \theta^P \right)$ is first-moment continuous at all $\theta^P \in \Theta^P$.

The GMM-type criterion in Equation (4.3.13) converges almost surely uniformly to a non-stochastic limit criterion

$$Q_\infty^P \left(G_*, \bar{\xi}^M, \xi_{1,0}^P, \bar{\xi}_2^P, \theta^P \right) = \frac{1}{2} \mathbb{E}\left[u_t^P \left(y_t^P, \theta^P \right) \right]^\top W^P \mathbb{E}\left[u_t^P \left(y_t^P, \theta^P \right) \right]$$

under Assumption 4.B.6 for stationary and ergodic time series. This limit criterion is uniquely minimized by θ_0^P, i.e. $\theta_0^P = \underset{\{\theta^P \in \Theta^P\}}{\operatorname{argmin}} \, Q_\infty^P \left(G_*, \bar{\xi}^M, \xi_{1,0}^P, \bar{\xi}_2^P, \theta^P \right)$ under Assumption 4.B.5. It follows that under Assumptions 4.3.2, 4.B.5, and 4.B.6,

$$\hat{\theta}_T^P = \underset{\{\theta^P \in \Theta^P\}}{\operatorname{argmin}} \, Q_T^P \left(y_t^P, \theta^P \right)$$

is a consistent estimator of θ_0^P.

For the misspecified structural model, the probability limit for the instrumental parameter θ^P based on the simulated time series is different than that of the observed time series. We also need additional assumptions for the misspecified simulator of the financial variables. We impose an identification assumption for the simulated time series, i.e.

Assumption 4.B.7.

$$\mathbb{E}^* \left[u_t^P \left(\tilde{y}_t^{P,s}(\xi^M, \xi^P), \theta^P \right) \right] \neq 0 \quad \text{for all} \quad \theta^P \neq \tilde{\theta}^P(\bar{\xi}^M, \xi_{1,0}^P, \bar{\xi}_2^P) \in \Theta^P,$$

and regularity conditions for the GMM moment conditions evaluated at the simulated financial time series, given by

Assumption 4.B.8. (i) Θ^P is a compact subset of $\mathbb{R}^{k^P}$,

(ii) $u_t^P(\cdot, \theta^P)$ is Borel measurable for each $\theta^P \in \Theta^P$,

(iii) $\mathbb{E}^* \left[u_t^P \left(\tilde{y}_t^{P,s}(\xi^M, \xi^P), \theta^P \right) \right]$ exists and is finite for all $\theta^P \in \Theta^P$,

(iv) $u_t^P \left(\tilde{y}_t^{P,s}(\xi^M, \xi^P), \theta^P \right)$ is first-moment continuous at all $\theta^P \in \Theta^P$.

Under Assumption 4.B.7 for stationary and ergodic simulated series, the GMM-type criterion in Equation (4.3.13) evaluated at the simulated data, $Q_T^P \left(\tilde{y}_t^{P,s}(\xi^M, \xi^P), \theta^P \right)$, converges almost surely uniformly to a non-stochastic limit criterion[13] that is uniquely minimized by $\tilde{\theta}^P(\bar{\xi}^M, \xi_0^P)$ under Assumption 4.B.8. Under Assumptions 4.3.2, 4.B.7, and 4.B.8,

$$\tilde{\theta}_T^{P,s}(\bar{\xi}^M, \xi_{1,0}^P, \bar{\xi}_2^P) = \underset{\{\theta^P \in \Theta^P\}}{\operatorname{argmin}} \, Q_T^P \left(\tilde{y}_T^{P,s}(\xi^M, \xi^P), \theta^P \right)$$

and $\tilde{\theta}_T^P(\bar{\xi}^M, \xi_{1,0}^P, \bar{\xi}_2^P) = \frac{1}{S} \sum_{s=1}^S \tilde{\theta}_T^{P,s}(\bar{\xi}^M, \xi_{1,0}^P, \bar{\xi}_2^P)$ is a consistent estimator of $\tilde{\theta}^P(\bar{\xi}^M, \xi_{1,0}^P, \bar{\xi}_2^P)$.

[13] This limit criterion is given by

$$Q_\infty^{*,P} \left(G_*, \bar{\xi}^M, \xi_{1,0}^P, \bar{\xi}_2^P, \theta^P \right) = \frac{1}{2} \mathbb{E}^* \left[u_t^P \left(\tilde{y}_t^{P,s}(\xi^M, \xi^P), \theta^P \right) \right]^\top W^P \mathbb{E}^* \left[u_t^P \left(\tilde{y}_t^{P,s}(\xi^M, \xi^P), \theta^P \right) \right].$$

Under Assumptions 4.B.5 to 4.3.6 it follows that the instrumental parameter estimator for the preference parameters converges uniformly in ξ^P to[14]

$$P_0 \lim_{T\to\infty} \hat{\theta}_T^P = \theta_0^P \qquad \text{and} \qquad P_* \lim_{T\to\infty} \tilde{\theta}_T^{P,s}(\xi^M, \xi^P) = \tilde{\theta}^P(\bar{\xi}^M, \xi_{1,0}^P, \bar{\xi}_2^P),$$

where $P_0 \lim_{T\to\infty}$ and $P_* \lim_{T\to\infty}$ denote the limit with respect to the probability distributions P_0 and P_* for the observed and simulated series.

Appendix 4.C Asymptotic distribution of the two-step partial indirect inference estimator

In a first step, we estimate the parameters of the macroeconomic or technology part, ξ^M using an AR(p) model on consumption growth or a HAR(p) model on consumption and dividend growth as instrumental models. The instrumental parameters for the observed and simulated time series are given by the extremum estimator with GMM moment conditions

$$\begin{aligned}
\hat{\theta}_T^M &= \underset{\{\theta^M \in \Theta^M\}}{\operatorname{argmin}} \; Q_T^M \left(y_t^M, \theta^M \right), \\
\tilde{\theta}_T^{M,s}(\xi^M) &= \underset{\{\theta^M \in \Theta^M\}}{\operatorname{argmin}} \; Q_T^M \left(\tilde{y}_t^{M,s}(\xi^M), \theta^M \right),
\end{aligned}$$

with $Q_T^M \left(\cdot, \theta^M \right) = \frac{1}{2} \, g_T^M(\cdot, \theta^M)^\top \widehat{W}_T^M g_T^M(\cdot, \theta^M)$ and $g_T^M(\cdot, \theta^M) = \frac{1}{T} \sum_{t=1}^T u_t^M \left(\cdot, \theta^M \right)$.

This is feasible as the partial equilibrium asset pricing models are separable in macroeconomic or technology parameters ξ^M and the preference parameters ξ^P. The simulation of the macroeconomic part, $\left\{ \tilde{y}_t^{M,s}(\xi^M) \right\}_{t=1}^T$, only depends on the technology parameters ξ^M. It is clear that the simulator here does not align with the true data generating process. The technology parameters can be estimated by partial indirect inference based on the (misspecified) simulator. They are obtained by minimizing the quadratic form

$$\hat{\xi}_T^M = \underset{\xi^M \in \Xi^M\}}{\operatorname{argmin}} \left[\hat{\theta}_T^M - \frac{1}{S} \sum_{s=1}^S \tilde{\theta}_T^{M,s}(\xi^M) \right]^\top \widehat{\Omega}_T^M \left[\hat{\theta}_T^M - \frac{1}{S} \sum_{s=1}^S \tilde{\theta}_T^{M,s}(\xi^M) \right],$$

where $\widehat{\Omega}_T^M$ is a consistent estimator of the positive definite non-stochastic weighting matrix Ω^M.

[14] For the partial encompassing cases the probability limits are given by $P_* \lim_{T\to\infty} \tilde{\theta}_T^{P,s}(\xi^M, \xi^P) = \tilde{\theta}^P(\bar{\xi}^M, \xi_{1,0}^P)$ and $P_* \lim_{T\to\infty} \tilde{\theta}_T^{P,s}(\xi^M, \xi^P) = \tilde{\theta}^P(\xi_{1,0}^P)$.

The asymptotic distribution of the first step estimator is derived as follows. The key point is that the model of the technology part is only a crude approximation of the reality. Hence the simulated time series are based on a misspecified simulator. The consequence of this is that the nuisance technology parameters converge to pseudo-true values, $\bar{\xi}^M$, rather than to true parameter values, ξ_0^M. The asymptotic distribution is derived in multiple steps. We start by deriving the first-order conditions (FOC) of the objective function, i.e.

$$\frac{1}{S}\sum_{s=1}^{S}\frac{\partial\tilde{\theta}_T^{M,s}}{\partial\xi^M}\left(\hat{\xi}_T^M\right)^\top\widehat{\Omega}_T^M\left[\hat{\theta}_T^M-\frac{1}{S}\sum_{s=1}^{S}\tilde{\theta}_T^{M,s}(\hat{\xi}_T^M)\right]=0.$$

Subsequently, we use a mean value extension around the expression in brackets

$$\left[\hat{\theta}_T^M-\frac{1}{S}\sum_{s=1}^{S}\tilde{\theta}_T^{M,s}(\hat{\xi}_T^M)\right]\simeq\left[\hat{\theta}_T^M-\frac{1}{S}\sum_{s=1}^{S}\tilde{\theta}_T^{M,s}(\bar{\xi}^M)\right]+\frac{1}{S}\sum_{s=1}^{S}\frac{\partial\tilde{\theta}_T^{M,s}}{\partial\xi^{M\top}}\left(\bar{\xi}^M\right)\cdot\left[\hat{\xi}_T^M-\bar{\xi}^M\right].$$

This result is plugged into the FOC and rearranged for the nuisance parameters

$$\begin{aligned}\sqrt{T}\left[\hat{\xi}_T^M-\bar{\xi}^M\right]\simeq&\left\{\frac{\partial\tilde{\theta}_T^M}{\partial\xi^M}\left(\bar{\xi}^M\right)^\top\Omega^M\frac{\partial\tilde{\theta}_T^M}{\partial\xi^{M\top}}\left(\bar{\xi}^M\right)\right\}^{-1}\\&\frac{\partial\tilde{\theta}_T^M}{\partial\xi^M}\left(\bar{\xi}^M\right)^\top\Omega^M\sqrt{T}\left[\hat{\theta}_T^M-\frac{1}{S}\sum_{s=1}^{S}\tilde{\theta}_T^{M,s}\left(\hat{\xi}_T^M\right)\right].\end{aligned}$$

The probability limit of the Jacobian is given by $\frac{\partial\tilde{\theta}^M}{\xi^M}(\bar{\xi}^M)^\top=P^*\lim_{T\to\infty}\frac{1}{S}\sum_{s=1}^{S}\frac{\partial\tilde{\theta}_T^{M,s}}{\partial\xi^M}(\hat{\xi}_t^M)^\top$, where P^* is the probability limit of the misspecified technology simulator and the pseudo-true value is denoted by $\bar{\xi}^M=P^*\lim_{T\to\infty}\hat{\xi}_T^M$. The Jacobian, hence, converges uniformly in ξ^M.

In a second step, we aim at estimating the preference parameters ξ^P. These parameters are separated into structural parameters of interest and nuisance parameters. In case of the habit persistence model the former are the time preference parameter, δ, and the risk aversion parameter, γ, and ϕ is a nuisance parameter that gives the autoregressive parameter of surplus consumption, i.e. $\xi^P=\left(\xi_1^P,\xi_2^P\right)^\top=(\delta,\gamma,\phi)^\top$. For that purpose, we use a second instrumental model that is based on moment conditions from asset pricing theory. The second step is based on simulated data of the complete partial equilibrium asset pricing model based on the technology and preference parameters, i.e. $\left\{\tilde{y}_t^{P,s}(\xi^M,\xi^P)\right\}_{t=1}^{T}$.

The instrumental parameters of the observed and simulated data are estimated by

$$\hat{\theta}_T^P = \underset{\{\theta^P \in \Theta^P\}}{\operatorname{argmin}} \; Q_T^P \left(y_t^P, \theta^P \right),$$
$$\tilde{\theta}_T^{P,s} \left(\xi^M, \xi^P \right) = \underset{\{\theta^P \in \Theta^P\}}{\operatorname{argmin}} \; Q_T^P \left(\tilde{y}_t^{P,s}(\xi^M, \xi^P), \theta^P \right),$$

with $Q_T^P \left(\cdot, \theta^P \right) = \frac{1}{2} \, g_T^P \left(\cdot, \theta^P \right)^\top \widehat{W}_T^P g_T^P \left(\cdot, \theta^P \right)$ and $g_T^P \left(\cdot, \theta^P \right) = \frac{1}{T} \sum_{t=1}^T u_t^P \left(\cdot, \theta^P \right)$. The partial indirect inference estimator of the preference parameters is given by

$$\hat{\xi}_T^P = \underset{\xi^P \in \Xi^P\}}{\operatorname{argmin}} \left[\hat{\theta}_T^P - \frac{1}{S} \sum_{s=1}^S \tilde{\theta}_T^{P,s}(\xi^M, \xi^P) \right]^\top \widehat{\Omega}_T^P \left[\hat{\theta}_T^P - \frac{1}{S} \sum_{s=1}^S \tilde{\theta}_T^{P,s}(\xi^M, \xi^P) \right],$$

where $\widehat{\Omega}_T^P$ is a consistent estimator of the positive definite weighting matrix Ω^P.

The asymptotic distribution of the two-step partial indirect inference in the full encompassing setting follows the logic of the two-step estimator of Grammig and Küchlin (2018) and the partial indirect inference estimator in Dridi et al. (2007). We write the two-step procedure in a generalized method of moments (GMM) setup

$$\begin{pmatrix} \hat{\xi}_T^M \\ \hat{\xi}_T^P \end{pmatrix} = \underset{\{(\xi^M, \xi^P) \in \Xi^M \times \Xi^P\}}{\operatorname{argmin}} \begin{bmatrix} \hat{\theta}_T^M - \frac{1}{S} \sum_{s=1}^S \tilde{\theta}_T^{M,s}(\xi^M) \\ \hat{\theta}_T^P - \frac{1}{S} \sum_{s=1}^S \tilde{\theta}_T^{P,s}(\xi^M, \xi^P) \end{bmatrix} \begin{bmatrix} \widehat{\Omega}_T^M & 0 \\ 0 & \widehat{\Omega}_T^P \end{bmatrix} \begin{bmatrix} \hat{\theta}_T^M - \frac{1}{S} \sum_{s=1}^S \tilde{\theta}_T^{M,s}(\xi^M) \\ \hat{\theta}_T^P - \frac{1}{S} \sum_{s=1}^S \tilde{\theta}_T^{P,s}(\xi^M, \xi^P) \end{bmatrix}$$

and the first order condition is given by

$$\begin{bmatrix} \frac{1}{S} \sum_{s=1}^S \frac{\partial \tilde{\theta}_T^{M,s}}{\partial \xi^M}(\hat{\xi}_T^M)^\top & 0 \\ 0 & \frac{1}{S} \sum_{s=1}^S \frac{\partial \tilde{\theta}_T^{P,s}}{\partial \xi^P}(\hat{\xi}_T^M, \hat{\xi}_T^P)^\top \end{bmatrix} \begin{bmatrix} \widehat{\Omega}_T^M & 0 \\ 0 & \widehat{\Omega}_T^P \end{bmatrix} \begin{bmatrix} \hat{\theta}_T^M - \frac{1}{S} \sum_{s=1}^S \tilde{\theta}_T^{M,s}(\hat{\xi}_T^M) \\ \hat{\theta}_T^P - \frac{1}{S} \sum_{s=1}^S \tilde{\theta}_T^{P,s}(\hat{\xi}_T^M, \hat{\xi}_T^P) \end{bmatrix} = 0.$$

We expand the last term around the true value of the preference parameters ξ_0^P[15] and the pseudo-true value of the technology parameters $\bar{\xi}^M$

$$\begin{bmatrix} \hat{\theta}_T^M - \frac{1}{S}\sum_{s=1}^S \tilde{\theta}_T^{M,s}(\hat{\xi}_T^M) \\ \hat{\theta}_T^P - \frac{1}{S}\sum_{s=1}^S \tilde{\theta}_T^{P,s}(\hat{\xi}_T^M, \hat{\xi}_T^P) \end{bmatrix} \simeq \begin{bmatrix} \hat{\theta}_T^M - \frac{1}{S}\sum_{s=1}^S \tilde{\theta}_T^{M,s}(\bar{\xi}^M) \\ \hat{\theta}_T^P - \frac{1}{S}\sum_{s=1}^S \tilde{\theta}_T^{P,s}(\bar{\xi}^M, \xi_0^P) \end{bmatrix} + \begin{bmatrix} \frac{1}{S}\sum_{s=1}^S \frac{\partial \tilde{\theta}_T^{M,s}}{\partial \xi^{M\top}}(\bar{\xi}^M) & 0 \\ \frac{1}{S}\sum_{s=1}^S \frac{\partial \tilde{\theta}_T^{P,s}}{\partial \xi^{M\top}}(\bar{\xi}^M, \xi_0^P) & \frac{1}{S}\sum_{s=1}^S \frac{\partial \tilde{\theta}_T^{P,s}}{\partial \xi^{P\top}}(\bar{\xi}^M, \xi_0^P) \end{bmatrix} \begin{bmatrix} \hat{\xi}_T^M - \bar{\xi}^M \\ \hat{\xi}_t^P - \xi_0^P \end{bmatrix}.$$

Plugging this in, we obtain

$$\begin{aligned} &\begin{bmatrix} \frac{1}{S}\sum_{s=1}^S \frac{\partial \tilde{\theta}_T^{M,s}}{\partial \xi^M}(\hat{\xi}_T^M)^\top & 0 \\ 0 & \frac{1}{S}\sum_{s=1}^S \frac{\partial \tilde{\theta}_T^{P,s}}{\partial \xi^P}(\hat{\xi}_T^M, \hat{\xi}_T^P)^\top \end{bmatrix} \begin{bmatrix} \widehat{\Omega}_T^M & 0 \\ 0 & \widehat{\Omega}_T^P \end{bmatrix} \begin{bmatrix} \hat{\theta}_T^M - \frac{1}{S}\sum_{s=1}^S \tilde{\theta}_T^{M,s}(\bar{\xi}^M) \\ \hat{\theta}_T^P - \frac{1}{S}\sum_{s=1}^S \tilde{\theta}_T^{P,s}(\bar{\xi}^M, \xi_0^P) \end{bmatrix} \\ &+ \begin{bmatrix} \frac{1}{S}\sum_{s=1}^S \frac{\partial \tilde{\theta}_T^{M,s}}{\partial \xi^M}(\hat{\xi}_T^M)^\top & 0 \\ 0 & \frac{1}{S}\sum_{s=1}^S \frac{\partial \tilde{\theta}_T^{P,s}}{\partial \xi^P}(\hat{\xi}_T^M, \hat{\xi}_T^P)^\top \end{bmatrix} \begin{bmatrix} \widehat{\Omega}_T^M & 0 \\ 0 & \widehat{\Omega}_T^P \end{bmatrix} \begin{bmatrix} \frac{1}{S}\sum_{s=1}^S \frac{\partial \tilde{\theta}_T^{M,s}}{\partial \xi^{M\top}}(\bar{\xi}^M) & 0 \\ \frac{1}{S}\sum_{s=1}^S \frac{\partial \tilde{\theta}_T^{P,s}}{\partial \xi^{M\top}}(\bar{\xi}^M, \xi_0^P) & \frac{1}{S}\sum_{s=1}^S \frac{\partial \tilde{\theta}_T^{P,s}}{\partial \xi^{P\top}}(\bar{\xi}^M, \xi_0^P) \end{bmatrix} \begin{bmatrix} \hat{\xi}_T^M - \bar{\xi}^M \\ \hat{\xi}_t^P - \xi_0^P \end{bmatrix} = 0 \end{aligned}$$

and by multiplying out the second summand

$$\begin{aligned} &\begin{bmatrix} \frac{1}{S}\sum_{s=1}^S \frac{\partial \tilde{\theta}_T^{M,s}}{\partial \xi^M}(\hat{\xi}_T^M)^\top \widehat{\Omega}_T^M & 0 \\ 0 & \frac{1}{S}\sum_{s=1}^S \frac{\partial \tilde{\theta}_T^{P,s}}{\partial \xi^P}(\hat{\xi}_T^M, \hat{\xi}_T^P)^\top \widehat{\Omega}_T^P \end{bmatrix} \sqrt{T} \begin{bmatrix} \hat{\theta}_T^M - \frac{1}{S}\sum_{s=1}^S \tilde{\theta}_T^{M,s}(\bar{\xi}^M) \\ \hat{\theta}_T^P - \frac{1}{S}\sum_{s=1}^S \tilde{\theta}_T^{P,s}(\bar{\xi}^M, \xi_0^P) \end{bmatrix} \\ &+ \begin{bmatrix} \frac{1}{S}\sum_{s=1}^S \frac{\partial \tilde{\theta}_T^{M,s}}{\partial \xi^M}(\hat{\xi}_T^M)^\top \widehat{\Omega}_T^M \frac{1}{S}\sum_{s=1}^S \frac{\partial \tilde{\theta}_T^{M,s}}{\partial \xi^{M\top}}(\bar{\xi}^M) & 0 \\ \frac{1}{S}\sum_{s=1}^S \frac{\partial \tilde{\theta}_T^{P,s}}{\partial \xi^P}(\hat{\xi}_T^M, \hat{\xi}_T^P)^\top \widehat{\Omega}_T^P \frac{1}{S}\sum_{s=1}^S \frac{\partial \tilde{\theta}_T^{P,s}}{\partial \xi^{M\top}}(\bar{\xi}^M, \xi_0^P) & \frac{1}{S}\sum_{s=1}^S \frac{\partial \tilde{\theta}_T^{P,s}}{\partial \xi^{P\top}}(\hat{\xi}_T^M, \hat{\xi}_T^P)^\top \widehat{\Omega}_T^P \frac{1}{S}\sum_{s=1}^S \frac{\partial \tilde{\theta}_T^{P,s}}{\partial \xi^{P\top}}(\bar{\xi}^M, \xi_0^P) \end{bmatrix} \sqrt{T} \begin{bmatrix} \hat{\xi}_T^M - \bar{\xi}^M \\ \hat{\xi}_t^P - \xi_0^P \end{bmatrix} = 0. \end{aligned}$$

[15] We simplify the notation here; ξ_0^P is composed of $\xi_{1,0}^P$ and $\bar{\xi}_2^P$, the true values of the structural parameters of interest, ξ_1^P and the pseudo-true values of the nuisance parameters ξ_2^P.

Finally, we rearrange the equation above and solve for the structural parameters. Asymptotically, the structural parameters can be written as

$$\sqrt{T}\begin{bmatrix}\hat{\xi}_T^M - \bar{\xi}^M \\ \hat{\xi}_t^P - \xi_0^P\end{bmatrix} \simeq -\begin{bmatrix}\frac{\partial\tilde{\theta}^M}{\partial\xi^M}(\bar{\xi}^M)^\top\Omega^M\frac{\partial\tilde{\theta}^M}{\partial\xi^{M\top}}(\bar{\xi}^M) & 0 \\ \frac{\partial\tilde{\theta}^P}{\partial\xi^P}(\bar{\xi}^M,\xi_0^P)^\top\Omega^P\frac{\partial\tilde{\theta}^P}{\partial\xi^{M\top}}(\bar{\xi}^M,\xi_0^P) & \frac{\partial\tilde{\theta}^P}{\partial\xi^P}(\bar{\xi}^M,\xi_0^P)^\top\Omega^P\frac{\partial\tilde{\theta}^P}{\partial\xi^{P\top}}(\bar{\xi}^M,\xi_0^P)\end{bmatrix}^{-1} \cdot \begin{bmatrix}\frac{\partial\tilde{\theta}^M}{\bar{\xi}^M}(\hat{\xi}_T^M)^\top\Omega^M & 0 \\ 0 & \frac{\partial\tilde{\theta}^P}{\partial\xi^P}(\bar{\xi}^M,\xi_0^P)^\top\Omega^P\end{bmatrix}\sqrt{T}\begin{bmatrix}\hat{\theta}_T^M - \frac{1}{S}\sum_{s=1}^S\tilde{\theta}_T^{M,s}(\bar{\xi}^M) \\ \hat{\theta}_T^P - \frac{1}{S}\sum_{s=1}^S\tilde{\theta}_T^{P,s}(\bar{\xi}^M,\xi_0^P)\end{bmatrix} \tag{4.C.1}$$

$$\text{where}\quad \frac{\partial\tilde{\theta}^M}{\partial\xi^M}(\bar{\xi}^M)^\top = P^*\lim_{T\to\infty}\frac{1}{S}\sum_{s=1}^S\frac{\partial\tilde{\theta}_T^{M,s}}{\partial\xi^M}(\bar{\xi}^M)^\top \quad\text{and}\quad \widehat{\Omega}_T^M \underset{p}{\to} \Omega^M,$$

$$\frac{\partial\tilde{\theta}^P}{\partial\xi^M}(\bar{\xi}^M,\xi_0^P)^\top = P^*\lim_{T\to\infty}\frac{1}{S}\sum_{s=1}^S\frac{\partial\tilde{\theta}_T^{P,s}}{\partial\xi^M}(\bar{\xi}^M,\xi_0^P)^\top \quad\text{and}\quad \widehat{\Omega}_T^P \underset{p}{\to} \Omega^P,$$

$$\frac{\partial\tilde{\theta}^P}{\partial\xi^P}(\bar{\xi}^M,\xi_0^P)^\top = P^*\lim_{T\to\infty}\frac{1}{S}\sum_{s=1}^S\frac{\partial\tilde{\theta}_T^{P,s}}{\partial\xi^P}(\bar{\xi}^M,\xi_0^P)^\top$$

The next subsection shows how the last term in Equation (4.C.1) is distributed.

4.C.1 Asymptotic distribution of the instrumental parameters

First, we show how $\sqrt{T}\begin{bmatrix}\hat{\theta}_T^M - \theta_0^M \\ \hat{\theta}_T^P - \theta_0^P\end{bmatrix}$ is distributed. Both instrumental models are estimated using GMM-type moment conditions. Note that we denote $\{y_t^M\}_{t=1}^T$ by y^M and $\{y_t^P\}_{t=1}^T$ by y^P.

$$Q_T^M\left(y_t^M, \theta^M\right) = \frac{1}{2} g_T^M\left(y_t^M, \theta^M\right)^\top \widehat{W}_T^M g_T^M\left(y_t^M, \theta^M\right) \qquad \text{and} \quad g_T^M\left(y_t^M, \theta^M\right) = \frac{1}{T}\sum_{t=1}^T u_t^M\left(y_t^M, \theta^M\right) \underset{p}{\to} \mathbb{E}\left[u_t^M(\theta^M)\right]$$

$$Q_T^P\left(y_t^P, \theta^P\right) = \frac{1}{2} g_T^P\left(y_t^P, \theta^P\right)^\top \widehat{W}_T^P g_T^P\left(y_t^P, \theta^P\right) \qquad \text{and} \quad g_T^P\left(y_t^P, \theta^P\right) = \frac{1}{T}\sum_{t=1}^T u_t^P\left(y_t^P, \theta^P\right) \underset{p}{\to} \mathbb{E}\left[u_t^P(\theta^P)\right]$$

The FOC for the two instrumental models are given by $\frac{\partial g_T^{M\top}}{\partial \theta^M}\left(\hat{\theta}_T^M\right)\widehat{W}_T^M g_T^M\left(\hat{\theta}_T^M\right) = 0$ and $\frac{\partial g_T^{P\top}}{\partial \theta^P}\left(\hat{\theta}_T^P\right)\widehat{W}_T^P g_T^P\left(\hat{\theta}_T^P\right) = 0$. The mean value expansions around the true instrumental parameters read $g_T^M\left(\hat{\theta}_T^M\right) \simeq g_T^M\left(\theta_0^M\right) + \frac{\partial g_T^M}{\partial \theta^{M\top}}\left(\theta_0^M\right)\left(\hat{\theta}_T^M - \theta_0^M\right)$ and $g_T^P\left(\hat{\theta}_T^P\right) \simeq g_T^P\left(\theta_0^P\right) + \frac{\partial g_T^P}{\partial \theta^{P\top}}\left(\theta_0^P\right)\left(\hat{\theta}_T^P - \theta_0^P\right)$. This is plugged into the FOC and solved for $\left(\hat{\theta}_T^M - \theta_0^M\right)$ and $\left(\hat{\theta}_T^P - \theta_0^P\right)$:

$$\left(\hat{\theta}_T^M - \theta_0^M\right) = \left\{\frac{\partial g_T^{M\top}}{\partial \theta^M}\left(\hat{\theta}_T^M\right)\widehat{W}_T^M \frac{\partial g_T^M}{\partial \theta^{M\top}}\left(\theta_0^M\right)\right\}^{-1} \frac{\partial g_T^{M\top}}{\partial \theta^M}\left(\hat{\theta}_T^M\right)\widehat{W}_T^M g_T^M\left(\theta_0^M\right),$$

$$\left(\hat{\theta}_T^P - \theta_0^P\right) = \left\{\frac{\partial g_T^{P\top}}{\partial \theta^P}\left(\hat{\theta}_T^P\right)\widehat{W}_T^P \frac{\partial g_T^P}{\partial \theta^{P\top}}\left(\theta_0^P\right)\right\}^{-1} \frac{\partial g_T^{P\top}}{\partial \theta^P}\left(\hat{\theta}_T^P\right)\widehat{W}_T^P g_T^P\left(\theta_0^P\right).$$

Asymptotically, the instrumental parameters can also be written as

$$\sqrt{T}\left(\hat{\theta}_T^M - \theta_0^M\right) \simeq \left\{\mathbb{E}\left[\frac{\partial u_t^{M\top}}{\partial \theta^M}\left(\theta_0^M\right)\right] W^M \mathbb{E}\left[\frac{\partial u_t^M}{\partial \theta^{M\top}}\left(\theta_0^M\right)\right]\right\}^{-1} \mathbb{E}\left[\frac{\partial u_t^{M\top}}{\partial \theta^M}\left(\theta_0^M\right)\right] W^M \sqrt{T} g_T^M\left(\theta_0^M\right) = C_0^M \sqrt{T} g_T^M\left(\theta_0^M\right)$$

$$\sqrt{T}\left(\hat{\theta}_T^P - \theta_0^P\right) \simeq \left\{\mathbb{E}\left[\frac{\partial u_t^{P\top}}{\partial \theta^P}\left(\theta_0^P\right)\right] W^P \mathbb{E}\left[\frac{\partial u_t^P}{\partial \theta^{P\top}}\left(\theta_0^P\right)\right]\right\}^{-1} \mathbb{E}\left[\frac{\partial u_t^{P\top}}{\partial \theta^P}\left(\theta_0^P\right)\right] W^P \sqrt{T} g_T^P\left(\theta_0^P\right) = C_0^P \sqrt{T} g_T^P\left(\theta_0^P\right)$$

or in compact notation

$$\sqrt{T}\begin{bmatrix}\hat{\theta}_T^M - \theta_0^M \\ \hat{\theta}_T^P - \theta_0^P\end{bmatrix} \simeq \begin{bmatrix} C_0^M & 0 \\ 0 & C_0^P\end{bmatrix} \sqrt{T}\begin{bmatrix} g_T^M\left(\theta_0^M\right) \\ g_T^P\left(\theta_0^P\right)\end{bmatrix}$$

where $\sqrt{T}\begin{bmatrix} g_T^M\left(\theta_0^M\right) \\ g_T^P\left(\theta_0^P\right)\end{bmatrix} = \sqrt{T}\begin{bmatrix} \frac{1}{T}\sum_{t=1}^T u_t^M\left(\theta_0^M\right) \\ \frac{1}{T}\sum_{t=1}^T u_t^P\left(\theta_0^P\right)\end{bmatrix}$.

Assumption 4.C.1. A multivariate central limit theorem (CLT) applies, such that under Assumptions 4.B.1 and 4.B.5

$$\sqrt{T}\begin{bmatrix} \frac{1}{T}\sum_{t=1}^T u_t^M\left(\theta_0^M\right) \\ \frac{1}{T}\sum_{t=1}^T u_t^P\left(\theta_0^P\right)\end{bmatrix} \xrightarrow{d} \mathcal{N}(0, V)$$

where

$$V = \Gamma_0 + \sum_{j=1}^{\infty}\left(\Gamma_j + \Gamma_j^\top\right) \qquad \text{and} \quad \Gamma_j = \mathbb{E}\begin{bmatrix} u_t^M\left(y^M, \theta_0^M\right) u_{t-j}^M\left(y^M, \theta_0^M\right)^\top & u_t^M\left(y^M, \theta_0^M\right) u_{t-j}^P\left(y^P, \theta_0^P\right)^\top \\ u_t^P\left(y^P, \theta_0^P\right) u_{t-j}^M\left(y^M, \theta_0^M\right)^\top & u_t^P\left(y^P, \theta_0^P\right) u_{t-j}^P\left(y^P, \theta_0^P\right)^\top\end{bmatrix}.$$

The distribution of the instrumental parameters evaluated at the observed data, we obtain

$$\sqrt{T}\begin{bmatrix}\hat{\theta}_T^M - \theta_0^M \\ \hat{\theta}_T^P - \theta_0^P\end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{bmatrix}C_0^M & 0 \\ 0 & C_0^P\end{bmatrix} V \begin{bmatrix}C_0^{M\top} & 0 \\ 0 & C_0^{P\top}\end{bmatrix}\right)$$

Second, we investigate how is $\sqrt{T}\begin{bmatrix}\tilde{\theta}_T^{M,s}\left(\bar{\xi}^M\right) - \theta_0^M \\ \tilde{\theta}_T^{P,s}\left(\bar{\xi}^M, \xi_0^P\right) - \theta_0^P\end{bmatrix}$ is distributed. Note that the simulators is misspecified and has a different probability limit, P^*, than the true data P_0. The probability limit of the instrumental parameters are given by $\tilde{\theta}^M(\bar{\xi}^M) = P^* \lim_{T\to\infty} \tilde{\theta}_T^{M,s}(\bar{\xi}^M)$ and $\tilde{\theta}^P(\bar{\xi}^M, \xi_0^P) = P^* \lim_{T\to\infty} \tilde{\theta}_T^{P,s}(\bar{\xi}^M, \xi_0^P)$. Moreover, all first-step quantities are evaluated by the technology simulator, i.e. $\left\{\tilde{y}_t^{M,s}(\xi^M)\right\}_{t=1}^T$ and all second-step quantities are valuated by both the technology and the preference parameter simulator, i.e. $\left\{\tilde{y}_t^{P,s}(\xi^M, \xi^P)\right\}_{t=1}^T$. In the following, the dependence on the simulated time series is omitted for brevity. The FOC for the two instrumental models are given by

$$\begin{aligned}
&\frac{\partial g_T^{M\top}}{\partial \theta^M}\left(\tilde{\theta}_T^{M,s}(\bar{\xi}^M)\right)\widehat{W}_T^M g_T^M\left(\tilde{\theta}_T^{M,s}(\bar{\xi}^M)\right) = 0 \qquad \text{and} \\
&\frac{\partial g_T^{P\top}}{\partial \theta^P}\left(\tilde{\theta}_T^{P,s}(\bar{\xi}^M, \xi_0^P)\right)\widehat{W}_T^P g_T^P\left(\tilde{\theta}_T^{P,s}(\bar{\xi}^M, \xi_0^P)\right) = 0
\end{aligned}$$

The mean value expansions around the true instrumental parameters read

$$\begin{aligned}
g_T^M\left(\tilde{\theta}_T^{M,s}(\bar{\xi}^M)\right) &\simeq g_T^M\left(\tilde{\theta}^M(\bar{\xi}^M)\right) + \frac{\partial g_T^M}{\partial \theta^{M\top}}\left(\tilde{\theta}^M(\bar{\xi}^M)\right)\left(\tilde{\theta}_T^{M,s}(\bar{\xi}^M) - \tilde{\theta}^M(\bar{\xi}^M)\right) \qquad \text{and} \\
g_T^P\left(\tilde{\theta}_T^{P,s}(\bar{\xi}^M, \xi_0^P)\right) &\simeq g_T^P\left(\tilde{\theta}^P(\bar{\xi}^M, \xi_0^P)\right) + \frac{\partial g_T^P}{\partial \theta^{P\top}}\left(\tilde{\theta}^P(\bar{\xi}^M, \xi_0^P)\right)\left(\tilde{\theta}_T^{P,s}(\bar{\xi}^M, \xi_0^P) - \tilde{\theta}^P(\bar{\xi}^M, \xi_0^P)\right).
\end{aligned}$$

This is plugged into the FOC and solved for $\left(\tilde{\theta}_T^{M,s}(\bar{\xi}^M) - \tilde{\theta}^M(\bar{\xi}^M)\right)$ and $\left(\tilde{\theta}_T^{P,s}(\bar{\xi}^M, \xi_0^P) - \tilde{\theta}^P(\bar{\xi}^M, \xi_0^P)\right)$:

$$\left(\tilde{\theta}_T^{M,s}(\bar{\xi}^M) - \tilde{\theta}^M(\bar{\xi}^M)\right) = \left\{\frac{\partial g_T^{M\top}}{\partial \theta^M}\left(\tilde{\theta}_T^{M,s}(\bar{\xi}^M)\right)\widehat{W}_T^M \frac{\partial g_T^M}{\partial \theta^{M\top}}\left(\tilde{\theta}^M(\bar{\xi}^M)\right)\right\}^{-1} \frac{\partial g_T^{M\top}}{\partial \theta^M}\left(\tilde{\theta}_T^{M,s}(\bar{\xi}^M)\right)\widehat{W}_T^M g_T^M\left(\tilde{\theta}^M(\bar{\xi}^M)\right),$$

$$\left(\tilde{\theta}_T^{P,s}(\bar{\xi}^M, \xi_0^P) - \tilde{\theta}^P(\bar{\xi}^M, \xi_0^P)\right) = \left\{\frac{\partial g_T^{P\top}}{\partial \theta^P}\left(\tilde{\theta}_T^{P,s}(\bar{\xi}^M, \xi_0^P)\right)\widehat{W}_T^P \frac{\partial g_T^P}{\partial \theta^{P\top}}\left(\tilde{\theta}^P(\bar{\xi}^M, \xi_0^P)\right)\right\}^{-1} \frac{\partial g_T^{P\top}}{\partial \theta^P}\left(\tilde{\theta}_T^{P,s}(\bar{\xi}^M, \xi_0^P)\right)\widehat{W}_T^P g_T^P\left(\tilde{\theta}^P(\bar{\xi}^M, \xi_0^P)\right).$$

Asymptotically, the instrumental parameters can also be written as

$$\begin{aligned}\sqrt{T}\left(\tilde{\theta}_T^{M,s}(\bar{\xi}^M) - \tilde{\theta}^M(\bar{\xi}^M)\right) \simeq & \left\{\mathbb{E}\left[\frac{\partial u_t^{M\top}}{\partial \theta^M}\left(\tilde{\theta}^M(\bar{\xi}^M)\right)\right] W^M \mathbb{E}\left[\frac{\partial u_t^M}{\partial \theta^{M\top}}\left(\tilde{\theta}^M(\bar{\xi}^M)\right)\right]\right\}^{-1} \\ & \cdot \mathbb{E}\left[\frac{\partial u_t^{M\top}}{\partial \theta^M}\left(\tilde{\theta}^M(\bar{\xi}^M)\right)\right] W^M \sqrt{T} g_T^M\left(\theta_0^M\right) = C_0^{M^*}\sqrt{T} g_T^M\left(\tilde{\theta}^M(\bar{\xi}^M)\right) \\ \sqrt{T}\left(\tilde{\theta}_T^{P,s}(\bar{\xi}^M, \xi_0^P) - \tilde{\theta}^P(\bar{\xi}^M, \xi_0^P)\right) \simeq & \left\{\mathbb{E}\left[\frac{\partial u_t^{P\top}}{\partial \theta^P}\left(\tilde{\theta}^P(\bar{\xi}^M, \xi_0^P)\right)\right] W^P \mathbb{E}\left[\frac{\partial u_t^P}{\partial \theta^{P\top}}\left(\tilde{\theta}^P(\bar{\xi}^M, \xi_0^P)\right)\right]\right\}^{-1} \\ & \cdot \mathbb{E}\left[\frac{\partial u_t^{P\top}}{\partial \theta^P}\left(\tilde{\theta}^P(\bar{\xi}^M, \xi_0^P)\right)\right] W^P \sqrt{T} g_T^P\left(\tilde{\theta}^P(\bar{\xi}^M, \xi_0^P)\right) = C_0^{P^*}\sqrt{T} g_T^P\left(\tilde{\theta}^P(\bar{\xi}^M, \xi_0^P)\right)\end{aligned}$$

or in compact notation

$$\sqrt{T}\begin{bmatrix}\tilde{\theta}_T^{M,s}(\bar{\xi}^M)-\tilde{\theta}^M(\bar{\xi}^M)\\ \tilde{\theta}_T^{P,s}(\bar{\xi}^M,\xi_0^P)-\tilde{\theta}^P(\bar{\xi}^M,\xi_0^P)\end{bmatrix}\simeq\begin{bmatrix}C_0^{M*} & 0\\ 0 & C_0^{P*}\end{bmatrix}\sqrt{T}\begin{bmatrix}g_T^M\left(\tilde{\theta}^M\left(\bar{\xi}^M\right)\right)\\ g_T^P\left(\tilde{\theta}^P(\bar{\xi}^M,\xi_0^P)\right)\end{bmatrix}\qquad\text{and}$$

$$\sqrt{T}\begin{bmatrix}g_T^M\left(\tilde{\theta}^M(\bar{\xi}^M)\right)\\ g_T^P\left(\tilde{\theta}^P(\bar{\xi}^M,\xi_0^P)\right)\end{bmatrix}=\sqrt{T}\begin{bmatrix}\frac{1}{T}\sum_{t=1}^T u_t^M\left(\tilde{\theta}^M(\bar{\xi}^M)\right)\\ \frac{1}{T}\sum_{t=1}^T u_t^P\left(\tilde{\theta}^P(\bar{\xi}^M,\xi_0^P)\right)\end{bmatrix}\xrightarrow{d}\mathcal{N}(0,V^*)$$

Assumption 4.C.2. A multivariate central limit theorem applies such that under Assumptions 4.B.3 and 4.B.7

$$\sqrt{T}\begin{bmatrix}\frac{1}{T}\sum_{t=1}^T u_t^M\left(\tilde{\theta}^M(\bar{\xi}^M)\right)\\ \frac{1}{T}\sum_{t=1}^T u_t^P\left(\tilde{\theta}^P(\bar{\xi}^M,\xi_0^P)\right)\end{bmatrix}\xrightarrow{d}\mathcal{N}(0,V^*)$$

where

$$V^*=\Gamma_0^*+\sum_{j=1}^{\infty}\left(\Gamma_j^*+\Gamma_j^{*\top}\right)\qquad\text{and}\quad\Gamma_j^*=\mathbb{E}^*\begin{bmatrix}u_t^M\left(\tilde{\theta}^M(\bar{\xi}^M)\right)u_{t-j}^M\left(\tilde{\theta}^M(\bar{\xi}^M)\right)^\top & u_t^M\left(\tilde{\theta}^M(\bar{\xi}^M)\right)u_{t-j}^P\left(\tilde{\theta}^P(\bar{\xi}^M,\xi_0^P)\right)^\top\\ u_t^P\left(\tilde{\theta}^P(\bar{\xi}^M,\xi_0^P)\right)u_{t-j}^M\left(\tilde{\theta}^M(\bar{\xi}^M)\right)^\top & u_t^P\left(\tilde{\theta}^P(\bar{\xi}^M,\xi_0^P)\right)u_{t-j}^P\left(\tilde{\theta}^P(\bar{\xi}^M,\xi_0^P)\right)^\top\end{bmatrix}.$$

The asymptotic distribution of the instrumental parameters evaluated at the simulated data is

$$\sqrt{T}\begin{bmatrix}\tilde{\theta}_T^{M,s}(\bar{\xi}^M)-\tilde{\theta}^M(\bar{\xi}^M)\\ \tilde{\theta}_T^{P,s}(\bar{\xi}^M,\xi_0^P)-\tilde{\theta}^P(\bar{\xi}^M,\xi_0^P)\end{bmatrix}\xrightarrow{d}\mathcal{N}\left(0,\begin{bmatrix}C_0^{M*} & 0\\ 0 & C_0^{P*}\end{bmatrix}V^*\begin{bmatrix}C_0^{M*\top} & 0\\ 0 & C_0^{P*\top}\end{bmatrix}\right)$$

Under the null hypotheses of full encompassing $H_0^M: \theta_0^M = \tilde{\theta}^M\left(\bar{\xi}^M\right)$ given in Assumption 4.3.4 and $H_0^P: \theta_0^P = \tilde{\theta}^P\left(\bar{\xi}^M, \xi_0^P\right)$ given in Assumption 4.3.6

$$\sqrt{T}\begin{bmatrix} \tilde{\theta}_T^{M,s}(\bar{\xi}^M) - \theta_0^M \\ \tilde{\theta}_T^{P,s}(\bar{\xi}^M, \xi_0^P) - \theta_0^P \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{bmatrix} C_0^{M*} & 0 \\ 0 & C_0^{P*} \end{bmatrix} V^* \begin{bmatrix} C_0^{M*\top} & 0 \\ 0 & C_0^{P*\top} \end{bmatrix}\right)$$

where all the quantities in the asymptotic covariance matrix are evaluated at the simulated time series and the limit values θ_0^M and θ_0^P, i.e.

$$\begin{aligned}
C_0^{M*} &= \left\{\mathbb{E}^*\left[\frac{\partial u_t^{M\top}}{\partial \theta^M}\left(\tilde{y}^M(\bar{\xi}^M), \theta_0^M\right)\right] W^M \mathbb{E}^*\left[\frac{\partial u_t^M}{\partial \theta^{M\top}}\left(\tilde{y}^M(\bar{\xi}^M), \theta_0^M\right)\right]\right\}^{-1} \mathbb{E}^*\left[\frac{\partial u_t^{M\top}}{\partial \theta^M}\left(\tilde{y}^M(\bar{\xi}^M), \theta_0^M\right)\right] W^M \\
C_0^{P*} &= \left\{\mathbb{E}^*\left[\frac{\partial u_t^{P\top}}{\partial \theta^P}\left(\tilde{y}^P(\bar{\xi}^M, \xi_0^P), \theta_0^P\right)\right] W^P \mathbb{E}^*\left[\frac{\partial u_t^P}{\partial \theta^{P\top}}(\tilde{y}^P\left(\bar{\xi}^M, \xi_0^P\right), \theta_0^P)\right]\right\}^{-1} \mathbb{E}^*\left[\frac{\partial u_t^{P\top}}{\partial \theta^P}\left(\tilde{y}^P(\bar{\xi}^M, \xi_0^P), \theta_0^P\right)\right] W^P \\
V^* &= \Gamma_0^* + \sum_{j=1}^{\infty}\left(\Gamma_j^* + \Gamma_j^{*\top}\right) \\
\Gamma_j^* &= \mathbb{E}^*\begin{bmatrix} u_t^M\left(\tilde{y}^M(\bar{\xi}^M), \theta_0^M\right) u_{t-j}^M\left(\tilde{y}^M(\bar{\xi}^M), \theta_0^M\right)^\top & u_t^M\left(\tilde{y}^M(\bar{\xi}^M), \theta_0^M\right) u_{t-j}^P\left(\tilde{y}^P(\bar{\xi}^M, \xi_0^P), \theta_0^P\right)^\top \\ u_t^P\left(\tilde{y}^P(\bar{\xi}^M, \xi_0^P), \theta_0^P\right) u_{t-j}^M\left(\tilde{y}^M(\bar{\xi}^M), \theta_0^M\right)^\top & u_t^P\left(\tilde{y}^P(\bar{\xi}^M, \xi_0^P), \theta_0^P\right) u_{t-j}^P\left(\tilde{y}^P(\bar{\xi}^M, \xi_0^P), \theta_0^P\right)^\top \end{bmatrix}.
\end{aligned}$$

Third, we evaluate the distribution of $\sqrt{T}\begin{bmatrix} \frac{1}{S}\sum_{s=1}^S \tilde{\theta}_T^{M,s}(\bar{\xi}^M) - \theta_0^M \\ \frac{1}{S}\sum_{s=1}^S \tilde{\theta}_T^{P,s}(\bar{\xi}^M, \xi_0^P) - \theta_0^P \end{bmatrix}$.

$$\sqrt{T}\begin{bmatrix} \frac{1}{S}\sum_{s=1}^S \tilde{\theta}_T^{M,s}(\bar{\xi}^M) - \theta_0^M \\ \frac{1}{S}\sum_{s=1}^S \tilde{\theta}_T^{P,s}(\bar{\xi}^M, \xi_0^P) - \theta_0^P) \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{S}\begin{bmatrix} C_0^{M*} & 0 \\ 0 & C_0^{P*} \end{bmatrix} V^* \begin{bmatrix} C_0^{M*\top} & 0 \\ 0 & C_0^{P*\top} \end{bmatrix}\right)$$

Finally, the previous results are combined to show the distribution of $\sqrt{T}\begin{bmatrix} \hat{\theta}_T^M - \frac{1}{S}\sum_{s=1}^S \tilde{\theta}_T^{M,s}(\bar{\xi}^M) \\ \hat{\theta}_T^P - \frac{1}{S}\sum_{s=1}^S \tilde{\theta}_T^{P,s}(\bar{\xi}^M, \xi_0^P) \end{bmatrix}$.

Proposition 4.C.1 (Asymptotic distribution of the instrumental parameters in the presence of exogenous time series). Under Assumptions 4.B.1, 4.B.5, 4.C.1, 4.B.3, 4.B.7, 4.C.2 and 4.3.4 and 4.3.6, the instrumental parameters are asymptotically distributed as

$$\sqrt{T}\begin{bmatrix} \hat{\theta}_T^M - \frac{1}{S}\sum_{s=1}^S \tilde{\theta}_T^{M,s}(\bar{\xi}^M) \\ \hat{\theta}_T^P - \frac{1}{S}\sum_{s=1}^S \tilde{\theta}_T^{P,s}(\bar{\xi}^M, \xi_0^P) \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \Phi_0\left(S, W^M, W^P\right)\right)$$

where the asymptotic covariance matrix is given by

$$\begin{aligned}\Phi_0\left(S, W^M, W^P\right) = &\begin{bmatrix} C_0^M & 0 \\ 0 & C_0^P \end{bmatrix} V \begin{bmatrix} C_0^{M\top} & 0 \\ 0 & C_0^{P\top} \end{bmatrix} + \frac{1}{S}\begin{bmatrix} C_0^{M*} & 0 \\ 0 & C_0^{P*} \end{bmatrix} V^* \begin{bmatrix} C_0^{M*\top} & 0 \\ 0 & C_0^{P*\top} \end{bmatrix} - \begin{bmatrix} C_0^M & 0 \\ 0 & C_0^P \end{bmatrix} K \begin{bmatrix} C_0^{M*\top} & 0 \\ 0 & C_0^{P*\top} \end{bmatrix} \\ &- \begin{bmatrix} C_0^{M*} & 0 \\ 0 & C_0^{P*} \end{bmatrix} K^* \begin{bmatrix} C_0^{M\top} & 0 \\ 0 & C_0^{P\top} \end{bmatrix} + \left(1 - \frac{1}{S}\right)\begin{bmatrix} C_0^{M*} & 0 \\ 0 & C_0^{P*} \end{bmatrix} K^* \begin{bmatrix} C_0^{M*\top} & 0 \\ 0 & C_0^{P*\top} \end{bmatrix}\end{aligned} \tag{4.C.2}$$

$$C_0^M = \left\{\mathbb{E}\left[\frac{\partial u_t^{M\top}}{\partial \theta^M}\left(y^M, \theta_0^M\right)\right] W^M \mathbb{E}\left[\frac{\partial u_t^M}{\partial \theta^{M\top}}\left(y^M, \theta_0^M\right)\right]\right\}^{-1} \mathbb{E}\left[\frac{\partial u_t^{M\top}}{\partial \theta^M}\left(y^M, \theta_0^M\right)\right] W^M, \tag{4.C.3}$$

$$C_0^P = \left\{\mathbb{E}\left[\frac{\partial u_t^{P\top}}{\partial \theta^P}\left\{y^P, \theta_0^P\right)\right] W^P \mathbb{E}\left[\frac{\partial u_t^P}{\partial \theta^{P\top}}\left(y^P, \theta_0^P\right)\right]\right\}^{-1} \mathbb{E}\left[\frac{\partial u_t^{P\top}}{\partial \theta^P}\left(y^P, \theta_0^P\right)\right] W^P, \tag{4.C.4}$$

$$C_0^{M*} = \left\{\mathbb{E}^*\left[\frac{\partial u_t^{M\top}}{\partial \theta^M}\left(\tilde{y}^M, \theta_0^M\right)\right] W^M \mathbb{E}^*\left[\frac{\partial u_t^M}{\partial \theta^{M\top}}\left(\tilde{y}^M, \theta_0^M\right)\right]\right\}^{-1} \mathbb{E}^*\left[\frac{\partial u_t^{M\top}}{\partial \theta^M}\left(\tilde{y}^M, \theta_0^M\right)\right] W^M, \tag{4.C.5}$$

$$C_0^{P*} = \left\{\mathbb{E}^*\left[\frac{\partial u_t^{P\top}}{\partial \theta^P}\left(\tilde{y}^P, \theta_0^P\right)\right] W^P \mathbb{E}^*\left[\frac{\partial u_t^P}{\partial \theta^{P\top}}\left(\tilde{y}^P, \theta_0^P\right)\right]\right\}^{-1} \mathbb{E}^*\left[\frac{\partial u_t^{P\top}}{\partial \theta^P}\left(\tilde{y}^P, \theta_0^P\right)\right] W^P, \tag{4.C.6}$$

$$V = \Gamma_0 + \sum_{j=1}^{\infty} \left(\Gamma_j + \Gamma_j^\top \right) \quad \text{and} \quad \Gamma_j = \mathbb{E} \begin{bmatrix} u_t^M \left(y^M, \theta_0^M \right) u_{t-j}^M \left(y^M, \theta_0^M \right)^\top & u_t^M \left(y^M, \theta_0^M \right) u_{t-j}^P \left(y^P, \theta_0^P \right)^\top \\ u_t^P \left(y^P, \theta_0^P \right) u_{t-j}^M \left(y^M, \theta_0^M \right)^\top & u_t^P \left(y^P, \theta_0^P \right) u_{t-j}^P \left(y^P, \theta_0^P \right)^\top \end{bmatrix}, \tag{4.C.7}$$

$$V^* = \Gamma_0^* + \sum_{j=1}^{\infty} \left(\Gamma_j^* + \Gamma_j^{*\top} \right) \quad \text{and} \quad \Gamma_j^* = \mathbb{E}^* \begin{bmatrix} u_t^M \left(\tilde{y}^M, \theta_0^M \right) u_{t-j}^M \left(\tilde{y}^M, \theta_0^M \right)^\top & u_t^M \left(\tilde{y}^M, \theta_0^M \right) u_{t-j}^P \left(\tilde{y}^P, \theta_0^P \right)^\top \\ u_t^P \left(\tilde{y}^P, \theta_0^P \right) u_{t-j}^M \left(\tilde{y}^M, \theta_0^M \right)^\top & u_t^P \left(\tilde{y}^P, \theta_0^P \right) u_{t-j}^P \left(\tilde{y}^P, \theta_0^P \right)^\top \end{bmatrix}, \tag{4.C.8}$$

$$K = \operatorname{Acov} \left[\sqrt{T} \begin{pmatrix} \frac{1}{T} \sum_{t=1}^T u_t^M \left(y^M, \theta_0^M \right) \\ \frac{1}{T} \sum_{t=1}^T u_t^P \left(y^P, \theta_0^P \right) \end{pmatrix}, \sqrt{T} \begin{pmatrix} \frac{1}{T} \sum_{t=1}^T u_t^M \left(\tilde{y}^M, \theta_0^M \right) \\ \frac{1}{T} \sum_{t=1}^T u_t^P \left(\tilde{y}^P, \theta_0^P \right) \end{pmatrix} \right], \tag{4.C.9}$$

$$K^* = \operatorname{Acov} \left[\sqrt{T} \begin{pmatrix} \frac{1}{T} \sum_{t=1}^T u_t^M \left(\tilde{y}^M, \theta_0^M \right) \\ \frac{1}{T} \sum_{t=1}^T u_t^P \left(\tilde{y}^P, \theta_0^P \right) \end{pmatrix}, \sqrt{T} \begin{pmatrix} \frac{1}{T} \sum_{t=1}^T u_t^M \left(\tilde{y}^M, \theta_0^M \right) \\ \frac{1}{T} \sum_{t=1}^T u_t^P \left(\tilde{y}^P, \theta_0^P \right) \end{pmatrix} \right]. \tag{4.C.10}$$

where y^M and y^P are short-hand notations for $\{y_t^M\}_{t=1}^T$ and $\{y_t^P\}_{t=1}^T$, respectively, and $\tilde{y}^M$ and $\tilde{y}^P$ are a short-hand notation for $\left\{\tilde{y}_t^M(\xi^M)\right\}_{t=1}^T$ and $\left\{\tilde{y}_t^P(\xi^M, \xi^P)\right\}_{t=1}^T$, respectively.

The proposition above provides the limiting distribution of the instrumental parameters in the presence of exogenous time series that are common in the empirical and simulated time series as porposed by Gourieroux et al. (1993). As there are no exogenous time series in our context, the asymptotic covariance matrices K and K^* are equal to zero. The limiting distribution collapses to the one given in Proposition 4.3.3.

4.C.2 Asymptotic distribution of the structural parameters

Using the results from the last subsection, the structural parameters can asymptotically be written as

$$\sqrt{T}\begin{bmatrix}\hat{\xi}_T^M - \bar{\xi}^M \\ \hat{\xi}_t^P - \xi_0^P\end{bmatrix} \simeq -\begin{bmatrix}\frac{\partial\tilde{\theta}^M}{\partial\xi^M}(\bar{\xi}^M)^\top\Omega^M\frac{\partial\tilde{\theta}^M}{\partial\xi^{M\top}}(\bar{\xi}^M) & 0 \\ \frac{\partial\tilde{\theta}^P}{\partial\xi^P}(\bar{\xi}^M,\xi_0^P)^\top\Omega^P\frac{\partial\tilde{\theta}^P}{\partial\xi^{M\top}}(\bar{\xi}^M,\xi_0^P) & \frac{\partial\tilde{\theta}^P}{\partial\xi^P}(\bar{\xi}^M,\xi_0^P)^\top\Omega^P\frac{\partial\tilde{\theta}^P}{\partial\xi^{P\top}}(\bar{\xi}^M,\xi_0^P)\end{bmatrix}^{-1}$$
$$\cdot\begin{bmatrix}\frac{\partial\tilde{\theta}^M}{\partial\xi^M}(\bar{\xi}^M)^\top\Omega^M & 0 \\ 0 & \frac{\partial\tilde{\theta}^P}{\partial\xi^P}(\bar{\xi}^M,\xi_0^P)^\top\Omega^P\end{bmatrix}\sqrt{T}\begin{bmatrix}\hat{\theta}_T^M - \frac{1}{S}\sum_{s=1}^S\tilde{\theta}_T^{M,s}(\bar{\xi}^M) \\ \hat{\theta}_T^P - \frac{1}{S}\sum_{s=1}^S\tilde{\theta}_T^{P,s}(\bar{\xi}^M,\xi_0^P)\end{bmatrix}$$

We invert the upper matrix by partitioned inversion. The idea goes as follows, if we want to invert a lower triangular bock matrix $A = \begin{bmatrix}A_{11} & 0 \\ A_{21} & A_{22}\end{bmatrix}$, the inverse is given by $A^{-1} = \begin{bmatrix}A_{11}^{-1} & 0 \\ -A_{22}^{-1}A_{21}A_{11}^{-1} & A_{22}^{-1}\end{bmatrix}$ where $A_{11} = \frac{\partial\tilde{\theta}^M}{\partial\xi^M}(\bar{\xi}^M)^\top\Omega^M\frac{\partial\tilde{\theta}^M}{\partial\xi^{M\top}}(\bar{\xi}^M)$, $A_{12} = 0$, $A_{21} = \frac{\partial\tilde{\theta}^P}{\partial\xi^P}(\bar{\xi}^M,\xi_0^P)^\top\Omega^P\frac{\partial\tilde{\theta}^P}{\partial\xi^{M\top}}(\bar{\xi}^M,\xi_0^P)$, $A_{22} = \frac{\partial\tilde{\theta}^P}{\partial\xi^P}(\bar{\xi}^M,\xi_0^P)^\top\Omega^P\frac{\partial\tilde{\theta}^P}{\partial\xi^{P\top}}(\bar{\xi}^M,\xi_0^P)$.

$$
\begin{aligned}
\sqrt{T}\begin{bmatrix}\hat{\xi}_T^M-\bar{\xi}^M\\ \hat{\xi}_t^P-\xi_0^P\end{bmatrix} &\simeq \begin{bmatrix} \left\{\frac{\partial\tilde{\theta}^M}{\partial\xi^M}(\bar{\xi}^M)^\top\Omega^M\frac{\partial\tilde{\theta}^M}{\partial\xi^{M^\top}}(\bar{\xi}^M)\right\}^{-1} & 0 \\ \cdot\frac{\partial\tilde{\theta}^M}{\partial\xi^M}(\bar{\xi}^M)^\top\Omega^M & \\ -\left\{\frac{\partial\tilde{\theta}^P}{\partial\xi^P}(\bar{\xi}^M,\xi_0^P)^\top\Omega^P\frac{\partial\tilde{\theta}^P}{\partial\xi^{P^\top}}(\bar{\xi}^M,\xi_0^P)\right\}^{-1}\frac{\partial\tilde{\theta}^P}{\partial\xi^P}(\bar{\xi}^M,\xi_0^P)^\top\Omega^P & \left\{\frac{\partial\tilde{\theta}^P}{\partial\xi^P}(\bar{\xi}^M,\xi_0^P)^\top\Omega^P\frac{\partial\tilde{\theta}^P}{\partial\xi^{P^\top}}(\bar{\xi}^M,\xi_0^P)\right\}^{-1} \\ \cdot\frac{\partial\tilde{\theta}^P}{\partial\xi^{M^\top}}(\bar{\xi}^M,\xi_0^P)\cdot\left\{\frac{\partial\tilde{\theta}^M}{\partial\xi^M}(\bar{\xi}^M)^\top\frac{\partial\tilde{\theta}^M}{\partial\xi^{M^\top}}(\bar{\xi}^M)\right\}^{-1}\frac{\partial\tilde{\theta}^M}{\partial\xi^M}(\bar{\xi}^M)^\top\Omega^M & \cdot\frac{\partial\tilde{\theta}^P}{\partial\xi^P}(\bar{\xi}^M,\xi_0^P)^\top\Omega^P \end{bmatrix} \\
&\quad\cdot\sqrt{T}\begin{bmatrix}\hat{\theta}_T^M-\frac{1}{S}\sum_{s=1}^S\tilde{\theta}_T^{M,s}(\bar{\xi}^M)\\ \hat{\theta}_T^P-\frac{1}{S}\sum_{s=1}^S\tilde{\theta}_T^{P,s}(\bar{\xi}^M,\xi_0^P)\end{bmatrix} \\
&\simeq \begin{bmatrix} A\left(\Omega^M\right) & 0 \\ -B\left(\Omega^P\right)\frac{\partial\tilde{\theta}^P}{\partial\xi^{M^\top}}\left(\bar{\xi}^M,\xi_0^P\right)A\left(\Omega^M\right) & B\left(\Omega^P\right)\end{bmatrix}\cdot\sqrt{T}\begin{bmatrix}\hat{\theta}_T^M-\frac{1}{S}\sum_{s=1}^S\tilde{\theta}_T^{M,s}(\bar{\xi}^M)\\ \hat{\theta}_T^P-\frac{1}{S}\sum_{s=1}^S\tilde{\theta}_T^{P,s}(\bar{\xi}^M,\xi_0^P)\end{bmatrix} \\
&\simeq \begin{bmatrix} A\left(\Omega^M\right) & 0 \\ B\left(\Omega^P\right)C\left(\Omega^M\right) & B\left(\Omega^P\right)\end{bmatrix}\cdot\sqrt{T}\begin{bmatrix}\hat{\theta}_T^M-\frac{1}{S}\sum_{s=1}^S\tilde{\theta}_T^{M,s}(\bar{\xi}^M)\\ \hat{\theta}_T^P-\frac{1}{S}\sum_{s=1}^S\tilde{\theta}_T^{P,s}(\bar{\xi}^M,\xi_0^P)\end{bmatrix}
\end{aligned}
$$

where

$$
\begin{aligned}
A\left(\Omega^M\right) &= \left\{\frac{\partial\tilde{\theta}^M}{\partial\xi^M}\left(\bar{\xi}^M\right)^\top\Omega^M\frac{\partial\tilde{\theta}^M}{\partial\xi^{M^\top}}\left(\bar{\xi}^M\right)\right\}^{-1}\frac{\partial\tilde{\theta}^M}{\partial\xi^M}\left(\hat{\xi}_T^M\right)^\top\Omega^M \\
B\left(\Omega^P\right) &= \left\{\frac{\partial\tilde{\theta}^P}{\partial\xi^P}\left(\bar{\xi}^M,\xi_0^P\right)^\top\Omega^P\frac{\partial\tilde{\theta}^P}{\partial\xi^{P^\top}}\left(\bar{\xi}^M,\xi_0^P\right)\right\}^{-1}\cdot\frac{\partial\tilde{\theta}^P}{\partial\xi^P}\left(\hat{\xi}_T^M,\hat{\xi}_T^P\right)^\top\Omega^P \\
C\left(\Omega^M\right) &= -\frac{\partial\tilde{\theta}^P}{\partial\xi^{M^\top}}\left(\bar{\xi}^M,\xi_0^P\right)A\left(\Omega^M\right)
\end{aligned}
$$

Hence, the two-step partial indirect inference is distributed asymptotically normal

$$\sqrt{T}\begin{bmatrix}\hat{\xi}_T^M - \bar{\xi}^M \\ \hat{\xi}_t^P - \xi_0^P\end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \Sigma\left(S, \Omega^M, \Omega^P\right)\right)$$

where

$$\Sigma\left(S, \Omega^M, \Omega^P\right) = \begin{bmatrix} A\left(\Omega^M\right) & 0 \\ B\left(\Omega^P\right) C\left(\Omega^M\right) & B\left(\Omega^P\right)\end{bmatrix} \Phi_0\left(S, W^M, W^P\right) \begin{bmatrix} A\left(\Omega^M\right)^\top & C\left(\Omega^M\right)^\top B\left(\Omega^P\right)^\top \\ 0 & B\left(\Omega^P\right)^\top\end{bmatrix}$$

$$\Phi_0\left(S, W^M, W^P\right) = \begin{bmatrix} C_0^M & 0 \\ 0 & C_0^P\end{bmatrix} V \begin{bmatrix} C_0^{M\top} & 0 \\ 0 & C_0^{P\top}\end{bmatrix} + \frac{1}{S}\begin{bmatrix} C_0^{M*} & 0 \\ 0 & C_0^{P*}\end{bmatrix} V^* \begin{bmatrix} C_0^{M*\top} & 0 \\ 0 & C_0^{P*\top}\end{bmatrix}.$$

This concludes the derivation of Proposition 4.3.4. It is given with all definitions in compact form below. Under Assumptions 4.3.1 to 4.3.6 and 4.C.1 and 4.C.2, the two-step partial indirect inference estimator is asymptotically distributed as

$$\sqrt{T}\begin{bmatrix}\hat{\xi}_T^M - \bar{\xi}^M \\ \hat{\xi}_T^P - \xi_0^P\end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \Sigma\left(S, \Omega^M, \Omega^P\right)\right)$$

$$\begin{aligned}\Sigma\left(S, \Omega^M, \Omega^P\right) = &\begin{bmatrix} A\left(\Omega^M\right) C_0^M & 0 \\ B\left(\Omega^P\right) C\left(\Omega^M\right) C_0^M & B\left(\Omega^P\right) C_0^P\end{bmatrix} \begin{bmatrix} V_M & V_{MP} \\ V_{PM} & V_P\end{bmatrix} \begin{bmatrix} C_0^{M\top} A\left(\Omega^M\right)^\top & C_0^{M\top} C\left(\Omega^M\right)^\top B\left(\Omega^P\right)^\top \\ 0 & C_0^{P\top} B\left(\Omega^P\right)^\top\end{bmatrix} \\ &+ \frac{1}{S}\begin{bmatrix} A\left(\Omega^M\right) C_0^{M*} & 0 \\ B\left(\Omega^P\right) C\left(\Omega^M\right) C_0^{M*} & B\left(\Omega^P\right) C_0^{P*}\end{bmatrix} \begin{bmatrix} V_M^* & V_{MP}^* \\ V_{PM}^* & V_P^*\end{bmatrix} \begin{bmatrix} C_0^{M*\top} A\left(\Omega^M\right)^\top & C_0^{M*\top} C\left(\Omega^M\right)^\top B\left(\Omega^P\right)^\top \\ 0 & C_0^{P*\top} B\left(\Omega^P\right)^\top\end{bmatrix}\end{aligned}$$

$$
\begin{aligned}
\text{with} \quad A\left(\Omega^M\right) &= \left\{\frac{\partial \tilde{\theta}^M}{\partial \xi^M}\left(\bar{\xi}^M\right)^\top \Omega^M \frac{\partial \tilde{\theta}^M}{\partial \xi^{M\top}}\left(\bar{\xi}^M\right)\right\}^{-1} \frac{\partial \tilde{\theta}^M}{\partial \xi^M}\left(\hat{\xi}_T^M\right)^\top \Omega^M \\
B\left(\Omega^P\right) &= \left\{\frac{\partial \tilde{\theta}^P}{\partial \xi^P}\left(\bar{\xi}^M, \xi_0^P\right)^\top \Omega^P \frac{\partial \tilde{\theta}^P}{\partial \xi^{P\top}}\left(\bar{\xi}^M, \xi_0^P\right)\right\}^{-1} \cdot \frac{\partial \tilde{\theta}^P}{\partial \xi^P}\left(\hat{\xi}_T^M, \hat{\xi}_T^P\right)^\top \Omega^P \\
C\left(\Omega^M\right) &= -\frac{\partial \tilde{\theta}^P}{\partial \xi^{M\top}}\left(\bar{\xi}^M, \xi_0^P\right) A\left(\Omega^M\right)
\end{aligned}
$$

$$
\begin{aligned}
\text{and} \quad C_0^M &= \left\{\mathbb{E}\left[\frac{\partial u_t^M}{\partial \theta^M}^\top\left(y^M, \theta_0^M\right)\right] W^M \mathbb{E}\left[\frac{\partial u_t^M}{\partial \theta^{M\top}}\left(y^M, \theta_0^M\right)\right]\right\}^{-1} \mathbb{E}\left[\frac{\partial u_t^M}{\partial \theta^M}^\top\left(y^M, \theta_0^M\right)\right] W^M \\
C_0^P &= \left\{\mathbb{E}\left[\frac{\partial u_t^P}{\partial \theta^P}^\top\left(y^P, \theta_0^P\right)\right] W^P \mathbb{E}\left[\frac{\partial u_t^P}{\partial \theta^{P\top}}\left(y^P, \theta_0^P\right)\right]\right\}^{-1} \mathbb{E}\left[\frac{\partial u_t^P}{\partial \theta^P}^\top\left(y^P, \theta_0^P\right)\right] W^P \\
C_0^{M^*} &= \left\{\mathbb{E}^*\left[\frac{\partial u_t^M}{\partial \theta^M}^\top\left(\tilde{y}^M, \theta_0^M\right)\right] W^M \mathbb{E}^*\left[\frac{\partial u_t^M}{\partial \theta^{M\top}}\left(\tilde{y}^M, \theta_0^M\right)\right]\right\}^{-1} \mathbb{E}^*\left[\frac{\partial u_t^M}{\partial \theta^M}^\top\left(\tilde{y}^M, \theta_0^M\right)\right] W^M \\
C_0^{P^*} &= \left\{\mathbb{E}^*\left[\frac{\partial u_t^P}{\partial \theta^P}^\top\left(\tilde{y}^P, \theta_0^P\right)\right] W^P \mathbb{E}^*\left[\frac{\partial u_t^P}{\partial \theta^{P\top}}\left(\tilde{y}^P, \theta_0^P\right)\right]\right\}^{-1} \mathbb{E}^*\left[\frac{\partial u_t^P}{\partial \theta^P}^\top\left(\tilde{y}^P, \theta_0^P\right)\right] W^P \\
V &= \Gamma_0 + \sum_{j=1}^{\infty}\left(\Gamma_j + \Gamma_j^\top\right) \\
\Gamma_j &= \mathbb{E}\begin{bmatrix} u_t^M\left(y^M, \theta_0^M\right) u_{t-j}^M\left(y^M, \theta_0^M\right)^\top & u_t^M\left(y^M, \theta_0^M\right) u_{t-j}^P\left(y^P, \theta_0^P\right)^\top \\ u_t^P\left(y^P, \theta_0^P\right) u_{t-j}^M\left(y^M, \theta_0^M\right)^\top & u_t^P\left(y^P, \theta_0^P\right) u_{t-j}^P\left(y^P, \theta_0^P\right)^\top \end{bmatrix}
\end{aligned}
$$

$$V^* = \Gamma_0^* + \sum_{j=1}^{\infty} \left(\Gamma_j^* + \Gamma_j^{*\top} \right)$$

$$\Gamma_j^* = \mathbb{E}^* \begin{bmatrix} u_t^M \left(\tilde{y}^M, \theta_0^M \right) u_{t-j}^M \left(\tilde{y}^M, \theta_0^M \right)^\top & u_t^M \left(\tilde{y}^M, \theta_0^M \right) u_{t-j}^P \left(\tilde{y}^P, \theta_0^P \right)^\top \\ u_t^P \left(\tilde{y}^P, \theta_0^P \right) u_{t-j}^M \left(\tilde{y}^M, \theta_0^M \right)^\top & u_t^P \left(\tilde{y}^P, \theta_0^P \right) u_{t-j}^P \left(\tilde{y}^P, \theta_0^P \right)^\top \end{bmatrix},$$

where y^M and y^P are short-hand notations for $\{y_t^M\}_{t=1}^T$ and $\{y_t^P\}_{t=1}^T$, respectively, and $\tilde{y}^M$ and $\tilde{y}^P$ are a short-hand notation for $\{\tilde{y}_t^M(\xi^M)\}_{t=1}^T$ and $\{\tilde{y}_t^P(\xi^M, \xi^P)\}_{t=1}^T$, respectively.

4.C.3 Comparison to the asymptotic distribution of Grammig and Küchlin (2018)

The main difference to Grammig and Küchlin (2018) is that Assumption 4.3.2 is perceived to describe both the data generating processes of the empirical data and the simulated time series from the structural model appropriately. This implies that the model is correctly specified and that the simulated time series are conceived as a realization of the same data generating process, i.e. that $y_t^M = \tilde{y}_t^M(\xi^M)$ and $y_t^P = \tilde{y}_t^P(\xi^M, \xi^P)$. This, in turn, has consequences on the limiting distribution of the instrumental parameters. This limiting distribution is given by

$$\sqrt{T}\begin{bmatrix} \hat{\theta}_T^M - \frac{1}{S}\sum_{s=1}^S \tilde{\theta}_T^{M,s}(\xi_0^M) \\ \hat{\theta}_T^P - \frac{1}{S}\sum_{s=1}^S \tilde{\theta}_T^{P,s}(\xi_0^M, \xi_0^P) \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \Phi_0^{GK}\left(S, W^M, W^P\right)\right)$$

where the asymptotic covariance matrix is given by

$$\Phi_0^{GK}\left(S, W^M, W^P\right) = \left(1 + \frac{1}{S}\right)\begin{bmatrix} C_0^M & 0 \\ 0 & C_0^P \end{bmatrix} V \begin{bmatrix} C_0^{M\top} & 0 \\ 0 & C_0^{P\top} \end{bmatrix}$$

where C_0^M, C_0^M and V are defined as above. This also implies that the probability limit associated with the empirical and simulated time series is the same. Hence in case of correct specification, the quantities for the empirical and simulated time series coincide, i.e. $C_0^M = C_0^{P^*}$, $C_0^P = C_0^{P^*}$ and $V = V^*$.

The structure of the limiting distribution of the structural parameters is similar in that the only difference is the limiting distribution of the instrumental parameters detailed above. Their limiting distribution is given by

$$\sqrt{T}\begin{bmatrix} \hat{\xi}_T^M - \xi_0^M \\ \hat{\xi}_T^P - \xi_0^P \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \Sigma^{GK}\left(S, \Omega^M, \Omega^P\right)\right)$$

where

$$\begin{aligned}
&\Sigma^{GK}\left(S, \Omega^M, \Omega^P\right) \\
&\quad = \left(1 + \frac{1}{S}\right)\begin{bmatrix} A\left(\Omega^M\right) C_0^M & 0 \\ B\left(\Omega^P\right) C\left(\Omega^M\right) C_0^M & B\left(\Omega^P\right) C_0^P \end{bmatrix} V \begin{bmatrix} C_0^{M\top} A\left(\Omega^M\right)^\top & C_0^{M\top} C\left(\Omega^M\right)^\top B\left(\Omega^P\right)^\top \\ 0 & C_0^{P\top} B\left(\Omega^P\right)^\top \end{bmatrix}
\end{aligned}$$

where the associated quantities are defined as in our case.

4.C.4 Estimation of the asymptotic covariance matrix

For the calculation of $\Sigma\left(S, \Omega^M, \Omega^P\right)$, we rely on Equations (4.3.22) to (4.3.25) in Section 4.3.2 and Equations (4.C.3) to (4.C.8) in Appendix 4.C.

The first building block is calculating V and V^*, the variance-covariance matrices of the moment conditions evaluated using the empirical data and the model simulated data described in Equations (4.C.7) and (4.C.8), respectively. These are computed using the Newey and West (1987) HAC estimator to account for serial correlation, up to the 10th order, between the moment conditions.

We estimate C_0^M and C_0^P shown in Equations (4.C.3) and (4.C.4) and $C_0^{M^*}$ and $C_0^{P^*}$ given in Equations (4.C.5) and (4.C.6) for the empirical and simulated data, respectively. Towards this purpose, we use numerical derivatives of the moment conditions with respect to the respective instrumental parameters. In our implementation, W^M and W^P are identity matrices. As 1000 ensembles are simulated from the respective partial equilibrium asset pricing model for each parameter value, the second part of Equation (4.3.20) is evaluated 1000 times for each of the ensembles, and we use average over the ensembles to obtain estimates of $C_0^{M^*}$, $C_0^{P^*}$ and V^*. We use the estimated quantities from above to compute $\Phi_0\left(S, W^M, W^P\right)$.

To calculate $\Sigma\left(S, \Omega^M, \Omega^P\right)$ given in Equation (4.3.22), we calculate the numerical derivatives of the respective moment conditions with respect to the structural parameters in question as in Equations (4.3.23), (4.3.24) and (4.3.25) over all simulated ensembles. As for the instrumental parameters, we specify the weighting matrices, Ω^M and Ω^P, as identity matrices to assign identical weights to the sensitivity of each instrumental parameter to changes in each of the structural parameters.

Appendix 4.D Moment conditions of the macroeconomic parameters of the LRR model

In this section, we elaborate on the first step moment conditions for the LRR model. These are given by

$$\mathbb{E}\left[u_t^M\left(\{y_t^M\}_{t=1}^T,\theta^M\right)\right]=\mathbb{E}\begin{bmatrix}\begin{bmatrix}\zeta_{1,t}\\ \zeta_{2,t}\end{bmatrix}\otimes\begin{bmatrix}1\\ g_{t-1}\\ \vdots\\ g_{t-\tau}\\ g_{d,t-1}\\ \vdots\\ g_{d,t-\tau}\\ g_{t-1}^{f(h_1)}\\ g_{d,t-1}^{f(h_1)}\\ g_{t-1}^{f(h_2)}\\ g_{d,t-1}^{f(h_2)}\end{bmatrix}\\ (\zeta_{1,t})^2-(\sigma_{\zeta_1})^2\\ (\zeta_{2,t})^2-(\sigma_{\zeta_2})^2\\ \zeta_{1,t}\cdot\zeta_{2,t}-\sigma_{\zeta_1\zeta_2}\\ g_t-\mu_c\\ g_{d,t}-\mu_d\\ g_t^{f(h_1)}-\mu_c^{f(h_1)}\\ g_{d,t}^{f(h_1)}-\mu_d^{f(h_1)}\\ g_t^{f(h_2)}-\mu_c^{f(h_2)}\\ g_{d,t}^{f(h_2)}-\mu_d^{f(h_2)}\\ (g_t)^2-(\mu_c)^2-(\sigma_c)^2\\ (g_{d,t})^2-(\mu_d)^2-(\sigma_d)^2\\ (g_t^{f(h_1)})^2-(\mu_c^{f(h_1)})^2-(\sigma_c^{f(h_1)})^2\\ (g_{d,t}^{f(h_1)})^2-(\mu_d^{f(h_1)})^2-(\sigma_d^{f(h_1)})^2\\ (g_t^{f(h_2)})^2-(\mu_c^{f(h_2)})^2-(\sigma_c^{f(h_2)})^2\\ (g_{d,t}^{f(h_2)})^2-(\mu_d^{f(h_2)})^2-(\sigma_d^{f(h_2)})^2\end{bmatrix}, \tag{4.D.1}$$

where

$$\begin{bmatrix}\zeta_{1,t}\\ \zeta_{2,t}\end{bmatrix}=\begin{bmatrix}g_t\\ g_{d,t}\end{bmatrix}-\begin{bmatrix}\kappa_1\\ \kappa_2\end{bmatrix}-\sum_{i=1}^{\tau}\Phi_i L^i\begin{bmatrix}g_t\\ g_{d,t}\end{bmatrix}-\Phi_{\tau+1}\begin{bmatrix}g_t^{f(h_1)}\\ g_{d,t}^{f(h_1)}\end{bmatrix}-\Phi_{\tau+2}\begin{bmatrix}g_t^{f(h_2)}\\ g_{d,t}^{f(h_2)}\end{bmatrix}. \tag{4.D.2}$$

We follow Grammig and Küchlin (2018) and use six lags in the HAR model. Moreover, the time aggregation of the consumption and dividend growth series is conducted as proposed by Calvet and Czellar (2015). Details can be found in the Appendix of Grammig and Küchlin (2018).

Appendix 4.E Data

In this section, we briefly describe the data used for the estimation of the habit persistence and the long-run risk model. To be consistent with the calibrated values and estimation results reported in Campbell and Cochrane (1999), Bansal and Yaron (2004) and Grammig and Küchlin (2018), we do not extend the sample period but stick with data that is more compatible with the data used in these papers.

For the habit persistence model, the data set is obtained from Beeler and Campbell (2012) and spans a period from 1930 to 2008. Descriptive statistics of the data set used for the habit persistence model are provided in Table 4.E.1. For the long-run risk model, we follow Grammig and Küchlin (2018) and use quarterly data from 1947Q2 to 2014Q4. Table 4.E.2 provides descriptive statistics for the quarterly data. Beeler and Campbell provide consumption growth and stock market data, i.e. aggregate returns and the price-dividend ratio. They obtain the consumption data from the Bureau of Economic Analysis and the stock market data from the Center for Research in Security Prices (CRSP). Furthermore, the Consumer Price Index (CPI) is retrieved from the Bureau of Labor Statistics. The base frequency in the habit persistence model is annually, whereas that of the LRR risk model is quarterly.

Consumption is defined as U.S. real durables and service consumption per capita. Using the stock market data obtained from CRSP, Beeler and Campbell construct a monthly price index for the market and calculate monthly dividends. Annual aggregate returns are formed by calculating monthly returns of the market index, taking logs and summing the monthly returns to quarterly or annual frequency. To obtain the risk-free rate proxy, we follow Beeler and Campbell (2012) and use the three months nominal T-bill yield from CRSP. The ex ante risk-free rate is approximated by a forecast of the ex post real rate using the quarterly T-bill yield and the average of quarterly log inflation over the past 12 months as predictors. To determine the annual price-dividend ratio, we obtain the end of year dividend by the sum of the within year dividends. The annual price-dividend ratio is given by the price in the last month of the year divided by the sum of the dividends paid over the previous 12 months. For the quarterly data, we follow Grammig and Küchlin (2018) and Hasseltoft (2012) and take the average of the current log dividend growth and the previous three quarters to achieve as smoothed dividend growth series.

Table 4.E.1: Descriptive statistics for the habit persistence model

	Beeler & Campbell (2012)	Campbell & Cochrane (1999)
period	1930 - 2008	1929 - 1998
	Consumption growth	
g^a (%)	1.93	1.89
g^m	0.0016	0.0016
σ_v^a (%)	2.16	1.22
σ_v^m	0.0077	0.0043
	Stock market data	
$\mathbb{E}(R^e)$ (%)	4.91	6.69
$\sigma(R^e)$ (%)	20.46	15.70
Sharpe ratio	0.24	0.43
$\mathbb{E}(R^f)$ (%)	0.56	0.94
$\exp[\mathbb{E}(p-d)]$	28.86	24.70
$\sigma(p-d)$	0.45	0.26

Notes: In this table, we report the descriptive statistics for the data used to estimate the habit persistence model. We provide the mean and standard deviation of (log) annual consumption growth and the expected excess return of the aggregrate market return along with other asset market observations. We describe the Beeler and Campbell and the Campbell and Cochrane (1999) data set in the first and second column, respectively.

Table 4.E.2: Descriptive statistics for the long-run risk model

	Grammig & Küchlin (2018)	
period	1947 Q2 - 2014 Q4	
Consumption and dividend growth		
	mean	std. dev.
g	0.0048	0.0051
g_d	0.0066	0.0247
Stock market data		
	mean	std. dev.
r_m	0.0176	0.0825
r_f	0.0017	0.0045
z_m	3.4979	0.4217

Notes: We report the descriptive statistics used to estimate the long-run risk model taken from Grammig and Küchlin (2018). We provide the mean and standard deviation of (log) quarterly consumption, g_t and dividend growth, $g_{d,t}$, and the log market return, r_m, the log risk-free rate, r_f, and the log price-dividend ratio, z_m.

Complete References

Abel, A. B. (1999). "Risk premia and term premia in general equilibrium". *Journal of Monetary Economics* 43. (1), 3–33.

Ahn, S. C. and A. R. Horenstein (2013). "Eigenvalue ratio test for the number of factors". *Econometrica* 81. (3), 1203–1227.

Aldrich, E. M. and A. R. Gallant (2011). "Habit, long-run risks, prospect? A statistical inquiry". *Journal of Financial Econometrics* 9, 589–618.

Altig, D., L. J. Christiano, M. Eichenbaum, and J. Linde (2011). "Firm-specific capital, nominal rigidities and the business cycle". *Review of Economic Dynamics* 14. (2), 225–247.

Altug, S. (1989). "Time-to-build and aggregate fluctuations: some new evidence". *International Economic Review* 30. (4), 889–920.

An, S. and F. Schorfheide (2007). "Bayesian analysis of DSGE models". *Econometric Reviews* 26. (2-4), 113–172.

Andrews, D. W. K. (1999). "Estimation when a parameter is on a boundary". *Econometrica* 67. (6), 1341–1383.

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). "Coherent measures of risk". *Mathematical Finance* 9. (3), 203–228.

Bai, J. and K. Li (Feb. 2012). "Statistical analysis of factor models of high dimension". *The Annals of Statistics* 40. (1), 436–465.

— (2016). "Maximum likelihood estimation and inference for approximate factor models of high dimension". *Review of Economics and Statistics* 98. (2), 298–309.

Bai, J., K. Li, and L. Lu (2016). "Estimation and inference of FAVAR models". *Journal of Business & Economic Statistics* 34. (4), 620–641.

Bai, J. and Y. Liao (2016). "Efficient estimation of approximate factor models via penalized maximum likelihood". *Journal of Econometrics* 191. (1), 1–18.

Bai, J. and S. Ng (2002). "Determining the number of factors in approximate factor models". *Econometrica* 70. (1), 191–221.

— (2006). "Confidence intervals for diffusion index forecasts and inference for factor-Augmented regressions". *Econometrica* 74. (4), 1133–1150.

— (2007). "Determining the number of primitive shocks in factor models". *Journal of Business & Economic Statistics* 25. (1), 52–60.

— (2013). "Principal components estimation and identification of static factors". *Journal of Econometrics* 176. (1), 18–29.

Bai, J. and P. Wang (2014). "Identification theory for high dimensional static and dynamic factor models". *Journal of Econometrics* 178. (2), 794–804.

Bansal, R., A. R. Gallant, and G. Tauchen (2007). "Rational pessimism, rational exuberance, and asset pricing models". *The Review of Economic Studies* 74, 1005–1033.

Bansal, R., D. Kiku, and A. Yaron (2012). "An empirical evaluation of the long-run risks model for asset prices". *Critical Finance Review* 1, 183–221.

Bansal, R. and A. Yaron (2004). "Risks for the long run: A potential resolution of asset pricing puzzles". *The Journal of Finance* 59, 1481–1509.

Barendse, S. (2020). "Efficiently weighted estimation of tail and interquartile expectations". Working Paper, available at https://drive.google.com/file/d/1nI0QAWbM_VchAZDVg79p2vJcKoCrQB8o/view.

Basel Committee (2013). *Fundamental review of the trading book: A revised market risk framework*. Tech. rep. Available at http://www.bis.org/publ/bcbs265.pdf. Bank for International Settlements.

— (2016). *Minimum capital requirements for Market Risk*. Tech. rep. Available at http://www.bis.org/bcbs/publ/d352.pdf. Bank for International Settlements.

— (2017). *Pillar 3 disclosure requirements – consolidated and enhanced framework*. Tech. rep. Available at http://www.bis.org/bcbs/publ/d400.pdf. Basel Committee on Banking Supervision.

Bayer, S. and T. Dimitriadis (2020). "Regression based Expected Shortfall backtesting". *Journal of Financial Econometrics (forthcoming)*. available at https://arxiv.org/abs/1801.04112, doi: 10.1093/jjfinec/nbaa013.

Beeler, J. and J. Y. Campbell (2012). "The long-run risks model and aggregate asset prices: An empirical assessment". *Critical Finance Review* 1, 141–182.

Bekker, P. A. (1986). "A note on the identification of restricted factor loading matrices". *Psychometrika* 51. (4), 607–611.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005). "Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach". *The Quarterly Journal of Economics* 120. (1), 387–422.

Bien, J. and R. J. Tibshirani (2011). "Sparse estimation of a covariance matrix". *Biometrika* 98. (4), 807–820.

Boivin, J. and M. P. Giannoni (2007). "Global forces and monetary policy effectiveness". In: *International Dimensions of Monetary Policy*. University of Chicago Press, 429–478.

Boivin, J., M. P. Giannoni, and I. Mihov (2009). "Sticky prices and monetary policy: Evidence from disaggregated US data". *American Economic Review* 99. (1), 350–384.

Bollerslev, T. (1986). "Generalized autoregressive conditional heteroskedasticity". *Journal of Econometrics* 31. (3), 307–327.

Burnside, C., M. Eichenbaum, and S. Rebelo (1993). "Labor hoarding and the business cycle". *The Journal of Political Economy* 101. (2), 245–273.

Calvet, L. E. and V. Czellar (2015). "Through the looking glass: Indirect inference via simple equilibria". *Journal of Econometrics* 185. (2), 343–358.

Campbell, J. Y. (2003). "Consumption-based asset pricing". *Handbook of the Economics of Finance* 1, 803–887.

Campbell, J. Y. and J. H. Cochrane (1999). "By force of habit: A consumption-based explanation of aggregate stock market behavior". *The Journal of Political Economy* 107, 205–251.

Campbell, J. Y. and R. J. Shiller (1988). "The dividend-price ratio and expectations of future dividends and discount factors". *The Review of Financial Studies* 1. (3), 195–228.

Caner, M. and X. Han (2014). "Selecting the correct number of factors in approximate factor models: The large panel case with group bridge estimators". *Journal of Business & Economic Statistics* 32. (3), 359–374.

Chamberlain, G. and M. Rothschild (1983). "Arbitrage, factor structure, and mean-variance analysis on large asset markets". *Econometrica* 51. (5), 1281–1304.

Chen, A. Y. (2017). "External habit in a production economy: A model of asset prices and consumption volatility risk". *The Review of Financial Studies* 30. (8), 2890–2932.

Chong, Y. Y. and D. Hendry (1986). "Econometric evaluation of linear macroeconomic models". *Review of Economic Studies* 53. (4), 671–690.

Christiano, L. J. and M. Eichenbaum (1992). "Current real-business-cycle theories and aggregate labor-market fluctuations". *The American Economic Review* 82. (3), 430–450.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (2005). "Nominal rigidities and the dynamic effects of a shock to monetary policy". *The Journal of Political Economy* 113. (1), 1–45.

Clark, T. E. and M. W. McCracken (2001). "Tests of equal forecast accuracy and encompassing for nested models". *Journal of Econometrics* 105. (1), 85–110.

Clements, M. and D. Harvey (2009). "Forecast combination and encompassing". In: *Palgrave Handbook of Econometrics: Volume 2: Applied Econometrics*. Ed. by Mills, T. C. and Patterson, K. London: Palgrave Macmillan UK, 169–198.

— (2010). "Forecast encompassing tests and probability forecasts". *Journal of Applied Econometrics* 25. (6), 1028–1062.

Cochrane, J. H. (2009). *Asset Pricing*. Princeton University Press.

— (2016). *The Habit Habit*. Economics Working Papers 16105. Hoover Institution, Stanford University.

— (2017). "Macro-finance". *Review of Finance* 21. (3), 945–985.

Constantinides, G. M. and A. Ghosh (2011). "Asset pricing tests with long-run risks in consumption growth". *The Review of Asset Pricing Studies* 1. (1), 96–136.

Cont, R., R. Deguest, and G. Scandolo (2010). "Robustness and sensitivity analysis of risk measurement procedures". *Quantitative Finance* 10. (6), 593–606.

Corsi, F. (2009). "A simple approximate long-memory model of realized volatility". *Journal of Financial Econometrics* 7. (2), 174–196.

Creal, D., S. J. Koopman, and A. Lucas (2013). "Generalized autoregressive score models with applications". *Journal of Applied Econometrics* 28. (5), 777–795.

Daniele, M., W. Pohlmeier, and A. Zagidullina (2018). *Sparse approximate factor estimation for high-dimensional covariance matrices*. Working Paper. University of Konstanz.

Danielsson, J., P. Embrechts, C. Goodhart, C. Keating, F. Muennich, O. Renault, and H. S. Shin (2001). "An academic response to Basel II". FMG Special Papers. Financial Markets Group Special Papers, available at https://EconPapers.repec.org/RePEc:fmg:fmgsps:sp130.

Diebold, F. X. and R. S. Mariano (1995). "Comparing predictive accuracy". *Journal of Business & Economic Statistics* 13. (3), 253–63.

Diebold, F. X. (1989). "Forecast combination and encompassing: Reconciling two divergent literatures". *International Journal of Forecasting* 5. (4), 589 –592.

Dimitriadis, T. and S. Bayer (2019). "A joint quantile and expected shortfall regression framework". *Electron. J. Statist.* 13. (1), 1823–1871.

Dimitriadis, T. and J. Schnaitmann (2020). "Forecast Encompassing Tests for the Expected Shortfall". *International Journal of Forecasting (forthcoming)*.

Doz, C., D. Giannone, and L. Reichlin (2011). "A two-step estimator for large approximate dynamic factor models based on Kalman filtering". *Journal of Econometrics* 164. (1), 188–205.

Dridi, R., A. Guay, and E. Renault (2007). "Indirect inference and calibration of dynamic stochastic general equilibrium models". *Journal of Econometrics* 136. (2), 397–430.

Dridi, R. and E. Renault (2000). "Semi-parametric indirect inference". *LSE STICERD Research Paper No. EM392*.

Duffie, D. and K. J. Singleton (1993). "Simulated moments estimation of Markov models of asset prices". *Econometrica* 61. (4), 929–952.

Elliott, G., I. Komunjer, and A. Timmermann (2005). "Estimation and testing of forecast rationality under flexible loss". *The Review of Economic Studies* 72. (4), 1107–1125.

Embrechts, P., H. Liu, and R. Wang (2018). "Quantile-based risk sharing". *Operations Research* 66. (4), 936–949.

Engle, R. and S. Manganelli (2004). "CAViaR: Conditional autoregressive value at risk by regression quantiles". *Journal of Business & Economic Statistics* 22. (4), 367–381.

Epstein, L. G. and S. E. Zin (1989). "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework". *Econometrica* 57. (4), 937–969.

Ericsson, N. R. (1993). "On the limitations of comparing mean square forecast errors: Clarifications and extensions". *Journal of Forecasting* 12. (8), 644–651.

Fan, J., Y. Liao, and M. Mincheva (2011). "High dimensional covariance matrix estimation in approximate factor models". *Annals of Statistics* 39. (6), 3320–3356.

— (2013). "Large covariance estimation by thresholding principal orthogonal complements". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75. (4), 603–680.

Fernald, J. (2014). "A quarterly, utilization-adjusted series on total factor productivity". In: Federal Reserve Bank of San Francisco.

Fernández-Villaverde, J. (2010). "The econometrics of DSGE models". *SERIEs* 1. (1-2), 3–49.

Fernández-Villaverde, J., J. F. Rubio-Ramírez, and F. Schorfheide (2016). "Solution and estimation methods for DSGE models". *Handbook of Macroeconomics* 2, 527–724.

Fissler, T. and J. F. Ziegel (2016). "Higher order elicitability and Osband's principle". *Annals of Statistics* 44. (4), 1680–1707.

— (2019). "Elicitability of range value at risk". arXiv:1902.04489 [math.ST].

Fissler, T., J. F. Ziegel, and T. Gneiting (2016). "Expected Shortfall is jointly elicitable with Value at Risk - Implications for backtesting". *Risk* January, 58–61.

Forni, M. and L. Gambetti (2010). "The dynamic effects of monetary policy: A structural factor model approach". *Journal of Monetary Economics* 57. (2), 203–216.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). "The generalized dynamic-factor model: Identification and estimation". *Review of Economics and Statistics* 82. (4), 540–554.

Gallant, A. R. and R. E. McCulloch (2009). "On the determination of general scientific models with application to asset pricing". *Journal of the American Statistical Association* 104. (485), 117–131.

Gallant, A. R. and G. Tauchen (1996). "Which moments to match?" *Econometric Theory* 12, 657–681.

Geweke, J. (1977). "The dynamic factor analysis of economic time series". *Latent variables in socio-economic models*.

Giacomini, R. and I. Komunjer (2005). "Evaluation and combination of conditional quantile forecasts". *Journal of Business & Economic Statistics* 23. (4), 416–431.

Giacomini, R. and H. White (2006). "Tests of conditional predictive ability". *Econometrica* 74. (6), 1545–1578.

Gikhman, I. I. and A. V. Skorokhod (2004). *The theory of stochastic processes I*. Vol. 210. Classics in Mathematics. Springer Berlin Heidelberg.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). "On the relation between the expected value and the volatility of the nominal excess return on stocks". *The Journal of Finance* 48. (5), 1779–1801.

Gneiting, T. (2011). "Making and evaluating point forecasts". *Journal of the American Statistical Association* 106. (494), 746–762.

Gonçalves, S. and B. Perron (2014). "Bootstrapping factor-augmented regression models". *Journal of Econometrics* 182. (1), 156–173.

Gourieroux, C., A. Monfort, and E. Renault (1993). "Indirect inference". *Journal of Applied Econometrics* 8, 85–118.

Grammig, J. and E.-M. Küchlin (2018). "A two-step indirect inference approach to estimate the long-run risk asset pricing model". *Journal of Econometrics* 205. (1), 6–33.

Guvenen, F. (2009). "A parsimonious macroeconomic model for asset pricing". *Econometrica* 77. (6), 1711–1750.

Halbleib, R. and W. Pohlmeier (2012). "Improving the value at risk forecasts: Theory and evidence from the financial crisis". *Journal of Economic Dynamics and Control* 36. (8), 1212–1228.

Hall, A. R. and A. Inoue (2003). "The large sample behaviour of the generalized method of moments estimator in misspecified models". *Journal of Econometrics* 114. (2), 361–394.

Hallin, M. and R. Liška (2011). "Dynamic factors in the presence of blocks". *Journal of Econometrics* 163. (1), 29–41.

Hansen, B. E. and S. Lee (2019). "Inference for iterated GMM under misspecification". Working Paper, available at https://www.ssc.wisc.edu/~bhansen/papers/IteratedGMM.html.

Hansen, L. P. (1982). "Large sample properties of generalized method of moments estimators". *Econometrica* 50. (4), 1029–1054.

Hansen, L. P. and J. J. Heckman (1996). "The empirical foundations of calibration". *The Journal of Economic Perspectives* 10. (1), 87–104.

Hansen, P. R., A. Lunde, and J. M. Nason (2011). "The model confidence set". *Econometrica* 79. (2), 453–497.

Harvey, D., S. Leybourne, and P. Newbold (1998). "Tests for forecast encompassing". *Journal of Business & Economic Statistics* 16. (2), 254–259.

Harvey, D. and P. Newbold (2000). "Tests for multiple forecast encompassing". *Journal of Applied Econometrics* 15. (5), 471–482.

Hasseltoft, H. (2012). "Stocks, bonds, and long-run consumption risks". *Journal of Financial and Quantitative Analysis* 47. (2), 309–332.

Heinrich, C. (2014). "The mode functional is not elicitable". *Biometrika* 101. (1), 245–251.

Hendricks, W. and R. Koenker (1992). "Hierarchical spline models for conditional quantiles and the demand for electricity". *Journal of the American Statistical Association* 87. (417), 58–68.

Hendry, D. and J.-F. Richard (1982). "On the formulation of empirical models in dynamic econometrics". *Journal of Econometrics* 20. (1), 3–33.

Holzmann, H. and M. Eulert (Mar. 2014). "The role of the information set for forecasting with applications to risk management". *Ann. Appl. Stat.* 8. (1), 595–621.

Huber, P. J. (1967). "The behavior of maximum likelihood estimates under nonstandard conditions". In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press, 221–233.

Ireland, P. N. (2004). "Technology shocks in the new Keynesian model". *Review of Economics and Statistics* 86. (4), 923–936.

Jermann, U. J. (1998). "Asset pricing in production economies". *Journal of Monetary Economics* 41. (2), 257–275.

Jermann, U. J. (2010). "The equity premium implied by production". *Journal of Financial Economics* 98. (2), 279–296.

Kaltenbrunner, G. and L. A. Lochstoer (2010). "Long-run risk through consumption smoothing". *The Review of Financial Studies* 23. (8), 3190–3224.

Kerssenfischer, M. (2019). "The puzzling effects of monetary policy in VARs: Invalid identification or missing information?" *Journal of Applied Econometrics* 34. (1), 18–25.

Kilian, L. and L. T. Lewis (2011). "Does the Fed respond to oil price shocks?" *The Economic Journal* 121. (555), 1047–1072.

Kim, K. and A. R. Pagan (1999). "The econometric analysis of calibrated macroeconomic models". In: *Handbook of Applied Econometrics*. Ed. by Pesaran, H. and Wickens, M. Vol. 1: Macroeconomics. Oxford: Blackwell Publishing Ltd. Chap. 7, 309–338.

Koenker, R. W. and G. Bassett (1978). "Regression quantiles". *Econometrica* 46. (1), 33–50.

Kogan, L. (2004). "Asset prices and real investment". *Journal of Financial Economics* 73. (3), 411–431.

Kogan, L. and D. Papanikolaou (2012). "Economic activity of firms and asset prices". *Annual Review Financial Economics* 4. (1), 361–384.

Lawley, D. and A. Maxwell (1971). *Factor analysis as a statistical method, second ed.* Butterworths, London.

Leeper, E. M. and C. A. Sims (1994). "Toward a modern macroeconomic model usable for policy analysis". *NBER Macroeconomics Annual* 9, 81–118.

Lucas, R. E. (1978). "Asset prices in an exchange economy". *Econometrica* 46. (6), 1429–1445.

McCracken, M. W. and S. Ng (2016). "FRED-MD: A monthly database for macroeconomic research". *Journal of Business & Economic Statistics* 34. (4), 574–589.

McFadden, D. (1989). "A method of simulated moments for estimation of discrete response models without numerical integration". *Econometrica* 57. (5), 995–1026.

McGrattan, E. R., R. Rogerson, and R. Wright (1997). "An equilibrium model of the business cycle with household production and fiscal policy". *International Economic Review* 38. (2), 267–290.

Mehra, R. and E. C. Prescott (1985). "The equity premium: A puzzle". *Journal of Monetary Economics* 15, 145–161.

Mizon, G. and J.-F. Richard (1986). "The encompassing principle and its application to testing non-nested hypotheses". *Econometrica* 54. (3), 657–78.

Mumtaz, H. and P. Surico (2009). "The transmission of international shocks: A factor-augmented VAR approach". *Journal of Money, Credit and Banking* 41. (s1), 71–100.

Newbold, P. and D. Harvey (2007). "Forecast combination and encompassing". In: *A Companion to Economic Forecasting*. John Wiley and Sons, Ltd. Chap. 12, 268–283.

Newey, W. and D. McFadden (1994). "Large sample estimation and hypothesis testing". In: *Handbook of Econometrics*. Ed. by Engle, R. and McFadden, D. Vol. 4. Elsevier. Chap. 36, 2111–2245.

Newey, W. K. and K. D. West (1987). "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix". *Econometrica* 55. (3), 703–708.

Nolde, N. and J. F. Ziegel (Dec. 2017). "Elicitability and backtesting: Perspectives for banking regulation". *The Annals of Applied Statistics* 11. (4), 1833–1874.

Onatski, A. (2012). "Asymptotics of the principal components estimator of large factor models with weakly influential factors". *Journal of Econometrics* 168. (2), 244–258.

Patton, A. J. (2011). "Data-based ranking of realised volatility estimators". *Journal of Econometrics* 161. (2), 284–303.

Patton, A. J. (2020). "Comparing possibly misspecified forecasts". *Journal of Business & Economic Statistics* 38. (4), 796–809.

Patton, A. J. and A. Timmermann (2007). "Testing forecast optimality under unknown loss". *Journal of the American Statistical Association* 102. (480), 1172–1184.

Patton, A. J., J. F. Ziegel, and R. Chen (2019). "Dynamic semiparametric models for expected shortfall (and Value-at-Risk)". *Journal of Econometrics* 211. (2), 388 –413.

Ramey, V. A. (2016). "Macroeconomic shocks and their propagation". In: *Handbook of Macroeconomics*. Vol. 2. Elsevier, 71–162.

Rotemberg, J. J. and M. Woodford (1997). "An optimization-based econometric framework for the evaluation of monetary policy". *NBER Macroeconomics Annual* 12, 297–346.

Rouwenhorst, G. K. (1995). *Asset returns and business cycles in frontiers of business cycle research*. Ed. by Cooley, T. F. Princeton University Press.

Rubinstein, M. (1976). "The valuation of uncertain income streams and the pricing of options". *The Bell Journal of Economics* 7. (2), 407–425.

Ruge-Murcia, F. (2012). "Estimating nonlinear DSGE models by the simulated method of moments: With an application to business cycles". *Journal of Economic Dynamics and Control* 36. (6), 914–938.

— (2014). *Indirect inference estimation of nonlinear dynamic general equilibrium models: with an application to asset pricing under skewness risk*. Tech. rep. Centre interuniversitaire de recherche en économie quantitative, CIREQ.

Ruge-Murcia, F. J. (2007). "Methods to estimate dynamic stochastic general equilibrium models". *Journal of Economic Dynamics and Control* 31. (8), 2599–2636.

Sargent, T. J., C. A. Sims, et al. (1977). "Business cycle modeling without pretending to have too much a priori economic theory". *New methods in business cycle research* 1, 145–168.

Schorfheide, F., D. Song, and A. Yaron (2018). "Identifying long-run risks: A Bayesian mixed-frequency approach". *Econometrica* 86. (2), 617–654.

Sims, C. A. (1992). "Interpreting the macroeconomic time series facts: The effects of monetary policy". *European Economic Review* 36. (5), 975–1000.

Singleton, K. J. (2009). *Empirical dynamic asset pricing: Model specification and econometric assessment*. Princeton University Press.

Smith, A. A. (1993). "Estimating nonlinear time-series models using simulated vector autoregressions". *Journal of Applied Econometrics* 8, 63–84.

Stock, J. H. and M. Watson (2011). "Dynamic factor models". *Oxford Handbook on Economic Forecasting*.

Stock, J. H. and M. W. Watson (2002a). "Forecasting using principal components from a large number of predictors". *Journal of the American Statistical Association* 97. (460), 1167–1179.

— (2002b). "Forecasting using principal components from a large number of predictors". *Journal of the American Statistical Association* 97. (460), 1167–1179.

— (2002c). "Macroeconomic forecasting using diffusion indexes". *Journal of Business & Economic Statistics* 20. (2), 147–162.

— (2005). *Implications of dynamic factor models for VAR analysis*. Tech. rep. w11467. National Bureau of Economic Research.

— (2016). "Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics". In: *Handbook of Macroeconomics*. Vol. 2. Elsevier, 415–525.

Taylor, J. W. (2019). "Forecasting value at risk and Expected Shortfall using a semiparametric approach based on the asymmetric Laplace distribution". *Journal of Business & Economic Statistics* 37. (1), 121–133.

— (2020). "Forecast combinations for value at risk and expected shortfall". *International Journal of Forecasting* 36. (2), 428–441.

Timmermann, A. (2006). "Forecast combinations". In: *Handbook of Economic Forecasting*. Ed. by Elliott, G., Granger, C., and Timmermann, A. 1st ed. Vol. 1. Elsevier. Chap. 04, 135–196.

Wachter, J. A. (2005). "Solving models with external habit". *Finance Research Letters* 2, 210–226.

Weil, P. (1989). "The equity premium puzzle and the risk-free rate puzzle". *Journal of Monetary Economics* 24, 401–421.

— (1990). "Nonexpected utility in macroeconomics". *The Quarterly Journal of Economics* 105. (1), 29–42.

Weiss, A. A. (1991). "Estimating nonlinear dynamic models using least absolute error estimation". *Econometric Theory* 7. (1), 46–68.

West, K. (2006). "Forecast evaluation". In: *Handbook of Economic Forecasting*. Ed. by Elliott, G., Granger, C., and Timmermann, A. 1st ed. Vol. 1. Elsevier. Chap. 03, 99–134.

White, H. (1994). *Estimation, inference and specification analysis*. Econometric Society Monographs. Cambridge University Press.

— (2001). *Asymptotic theory for econometricians*. San Diego: Academic Press.

Wu, J. C. and F. D. Xia (2016). "Measuring the macroeconomic impact of monetary policy at the zero lower bound". *Journal of Money, Credit and Banking* 48. (2-3), 253–291.

Eigenabgrenzung

Das erste Kapitel, *A Regularized Factor-Augmented Vector Autoregressive Model*, ist in Zusammenarbeit mit Maurizio Daniele entstanden, welcher ebenfalls Doktorand an der Graduate School of Decision Science war. Die Leistungsaufteilung betrug 50% pro Person.

Das zweite Kapitel, *Forecast Encompassing Tests for the Expected Shortfall*, habe ich zusammen mit dem ehemaligen Konstanzer Doktoranden Timo Dimitriadis erstellt, wobei meine individuelle Leistung bei 50% liegt.

Das dritte Kapitel, *Empirical Asset Pricing in a DGSE Framework: Reconciling Calibration and Econometrics using Partial Indirect Inference*, ist in Kooperation mit Joachim Grammig, Professor für Ökonometrie, Statistik und empirische Wirtschaftsforschung an der Eberhard Karls Universität Tübingen und seiner Doktorandin Dalia Elshiaty entstanden. Mein Anteil an diesem Projekt beträgt 70%.

Das viertel Kapitel, *Econometric Analysis of Partial Equilibrium Asset Pricing Models: A Two-Step Partial Indirect Inference Approach*, habe ich selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel erstellt.