Three Essays on Identification and Dimension Reduction in Vector Autoregressive Models

Dissertation zur Erlangung des

akademischen Grades eines Doktors der Wirtschaftswissenschaften

(Dr.rer.pol.)

vorgelegt von

Dominik Bertsche

an der

Universität Konstanz

Sektion Politik – Recht - Wirtschaft

Fachbereich Wirtschaftswissenschaften

Konstanz, 2020

Tag der mündlichen Prüfung: 30.04.2020

1. Referent/Referentin: Prof. Dr. Ralf Brüggemann

2. Referent/Referentin: Prof. Dr. Winfried Pohlmeier

Danksagung

Zunächst möchte ich mich bei meinem Doktorvater Prof. Dr. Ralf Brüggemann bedanken, ohne dessen Initiative ich diese Promotion nie begonnen hätte. Seine unaufgeregte Art half mir vor allem in der frühen Phase des Promotionsstudiums dabei, zu akzeptieren, dass ich nicht alles auf einmal verstehen kann und ich auch das Leben außerhalb der Universität genießen sollte. Ich wurde in den letzten Jahren nie unter Druck gesetzt und mir wurde immer die Freiheit gelassen, meine Forschungsgebiete selbst zu wählen. Dabei nahm sich Prof. Brüggemann stets die Zeit, Entwürfe zu lesen oder Ideen zu diskutieren.

Des Weiteren bin ich meinem Zweitbetreuer Prof. Dr. Winfried Pohlmeier - auch für seine hilfreichen Ratschläge in zahlreichen Vorträgen und Seminaren - dankbar. Für die Möglichkeit, einen Forschungsaufenthalt an der University of Melbourne zu absolvieren und die inspirierenden Diskussionen vor Ort möchte ich außerdem Dr. Tomasz Wozniak meinen Dank aussprechen. Darüber hinaus bedanke ich mich bei meinem Koautor Dr. Christian Kascha und bei Prof. Dr. Axel Kind für die bereitwillige Übernahme des Prüfungsvorsitzes im Bewertungskomitee.

Über die Jahre habe ich die Arbeitsatmosphäre am Lehrstuhl für Statistik und Ökonometrie, zu der neben Prof. Brüggemann auch Julie Schnaitmann, Verena Kretz, Maurizio Daniele und Dr. Robin Braun beigetragen haben, sehr geschätzt. Daher möchte ich mich bei ihnen für die gemeinsame Zeit bedanken und wünsche ihnen für die Zukunft nur das Beste.

Unterhalb des Gießbergs haben mir in den letzten Jahren meine Freunde aus dem Tennisclub und meine Mitbewohner in der Menzelstraße die nötige Ablenkung von der Forschungsarbeit verschafft. Nicht unerwähnt bleiben sollte auch meine große Familie, die mir früh beigebracht hat, meine eigenen Entscheidungen zu treffen und mir damit den Weg in diese Ausbildung ermöglicht hat.

Der letzte Dank geht wie üblich an die wichtigste Person. Es ist wahrscheinlich nicht immer einfach, mit jemandem zusammen zu leben, der nachts aufsteht um Simulationsergebnisse zu überprüfen oder Beweisideen nachzugehen und dessen Laune mitunter vom Vorzeichen einer Impuls-Antwort-Funktion abhängt. Ich hätte mir keine bessere als die bedingungslose Unterstützung von Jasmin wünschen können und bin ihr dafür sehr dankbar.

Table of Contents

List of Tables

List of Figures

Summary

How can we describe the dynamics of key macroeconomic time series in an accurate and economically meaningful way? In contemporaneous macroeconomics, there are two complementing modeling strategies for this purpose. The first way is to set up a fully-fledged dynamic macroeconomic model based on well-known microeconomic foundations, also known as dynamic stochastic general equilibrium (DSGE) models. While these models may offer a fully structural view of the economy, they are usually overly restricted. Hence, when put to the data, these models can only offer a limited fit and usually perform poorly in forecasting. Therefore, a complementing approach is to rely on less restricted multiple time series models. Here, the workhorse is the vector autoregressive (VAR) model which assumes that current vector realizations of time series can be described as a linear function of past realizations. In this thesis, I work on novel methodology related to VAR models.

Plain VAR models are reduced form, meaning that they can only serve for descriptive purposes and unconditional forecasting. If the model is to be used for structural analysis, a structural VAR can be employed to back out orthogonal driving forces that carry an economic interpretation. Such exogenous driving forces are also known as structural shocks and typical examples include fiscal policy shocks, monetary policy shocks, uncertainty shocks or shocks related to the oil market. Decomposing VAR innovations into uncorrelated errors is a straightforward exercise. However, without identifying restrictions, infinitely many structural models are observationally equivalent, i.e. they imply the same reduced form dynamics. Since the pioneering work of Sims (1980), several approaches have emerged to incorporate meaningful economic information to identify structural shocks. Earlier literature uses macroeconomic theory to justify short- and long-run restrictions on the effects of structural shocks. In this thesis, I work on more recent approaches that try to be less restrictive about the effects of the shocks to be identified. Among those, I use external information provided by instrumental variables or I exploit variation in the second moment of structural shocks for identification purposes. The second theme I work on is ensuring that all relevant information is captured in the model. If this is not the case, a VAR model can suffer from an omitted

variable bias that can cause misleading conclusions. Furthermore, informational deficiency can also harm the validity of an identification scheme for structural shocks. One way for tackling this challenge is to provide a variable selection tool that decides which variables should be included in a VAR system. I also consider augmenting the VAR with factors that are designed to summarize the dynamics of a large panel of time series, effectively yielding a dynamic factor model (Stock and Watson, 2016). This thesis comprises three chapters that work on novel methodology related to these issues: identification of structural shocks and sufficient information.

Chapter 1, written in collaboration with Robin Braun, concentrates on the identification of structural shocks. Hereby, we develop a novel econometric methodology to exploit heteroskedasticity in the structural innovations for identification. Following the idea of Rigobon (2003), several parametric approaches to capture the time variation in the second moment of structural shocks are applied for that purpose: an exogenous breakpoint model, a generalized autoregressive conditional heteroskedasticity (GARCH) model, a Smooth Transition model and a Markov Switching model. Instead, we suggest to employ the concept of stochastic volatility for statistical identification of Structural Vector Autoregressive models (SV-SVAR). While inference is more challenging than for the competitor models mentioned before, stochastic volatility models are known to provide a superior model fit. We discuss full and partial identification of the model and develop efficient Expectation-Maximization algorithms for Maximum Likelihood inference. Simulation evidence suggests that the SV-SVAR works well in identifying structural parameters also under misspecification of the variance process, particularly if compared to the alternative heteroskedastic SVARs listed above. We apply the model to study the interdependence between monetary policy and stock markets and find that the stochastic volatility model provides by far the best model fit and captures heteroskedasticity in the structural shocks much better than its competitor models. Since shocks identified by heteroskedasticity may not be economically meaningful, we exploit the framework to test conventional exclusion restrictions as well as Proxy-SVAR restrictions which are overidentifying in the heteroskedastic model. The tests indicate statistical evidence against employing conventional short-run restrictions to identify a monetary policy shock. However, there is no evidence against using a certain long-run restriction and no evidence against identification by one of two external instruments included into the analysis.

In Chapter 2, which is joint work with Ralf Brüggemann and Christian Kascha, we deal with the variable selection problem in VAR models. To do so, we represent the dynamic relations among variables in vector autoregressive models as directed graphs. Based on this, graph theoretical concepts are transferred to the VAR model. More specifically, we identify so-called strongly connected components (SCCs). In a VAR setting, these SCCs can be interpreted as classes of variables that are related to each other. The links among the strongly

connected components finally yield a variable selection tool to filter the set of time series that needs to be included in a VAR system if interest is in forecasting or impulse response analysis of a given set of variables. Although the usage of graph theoretical concepts is not completely novel to the literature in multivariate time series (see e.g. Eichler (2006)), to the best of our knowledge they have not been employed for model selection purposes so far. Moreover, SCCs have not been considered in econometrics at all. In a theoretical part, we show that the set of selected variables from the graphical method coincides with the set of variables that is multi-step causal for the variables of interest by relating the paths in the graph to the coefficients of the 'direct' VAR representation. An empirical application illustrates the usefulness of the suggested approach: Including the selected variables into a small US monetary VAR is useful for impulse response analysis as it avoids the well-known 'price-puzzle'. Therefore, adding only six variables to the small monetary system selected by our proposed variable selection tool, causes similar changes to the estimated responses as including many more variables (Banbura et al., 2010).

Chapter 3 addresses the identification of structural shocks as well as the inclusion of all relevant information in the model. Hence, in a way it combines the main goals that are pursued separately in the previous chapters. I connect identification via external instruments and factor models in an oil market application dealing with oil supply shocks. These have been found to be important drivers of US business cycles. The literature studies the effects of two types of shortfalls in supply separately: a supply 'flow' shock revealing an immediate drop in production as well as a 'news' shock associated with unanticipated shifts in future oil production. In this chapter, I simultaneously identify both kinds of supply shocks allowing me to assess their relative importance within one model. For this purpose, I develop a factor-augmented vector autoregressive model that is identified by external instruments (Proxy-FAVAR). Compared to alternative models, the framework allows to identify multiple structural shocks by external instruments in a factor model without requiring additional exclusion restrictions to disentangle the shocks of interest. Thereby, this approach ensures that the identified shocks are orthogonal to each other while also using a rich information set and a credible identification scheme. From an empirical point of view, an oil supply flow shock is typically identified in a small VAR model (see e.g. Kilian and Murphy (2014)). However, I show that a large information set is necessary to reasonably identify a flow shock by a proxy variable. My results suggest that the two shocks of interest have substantially distinguishable effects: While news shocks clearly dominate the reaction of the oil price, supply flow surprises have more pronounced effects on many macroeconomic indicators and financial variables. The latter finding is more striking when both oil supply shocks are identified simultaneously within one model compared to shocks stemming from separate models that identify exactly one structural shock by external instruments.

References

Banbura, M., D. Giannone, and L. Reichlin (2010). "Large Bayesian vector autoregressions". *Journal of Applied Econometrics* 25, 71–92.

Eichler, M. (2006). "Graphical modeling of dynamic relationships in multivariate time series". In: *Handbook of Time Series Analysis: Recent Developments and Applications*. Ed. by Schelter, B., Winterhalder, M., and Timmer, J. Wiley.

Kilian, L. and D. P. Murphy (2014). "The role of inventories and speculative trading in the global market for crude oil". *Journal of Applied Econometrics* 29 (3), 454–478.

Rigobon, R. (2003). "Identification through heteroskedasticity". *Review of Economics and Statistics* 85 (4), 777–792.

Sims, C. A. (1980). "Macroeconomics and reality". *Econometrica* 48 (1), 1–48.

Stock, J. H. and M. W. Watson (2016). "Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics". In: *Handbook of Macroeconomics*. Ed. by Taylor, J. B. and Uhlig, H. Vol. 2. Elsevier. Chap. 8, 415–525.

Zusammenfassung

Wie kann man volkswirtschaftliche Zusammenhänge möglichst genau modellieren und ökonomisch interpretierbar beschreiben? Zu diesem Zweck gibt es in der Makroökonomie zwei grundverschiedene Strategien. Zum einen kann man ein dynamisches stochastisches allgemeines Gleichgewichtsmodell (DSGE) aufsetzen, das sich ausschließlich aus Optimierungsansätzen eines repräsentativen Haushalts und eines repräsentativen Unternehmens ableitet. Diese Modelle ermöglichen es zwar, den Einfluss bestimmter Faktoren auf die Wirtschaft isoliert zu betrachten, sind jedoch normalerweise recht stark restringiert. Daher lassen sie sich meist nur mit mäßigem Erfolg auf Daten anwenden und liefern oft äußerst fehlerhafte Prognosen. Deshalb gibt es den weniger restriktiven Ansatz der multivariaten Zeitreihenmodelle. Das vektorautoregressive (VAR) Modell beruht auf der Annahme, dass ein Zeitreihenvektor anhand linearer Funktionen seiner vergangenen Beobachtungen beschrieben werden kann. Diese Dissertation befasst sich mit neuartigen Methoden im Bereich der VAR-Modelle.

Einfache VAR-Modelle sind in reduzierter Form und eignen sich daher lediglich zur Beschreibung volkswirtschaftlicher Variablen oder für unbedingte Prognosen jener Zeitreihen. Wenn hingegen exogene Antriebskräfte mit einer ökonomischen Interpretation herausgefiltert werden sollen, kann ein strukturelles VAR-Modell benutzt werden. Solche exogenen Antriebskräfte werden auch strukturelle Schocks genannt und sind beispielsweise steuerpolitische Schocks, geldpolitische Schocks, Unsicherheitsschocks oder Schocks in Zusammenhang mit dem Rohölmarkt. Prinzipiell ist es sehr einfach, die VAR-Fehlerterme in kontemporär unkorrelierte Komponenten zu zerlegen. Ohne zusätzliche Annahmen zur Identifikation implizieren allerdings unendlich viele strukturelle Modelle ein und dasselbe Modell in reduzierter Form. Daher wurden seit der wegweisenden Arbeit von Sims (1980) zahlreiche Methoden entwickelt, um zusätzliche Information zur Identifikation struktureller Schocks in das VAR-Modell zu integrieren. Die frühe Literatur stützt sich auf volkswirtschaftliche Theorie, um kurz- und langfristige Restriktionen auf die Auswirkungen der strukturellen Schocks sinnvoll zu begründen. In dieser Arbeit befasse ich mich mit aktuelleren Metho-

den, die auf weniger restriktive Annahmen zurückgreifen. Hier konzentriere ich mich auf Information aus Instrumentvariablen und auf zweite Momente der strukturellen Schocks, die sich über die Zeit ändern, um die strukturellen Schocks zu identifizieren. Das zweite Thema, mit dem ich mich beschäftige, ist die Berücksichtigung aller relevanter Information. Wenn nicht alle relevanten Informationen benutzt werden, kann das VAR-Modell aufgrund fehlender Variablen eine Verzerrung aufweisen, die irreführende Schlussfolgerungen nach sich ziehen kann. Außerdem kann unzureichende Information ein ansonsten gültiges Schema zur Identifikation struktureller Schocks unbrauchbar machen. Eine Möglichkeit, dieses Problem anzugehen, ist die Entwicklung eines Mechanismus zur Variablenselektion, der entscheidet, welche Zeitreihen Teil eines VAR-Systems sein sollten. Außerdem beschäftige ich mich mit der Erweiterung des VAR-Modells um Faktoren, welche die Bewegungen einer großen Menge von Zeitreihen zusammenfassen sollen, was letztendlich zu einem dynamischen Faktormodell (Stock und Watson, 2016) führt. Die Arbeit besteht aus drei Kapiteln, die sich mit neuartigen Lösungsansätzen zu folgenden Herausforderungen befassen: die Identifikation struktureller Schocks und die Nutzung hinreichender Information.

Kapitel 1, das in Zusammenarbeit mit Robin Braun entstanden ist, bezieht sich ausschließlich auf die Identifikation struktureller Schocks. Wir entwickeln eine neuartige ökonometrische Methode, die Heteroskedastizität in den strukturellen Neuerungen zu Identifikationszwecken nutzt. Zahlreiche parametrische Ansätze modellieren die zweiten Momente der strukturellen Schocks als Prozess über die Zeit und gehen so auf die Idee von Rigobon (2003) zurück: ein Strukturbruch-Modell, ein verallgemeinertes autoregressives Modell mit bedingter Heteroskedastizität (GARCH), ein Smooth-Transition-Modell und ein Markov-Switching-Modell. Stattdessen benutzen wir das stochastische Volatilitätsmodell zur statistischen Identifikation des strukturellen vektorautoregressiven Modells (SV-SVAR). Zwar gestaltet sich die Inferenz schwieriger als bei den zuvor aufgezählten Modellen, jedoch eignet sich das stochastische Volatilitätsmodell im Allgemeinen besser, um volkswirtschaftliche Variablen möglichst akkurat zu beschreiben. Wir befassen uns sowohl mit vollständiger, als auch mit partieller Identifikation des eingeführten Modells und entwickeln effiziente Erwartungs-Maximierungs-Algorithmen zur Maximum-Likelihood-Schätzung. Simulationsergebnisse deuten darauf hin, dass das SV-SVAR-Modell die strukturellen Parameter auch dann akkurat identifiziert, wenn der tatsächliche Varianzprozess ein anderer ist. Dies zeigt sich insbesondere, wenn man das stochastische Volatilitätsmodell mit den oben genannten heteroskedastischen SVAR-Modellen vergleicht. In einer empirischen Anwendung benutzen wir das Modell, um die Zusammenhänge zwischen der Geldpolitik und den Aktienmärkten zu untersuchen. Dabei finden wir heraus, dass das stochastische Volatilitätsmodell die Daten mit Abstand am besten beschreibt und die Zeitvariation in den strukturellen Schocks deutlich besser modelliert als die alternativen Modelle. Die mit der vorgeschlagenen Methode identi-

fizierten Schocks haben nicht zwingend eine ökonomische Interpretation. Herkömmliche Ausschluss- und Instrumentvariablenrestriktionen werden in heteroskedastischen Modellen nicht zur Identifikation gebraucht. Daher können wir deren Gültigkeit anhand von Hypothesentests überprüfen: Einerseits raten die Tests von kurzfristigen Restriktionen zur Identifikation eines geldpolitischen Schocks ab, andererseits lässt sich keine Evidenz gegen eine bestimmte langfristige Restriktion und gegen eines der beiden untersuchten Instrumente finden.

In Kapitel 2, das ich gemeinsam mit Ralf Brüggemann und Christian Kascha verfasst habe, beschäftigen wir uns mit der Variablenselektion in VAR-Modellen. Zu diesem Zweck stellen wir die Beziehungen zwischen den Variablen in vektorautoregressiven Modellen als gerichteten Graphen dar. Darauf aufbauend wenden wir graphentheoretische Konzepte an und teilen so die Variablen in Gruppen - sogenannte starke Zusammenhangskomponenten (SCCs) - ein. Diese SCCs können als miteinander verbundene Variablen betrachtet werden. Grundsätzlich werden VAR-Modelle zu Prognosezwecken oder zur strukturellen Analyse einer zuvor definierten Menge von Variablen - den Schlüsselvariablen - eingesetzt. Die Verbindungen zwischen den SCCs wiederum definieren unsere Methode zur Auswahl jener Zeitreihen, die in einem VAR-System enthalten sein sollten, welches die Prognose oder die strukturelle Analyse dieser Schlüsselvariablen zum Ziel hat. Nach unserem Kenntnisstand wurde die Graphentheorie bislang noch nicht zur Variablenselektion eingesetzt, auch wenn deren Anwendung in der multivariate Zeitreihenanalyse kein komplett neues Vorgehen ist (siehe z.B. Eichler (2006)). Darüber hinaus wurden SCCs in der Ökonometrie bisher noch gar nicht in Betracht gezogen. Im theoretischen Teil dieses Kapitels zeigen wir, dass die von unserer Methode ausgewählten Zeitreihen mit jenen Zeitreihen übereinstimmen, die zu mindestens einem Horizont kausal für mindestens eine dieser Schlüsselvariablen sind. Eine empirische Anwendung unterstreicht den Nutzen unserer Methode: Wenn wir die ausgewählten Zeitreihen zu einem dreidimensionalen VAR-System, das Konjunkturdaten, Preise und Zinssätze enthält, hinzufügen, verschwindet das altbekannte Problem, dass die Preise nach einem geldpolitischen Schock kurzzeitig ansteigen. Also hat das Hinzufügen von nur sechs Variablen, die mit unserer Methode ausgewählt wurden, ähnliche Auswirkungen auf die geschätzten Impuls-Antwort-Funktionen wie das Hinzufügen einer deutlich größeren Anzahl von Zeitreihen (Banbura u. a., 2010).

Kapitel 3 beschäftigt sich sowohl mit der Identifikation struktureller Schocks, als auch mit der Einbeziehung aller relevanter Information. Daher verbindet es die beiden Ziele, derer sich diese Arbeit in den vorangegangenen Kapiteln getrennt annimmt. Ich kombiniere die Identifikation struktureller Schocks mithilfe von Instrumentvariablen und Faktormodelle in einer empirischen Anwendung, die sich mit Angebotsschocks auf dem Rohölmarkt befasst. Diese gelten als wichtige Treiber für die Konjunktur in den USA. Bislang werden in der

Literatur zwei Arten von Angebotsschocks getrennt behandelt: direkte Angebotsschocks, die ein unmittelbares Absacken der Erdölförderung mit sich bringen und Nachrichtenschocks, welche die Erdölförderung lediglich in der Zukunft beeinflussen. In diesem Kapitel identifiziere ich diese beiden Angebotsschocks auf einmal, sodass ich deren Auswirkungen auf die Wirtschaft in einem Modell bewerten kann. Dazu entwickle ich ein durch Faktoren erweitertes vektorautoregressives Modell, das durch Instrumentvariablen identifiziert wird (Proxy-FAVAR). Im Gegensatz zu anderen Modellen ermöglicht es dieser Ansatz, mehrere strukturelle Schocks in einem Faktormodell mithilfe von Instrumentvariablen zu identifizieren, ohne dabei zusätzliche Ausschlussrestriktionen zu erfordern. Somit garantiert diese Methode die Orthogonalität der zu identifizierenden Schocks und basiert zusätzlich auf einem breiten Informationsspektrum sowie auf einem sinnvollen Identifikationsansatz. In der Regel wird ein direkter Ölangebotsschock in kleinen VAR-Modellen identifiziert (siehe z.B. Kilian und Murphy (2014)). Ich zeige jedoch, dass eine große Menge an Informationen erforderlich ist, um einen direkten Angebotsschock sinnvoll durch eine Instrumentvariable zu identifizieren. Die Auswirkungen der beiden Ölangebotsschocks lassen sich deutlich unterscheiden: Während Nachrichtenschocks die Reaktion des Rohölpreises dominieren, haben direkte Angebotsschocks wesentlich stärker ausgeprägte Effekte auf volkswirtschaftliche Indikatoren und Finanzvariablen. Das letztere Ergebnis wird umso deutlicher, wenn man beide Angebotsschocks in einem Modell identifiziert. Wenn man die Schocks stattdessen mithilfe von Instrumentvariablen in getrennten Modellen identifiziert, sind diese Unterschiede weniger stark ausgeprägt.

Literatur

Banbura, M., D. Giannone und L. Reichlin (2010). “Large Bayesian vector autoregressions”. *Journal of Applied Econometrics* 25, 71–92.

Eichler, M. (2006). “Graphical modeling of dynamic relationships in multivariate time series”. In: *Handbook of Time Series Analysis: Recent Developments and Applications*. Hrsg. von Schelter, B., Winterhalder, M. und Timmer, J. Wiley.

Kilian, L. und D. P. Murphy (2014). “The role of inventories and speculative trading in the global market for crude oil”. *Journal of Applied Econometrics* 29 (3), 454–478.

Rigobon, R. (2003). “Identification through heteroskedasticity”. *Review of Economics and Statistics* 85 (4), 777–792.

Sims, C. A. (1980). “Macroeconomics and reality”. *Econometrica* 48 (1), 1–48.

Stock, J. H. und M. W. Watson (2016). “Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics”. In: *Handbook of Macroeconomics*. Hrsg. von Taylor, J. B. und Uhlig, H. Bd. 2. Elsevier. Kap. 8, 415–525.

Chapter 1

Identification of structural vector autoregressions by stochastic volatility

1.1 Introduction

Following Sims (1980), structural vector autoregressive (SVAR) models have been used extensively in empirical macroeconomics. Based on a reduced form VAR, identifying restrictions are imposed to back out a unique set of structural shocks and estimate their dynamic effects on the endogenous variables. Popular approaches for identification include short- and long-run restrictions on the effects of structural shocks (Bernanke and Mihov, 1998; Blanchard and Quah, 1989; Sims, 1980), sign restrictions (Canova and De Nicoló, 2002; Faust, 1998; Uhlig, 2005) and identification via external instruments, also known as Proxy-SVARs (Mertens and Ravn, 2013; Montiel-Olea et al., 2016; Stock and Watson, 2012a). Furthermore, a growing body of literature exploits statistical properties of the data to identify SVAR models, assuming non-Gaussianity (Gourieroux et al., 2017; Lanne et al., 2017) or heteroskedasticity of the structural shocks (see Lütkepohl and Netšunajev (2017b) for a review).[1]

In this paper, we discuss the identification and estimation of SVARs by a stochastic volatility (SV) model. Specifically, we assume that the log-variances of structural shocks are latent, each following independent AR(1) processes. Drawing on recent methodology of Lewis (2018), we show that in conjunction with a fixed impact matrix, our model yields additional restrictions that allow to pin down a unique set of orthogonal shocks. Besides identification, we extensively discuss classical Maximum Likelihood inference and provide fast algorithms for estimation purposes.

Our paper fits into the literature of identifying structural shocks in SVARs by heteroskedasticity. A variety of models have been proposed in the literature so far, including a simple breakpoint model (Rigobon, 2003), a Markov Switching model (Lanne et al., 2010), a GARCH model (Normandin and Phaneuf, 2004) and a Smooth Transition model (Lütkepohl and Netšunajev, 2017a). Furthermore, Lewis (2018) discusses identification and estimation of heteroskedastic SVARs in a GMM framework without specifying any functional form for the variances. We complement this literature by adding the SV specification to this list of models.

Closely related to our paper is the work of Carriero et al. (2019) who also exploit a SV model to identify uncertainty shocks in SVARs. However, while we focus on the identification of the traditional SVAR model and implement classical inference, the model of Carriero et al. (2019) is more structural adapted to their application, and their inference is conducted within a Bayesian framework.

Using a stochastic volatility model for the variance of structural shocks is an attractive specification for many reasons. First, SV models enjoy increasing popularity in theoretical

[1]For a textbook treatment of identification in SVARs we refer to Kilian and Lütkepohl (2017).

and empirical macroeconomics. For example, Justiniano and Primiceri (2008) and Fernández-Villaverde and Rubio-Ramírez (2007) allow for SV within fitted DSGE models, finding substantial time variation in the second moments of their structural shocks. Furthermore, SV models are often used to complement time-varying parameter VARs and have been found to provide a good description of volatility patterns in macroeconomic data (Koop and Korobilis, 2010; Primiceri, 2005). Given this context, it seems natural to exploit the model also for identification purposes of SVARs. Second, a stochastic volatility specification is known to be more flexible than models with deterministic variance processes. This is because the SV model, in contrast to the alternative specifications, includes shocks in the volatility equation that do not depend on the innovations in the VAR equation. As pointed out in Kim et al. (1998), this additional flexibility typically translates into superior fit in comparison to equally parameterized models from the GARCH family. This is an important aspect, given that recent evidence of Lütkepohl and Schlaak (2018) suggests to choose the heteroskedasticity model of SVARs by information criteria.

Since the SV specification implies a nonlinear state space model, standard linear filtering algorithms cannot be applied to evaluate the likelihood function which makes estimation of the SV-SVAR model relatively challenging. However, many estimation methods have been proposed in the literature to overcome this difficulty starting with Generalized Methods of Moments (Melino and Turnbull, 1990), Quasi Maximum Likelihood (Harvey et al., 1994; Ruiz, 1994), Simulated Likelihood (Danielsson and Richard, 1993) and Bayesian methods (Kim et al., 1998) based on Markov Chain Monte Carlo (MCMC) simulation. For this paper, we choose a full Gaussian Maximum Likelihood framework. This is essential such that we are able to assess economic theory using classical hypothesis tests. We consider likelihood evaluation and its maximization separately. To evaluate the likelihood, we follow Durbin and Koopman (1997) and use an importance sampling approach. In order to maximize the likelihood function, we develop two versions of an Expectation Maximization (EM) algorithm. The first is based on a second order Taylor approximation of the intractable smoothing distribution necessary in the E-step and relies on sparse matrix algorithms developed for Gaussian Markov random fields (Chan, 2017; Rue et al., 2009). Therefore, the algorithm is extremely fast and typically converges reliably within seconds. Our second EM algorithm approximates the E-step by Monte Carlo integration, exploiting that the error term of a log-linearized state equation can be accurately approximated by a mixture of normal distributions (Kim et al., 1998). Conditional on simulated mixture indicators, the model has a normal linear state space representation allowing to compute the expectations necessary in the E-step by standard Kalman smoothing recursions. Thereby, the second order approximation can be avoided at the cost of higher computational effort. Note that both EM algorithms provide very reliable and stable estimates. This is a clear advantage

over other variance specifications as e.g. a Markov Switching model with three regimes which requires a huge amount of initial values to converge to a stable global maximum.

In a simulation exercise we provide evidence that, in comparison to alternative heteroskedastic SVARs, the SV-SVAR model works well in estimating the structural parameters under misspecification of the variance process, proofing itself capable to capture volatility patterns generated by very different data generating processes (DGPs). More specifically, by simulating data from SVAR models subject to four distinct variance specifications we find that the SV model performs superior in terms of the mean squared error of estimated impulse response functions.

In an empirical application we apply the proposed model to identify the structural parameters in a SVAR specified by Bjørnland and Leitemo (2009). Relying on a combination of short- and long-run restrictions, they study the interdependence between monetary policy and the stock market. We find that for this application the SV model provides superior fit and is favored by all conventional information criteria, if compared to other heteroskedastic SVAR models. Since structural shocks identified by heteroskedasticity are not guaranteed to be economically meaningful, we follow Lütkepohl and Netšunajev (2017b) and test the exclusion restrictions used by Bjørnland and Leitemo (2009). In addition, we also test Proxy-SVAR restrictions which arise if the narrative series of Romer and Romer (2004) and Gertler and Karadi (2015) are used as external instruments to identify a monetary policy shock. Our results indicate that the short-run restrictions of Bjørnland and Leitemo (2009) and Proxy-SVAR restrictions based on the shock of Gertler and Karadi (2015) are rejected by the data. However, we do neither find evidence against imposing the long-run restriction of Bjørnland and Leitemo (2009) nor against identifying a monetary policy shock by the Romer and Romer (2004) series.

The paper is structured as follows. Section 1.2 introduces the SVAR model with stochastic volatility and discusses under which conditions the structural parameters are identified. Section 1.3 considers Maximum Likelihood estimation and reviews a procedure to test for identification. In Section 1.4, we present simulation evidence while in Section 1.5 we apply the proposed model to study the interdependence between US monetary policy and stock markets. Section 1.6 concludes.

1.2 Identification of SVARs by stochastic volatility

Let y_t be a $K \times 1$ vector of endogenous variables. We consider the heteroskedastic SVAR model reading:

$$y_t = \nu + \sum_{j=1}^{p} A_j y_{t-j} + u_t, \tag{1.2.1}$$

$$u_t = B V_t^{\frac{1}{2}} \eta_t, \tag{1.2.2}$$

where $\eta_t \sim (0, I_K)$ is assumed to be a white noise error term. Equation (1.2.1) corresponds to a standard reduced form VAR(p) model for y_t, capturing common dynamics across the time series data by a linear specification. Here, A_j for $j = 1, \ldots, p$ are $K \times K$ matrices of autoregressive coefficients and ν is a $K \times 1$ vector of intercepts. Since we only consider stable time series throughout the paper, we assume:

$$\det A(z) = \det(I_K - A_1 z - \ldots - A_p z^p) \neq 0, \qquad \text{for } |z| \leq 1.$$

Equation (1.2.2) models the structural part and is set up as a B-model in the terminology of Lütkepohl (2005). The reduced form error terms u_t are decomposed into a linear function of K structural shocks $\varepsilon_t = V_t^{\frac{1}{2}} \eta_t$, with B a $K \times K$ invertible contemporaneous impact matrix and $V_t^{\frac{1}{2}}$ a stochastic diagonal matrix with strictly positive elements capturing potential heteroskedasticity and/or non-normality in each structural shock. This specification yields a time-varying covariance matrix of the reduced form errors u_t given as $\Sigma_t = \mathrm{E}(u_t u_t') = B V_t B'$. Throughout the paper, we assume that there are $r \leq K$ heteroskedastic shocks which are ordered such that they appear first in the vector ε_t. To model the time-varying second moments of these shocks, we specify an independent Gaussian AR(1) log stochastic volatility model for each of the r heteroskedastic components:

$$V_t = \begin{pmatrix} \mathrm{diag}(\exp([h_{1t}, \ldots, h_{rt}]')) & 0 \\ 0 & I_{K-r} \end{pmatrix}, \tag{1.2.3}$$

$$h_{it} = \mu_i + \phi_i (h_{i,t-1} - \mu_i) + \sqrt{s_i}\omega_{it}, \qquad \text{for } i = 1, \ldots, r, \tag{1.2.4}$$

where $\omega_{it} \sim \mathcal{N}(0,1)$ and $\mathrm{E}(\varepsilon_t' \omega_t) = 0$ for $\omega_t = (\omega_{1t}, \ldots, \omega_{rt})'$. Furthermore, the initial states are assumed to be initialized from the unconditional distribution $h_{i1} \sim \mathcal{N}(\mu_i, s_i/(1-\phi_i^2))$. Note that the proposed model for equation (1.2.2) is very similar to the Generalized Orthogonal GARCH (GO-GARCH) model of Van der Weide (2002) and Lanne and Saikkonen (2007), with the major difference in the specification (1.2.3)-(1.2.4) of V_t. While for the GO-GARCH the first r diagonal components are modeled by deterministic GARCH(1,1) processes,

we model their logarithms as latent AR(1)'s. We will assume that the underlying AR(1) processes of the log-volatilities are stable with finite variance implying that for $i = 1, \ldots, r$, $|\phi_i| < 1$ and $0 < s_i < \infty$. It immediately follows that ε_t is assumed to be a strictly stationary stochastic process with finite second moment, which will aid in the identification analysis. In particular, the following basic properties can be derived for the model in a straightforward manner (see e.g. Jacquier et al. (1994)):

$$\gamma_i(\tau) = \mathrm{Cov}(\varepsilon_{it}^2, \varepsilon_{i,t+\tau}^2) = \exp(2\mu_i + \sigma_{h_i}^2)(\exp(\sigma_{h_i}^2 \phi_i^\tau) - 1) \tag{1.2.5}$$

$$\kappa_i = \frac{\mathrm{E}(\varepsilon_{it}^4)}{\mathrm{E}(\varepsilon_{it}^2)^2} = \mathrm{E}(\eta_{it}^4) \exp\left(\sigma_{h_i}^2\right), \tag{1.2.6}$$

$$\mathrm{E}(\varepsilon_{it}^2) = \mathrm{E}(\exp(h_{it})\eta_{it}^2) = \mathrm{E}(\exp(h_{it}))\mathrm{E}(\eta_{it}^2) = \mathrm{E}(\exp(h_{it})) = \exp\left(\mu_i + \sigma_{h_i}^2/2\right), \tag{1.2.7}$$

for $i = 1, \ldots, r$, where $\sigma_{h_i}^2 = s_i/(1 - \phi_i^2)$ is the unconditional variance of the underlying log-volatility process.

The model is able to capture the main stylized facts of structural shocks that are typically encountered in empirical SVAR analysis. First, heteroskedasticity can be modeled by setting $\phi_i > 0$. The respective autocovariance function in the second moment of ε_{it} is given by equation (1.2.5), displaying an exponential decay in ϕ_i. This autocovariance function has been found to be very flexible enabling to capture a large variety of heteroskedasticity patterns, an argument that we can confirm based on our simulation evidence. Second, the model can capture heavy tailed errors and the respective kurtosis function κ_i can be decomposed into a part that is due to the kurtosis of the standardized structural shocks η_{it} and a component which inflates the value depending on the underlying SV parameters. That is, given a conditional Gaussian error distribution in ε_{it}, excess kurtosis kicks in as soon as the SV process is nontrivial, that is $s_i > 0$. This means that even if the shock is homoskedastic ($\phi_i = 0$), the model is still able to capture heavy tails under conditional Gaussianity. In this particular case, the structural error would be independent and identically distributed following a mixture of log-normal and Gaussian errors.[2] We argue that this is a key advantage with respect to a model from the GARCH family, which are generally unable to generate homoskedastic shocks featuring excess kurtosis given the assumption of conditional Gaussianity. Finally, equation (1.2.7) gives the unconditional scale of the structural shocks as a function of the underlying SV parameters.

In the following, we will use equations (1.2.5)-(1.2.7) to discuss identification in detail. First, note that the structural shocks are latent variables and a unique scaling must be obtained. For this purpose, we follow the widely used normalization of setting the scale to $\mathrm{E}(\varepsilon_t \varepsilon_t') = I_K$.

[2]Note the similarity to a t-distribution, which can be represented as a product of an independent Gamma and Gaussian random variable.

Using equation (1.2.7), this can be achieved by restricting the mean of the AR(1) processes to $\mu_i = -0.5s_i/(1-\phi_i^2)$. It follows that the structural parameters in B are related to the unconditional reduced form covariance matrix by:

$$\mathrm{E}(u_t u_t') = \Sigma_u = BB'. \tag{1.2.8}$$

Given this normalization, a standard interpretation applies in that the j-th column of B corresponds to the average contemporaneous response of the endogenous variables y_t to shock ε_{jt} of size 'one standard deviation'.

Due to the symmetry of the covariance matrix, identification in the SV-SVAR model cannot be discussed based on equation (1.2.8) solely. For that purpose, we rely on Lewis (2018) who treats identification by time-varying volatility in a more general context requiring no specific functional form. In particular, identification can be analyzed based on the lag τ autocovariance in the squared reduced form residuals $\xi_t = \mathrm{vech}(u_t u_t')$. This function takes the following form (Lewis, 2018):

$$\mathrm{Cov}(\xi_t, \xi_{t+\tau}) = L_K(B \otimes B)G_K M_\tau G_K'(B \otimes B)' L_K', \tag{1.2.9}$$

where L_K is the elimination matrix such that $\mathrm{vech}(A) = L_K \,\mathrm{vec}(A)$, G_K is a selection matrix with zeros and ones such that $\mathrm{vec}(D) = G_K d$ for $D = \mathrm{diag}(d)$ and $M_\tau = \mathrm{diag}(\gamma_1(\tau), \ldots, \gamma_r(\tau), 0_{K-r})$.[3] Note that one autocovariance has $\sum_{i=1}^{5} \binom{i+K-3}{K-2}$ unique elements ($K \geq 2$), while the structural model contains K^2 entries in B and r autocovariances in $\gamma_i(\tau)$'s, implicitly parameterized nonlinearly by the underlying SV processes. Lewis (2018) proofs general identification of the elements in M_τ and B under the restriction that the diagonal of B is fixed at unity. In order to account for the standard deviation normalization implied by (1.2.8), a modification is necessary. In Proposition 1.1, we summarize identification of B for any $r \leq K$ for our setting.

Proposition 1.1. *Let $B = (B_1, B_2)$ with $B_1 \in \mathbb{R}^{K\times r}$ and $B_2 \in \mathbb{R}^{K\times(K-r)}$. Assume the stable SV-SVAR model presented above with $|\phi_i| < 1$, $\phi_i \neq 0$ and $0 < s_i < \infty$ for $i = 1, \ldots, r$, implying that equations* (1.2.8) *and* (1.2.9) *hold. Then, matrix B_1 is unique up to permutation and sign switches.*

Proof. See Appendix 1.A.2.

In fact, it is not necessary that $r = K$ shocks are heteroskedastic in order that the impact matrix is identified. The orthogonality constraints implied by equation (1.2.8) yield enough structure to fully identify the model in case of $r = K - 1$, which is summarized in Corollary 1.1.

[3] See also Appendix 1.A.1 for a derivation of this function for the SV-SVAR model.

Corollary 1.1. *Assume the setting from Proposition 1.1 for the special case $r = K - 1$. Then, the entire matrix $B \in \mathbb{R}^{K\times K}$ is unique up to multiplication of its columns by -1 and permutation of its first $K - 1$ columns.*

Proof. See Appendix 1.A.3.

The presented results are broadly in line with those provided by Lewis (2018). However, our results deviate in the sense that identification is given also under $r = K-1$ heteroskedastic shocks, based on the additional information provided by (1.2.8). Furthermore, the simple structure assumed for the SV-SVAR allows for a much simpler proof.

At this point we highlight that identification of the model can also be discussed based on non-Gaussianity implied by the SV model. If one is willing to assume mutual independence in ε_{it}, the SV-SVAR model as discussed in this paper is covered by the general framework of Lanne et al. (2017). Specifically, the structural parameters in B are identified up to permutation and sign if the structural shocks are strictly stationary with finite second moments, mutually independent and with at most one Gaussian component. For the SV-SVAR model, this means that in order to achieve strict stationarity and finite second moments, we need $s_i < \infty$ and $|\phi_i| < 1$ for all i as discussed above. Furthermore, under conditionally Gaussian errors, at most one structural shock can display a degenerate SV process with $s_i = 0$, implying a Gaussian marginal. Analogous results regarding partial identification are available in Maxand (2018). As in Proposition 1.1, the structural parameters associated with the non-Gaussian shocks are locally identified up to permutation and sign-changes.

An additional interesting feature of the model is that the continuously changing variances imply that the impact matrix B is strongly overidentified. This implies that the above presented framework could be used to test for parameter instability in B without imposing any further restrictions, e.g. by a Chow or Sup LR type of test.

Before we continue with estimation of the model, we discuss an additional constraint that we impose on the log-volatilities. Note that we identify the scale of the structural shocks by setting $\mu_i = -0.5s_i(1 - \phi_i^2)^{-1}$, implying that $\mathrm{E}(\varepsilon_t\varepsilon_t') = I_K$. However, this constraint holds only in expectation and for very persistent heteroskedasticity patterns, the sample moment can be very distinct in finite samples. In such cases, restricting μ_i is not too informative for the scale and one can potentially suffer from weak identification. Therefore, throughout this paper we will in addition consider the sample constraint for $i = 1, \ldots, r$:

$$A_h h_i = \mu_i, \tag{1.2.10}$$

where $A_h = \mathbf{1_T}'/T$ and $h_i = (h_{i1}, \ldots, h_{iT})'$. Note that this constraint leads to a rank reduction of the covariance matrix implied for h_i by the Gaussian AR(1) model. This is similar in spirit to imposing the alternative normalizing constraint that $\mathrm{E}(h_{i1}) = \mathrm{Var}(h_{i1}) = 0$, implying that

$\mathrm{E}(u_1 u_1') = BB'$ which is typically used to identify the scaling in Markov Switching SVAR models (Herwartz and Lütkepohl, 2014; Lanne et al., 2010). However, this would require that we leave μ_i unrestricted implying an additional parameter to estimate, which is why we prefer restriction (1.2.10).

1.3 Maximum likelihood estimation

In order to estimate the model, we propose a full Maximum Likelihood approach. Let $\theta = [\mathrm{vec}([\nu, A_1, \ldots, A_p])', \mathrm{vec}(B)', \phi', s']'$ denote the full vector of parameters in the SV-SVAR model where $\phi = [\phi_1, \ldots, \phi_r]'$ and $s = [s_1, \ldots, s_r]'$. Assuming normality of the standardized structural shocks η_t, the log-likelihood function based on the prediction error decomposition is given as follows:

$$\mathcal{L}(\theta) = \sum_{t=1}^{T} \left[-\frac{K}{2} \log(2\pi) - \frac{1}{2} \log |BV_{t|t-1}B'| - \frac{1}{2} u_t' (BV_{t|t-1}B')^{-1} u_t \right],$$

where $u_t = y_t - \nu - \sum_{j=1}^{p} A_j y_{t-j}$ and $V_{t|t-1} = \mathrm{E}[V_t | \mathcal{F}_{t-1}]$ are one-step ahead predicted variances conditional on the information set at time $t-1$. Since the SV model implies a nonlinear state space model, the predictive distributions $p(h_t|\theta, y_{t-1})$ necessary to compute $V_{t|t-1}$ are not available in closed form. That is, the likelihood is intractable and standard Kalman filter algorithms cannot be applied. To overcome this difficulty, we follow Durbin and Koopman (1997) and Chan and Grant (2016) in evaluating the likelihood function by importance sampling in a computationally efficient way. Furthermore, to maximize the likelihood, we develop two versions of an Expectation Maximization algorithm which lead to fast and reliable results.

1.3.1 Evaluation of the likelihood

To show how the likelihood can be evaluated by importance sampling, we slightly manipulate the log-likelihood function. For that purpose, let $\varepsilon_t = B^{-1} u_t$ and $v_{i,t|t-1}$ the i-th diagonal element of $V_{t|t-1}$, then:

$$\begin{aligned}
\mathcal{L}(\theta) &= -T \log |B| + \sum_{i=1}^{K} \sum_{t=1}^{T} \left[-\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(v_{i,t|t-1}) - \frac{1}{2} \varepsilon_{it}^2 / v_{i,t|t-1} \right] \\
&= -T \log |B| + \sum_{i=1}^{K} \log p(\varepsilon_i | \theta),
\end{aligned}$$

where we have used that $\log |BV_{t|t-1}B'| = 2 \log |B| + \sum_{i=1}^{K} \log \left(v_{i,t|t-1} \right)$. Therefore, given autoregressive coefficients and contemporaneous impact matrix, likelihood evaluation of

the SV-SVAR model reduces to the evaluation of K univariate densities for each structural shock. For $i = r+1, \dots, K$ these densities are trivial to compute since $v_{i,t|t-1} = 1$. However, the densities $\log p(\varepsilon_i|\theta)$ for $i \leq r$ are not tractable. Their evaluation equals computing the following high-dimensional integral for $i = 1, \dots, r$:

$$p(\varepsilon_i|\theta) = \int p(\varepsilon_i|\theta, h_i) p^c(h_i|\theta) dh_i. \tag{1.3.1}$$

where $p(\varepsilon_i|\theta, h_i)$ is a Gaussian distribution and $p^c(h_i|\theta)$ the prior density implied by the Gaussian AR(1) model subject to the constraint $A_h h_i = \mu_i$.

To evaluate this integral, we use an importance sampling estimator. Therefore, let $q(h_i)$ be a proposal distribution from which independent random draws $h_i^{(1)}, \dots, h_i^{(R)}$ can be generated, and further let $q(h_i)$ dominate $p(\varepsilon_i|\theta, h_i)p^c(h_i|\theta)$. An unbiased importance sampling estimator of the integral in equation (1.3.1) is:

$$\widehat{p(\varepsilon_i|\theta)} = \frac{1}{R} \sum_{j=1}^{R} \frac{p(\varepsilon_i|\theta, h_i^{(j)}) p^c(h_i^{(j)}|\theta)}{q(h_i^{(j)})}. \tag{1.3.2}$$

Plugging (1.3.2) into the SV-SVAR log-likelihood yields an IS estimator of the SV-SVAR log-likelihood function:

$$\widehat{\mathcal{L}(\theta)} = -T \log |B| + \sum_{i=1}^{r} \log \widehat{p(\varepsilon_i|\theta)} + \sum_{i=r+1}^{K} \log p(\varepsilon_i|\theta). \tag{1.3.3}$$

The accuracy of the IS estimator crucially depends on our choice for the importance densities $q(h_i)$ which we discuss in the following. First, note that the optimal (zero variance) importance density is given by the smoothing distribution $p(h_i|\theta, \varepsilon_i) \propto p(\varepsilon_i|\theta, h_i)p(h_i|\theta)$. However, since the likelihood of the measurement equation is nonlinear in h_i, the normalizing constant is unknown which is why we rely on IS in the first place. We follow Durbin and Koopman (1997, 2000) and use a Gaussian importance density denoted by $\pi_G(h_i|\theta, \varepsilon_i)$, which is centered at the mode of $p(h_i|\theta, \varepsilon_i)$ with precision equal to the curvature at this point. For computational reasons, we rely on fast algorithms that exploit the sparse precision matrices of Gaussian Markov random fields as used e.g. in Rue et al. (2009) for a broad class of models and Chan and Grant (2016) for stochastic volatility models in particular.

To derive $\pi_G(h_i|\theta, \varepsilon_i)$, we follow the exposition of Chan and Grant (2016). For a moment, assume that there was no linear constraint on h_i. Then, normality implies the following explicit form of the zero variance IS density:

$$p(h_i|\theta, \varepsilon_i) \propto \exp\left(-\frac{1}{2}(h_i - \delta_i)' Q_i (h_i - \delta_i) + \log p(\varepsilon_i|\theta, h_i)\right),$$

where $Q_i = H_i' \Sigma_{h_i}^{-1} H_i$ with

$$H_i = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -\phi_i & 1 & 0 & \dots & 0 \\ 0 & -\phi_i & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -\phi_i & 1 \end{pmatrix},$$

and $\Sigma_{h_i} = \text{diag}([\frac{s_i}{1-\phi_i}, s_i, \dots, s_i]')$. Furthermore, $\delta_i = H_i^{-1}\tilde{\delta}_i$ with $\tilde{\delta}_i = (\mu_i, (1-\phi_i)\mu_i, \dots, (1-\phi_i)\mu_i)'$. The Gaussian approximation is based on a second order Taylor expansion of the nonlinear density $\log p(\varepsilon_i|\theta, h_i)$ around some properly chosen $\tilde{h}_i^{(0)}$:

$$\log p(\varepsilon_{it}|\theta, h_{it}) \approx \log p(\varepsilon_{it}|\theta, \tilde{h}_{it}^{(0)}) + b_{it}h_{it} - \frac{1}{2}c_{it}h_{it}^2, \tag{1.3.4}$$

where b_{it} and c_{it} depend on $\tilde{h}_{it}^{(0)}$. Based on the linearized kernel, an approximate smoothing distribution $\pi_G(h_i|\theta, \varepsilon_i)$ takes the form of a Normal distribution with precision matrix $\bar{Q}_i = Q_i + C_i$ and mean $\bar{\delta}_i = \bar{Q}_i^{-1}(b_i + Q_i\delta_i)$, where $C_i = \text{diag}([c_{i1}, \dots, c_{iT}]')$ and $b_i = (b_{i1}, \dots, b_{iT})'$. The T-dimensional density has a tridiagonal precision matrix which allows for fast generation of random samples and likelihood evaluation. The approximation is evaluated at the mode of the smoothing distribution obtained by a Newton-Raphson method that typically converges in few iterations. Details on the Newton-Raphson method and on explicit expressions for b_{it} and c_{it} are given in Appendix 1.B.1.

As discussed in Section 1.2, our prior density for h_i is subject to the normalizing constraint $A_h h_i = \mu_i$. Therefore, the IS density $\pi_G(h_i|\theta, \varepsilon_i)$ needs a slight modification to account for this linear constraint. In particular, an application of Bayes' theorem yields a constraint density $\pi_G^c(h_i|\theta, \varepsilon_i)$ which is also Gaussian but has mean and covariance:

$$\bar{\delta}_i^c = \bar{\delta}_i - \bar{Q}_i^{-1}A_h'(A_h\bar{Q}_i^{-1}A_h')^{-1}(A_h\bar{\delta}_i - \mu_i), \tag{1.3.5}$$

$$\text{Cov}(h_i|\theta, \varepsilon_i, A_h h_i{=}\mu_i) = \bar{Q}_i^{-1} - \bar{Q}_i^{-1}A_h'(A_h\bar{Q}_i^{-1}A_h')^{-1}A_h\bar{Q}_i^{-1}. \tag{1.3.6}$$

Note that imposing the linear restriction yields a non-sparse precision and a reduced rank covariance which impedes direct efficient sampling and density evaluation. Following Rue et al. (2009), sampling and evaluation of $\pi_G^c(h_i|\theta, \varepsilon_i)$ can still be implemented at trivial extra costs by what is known as 'conditioning by kriging'. Specifically, a random sample $\tilde{h}_i^{(j)}$ is first generated from $\pi_G(h_i|\theta, \varepsilon_i)$, exploiting the sparse precision $\bar{Q}_i^{-1}$. In a second step, the draw is corrected for the linear constraint by setting $h_i^{(j)} = \tilde{h}_i^{(j)} - \bar{Q}_i^{-1}A_h'(A_h\bar{Q}_i^{-1}A_h')^{-1}(A_h\tilde{h}_i^{(j)} - \mu_i)$.

Also evaluation of the adjusted IS density can be achieved efficiently by applying Bayes' theorem:

$$\pi_G^c(h_i|\theta,\varepsilon_i) = \frac{\pi_G(h_i|\theta,\varepsilon_i)\pi(A_h h_i|h_i)}{\pi_1(A_h h_i)}, \tag{1.3.7}$$

where $\log \pi(A_h h_i|h_i) = -\frac{1}{2}\log|A_h A_h'|$ and $\pi_1(A_h h_i) \sim \mathcal{N}(A_h\bar{\delta}_i, A_h\bar{Q}_i^{-1}A_h')$. Note that the same routine can be used to evaluate the prior density $p^c(h_i|\theta)$ which displays the same constraint. That is, the constraint prior density is evaluated as follows:

$$p^c(h_i|\theta) = \frac{p(h_i|\theta)\pi(A_h h_i|h_i)}{\pi_2(A_h h_i)}, \tag{1.3.8}$$

where $p(h_i|\theta) \sim \mathcal{N}(\delta_i, Q_i)$, $\pi_2(A_h h_i) \sim \mathcal{N}(A_h\delta_i, A_h Q_i^{-1}A_h')$ and $\pi(A_h h_i|h_i)$ is as above.

Finally, we recommend to assess the quality of the estimator (1.3.3) by reporting its standard error which can be computed e.g. by the batch means method. Furthermore, for the validity of the standard error and $\sqrt{R}$-convergence of the IS estimator, the variance of the importance weights has to exist. Since for the high-dimensional integral (1.3.1) this is not clear a priori, we advise to test for the existence of the variance using e.g. the test of Koopman et al. (2009). However, for sample sizes typically used in macroeconomics we do not expect this to be a serious issue.

1.3.2 EM algorithm

In order to optimize the likelihood function, we exploit the Expectation Maximization algorithm first introduced by Dempster et al. (1977). The EM procedure is particularly suitable for maximization problems under the presence of hidden variables. In our setting, the hidden variables are the set of r log variances denoted by $h = (h_1, \ldots, h_r)$. Our goal is to maximize:

$$\mathcal{L}(\theta) = \log p(y|\theta) = \log \int p(y|\theta,h)p(h|\theta)dh.$$

Following Neal and Hinton (1998) and Roweis and Ghahramani (2001), let $\tilde{p}(h)$ be any distribution of the hidden variables, possibly depending on θ and y. Then, a lower bound on $\mathcal{L}(\theta)$ can be obtained by an application of Jensen's inequality:

$$\begin{aligned}
\mathcal{L}(\theta) &= \log \int p(y|\theta,h)p(h|\theta)dh && (1.3.9)\\
&= \log \int \frac{p(y|\theta,h)p(h|\theta)}{\tilde{p}(h)}\tilde{p}(h)dh && (1.3.10)\\
&\geq \int \log\left(\frac{p(y|\theta,h)p(h|\theta)}{\tilde{p}(h)}\right)\tilde{p}(h)dh && (1.3.11)\\
&= \int \log\left(p(y|\theta,h)p(h|\theta)\right)\tilde{p}(h)dh - \int \log\left(\tilde{p}(h)\right)\tilde{p}(h)dh && (1.3.12)\\
&=: F(\tilde{p},\theta). && (1.3.13)
\end{aligned}$$

The EM algorithm starts with some initial parameter vector $\theta^{(0)}$ and proceeds by iteratively maximizing:

$$\text{E-step:} \qquad \tilde{p}^{(l)} = \arg\max_{\tilde{p}} F(\tilde{p},\theta^{(l-1)}), \tag{1.3.14}$$

$$\text{M-step:} \qquad \theta^{(l)} = \arg\max_{\theta} F(\tilde{p}^{(l)},\theta). \tag{1.3.15}$$

Under mild regularity conditions the EM algorithm converges reliably towards a local optimum.[4] It is easy to show that the E-step in (1.3.14) is given by setting $\tilde{p}^{(l)}$ equal to the smoothing distribution $p(h|\theta^{(l-1)},y)$. This can be seen by noting that for this choice, equation (1.3.11) holds with equality which means that the lower bound $F(\tilde{p},\theta)$ exactly equals the log-likelihood $\mathcal{L}(\theta)$. Furthermore, the M-step in equation (1.3.15) is given by maximizing the criterion function:

$$\begin{aligned}
Q(\theta;\theta^{(l-1)}) &= \int \log\left(p(y|\theta,h)p(h|\theta)\right)\tilde{p}^{(l)}(h)dh && (1.3.16)\\
&= \mathrm{E}_{\theta^{(l-1)}}\left(\mathcal{L}_c(\theta)\right), && (1.3.17)
\end{aligned}$$

where the expectation is taken with respect to $\tilde{p}^{(l)}(h)$ and $\mathcal{L}_c(\theta) = \log\left(p(y|\theta,h)p(h|\theta)\right)$ is the complete data log-likelihood.

For the SV-SVAR model, the complete data log-likelihood is rather simple and we refer to Appendix 1.B.3 for an explicit expression. It follows that for a given choice of $\tilde{p}^{(l)}$, computing the M-Step is straightforward. However, since the smoothing distribution in SV models is generally not tractable, we cannot simply set $\tilde{p}^{(l)} = p(h|\theta^{(l-1)},y)$. Instead, we develop two algorithms which approximate this density to a different extent, one based on an

[4]For details on convergence, we refer to the textbook treatment in McLachlan and Krishnan (2007).

analytical approximation and the other based on Monte Carlo integration. In the following, we use that independence among the structural errors implies that the smoothing distribution can be factored as: $p(h|\theta^{(l-1)}, y) = \prod_{i=1}^{r} p(h_i|\theta^{(l-1)}, y)$.

Analytical approximation

Our analytical approximation is based on the following E-step:

$$\tilde{p}^{(l)}(h) = \prod_{i=1}^{r} \pi_G^c(h_i|\theta^{(l-1)}, \varepsilon_i), \tag{1.3.18}$$

which is the Gaussian approximation of the smoothing distribution that we already introduced as importance density. This E-step corresponds to maximizing $F(\tilde{p}, \theta^{(l-1)})$ with respect to $\tilde{p}$ considering only the family of Gaussian distributions. To motivate this approach, we follow the arguments of Neal and Hinton (1998) who argue that it is not necessary to work with the exact smoothing distributions in the EM algorithm to get monotonic increases in the log-likelihood function $\mathcal{L}(\theta)$. In fact, it can be shown that $F(\tilde{p}, \theta) = \mathcal{L}(\theta) - D_{KL}\left(\tilde{p}(h)||p(h|y, \theta)\right)$ where $D_{KL}(\cdot||\cdot)$ is the Kullback - Leibler (KL) divergence measure. Therefore, if the Gaussian approximation is close to the smoothing density in a KL sense, iteratively optimizing $F(\tilde{p}, \theta)$ yields convergence to a point very close to the corresponding local maximum of $\mathcal{L}(\theta)$. In the following, we refer to this algorithm as EM-1 and provide details in Appendix 1.B.3.

Monte Carlo approximation

The second approach is based on Markov Chain Monte Carlo (MCMC) integration and draws on the results of Kim et al. (1998).[5] The idea is to consider the linearized state space representation of the r independent SV equations:

$$\log(\varepsilon_{it}^2) = h_{it} + \log(\eta_{it}^2), \tag{1.3.19}$$

$$h_{it} = \mu_i + \phi_i(h_{i,t-1} - \mu_i) + \sqrt{s_i}\omega_{it}, \tag{1.3.20}$$

where $\eta_{it} \sim N(0,1)$ and $\omega_{it} \sim N(0,1)$. Kim et al. (1998) propose to closely approximate the log-χ^2 error distribution in (1.3.19) by a mixture of seven normals. In particular, they specify:

$$p(\log(\eta_t^2)|z_{it} = k) \sim \mathcal{N}(\log(\varepsilon_{it}^2); m_k, v_k^2), \tag{1.3.21}$$

$$p(z_{it} = k) = p_k, \tag{1.3.22}$$

[5] See also Mahieu and Schotman (1998) for a similar Monte Carlo EM algorithm to estimate a univariate SV model.

with mixture parameters p_k, m_k, v_k^2 for $k = 1, \ldots, 7$ tabulated in Table 1.B.1 in Appendix 1.B.3. The advantage of representing the transformed measurement error with a normal mixture is that conditional on a realization of the indicators $z_i = (z_{i1}, \ldots, z_{iT})'$, the state space model is both, linear and Gaussian which allows for closed form computations of $p(h_{it}|\theta, z_{it}, y)$ by Kalman smoothing recursions.

We exploit this property in our Monte Carlo EM algorithm in the following way. First, consider the mixture representation of the intractable smoothing distribution:

$$p(h|\theta^{(l-1)}, y) \approx \int p(h|\theta^{(l-1)}, z, y) p(z|\theta^{(l-1)}, y) dz.$$

Using this distribution in the EM algorithm yields the following objective function in the M-step:

$$Q(\theta; \theta^{(l-1)}) \approx \int \int \log\left[p(y|\theta, h)p(h|\theta)\right] p(h|\theta^{(l-1)}, z, y) p(z|\theta^{(l-1)}, y) dz dh.$$

To approximatively solve this high-dimensional integral, we simulate a large number of mixture indicators z from $p(z|\theta^{(l-1)}, y)$ by MCMC methods and consider the Monte Carlo counterpart:

$$Q(\theta, \theta^{(l-1)}) \approx \frac{1}{R} \sum_{j=1}^{R} \mathrm{E}^{(j)}_{\theta^{(l-1)}}[\mathcal{L}(\theta)],$$

where the expectation is now taken with respect to the tractable Gaussian distribution $p(h|\theta^{(l-1)}, z^{(j)}, y)$ which can be computed by Kalman smoothing recursions.[6]

In order to generate random draws of the mixture indicators we follow the MCMC scheme of Kim et al. (1998) which involves iteratively drawing from the conditional distributions $p(h_i|\theta^{(l-1)}, z_i, y)$ and $p(z_i|\theta^{(l-1)}, h_i, y)$. For computational reasons we rely on the precision sampler of Chan and Jeliazkov (2009) which exploits the sparsity in the precision matrix. Furthermore, it allows for a straightforward extension to implement the linear normalizing constraint on h_i. In the remainder, we call the Monte Carlo based algorithm EM-2 and for details on the MCMC algorithm and respective M-steps, we refer to Appendix 1.B.3.

1.3.3 Properties of the estimator

Because the SV-SVAR model is a special case of a Hidden Markov Model, the asymptotic properties of the Maximum Likelihood estimator can be inferred from Cappé et al. (2005).

[6] If desired, one could correct for the minor approximation error by applying an importance reweighting procedure (Kim et al., 1998). However, this would slow down the algorithm and would have only marginal effects on its accuracy.

Let $\hat{\theta}$ denote the ML estimator, under appropriate regularity conditions, $\hat{\theta}$ is consistent and asymptotically normally distributed:

$$T^{1/2}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}(\theta)^{-1}), \tag{1.3.23}$$

where $\mathcal{I}(\theta) = -\mathrm{E}\left(\frac{\partial^2 \log p(y|\theta)}{\partial\theta\partial\theta'}\right)$ is the information matrix. Furthermore, a strongly consistent estimator for the asymptotic variance is given as:

$$\widehat{\mathcal{I}(\theta)} = T^{-1}\mathcal{J}(\hat{\theta}) \tag{1.3.24}$$

where $\mathcal{J}(\hat{\theta}) = -\frac{\partial^2 \mathcal{L}(\theta)}{\partial\theta\partial\theta'}\Big|_{\theta=\hat{\theta}}$ is the observed information matrix evaluated at the ML estimator.

To compute estimator (1.3.24) in algorithm EM-1, note that we can evaluate an approximate log-likelihood in closed form based on the Gaussian approximation which we rely on in the E-step. In particular, based on Bayes' theorem:

$$\log p(\varepsilon_i|\theta) \approx \log p(\varepsilon_i|\theta, h_i) + \log p^c(h_i|\theta) - \log \pi_G^c(h_i|\theta, \varepsilon_i), \tag{1.3.25}$$

which can be evaluated for any h_i. For convenience, the r likelihoods for the heteroskedastic structural shocks are evaluated at the mean $h_i = \bar{\delta}_i^c$, such that the exponential term in $\pi_G^c(h_i|\theta, \varepsilon_i)$ drops out. Therefore, based on (1.3.25) an approximate complete log-likelihood is given as:

$$\mathcal{L}_a(\theta) = -T\log|B| + \sum_{i=1}^{r}\left[\log p(\varepsilon_i|\theta, h_i) + \log p^c(h_i|\theta) - \log \pi_G^c(h_i|\theta, \varepsilon_i)\right] + \sum_{i=r+1}^{K}\log p(\varepsilon_i|\theta).$$

We take the second derivative of this approximation with respect to the parameter vector θ using numerical differentiation to obtain an approximation of the observed information matrix $\mathcal{J}_1(\hat{\theta}) = -\frac{\partial^2 \mathcal{L}_a(\theta)}{\partial\theta\partial\theta'}\Big|_{\theta=\hat{\theta}}$.

For the Monte Carlo based algorithm EM-2, no closed form approximation of the likelihood is available which makes the computation of the information matrix estimator more involved. We apply Louis Identity (Louis, 1982) to the observed information matrix:

$$\mathcal{J}_2(\hat{\theta}) = \mathrm{E}\left[\mathcal{J}_c(\hat{\theta})|y\right] - \mathrm{Cov}(S_c(\hat{\theta})|y), \tag{1.3.26}$$

where $\mathcal{J}_c(\hat{\theta}) = -\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial\theta\partial\theta'}\Big|_{\theta=\hat{\theta}}$, $S_c(\hat{\theta}) = \frac{\partial \mathcal{L}_c(\theta)}{\partial\theta}\Big|_{\theta=\hat{\theta}}$ are the observed information matrix and score of the complete data log-likelihood $\mathcal{L}_c$, respectively. The integrals necessary to compute expected value and variance are with respect to the smoothing distribution at the

ML estimator $p(h|\hat{\theta}, y)$ which is intractable for the SV model. However, based on simulated values of the mixture indicators $z^{(j)}(j = 1, \ldots, R)$, Monte Carlo integration is feasible with:

$$\begin{aligned}
\mathrm{E}\left[\mathcal{J}_c(\hat{\theta})|y\right] \approx & \frac{1}{R}\sum_{j=1}^{R} -\mathrm{E}\left[\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial\theta\partial\theta'} \;\middle|\; z^{(j)}, y\right]_{\theta=\hat{\theta}}, \\
\mathrm{Cov}(S_c(\hat{\theta})) \approx & \frac{1}{R}\sum_{j=1}^{R} \mathrm{E}\left[\frac{\partial \mathcal{L}_c(\theta)}{\partial\theta}\frac{\partial \mathcal{L}_c(\theta)}{\partial\theta'} \;\middle|\; z^{(j)}, y\right]_{\theta=\hat{\theta}},
\end{aligned}$$

where the second approximation holds since $\mathrm{E}(S_c(\hat{\theta})|y) = 0$. The integrals required to compute the expected values are with respect to the tractable Gaussian distributions $p(h|\hat{\theta}, z^{(j)}, y)$. The derivatives necessary to apply the Louis Method are available in closed form and given in Appendix 1.B.4.

Identification of the SVAR model is ultimately useful to conduct structural analysis. Since Impulse Response Functions (IRFs) are likely to be the most widely used tool for that purpose, we outline in Appendix 1.B.5 how to conduct inference on these quantities within our model. In particular, we describe a Delta Method approach to quantify uncertainty of the identified IRFs.

1.3.4 Testing for identification

For valid likelihood inference on the structural parameters including the impact matrix B, the model must be identified. As highlighted in Section 1.2, at most one component of ε_t is allowed to be homoskedastic if the model is to be identified solely by heteroskedasticity. To determine the number of heteroskedastic shocks in a given application, we recommend to follow a procedure considered by Lanne and Saikkonen (2007) and Lütkepohl and Milunovich (2016) within SVAR-GARCH models. The idea is to conduct the following sequence of tests:

$$H_0 : r = r_0 \qquad \text{vs} \qquad H_1 : r > r_0, \tag{1.3.27}$$

for $r_0 = 0, \ldots, K-1$. If all null hypotheses up to $r_0 = K-2$ can be rejected, there is evidence for sufficient heteroskedasticity in the data to fully identify B.

The testing problem given in (1.3.27) is nonstandard since parts of the parameter space differ between null and alternative hypothesis. Therefore, Lanne and Saikkonen (2007) suggest test statistics which require estimation under H_0 only. In particular, suppose that r_0 is the true number of heteroskedastic errors, and separate the structural shocks $\varepsilon_t = B^{-1}u_t = (\varepsilon_{1t}', \varepsilon_{2t}')'$ into a heteroskedastic part $\varepsilon_{1t} \in \mathbb{R}^{r_0}$ and homeskedastic innovations $\varepsilon_{2t} \in \mathbb{R}^{K-r_0}$. Note that if the null is true ($r = r_0$), $\varepsilon_{2t} \sim (0, I_{K-r_0})$ is white noise. To test for

remaining heteroskedasticity in ε_{2t}, Lanne and Saikkonen (2007) propose to use Portmanteau types of statistics on the second moment of ε_{2t}. In particular, they construct the following time series:

$$\xi_t = \varepsilon_{2t}'\varepsilon_{2t} - T^{-1}\sum_{t=1}^{T}\varepsilon_{2t}'\varepsilon_{2t}, \tag{1.3.28}$$

$$\vartheta_t = \text{vech}(\varepsilon_{2t}\varepsilon_{2t}') - T^{-1}\sum_{t=1}^{T}\text{vech}(\varepsilon_{2t}\varepsilon_{2t}'), \tag{1.3.29}$$

with vech$(\cdot)$ being the half-vectorization operator as defined e.g. in Lütkepohl (2005). Based on these time series, autocovariances up to a prespecified horizon H are tested considering the following statistics:

$$Q_1(H) = T\sum_{h=1}^{H}\left(\frac{\tilde{\gamma}(h)}{\tilde{\gamma}(0)}\right)^2, \tag{1.3.30}$$

$$Q_2(H) = T\sum_{h=1}^{H}\text{tr}\left[\tilde{\Gamma}(h)'\tilde{\Gamma}(0)^{-1}\tilde{\Gamma}(h)\tilde{\Gamma}(0)^{-1}\right], \tag{1.3.31}$$

where $\tilde{\gamma}(h) = T^{-1}\sum_{t=h+1}^{T}\xi_t\xi_{t-h}$ and $\tilde{\Gamma}(h) = T^{-1}\sum_{t=h+1}^{H}\vartheta_t\vartheta_{t-h}'$. It is shown that under the null, $Q_1(H) \xrightarrow{d} \chi^2(H)$ and $Q_2(H) \xrightarrow{d} \chi^2\left(\frac{1}{4}H(K-r_0)^2(K-r_0+1)^2\right)$.

To apply these tests, we must be able to estimate the model under H_0 which requires additional restrictions on B if $r_0 < K-1$. To uniquely disentangle the shocks in ε_{2t}, it turns out that it is sufficient to impose a lower triangular structure on the lower right $(K-r)\times(K-r)$ block of B:

Corollary 1.2. *Assume the setting from Proposition 1.1 for $r \leq K-2$. Moreover, separate $B = \begin{pmatrix} B_{11} & B_{21} \\ B_{12} & B_{22} \end{pmatrix}$, $B_{11} \in \mathbb{R}^{r\times r}$, $B_{12} \in \mathbb{R}^{(K-r)\times r}$, $B_{21} \in \mathbb{R}^{r\times(K-r)}$ and $B_{22} \in \mathbb{R}^{(K-r)\times(K-r)}$. Let B_{22} be restricted to be a lower triangular matrix. Then, the full matrix B is unique up to multiplication of its columns by -1 and permutation of its first r columns.*

Proof. See Appendix 1.A.4.

We conclude with a remark regarding the small sample properties of the tests. Based on extensive simulation studies, Lütkepohl and Milunovich (2016) find a substantial lack in power for sample sizes typically available in macroeconomics. Hence, if the null hypothesis can be rejected for all r_0's up to $K-2$, this can be interpreted as strong evidence in favor of model identification.

1.4 Monte Carlo study

An important question for practitioners is how a heteroskedastic SVAR model performs in estimating structural parameters under inherent misspecification of the variance process. To shed some light on this question, we conduct a small scale Monte Carlo (MC) study. Specifically, we compare the estimation performance of the SV-SVAR model under misspecification to that of alternative heteroskedastic SVARs, namely a simple Breakpoint model (BP-SVAR), Markov Switching models (MS-SVAR) and a GARCH model (GARCH-SVAR).

Our analysis involves generating a large number of datasets from the four stated heteroskedastic SVARs. Then, we estimate each model and compare the relative estimation performance of the misspecified to the correctly specified model. We focus on estimation of structural IRFs which are probably the most widely used tool in SVAR analysis. Furthermore, they are nonlinear functions of both, the structural impact matrix and reduced form autoregressive parameters. Thus, they are particularly suited to summarize the overall estimation performance of a SVAR model. As a metric of comparison, we use cumulated Mean Squared Errors (MSEs) of the IRF estimates.

The following data generating processes (DGPs) are specified to simulate the datasets, closely resembling the MC design of Lütkepohl and Schlaak (2018).[7] Time series of lengths $T \in \{200, 500\}$ are generated by the following bivariate VAR(1) process:

$$y_t = A_1 y_{t-1} + u_t,$$

with $u_t \sim \mathcal{N}(0, B\Lambda_t B')$ for $t = 1, \ldots, T$ and:

$$A_1 = \begin{pmatrix} 0.6 & 0.35 \\ -0.1 & 0.7 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 \\ 0.5 & 2 \end{pmatrix}.$$

For the diagonal matrix Λ_t, the following DGPs are specified:

1. **BP-SVAR**: The BP-SVAR is subject to a one time change in the variance. We set $\Lambda_t = I_2$ for $t = 1, \ldots T/2$ and $\Lambda_t = \text{diag}([2, 7]')$ for $t = T/2 + 1, \ldots T$.

2. **MS(2)-SVAR**: The specified MS-SVAR involves a switching variance with the same regimes than the BP-SVAR. We specify the transition probability matrix:

$$P = \begin{pmatrix} .95 & .05 \\ .1 & .9 \end{pmatrix}.$$

[7] Some difference to their design comes from our choice of the impact matrix. In particular, we use what we think are more realistic values of the impact matrix in a sense that they lead to less dramatic changes in the VAR error variance.

Based on simulated states $s_1, \ldots, s_T \in \{1, 2\}$, $\Lambda_{s_t=1} = I_2$ and $\Lambda_{s_t=2} = \text{diag}([2, 7]')$.

3. **GARCH-SVAR**: For this specification, the diagonal elements of $\Lambda_t = \text{diag}([\lambda_{1t}, \lambda_{2t}]')$ follow univariate GARCH(1,1) processes with unit unconditional variance:

$$\lambda_{it} = (1 - \alpha_i - \beta_i) + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i \lambda_{i,t-1}, \qquad i \in \{1, 2\},$$

where $\varepsilon_t = B^{-1} u_t$ is the vector of structural shocks at time t. We set $\alpha_i = 0.15$ and $\beta_i = 0.8$ $(i = 1, 2)$ which correspond to values typically estimated for empirical data.

4. **SV-SVAR**: For this DGP, $\Lambda_t = \text{diag}([\exp(h_{1t}), \exp(h_{2t})]')$ with:

$$h_{it} = \mu_i + \phi_i (h_{i,t-1} - \mu_i) + \sqrt{s_i}\omega_{it},$$

where $\omega_{it} \sim \mathcal{N}(0, 1)$. We set $\mu_i = -0.5 s_i / (1 - \phi_i^2)$ such that $\text{E}(\varepsilon_{it}^2) = 1$. Furthermore, we set $\phi_i = 0.95$ and $s_i = 0.04$ $(i = 1, 2)$ what corresponds to fairly persistent processes in the variance often observed in macroeconomic and financial data.

To avoid that our results are driven by issues regarding to weak identification, we only accept datasets in the MS(2)-SVAR DGP if at least 25% of the observations are associated with either of the regimes. Likewise, for the GARCH and SV DGPs, only datasets with an empirical kurtosis of the simulated structural shocks of at least 3.6 are accepted.

A total of M=1000 datasets are simulated for each variance specification. In the following, let $\hat{\theta}_{jk,i}(m)$ for $(j, k \in \{1, 2\})$ denote the estimated impulse response function in variable j caused by structural shock k after i periods based on estimates for the m-th dataset. Our metric of comparison is then given as:

$$\text{MSE}\left(\theta_{jk}\right)_h = \frac{1}{M} \sum_{m=1}^{M} \left(\sum_{i=0}^{h} \left(\hat{\theta}_{jk,i}(m) - \theta_{jk,i} \right)^2 \right). \tag{1.4.1}$$

We choose horizon h=5 as in Lütkepohl and Schlaak (2018). To compute parameter estimates, we use algorithm EM-1 for the SV-SVAR model. For the BP-SVAR we maximize a Gaussian likelihood over a grid of possible break-dates. Furthermore, for the MS-SVARs we use the EM algorithm outlined in Herwartz and Lütkepohl (2014). Finally, for the GARCH-SVAR we compute ML estimates based on the procedure of Lanne and Saikkonen (2007). Note that the estimated models rely on different normalizing constraints for the structural shocks which is why we rescale all impulse response functions to unit shock size.

The results of the simulation study are provided in Table 1.1. For improved readability, we report relative MSEs in comparison to the correctly specified model. Overall, we find that the SV-SVAR model performs very well regardless of the true DGP or the sample size

Table 1.1: Cumulated MSEs at horizon $h = 5$

		T=200				T=500			
		θ_{11}	θ_{12}	θ_{21}	θ_{22}	θ_{11}	θ_{12}	θ_{21}	θ_{22}
BP-DGP	BP	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MS(2)	1.00	1.01	1.01	1.00	1.00	1.00	1.00	1.00
	GARCH	1.61	1.79	1.58	1.14	1.20	1.24	1.19	1.04
	SV	1.22	1.32	1.21	1.06	1.09	1.11	1.09	1.03
MS-DGP	BP	3.23	3.72	4.71	1.37	7.98	9.75	12.01	1.79
	MS(2)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	GARCH	3.89	4.43	3.45	1.26	3.52	4.14	3.90	1.28
	SV	1.74	1.94	1.54	1.08	1.23	1.30	1.29	1.08
GARCH-DGP	BP	3.88	4.23	2.56	1.26	11.58	12.67	4.99	1.47
	MS(2)	8.18	9.01	3.67	1.29	21.71	24.52	7.14	1.38
	MS(3)	3.95	4.23	1.98	1.13	5.19	5.60	2.22	1.19
	GARCH	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	SV	1.15	1.16	1.04	1.03	1.10	1.10	1.06	1.04
SV-DGP	BP	3.35	3.53	2.26	1.18	8.52	9.52	4.36	1.35
	MS(2)	5.62	6.10	3.28	1.19	13.60	15.22	5.72	1.30
	MS(3)	4.20	4.58	2.02	1.12	3.12	3.34	1.74	1.14
	GARCH	2.41	2.60	1.77	1.15	1.50	1.54	1.23	1.07
	SV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Note: MSEs of impulse response functions calculated as in (1.4.1) and displayed relative to true model MSEs.

for each of the impulse responses θ_{jk}. In fact, the largest deterioration that we register in terms of MSE is found to be 94% in θ_{12} of the Markov Switching DGP. This contrasts all other models included into the Monte Carlo study which are subject to a very heterogeneous performance. Whenever they are inherently misspecified, we find relative MSE of much higher orders of magnitude. For example, with detoriations of up to 24 times, estimates based on a MS(2)-SVAR seem completely unreliable for data generated by the SV and GARCH DGPs. Admittably, the complexity of a MS model can be increased by adding additional states. Therefore, we also report estimates based on a MS(3) for the SV and GARCH DGPs. While indeed this yields substantial improvements, we still register detoriations in MSE up to 460%.

If we compare the IRF estimates of the SV-SVAR to all other misspecified models in a certain DGP, we find it to perform strictly better in two out of three DGPs. Specifically, for residuals generated by a MS(2) and GARCH model, all impulse responses estimated by the SV-SVAR have lower cumulative MSEs than the other misspecified models. Only if the structural errors are simulated with a one time shift in the variance there is no clear

advantage of the SV model over the MS model. However, this is not surprising given that the latter is perfectly able to capture such sudden shifts in the variance.

Finally, we find that the SV-SVAR model also compares favorable if its performance is directly matched to the most related model, the GARCH-SVAR. In particular, the SV-SVAR model always performs better when both models are misspecified. Furthermore, while there is almost no deterioration in the MSE of the SV-SVAR estimates in a GARCH-DGP, the other way around we record substantially higher relative MSEs.

Summing up, our small simulation study yields promising results indicating that the SV-SVAR may be a safe choice to identify structural shocks for different types of heteroskedasticity patterns and to estimate the corresponding impulse response functions.

1.5 Interdependence between monetary policy and stock markets

SVAR models are a widely used tool to investigate the dynamic effects of monetary policy, see e.g. Ramey (2016) for an extensive overview of the literature. To identify the structural shocks, the most simple way uses a Cholesky decomposition of the covariance matrix in a reduced form VAR with the policy variable ordered last (Bernanke et al., 2005; Christiano et al., 1999). In accordance with theoretical economic models featuring nominal rigidities (Christiano et al., 2005), this implies that only the central bank is allowed to respond to all movements in the economy on impact, while all variables in the system ordered above react with at least one lag to a monetary policy shock. While this seems reasonable for slowly moving real macroeconomic aggregates, such a recursivity assumption becomes unrealistic once fast moving financial variables are included into the SVAR analysis.

Over the last years, many other identification schemes have been developed to study the effects of monetary policy shocks avoiding the use of a recursiveness assumption. Bjørnland and Leitemo (2009) propose to identify a monetary policy shock under the presence of stock market returns by a combination of short- and long-run restrictions. Besides zero impact restrictions on real variables, a monetary policy shock is furthermore restricted to have a zero long-term impact on stock markets. This additional restriction allows the authors to disentangle monetary policy innovations from financial shocks.

Another promising way to address identification in presence of fast moving variables are Proxy-SVARs based on external instruments. If there is an external time series that is correlated with the structural shock to be identified and uncorrelated with all other shocks in the system, no exclusion restrictions are necessary at all. Recently, many narrative measures have been proposed to identify monetary policy shocks. Widely used are proxies constructed based on either readings of Federal Open Market Committee (FOMC) minutes (e.g. Coibion

(2012) and Romer and Romer (2004)) or changes in high frequency future prices in a narrow window around FOMC meetings (Faust et al., 2004; Gertler and Karadi, 2015; Nakamura and Steinsson, 2018).[8]

Finally, heteroskedasticity can be exploited to identify the interdependence between monetary policy and financial variables. For example, Rigobon (2003) combines identification via heteroskedasticity and economic narratives to estimate the reaction of monetary policy to stock market returns. Also Wright (2012) links economic and statistical identification within a daily SVAR, assuming that monetary policy shocks have a higher variance around FOMC meetings. Even if no economic narrative is available for the statistically identified structural parameters, the heteroskedastic SVAR model can be used to formally test conventional identifying restrictions. For example, Lütkepohl and Netšunajev (2017b) review various heteroskedastic SVAR models and use them to test the combination of exclusion restrictions employed by Bjørnland and Leitemo (2009).[9] Their analysis includes a GARCH-SVAR, two specifications of a MS-SVAR and a SVAR featuring a Smooth Transition model for the variance (STVAR).

To illustrate the use of our methods, we repeat the analysis of Lütkepohl and Netšunajev (2017b) complemented by the SV-SVAR model. Besides testing the short- and long-run restrictions used by Bjørnland and Leitemo (2009), we additionally test Proxy-SVAR restrictions that arise if the narrative series of Romer and Romer (2004) and Gertler and Karadi (2015) are used as instruments for a monetary policy shock.

1.5.1 Model and identifying constraints

The VAR model of Bjørnland and Leitemo (2009) is based on the following variables: $y_t = (q_t, \pi_t, c_t, \Delta s_t, r_t)'$, where q_t is a linearly detrended index of log industrial production, π_t the annualized inflation rate based on consumer prices, c_t the annualized change in log commodity prices as measured by the World Bank, Δs_t S&P500 real stock returns and r_t the federal funds rate. For detailed description of the data sources, transformations and time series plots see Appendix 1.C. As in Lütkepohl and Netšunajev (2017b), we use an extended sample period including data from 1970M1 until 2007M6, summing up to a total of 450 observations. To make our results comparable, we also choose $p = 3$ lags which is supported by the AIC applied within a linear VAR model.

[8] Yet another branch of the literature relies on sign restrictions of the impulse response functions (Canova and De Nicoló, 2002; Faust, 1998; Uhlig, 2005) or on a combination of sign restrictions and information in proxy variables (Braun and Brüggemann, 2019).

[9] See also Lütkepohl and Netšunajev (2017a) for a similar analysis based on a Smooth Transition SVAR model only.

In our analysis, we test the following set of short- and long-run constraints used by Bjørnland and Leitemo (2009):

$$B = \begin{pmatrix} * & 0 & 0 & 0 & 0 \\ * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & * \\ * & * & * & * & * \end{pmatrix}, \qquad \text{and} \qquad \Xi_\infty = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & 0 \\ * & * & * & * & * \end{pmatrix}, \tag{1.5.1}$$

where $\Xi_\infty = (I_K - A_1 - \ldots - A_p)^{-1}B$ is the long-run impact matrix of the structural shocks on y_t. Note than an asterisk means that the corresponding entry in B and Ξ_∞ is left unrestricted. The last columns of B and Ξ_∞ correspond to the reaction of y_t to a monetary policy shock. Economic activity, consumer prices and commodity prices are only allowed to respond with a delay of one month to a monetary policy shock, while stock markets are allowed to react contemporaneously. However, in the long run, a monetary policy shock is assumed to have a zero effect on the stock market. The fourth column of B corresponds to a stock price shock which is constrained to have no contemporaneous impact on activity and prices while the central bank is allowed to adjust the interest rates within the same period. The remaining shocks do not have an economic interpretation. To identify the model, Bjørnland and Leitemo (2009) simply disentangle these shocks by imposing a recursivity assumption. As outlined before, restrictions (1.5.1) are overidentifying in heteroskedastic SVAR models and can be tested against the data. In line with Lütkepohl and Netšunajev (2017b), the following set of restrictions is tested:

R1: Both, B and Ξ_∞ restricted as in (1.5.1).

R2: Only the last two columns of B and Ξ_∞ are restricted as in (1.5.1).

R3: Only B is restricted as in (1.5.1).

We further contribute to the literature by testing Proxy-SVAR restrictions that arise if an external instrument z is used for identification of a structural shock. The identifying assumptions are that the instrument is correlated with the structural shock it is designed for (relevance) and uncorrelated with all remaining shocks (exogeneity). Without loss of generality, assume that the first shock is identified by the instrument. Then, Mertens and Ravn (2013) show that the relevance and exogeneity assumption can be translated into the following set of linear restrictions on β_1, denoting the first column of B:

$$\beta_{21} = (\Sigma_{zu_1'}^{-1}\Sigma_{zu_2'})'\beta_{11}. \tag{1.5.2}$$

where $\beta_1 = (\beta_{11}, \beta_{21}')'$ with β_{11} scalar and $\beta_{21} \in \mathbb{R}^{K-1}$. Furthermore, $\Sigma_{zu'} = \text{Cov}(z, u') = [\Sigma_{zu_1'}, \Sigma_{zu_2'}]$ with $\Sigma_{zu_1'}$ scalar and $\Sigma_{zu_2'}' \in \mathbb{R}^{K-1}$. In practice, elements of $\Sigma_{zu'}$ are estimated by the corresponding sample moments.[10] To identify a monetary policy shock, we use the narrative series constructed by Romer and Romer (2004) (RR henceforth) and Gertler and Karadi (2015) (GK henceforth). We test the following Proxy-SVAR restrictions that arise when the first column of B is identified via either RR's or GK's instrument:

R4rr: IV moment restrictions (1.5.2) based on the RR shock.

R4gk: IV moment restrictions (1.5.2) based on the GK shock.

We use the RR series extended by Wieland and Yang (2016) which is available for the whole sample. The GK shock is only available for a subsample starting in 1990M1. We use their baseline series which is constructed based on the three months ahead monthly fed funds futures.[11] Time series plots of both series are available in Appendix 1.C.

1.5.2 Statistical analysis

Before we start testing the aforementioned restrictions, we conduct formal model selection for the variance specification of the structural shocks. By means of information criteria and residual plots, we compare the SV model to those models included in Lütkepohl and Netšunajev (2017b): a GARCH, a Smooth Transition (ST) and different specifications of a Markov Switching model. This allows us to directly compare our results.

Table 1.2 reports log-likelihood values, Akaike information criteria (AIC) and Bayesian information criteria (BIC) for a linear VAR and all heteroskedastic models. First of all, we highlight that there is only a small gain in terms of likelihood value of the SV model using the Monte Carlo based algorithm (EM-2) compared to the deterministic approximation (EM-1). To assess the Monte Carlo error of the estimates, we also report approximate 95%-confidence intervals based on an application of the batch means method and R =100,000 draws of the importance density.[12] Comparing the different models, our results suggest that including time-variation in the second moment is strongly supported by both information criteria. Moreover, among the heteroskedastic models we find that particularly models designed for financial variables are favored, that is the GARCH model and the SV model. This may be not surprising given that stock market returns are included in the system.

[10] In particular, at each M-step we compute $\hat{\Sigma}_{zu'} = N_z^{-1} \sum_{t=1}^{T} D_t \hat{u}_t z_t'$ where D_t is a dummy indicating whether the instrument is available at time t and $N_z = \sum_{t=1}^{T} D_t$.

[11] We repeat our analysis for the other instruments available in Gertler and Karadi (2015). The results do not change qualitatively.

[12] A formal test of Koopman et al. (2009) indicates that the variance of the importance weights is finite which further supports the validity of our likelihood estimates.

Table 1.2: Model selection by information criteria

	Linear	SV-EM1	SV-EM2	GARCH	STVAR	MS(2)	MS(3)
$\ln L$	−3159.3	−2680.4	−2677.9	−2763.6	−2878.3	−2827.4	−2775.3
AIC	6508.7	5590.9	5585.8	5757.2	5980.5	5878.8	5792.6
BIC	6898.4	6062.6	6057.6	6229.0	6440.0	6338.3	6289.0

Note: $\ln L$ - log-likelihood function, AIC=$-2\ln L + 2 \times n_p$ and BIC=$-2\ln L + \ln(T) \times n_p$ with n_p the number of free parameters. For SV-EM1 and SV-EM2, application of the batch means method yields approximate 95%-confidence intervals of [-2680.48,-2680.33] and [-2678.11,-2677.68], respectively.

Among all models considered, we find that the SV model performs best in terms of information criteria. In this regard, our results deviate from those of Lütkepohl and Netšunajev (2017b) who find that the MS(3) model provides the best description for this dataset.[13]

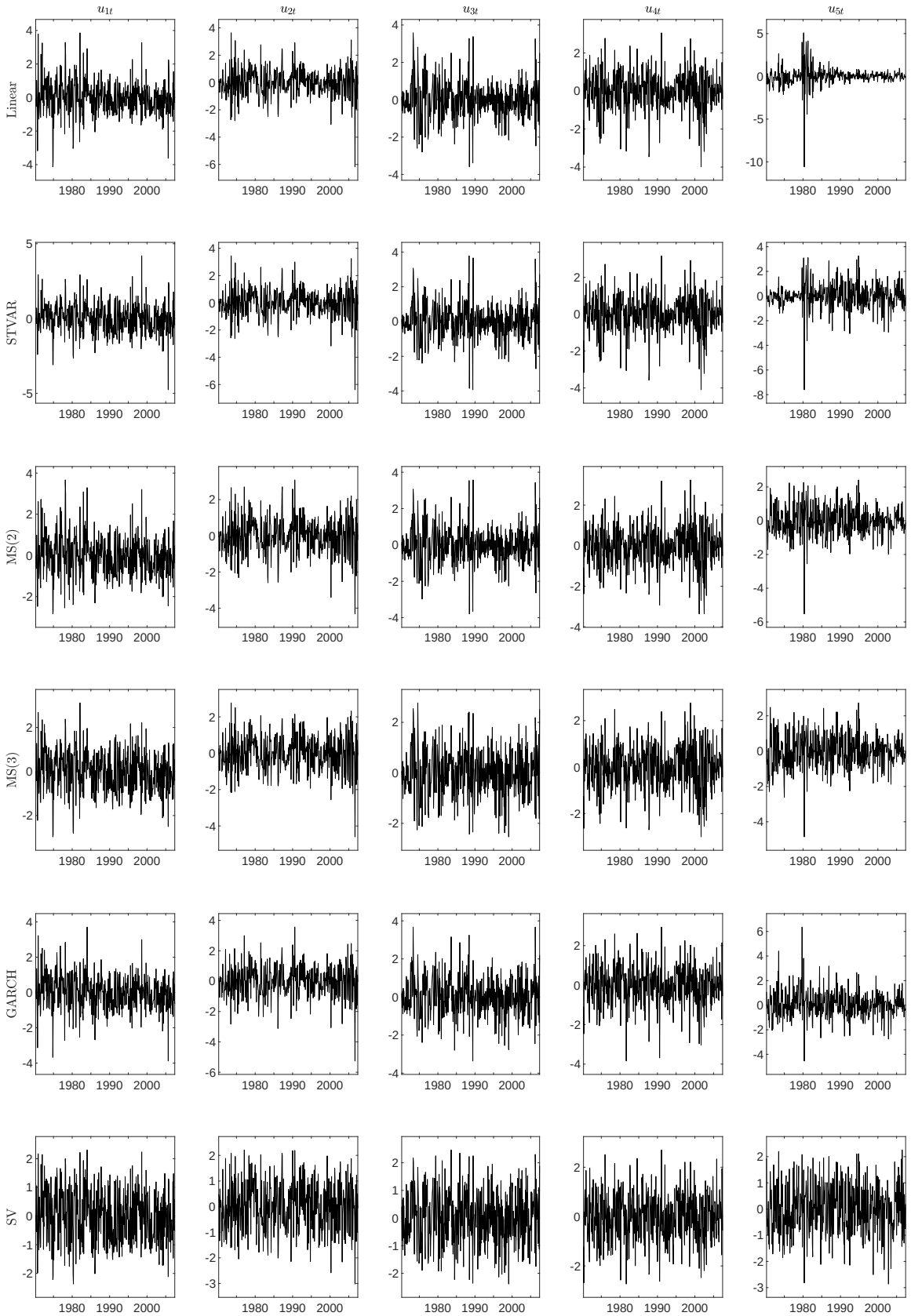
In accordance with Lütkepohl and Netšunajev (2017b), we also consider standardized residuals as an additional model checking device. Figure 1.1 provides a plot for the standardized residuals of all models computed as $\hat{u}_{it}/\hat{\sigma}_{ii,t}$ where $\hat{\sigma}^2_{ii,t}$ is the i-th diagonal entry of the estimated VAR covariance matrix $\hat{\Sigma}_t$. These plots clearly suggest that none of the other methods is fully satisfactory in yielding standardized residuals that seem to be homoskedastic and approximately normally distributed. However, for the SV-SVAR model, standardized residuals seem well behaved with no apparent heteroskedasticity and virtually no outliers. To confirm this impression, we provide complementary test results in Appendix 1.C.1 concerned with remaining heteroskedasticity and non-normality in standardized structural shocks. We find that only for the shocks of the SV-SVAR model, there is no evidence against both normality and homoskedasticity. To conclude, statistical analysis suggests that the proposed SV-SVAR is the most adequate for this application and we continue our analysis based on this model.

In order to test restrictions R1-R4 as overidentifying, it is necessary to count with enough heteroskedastic shocks ($r \geq K - 1$) to fully identify the impact matrix B. As described in Section 1.3.4, we apply a sequence of tests with $H_0 : r = r_0$ against $H_1 : r > r_0$ for $r_0 = 0, 1, \ldots K - 1$. The results are reported in Table 1.3. We find strong evidence that $r = K$ in our model, implying that the model can be fully identified by heteroskedasticity.

We continue our analysis and test the economically motivated restrictions R1-R4 as overidentifying. In Table 1.4 we provide Likelihood Ratio (LR) test statistics for the

[13]We also find a better ranking for the GARCH model compared to MS(3). Most likely, this is caused by a different estimation procedure. Specifically, Lütkepohl and Netšunajev (2017b) do only approximatively maximize the likelihood by a sequential estimation procedure.

Figure 1.1: Standardized residuals



Standardized residuals of linear model, ST model, MS(2) model, MS(3) model, GARCH model and SV-SVAR model.

Table 1.3: Tests of identification

	$Q_1(1)$	dof	p-value	$Q_2(1)$	dof	p-value
$r_0 = 0$	15.02	1	0.00	596.60	225	0.00
$r_0 = 1$	23.89	1	0.00	250.09	100	0.00
$r_0 = 2$	29.53	1	0.00	141.07	36	0.00
$r_0 = 3$	18.43	1	0.00	43.70	9	0.00
$r_0 = 4$	17.30	1	0.00	17.30	1	0.00
	$Q_1(3)$	dof	p-value	$Q_2(3)$	dof	p-value
$r_0 = 0$	52.34	3	0.00	1433.70	675	0.00
$r_0 = 1$	39.75	3	0.00	528.72	300	0.00
$r_0 = 2$	32.85	3	0.00	222.01	108	0.00
$r_0 = 3$	20.36	3	0.00	60.93	27	0.00
$r_0 = 4$	19.86	3	0.00	19.86	3	0.00

Note: Sequence of tests to check the number of heteroskedastic shocks in the system as introduced in Section 1.3.4 (Lanne and Saikkonen, 2007).

restrictions introduced previously.[14] Note that if B is identified under H_0, they have a standard asymptotic $\chi^2(n_r)$-distribution with n_r being the number of restrictions tested. Since we estimate the likelihood values with the help of importance sampling, we account for the Monte Carlo error by applying the batch means method and reporting approximate 95%-confidence intervals for the p-values.

In line with the findings of Lütkepohl and Netšunajev (2017b), our results suggest that R1, the restrictions of Bjørnland and Leitemo (2009), are rejected by the data. To make sure that this result does not come from the lower triangular block corresponding to the economically meaningless shocks, Lütkepohl and Netšunajev (2017b) also propose to test R2, which are the restrictions in B corresponding to the impact of monetary policy and stock market shocks. Within the SV model, these restrictions are also rejected. Testing for the zero restrictions in B in isolation (R3) also results in a rejection. However, in contrast to Lütkepohl and Netšunajev (2017b), we find that the long-run restriction is not rejected at any conventional significance level if R1 is tested against R3. This indicates that the long-run restriction is less of a problem, but rather are those in the short run. This key difference in the empirical analysis might arise due to more precisely estimated IRFs by the SV-SVAR model, strongly supported by statistical evidence. The fact that we are able to draw a different empirical conclusion emphasizes the importance of model selection in the context of heteroskedastic SVARs.

[14]This table is based on parameter estimates provided by EM-1. A corresponding Table based on EM-2 can be found in Appendix 1.C.1 and does not differ qualitatively.

Table 1.4: Test for overidentifying restrictions (EM-1)

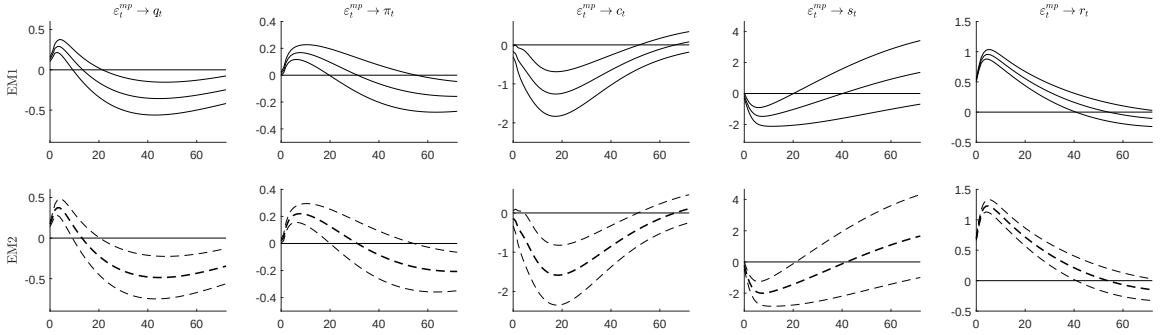
H_0	H_1	LR	dof	p-value	$p_{.025}$	$p_{.975}$
R1	UC	25.649	10	0.0042	0.0039	0.0046
R2	UC	22.750	7	0.0019	0.0017	0.0020
R3	UC	24.004	9	0.0043	0.0040	0.0046
R1	R3	1.653	1	0.1986	0.1957	0.2016
R4rr	UC	7.169	4	0.1272	0.0943	0.1705
R4gk	UC	256.480	4	0.0000	0.0000	0.0000

Note: For details about overidentifying restrictions see Section 1.5.1. Likelihood ratio test statistics are computed as $2\left(\ln L_{H_1} - \ln L_{H_0}\right)$ and are approximatively χ^2-distributed under H_0. Right columns report approximate 95%-confidence intervals for the p-value resulting from an application of the batch means method to the LR test statistic.

With respect to the Proxy-SVAR restrictions, we find that identifying a monetary policy shock with the shock series of Gertler and Karadi (2015) is strongly rejected by the data with a likelihood ratio test statistic exceeding 250. In turn, identification via the narrative series of Romer and Romer (2004) cannot be rejected at any conventional significance level. To further understand these results, we compute sample correlations of the instruments z with $\hat{\varepsilon}$, the estimated structural shocks of the unconstrained SV-SVAR model. For GK, we find $\text{Corr}(z^{GK}, \hat{\varepsilon}) = (0.039, -0.067, 0.050, -0.242, 0.419)$, while for RR, $\text{Corr}(z^{RR}, \hat{\varepsilon}) = (0.042, 0.005, 0.031, -0.021, 0.453)$. While both shocks are subject to a strong correlation with one of the statistically identified shocks, the instrument of GK is highly correlated with at least one additional shock. This clearly violates the exogeneity condition on the instrument. Thereby, our results support the argument of Ramey (2016) who questions the exogeneity of the GK instrument finding that it is autocorrelated and predictable by Greenbook variables. In turn, for the RR shock we find that there is little correlation with the remaining structural residuals of the SVAR. This clearly explains why identification via the RR shock is not rejected. Since the Proxy-SVAR restrictions based on RR cannot be rejected, we can interpret the last shock of the unconstrained model as a monetary policy shock for which $\text{Corr}(z^{RR}, \hat{\varepsilon}_5) = 0.45$. In Figure 1.2 we plot impulse response functions (IRFs) up to 72 months (6 years) of the system variables in response to a monetary policy shock. Besides point estimates, we provide 68% asymptotic confidence intervals. Again, we note that there is qualitatively no difference in using EM-1 or EM-2 to compute the estimates and corresponding standard errors.[15] The IRFs and their asymptotic confidence intervals coincide for all variables at all horizons. In line with the IRFs computed by Lütkepohl and

[15]There is only a slight difference in scaling of impulse responses because of a slightly rescaled monetary policy shock in EM-2.

Figure 1.2: Impulse responses in monetary SVAR



IRFs up to a horizon of 72 months of a monetary policy shock with 68% confidence bounds. Figures compare estimates based on EM-1 (solid line) and EM-2 (dashed line) with corresponding asymptotic confidence intervals.

Netšunajev (2017b) based on other heteroskedastic models, an unexpected tightening in monetary policy is associated with a puzzling short-term increase in activity and prices before they reach negative values on the medium and long term. In turn, commodity prices as well as stock market returns are found to react significantly negative in the short run. This fact seems reasonable given that one would expect a shift in demand towards risk free assets.

1.6 Conclusion

In this paper, we have considered stochastic volatility to identify structural parameters of SVAR models. The resulting model (SV-SVAR) can generate patterns of heteroskedasticity which are very typical in VAR analysis and therefore, we expect it to be useful in a wide range of applications.

We discussed conditions for full and partial identification and proposed to estimate the model by Gaussian Maximum Likelihood. For this purpose, we developed two EM algorithms which approximate the intractable E-step to a different extent. One algorithm is based on a Laplace approximation while the other relies on MCMC integration. We leave the choice of algorithm to individual preferences, but find that in practice little is gained by using the computationally more burdensome Monte Carlo EM. Besides discussing optimization, we stated the main properties of the estimator and present tools to approximate the asymptotic covariance matrix. Tests considered by Lanne and Saikkonen (2007) can be used to determine the number of heteroskedastic shocks and to test for identification.

To demonstrate the flexibility of the SV-SVAR model, we conducted a Monte Carlo study investigating how precise Impulse Response Functions are estimated under misspecification of the variance process. In contrast to alternative heteroskedastic SVARs, we find that the proposed model performs very well regardless of the DGP specified for the variance.

In an empirical application, we have revisited the model of Bjørnland and Leitemo (2009) who rely on a combination of short- and long-run restrictions to disentangle monetary policy from stock market shocks. Formal model selection strongly supports a SV specification in the variance if compared to other heteroskedastic SVARs used by Lütkepohl and Netšunajev (2017b) in this context. The SV-SVAR is used to test the exclusion restrictions of Bjørnland and Leitemo (2009) as overidentifying, and additionally test Proxy-SVAR restrictions that arise if external instruments are used to identify a monetary policy shock. We find no evidence against identification via the instrument of Romer and Romer (2004) and using a certain long-run restriction to disentangle monetary policy shocks.

References

Bernanke, B. S. and I. Mihov (1998). "Measuring monetary policy". *Quarterly Journal of Economics* 113 (3), 869–902.

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.

Bjørnland, H. C. and K. Leitemo (2009). "Identifying the interdependence between US monetary policy and the stock market". *Journal of Monetary Economics* 56 (2), 275–282.

Blanchard, O. J. and D. Quah (1989). "The dynamic effects of aggregate demand and supply disturbances". *The American Economic Review* 79 (4), 655–673.

Braun, R. and R. Brüggemann (2019). "Identification of SVAR models by combining sign restrictions with external instruments". *Working Paper, University of Konstanz*.

Brüggemann, R., C. Jentsch, and C. Trenkler (2016). "Inference in VARs with conditional heteroskedasticity of unknown form". *Journal of Econometrics* 191 (1), 69–85.

Caffo, B. S., W. Jank, and G. L. Jones (2005). "Ascent-based Monte Carlo expectation-maximization". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67 (2), 235–251.

Canova, F. and G. De Nicoló (2002). "Monetary disturbances matter for business fluctuations in the G-7". *Journal of Monetary Economics* 49, 1131–1159.

Cappé, O., E. Moulines, and T. Ryden (2005). *Inference in Hidden Markov Models*. Springer.

Carriero, A., T. E. Clark, and M. Marcellino (2019). "The identifying information in vector autoregressions with time-varying volatilities: An application to endogenous uncertainty". *Working Paper*.

Chan, J. C. C. (2017). "The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling". *Journal of Business & Economic Statistics* 35 (1), 17–28.

Chan, J. C. C. and A. L. Grant (2016). "On the observed-data deviance information criterion for volatility modeling". *Journal of Financial Econometrics* 14 (4), 772–802.

Chan, J. C. C. and I. Jeliazkov (2009). "Efficient simulation and integrated likelihood estimation in state space models". *International Journal of Mathematical Modelling and Numerical Optimisation* 1 (1-2), 101–120.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999). "Monetary policy shocks: What have we learned and to what end?" In: *Handbook of Macroeconomics*. Ed. by Taylor, J. B. and Woodford, M. Vol. 1. Handbook of Macroeconomics. Elsevier. Chap. 2, 65–148.

– (2005). "Nominal rigidities and the dynamic effects of a shock to monetary policy". *Journal of Political Economy* 113 (1), 1–45.

Coibion, O. (2012). "Are the effects of monetary policy shocks big or small?" *American Economic Journal: Macroeconomics* 4 (2), 1–32.

Danielsson, J. and J. F. Richard (1993). "Accelerated Gaussian importance sampler with application to dynamic latent variable models". *Journal of Applied Econometrics* 8, 153–173.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). "Maximum likelihood from incomplete data via the EM algorithm". *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 39, 1–38.

Durbin, J. and S. J. Koopman (1997). "Monte Carlo maximum likelihood estimation for non-Gaussian state space models". *Biometrika* 84 (3), 669–684.

– (2000). "Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 62 (1), 3–56.

Faust, J. (1998). "The robustness of identified VAR conclusions about money". *Carnegie-Rochester Conference Series on Public Policy* 49, 207–244.

Faust, J., E. T. Swanson, and J. H. Wright (2004). "Identifying VARs based on high frequency futures data". *Journal of Monetary Economics* 51 (6), 1107–1131.

Fernández-Villaverde, J. and J. F. Rubio-Ramírez (2007). "Estimating macroeconomic models: A likelihood approach". *Review of Economic Studies* 74 (4), 1059–1087.

Gertler, M. and P. Karadi (2015). "Monetary policy surprises, credit costs, and economic activity". *American Economic Journal-Macroeconomics* 7 (1), 44–76.

Gourieroux, C., A. Monfort, and J. P. Renne (2017). "Statistical inference for independent component analysis: Application to structural VAR models". *Journal of Econometrics* 196 (1), 111–126.

Harvey, A., E. Ruiz, and N. Shephard (1994). "Multivariate stochastic variance models". *Review of Economic Studies* 61 (2), 247–264.

Herwartz, H. and H. Lütkepohl (2014). "Structural vector autoregressions with Markov switching: Combining conventional with statistical identification of shocks". *Journal of Econometrics* 183 (1), 104–116.

Jacquier, E., N. G. Polson, and P. E. Rossi (1994). "Bayesian analysis of stochastic volatility models". *Journal of Business & Economic Statistics* 12 (4), 371–389.

Justiniano, A. and G. E. Primiceri (2008). "The time-varying volatility of macroeconomic fluctuations". *American Economic Review* 98 (3), 604–641.

Kilian, L. and H. Lütkepohl (2017). *Structural Vector Autoregressive Analysis*. Themes in Modern Econometrics. Cambridge University Press.

Kim, S., N. Shephard, and S. Chib (1998). "Stochastic volatility: Likelihood inference and comparison with ARCH models". *Review of Economic Studies* 65 (3), 361–393.

Koop, G. and D. Korobilis (2010). "Bayesian multivariate time series methods for empirical macroeconomics". *Foundations and Trends in Econometrics* 3 (4), 267–358.

Koopman, S. J., N. Shephard, and D. Creal (2009). "Testing the assumptions behind importance sampling". *Journal of Econometrics* 149 (1), 2–11.

Lanne, M., H. Lütkepohl, and K. Maciejowska (2010). "Structural vector autoregressions with Markov switching". *Journal of Economic Dynamics & Control* 34 (2), 121–131.

Lanne, M., M. Meitz, and P. Saikkonen (2017). "Identification and estimation of non-Gaussian structural vector autoregressions". *Journal of Econometrics* 196 (2), 288–304.

Lanne, M. and P. Saikkonen (2007). "A multivariate generalized orthogonal factor GARCH model". *Journal of Business & Economic Statistics* 25 (1), 61–75.

Lewis, D. J. (2018). "Identifying shocks via time-varying volatility". *Federal Reserve Bank of New York Staff Report* (871).

Louis, T. A. (1982). "Finding the observed information matrix when using the EM algorithm". *Journal of the Royal Statistical Society Series B-Methodological* 44 (2), 226–233.

Lütkepohl, H. and A. Netšunajev (2017a). "Structural vector autoregressions with smooth transition in variances". *Journal of Economic Dynamics & Control* 84, 43–57.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer Science & Business Media.

Lütkepohl, H. and G. Milunovich (2016). "Testing for identification in SVAR-GARCH models". *Journal of Economic Dynamics and Control* 73, 241–258.

Lütkepohl, H. and A. Netšunajev (2017b). "Structural vector autoregressions with heteroskedasticity: A review of different volatility models". *Econometrics and Statistics* 1, 2–18.

Lütkepohl, H. and T. Schlaak (2018). "Choosing between different time-varying volatility models for structural vector autoregressive analysis". *Oxford Bulletin of Economics and Statistics* (forthcoming).

Mahieu, R. J. and P. C. Schotman (1998). "An empirical application of stochastic volatility models". *Journal of Applied Econometrics* 13 (4), 333–359.

Maxand, S. (2018). "Identification of independent structural shocks in the presence of multiple Gaussian components". *Econometrics and Statistics*.

McLachlan, G. and T. Krishnan (2007). *The EM Algorithm and Extensions*. Vol. 382. John Wiley & Sons.

Melino, A. and S. M. Turnbull (1990). "Pricing foreign-currency options with stochastic volatility". *Journal of Econometrics* 45 (1-2), 239–265.

Mertens, K. and M. O. Ravn (2013). "The dynamic effects of personal and corporate income tax changes in the United States". *American Economic Review* 103 (4), 1212–1247.

Montiel-Olea José, L, J. H. Stock, and M. W. Watson (2016). "Uniform inference in SVARs identified with external instruments". *Harvard Manuscript*.

Nakamura, E. and J. Steinsson (2018). "High frequency identification of monetary non-neutrality: The information effect". *Quarterly Journal of Economics* (forthcoming).

Neal, R. M. and G. E. Hinton (1998). "A view of the EM algorithm that justifies incremental, sparse, and other variants". In: *Learning in graphical models*. Springer, 355–368.

Normandin, M. and L. Phaneuf (2004). "Monetary policy shocks: Testing identification conditions under time-varying conditional volatility". *Journal of Monetary Economics* 51 (6), 1217–1243.

Primiceri, G. E. (2005). "Time varying structural vector autoregressions and monetary policy". *Review of Economic Studies* 72 (3), 821–852.

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

Rigobon, R. (2003). "Identification through heteroskedasticity". *Review of Economics and Statistics* 85 (4), 777–792.

Romer, C. D. and D. H. Romer (2004). "A new measure of monetary shocks: Derivation and implications". *American Economic Review* 94 (4), 1055–1084.

Roweis, S. and Z. Ghahramani (2001). "Learning nonlinear dynamical systems using the Expectation–Maximization algorithm". *Kalman filtering and neural networks* 6, 175–220.

Rue, H., S. Martino, and N. Chopin (2009). "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations". *Journal of the Royal Statistical Society Series B-Statistical Methodology* 71 (2), 319–392.

Ruiz, E. (1994). "Quasi-maximum likelihood estimation of stochastic volatility models". *Journal of Econometrics* 63 (1), 289–306.

Sims, C. A. (1980). "Macroeconomics and reality". *Econometrica* 48 (1), 1–48.

Stock, J. H. and M. W. Watson (2012a). "Disentangling the channels of the 2007-09 recession". *Brookings Papers on Economic Activity* 43 (1), 81–156.

Uhlig, H. (2005). "What are the effects of monetary policy on output? Results from an agnostic identification procedure". *Journal of Monetary Economics* 52, 381–419.

Van der Weide, R. (2002). "GO-GARCH: A multivariate generalized orthogonal GARCH model". *Journal of Applied Econometrics* 17 (5), 549–564.

Wieland, J. F. and M.-J. Yang (2016). "Financial dampening". *NBER Working Paper 22141*.

Wright, J. H. (2012). "What does monetary policy do to long-term interest rates at the zero lower bound?" *The Economic Journal* 122 (564), 447–466.

Appendix 1.A Derivations and proofs

To ensure identification of impact matrix B in model (1.2.1)-(1.2.4) we show that under sufficient heterogeneity in the second moments of the structural shocks, i.e. $r \geq K-1$, there is no B^* different from B except for column permutations and sign changes which yields an observationally equivalent model with the same time-varying second moment properties in reduced form errors u_t for all $t = 1,\dots,T$. Furthermore, for $r < K-1$, we show which parameters in impact matrix B are identified and which are not. This also includes one possible identification scheme for this scenario. We start with the derivation of the autocovariance function of the second moments of reduced form residuals u.

1.A.1 Autocovariance function of the second moments

The autocovariance function of the second moments of the structural shocks is:

$$\operatorname{Cov}\left(\operatorname{vec}\left(\varepsilon_t\varepsilon_t'\right), \operatorname{vec}\left(\varepsilon_{t+\tau}\varepsilon_{t+\tau}'\right)\right) = \left[E\left(\varepsilon_{it}\varepsilon_{jt}\varepsilon_{k,t+\tau}\varepsilon_{l,t+\tau}\right) - E\left(\varepsilon_{it}\varepsilon_{jt}\right)E\left(\varepsilon_{k,t+\tau}\varepsilon_{l,t+\tau}\right)\right]_{ijkl}.$$

The entries of this expression are only non-zero if both $i = j = k = l$ and $i \leq r$ hold for $i,j,k,l \in \{1,\dots,K\}$ due to the structure of the SV-SVAR model (1.2.1)-(1.2.4). Thus, it is

$$\operatorname{Cov}\left(\operatorname{vec}\left(\varepsilon_t\varepsilon_t'\right), \operatorname{vec}\left(\varepsilon_{t+\tau}\varepsilon_{t+\tau}'\right)\right) = G_K M_\tau G_K',$$

with G_K being a selection matrix and M_τ as defined in Section 1.2. Briefly recap that we define (Lewis, 2018):

$$\xi_t = \operatorname{vech}\left(u_t u_t'\right) = L_K \operatorname{vec}\left(u_t u_t'\right).$$

Consequently, the autocovariance function in ξ reads:

$$\begin{aligned}
\operatorname{Cov}\left(\xi_t, \xi_{t+\tau}\right) &= L_K \operatorname{Cov}\left(\operatorname{vec}\left(u_t u_t'\right), \operatorname{vec}\left(u_{t+\tau}u_{t+\tau}'\right)\right) L_K' \\
&= L_K \left(B \otimes B\right) \operatorname{Cov}\left(\operatorname{vec}\left(\varepsilon_t\varepsilon_t'\right), \operatorname{vec}\left(\varepsilon_{t+\tau}\varepsilon_{t+\tau}'\right)\right) \left(B \otimes B\right)' L_K' \\
&= L_K \left(B \otimes B\right) G_K M_\tau G_K' \left(B \otimes B\right)' L_K'.
\end{aligned}$$

1.A.2 Proof of Proposition 1.1

Proof. Suppose $\tilde{B} = BQ$ and $\tilde{\varepsilon}_t = Q^{-1}\varepsilon_t$ with $Q = \begin{pmatrix} Q_1 & Q_3 \\ Q_2 & Q_4 \end{pmatrix}$, where $Q_1 \in \mathbb{R}^{r\times r}$, $Q_2, Q_3' \in \mathbb{R}^{(K-r)\times r}$ and $Q_4 \in \mathbb{R}^{(K-r)\times(K-r)}$ define an observationally equivalent model satisfying (1.2.8) and (1.2.9). Due to (1.2.8), it is:

$$\Sigma_u = BB' = \tilde{B}\tilde{B}' = BQQ'B.$$

Hence, Q has to be an orthogonal matrix, i.e. $QQ' = I_K$. To keep the autocovariance function in the second moment of the reduced form errors, it is:

$$\begin{aligned}\operatorname{Cov}(\xi_t, \xi_{t+\tau}) &= L_K \left(\tilde{B} \otimes \tilde{B}\right) \operatorname{Cov}\left(\operatorname{vec}\left(\tilde{\varepsilon}_t\tilde{\varepsilon}_t'\right), \operatorname{vec}\left(\tilde{\varepsilon}_{t+\tau}\tilde{\varepsilon}_{t+\tau}'\right)\right) \left(\tilde{B} \otimes \tilde{B}\right)' L_K' \\ &= L_K \left(\tilde{B} \otimes \tilde{B}\right) (Q \otimes Q)' G_K M_\tau G_K' (Q \otimes Q) \left(\tilde{B} \otimes \tilde{B}\right)' L_K'.\end{aligned}$$

As we still have a SV-SVAR model, $(Q \otimes Q)' G_K M_\tau G_K' (Q \otimes Q)$ must have the same form as $G_K M_\tau G_K'$, i.e. it is a diagonal matrix with exactly r non-zero entries $\tilde{\gamma}_i(\tau)$ located at elements $(i-1)K + i$ for $i = 1, \ldots, r$ on the diagonal. Thus, it is:

$$G_K \begin{pmatrix} \tilde{\gamma}_1(\tau) & & & \\ & \ddots & & \\ & & \tilde{\gamma}_r(\tau) & \\ & & & 0_{K-r} \end{pmatrix} G_K' = (Q \otimes Q)' G_K \begin{pmatrix} \gamma_1(\tau) & & & \\ & \ddots & & \\ & & \gamma_r(\tau) & \\ & & & 0_{K-r} \end{pmatrix} G_K' (Q \otimes Q).$$

This yields the following conditions:

$$\forall i = 1, \ldots, r : \sum_{l=1}^{r} q_{li}^4 \gamma_l(\tau) = \tilde{\gamma}_l(\tau) \neq 0, \tag{1.A.1}$$

$$\forall a_j \in \{0,1,2,3\} : \sum_{j=1}^{K} a_j = 4 : \sum_{l=1}^{r} \left(\prod_{j=1}^{K} q_{lj}^{a_j} \right) \gamma_l(\tau) = 0. \tag{1.A.2}$$

Because of (1.A.2), it is $\sum_{l=1}^{r} q_{l\bullet}' \underbrace{q_{li}^2 q_{lj}}_{=:\lambda_{lij}} = 0$ for all $i, j \in \{1, \ldots, K\}$ with $i \neq j$. As it is an orthogonal matrix, row vectors $q_{l\bullet}$ of matrix Q are linearly independent such that $\lambda_{lij} = 0$ for all $l \in \{1, \ldots, r\}, i, j \in \{1, \ldots, K\} : i \neq j$. Consequently, considering the first r rows of Q, i.e. matrix (Q_1, Q_3), only one element per row can be different from zero. Due to the orthogonality, this element has to be ± 1.

Because of (1.A.1), in each column of $r \times r$ matrix Q_1 at least one element has to be non-zero. Following the previous argument, these r non-zero entries correspond to the r ± 1 entries

in (Q_1, Q_3). This directly implies that Q_3 is a zero matrix and Q_1 has exactly one element different from zero per row and column which is ± 1. Thus, Q_1 can be decomposed in DP where D is a diagonal matrix with ± 1 entries and P is a permutation matrix.
In addition, orthogonality of Q yields that Q_2 has to be a zero matrix. Finally, Q_4 has to be a $(K-r) \times (K-r)$ orthogonal matrix to satisfy $QQ' = I_K$. Therefore, block B_1 is unique up to permutation and sign changes.

□

1.A.3 Proof of Corollary 1.1

Using Proposition 1.1 shows that an observationally equivalent model with the same autocovariance function in the second moment of the reduced form errors can be obtained by $\tilde{B} = BQ$ if and only if Q has the structure $\begin{pmatrix} Q_1 & 0 \\ 0 & Q_4 \end{pmatrix}$, $Q_1 = DP$ with D a diagonal matrix with ± 1 entries on the diagonal, P a permutation matrix and $Q_4 \in \mathbb{R}^{(K-r)\times(K-r)}$ any orthogonal matrix. Thus, the decomposition $B = (B_1, B_2)$ with $B_1 \in \mathbb{R}^{K\times r}$ and $B_2 \in \mathbb{R}^{K\times(K-r)}$ yields uniqueness of B_1 apart from multiplication of its columns by -1 and permutation. Moreover, in case that $r = K - 1$, column vector B_2 is also unique up to multiplication with -1:

Proof. For $r = K - 1$, matrix Q_4 is a scalar with $Q_4^2 = 1 \Rightarrow Q_4 = \pm 1$. So, full matrix Q can be decomposed in a diagonal matrix with ± 1 entries and a permutation matrix having a one in the very last element. This proves the uniqueness of the full matrix B apart from sign reversal of its columns and permutation of its first r columns. □

1.A.4 Proof of Corollary 1.2

Proof. Let $Q = \begin{pmatrix} Q_1 & 0 \\ 0 & Q_4 \end{pmatrix}$ be a $K \times K$ matrix such that $BQ = \begin{pmatrix} B_{11}Q_1 & B_{21}Q_4 \\ B_{12}Q_1 & B_{22}Q_4 \end{pmatrix}$ has the same structure as B, i.e. $B_{22}Q_4$ is still a lower triangular matrix. Thereby, it directly follows that Q_4 is a lower triangular matrix itself. Moreover, because Q_4 is orthogonal, it is also normal and therefore diagonal. Any diagonal and orthogonal matrix has ± 1 entries on the diagonal. So, full matrix Q can be decomposed in a diagonal matrix D having ± 1 entries and a permutation matrix P having an identity block in the lower right $(K-r) \times (K-r)$ block. Thus, matrix B is unique up to multiplication of its columns with -1 and permutation of its first r columns. □

Appendix 1.B Estimation

1.B.1 Importance density

To derive the Gaussian approximation of the (unrestricted) IS density $\pi_G(h_i|\theta,\varepsilon_i)$ for $i = 1,\dots,r$, we closely follow the exposition of Chan and Grant (2016). We start with an application of Bayes' theorem which gives the zero variance importance density:

$$\log p(h_i|\theta,\varepsilon_i) \propto \log p(\varepsilon_i|\theta,h_i) + \log p(h_i). \tag{1.B.1}$$

The assumption of normality in both the transition and measurement equation gives:

$$\log p(h_i) \propto -\frac{1}{2}\left(h_i-\delta_i\right)' Q_i \left(h_i-\delta_i\right), \tag{1.B.2}$$

$$\log p(\varepsilon_{it}|\theta,h_{it}) \propto -\frac{1}{2}\left(h_{it}+\varepsilon_{it}^2 e^{-h_{it}}\right). \tag{1.B.3}$$

Since the measurement equation is nonlinear in h_i, the normalizing constant of the smoothing distribution in equation (1.B.1) is not known. An approximate distribution, however, can be obtained by a second order Taylor approximation of the measurement equation (1.B.3). The corresponding partial derivatives are given as:

$$\begin{aligned}
\frac{\partial \log p(\varepsilon_{it}|\theta,h_{it})}{\partial h_{it}} &= -\frac{1}{2}+\frac{1}{2}\varepsilon_{it}^2 e^{-h_{it}} =: f_{it} \quad \Rightarrow \quad f_i = (f_{i1},\dots,f_{iT})', \\
-\frac{\partial^2 \log p(\varepsilon_{it}|\theta,h_{it})}{\partial h_{it}^2} &= \frac{1}{2}\varepsilon_{it}^2 e^{-h_{it}} =: c_{it} \quad \Rightarrow \quad C_i = \operatorname{diag}\left([c_{i1},\dots,c_{iT}]'\right).
\end{aligned}$$

A second order Taylor approximation around $\tilde{h}_i^{(0)}$ then yields:

$$\begin{aligned}
\log p(\varepsilon_i|\theta,h_i) &\approx \log p(\varepsilon_i|\theta,\tilde{h}_i^{(0)}) + \left(h_i-\tilde{h}_i^{(0)}\right)' f_i - \frac{1}{2}\left(h_i-\tilde{h}_i^{(0)}\right)' C_i \left(h_i-\tilde{h}_i^{(0)}\right) \\
&= -\frac{1}{2}[h_i' C_i h_i - 2h_i' \underbrace{(f_i + C_i\tilde{h}_i^{(0)})}_{=:b_i}] + \text{constant}.
\end{aligned} \tag{1.B.4}$$

Combining (1.B.1), (1.B.2) and (1.B.4) provides an approximation of the smoothing distribution which takes the form of a normal kernel:

$$\log p(h_i|\theta,\varepsilon_i) \propto -\frac{1}{2}[h_i' \underbrace{(C_i+Q_i)}_{=:\bar{Q}_i} h_i - 2h_i'(b_i+Q_i\delta_i)].$$

Consequently, the approximate smoothing density is:

$$\pi_G\left(h_i|\theta,\varepsilon_i\right) \sim \mathcal{N}\left(\bar{\delta}_i, \bar{Q}_i^{-1}\right), \qquad \text{with} \quad \bar{\delta}_i = \bar{Q}_i^{-1}\left(b_i + Q_i\delta_i\right).$$

The restricted density $\pi_G^c\left(h_i|\theta,\varepsilon_i\right)$ is constructed as outlined in Section 1.3. Note that $\pi_G^c\left(h_i|\theta,\varepsilon_i\right)$ yields a good approximation only if $\tilde{h}_i^{(0)}$ is chosen appropriately. In the following, we sketch how the Newton Raphson method is used to evaluate the IS density at the mode of the smoothing distribution (1.B.1).

1.B.2 Newton Raphson method

The Newton-Raphson method is implemented as follows: h_i is initialized by some vector $h_i^{(0)}$ satisfying the linear constraint, i.e. $A_h h_i^{(0)} = \mu_i$. Then, $h_i^{(l)}$ is used to evaluate $\bar{Q}_i$, $\bar{\delta}_i$ and to iterate:

$$\begin{aligned}
\tilde{h}_i^{(l+1)} &= h_i^{(l)} + \bar{Q}_i^{-1}\left(-\bar{Q}_i h_i^{(l)} + \bar{\delta}_i\right) = \bar{Q}_i^{-1}\bar{\delta}_i, \\
h_i^{(l+1)} &= \tilde{h}_i^{(l+1)} - \bar{Q}_i^{-1}A_h'\left(A_h\bar{Q}_i^{-1}A_h'\right)^{-1}\left(A_h\tilde{h}_i^{(l+1)} - \mu_i\right),
\end{aligned}$$

for $l \geq 0$ until convergence, i.e. until $\left\|h_i^{(l+1)} - h_i^{(l)}\right\| < \varepsilon$ holds for a specified tolerance level ε.

1.B.3 EM algorithm

To fix notation, define the following quantities:

$$\begin{aligned}
Y^0 &:= (y_1, \ldots, y_T) && K \times T, \\
A &:= (\nu, A_1, \ldots, A_p) && K \times (Kp+1), \\
Y_t^0 &:= \left(y_{t-1}', \ldots, y_{t-p}'\right)' && Kp \times 1, \\
x_t &:= \left(1, \left(Y_t^0\right)'\right)' && (Kp+1) \times 1, \\
X &:= (x_1, \ldots, x_T) && (Kp+1) \times T, \\
y^0 &:= \mathrm{vec}(Y^0) && KT \times 1, \\
\alpha &:= \mathrm{vec}(A) && [K(Kp+1)] \times 1, \\
U &:= (u_1, \ldots, u_T) && K \times T, \\
u &:= \mathrm{vec}(U) && KT \times 1, \\
V_{(-1)} &:= (\exp(-h_1), \ldots, \exp(-h_T)) && K \times T.
\end{aligned}$$

Using this, VAR equation (1.2.1) can be compactly written as:

$$y^0 = Z\alpha + u,$$

with $Z = (X' \otimes I_K)$, $\mathrm{E}(uu') = \tilde{\Sigma}_u$. Note that its inverse is given by $\tilde{\Sigma}_u^{-1} = (\left[B^{-1}\right]' \otimes I_T)\Sigma_e^{-1}(B^{-1} \otimes I_T)$ where $\Sigma_e^{-1} = \mathrm{diag}(\mathrm{vec}(V_{(-1)}))$.

This yields the following compact representation of the complete data log-likelihood:

$$\begin{aligned}\mathcal{L}_c(\theta) \propto & -T\ln|B| - \frac{1}{2}\left(y^0 - Z\alpha\right)'\left(\left[B^{-1}\right]' \otimes I_T\right)\Sigma_e^{-1}\left(B^{-1} \otimes I_T\right)\left(y^0 - Z\alpha\right) \\ & + \sum_{i=1}^{r}\left\{-\frac{T}{2}\ln(s_i) + \frac{1}{2}\ln\left(1-\phi_i^2\right)\right. \\ & \left. -\frac{1}{2s_i}\left(\left[1-\phi_i^2\right]\left[h_{i1}-\mu_i\right]^2 + \sum_{t=2}^{T}\left(\left[h_{it}-\mu_i\right] - \phi_i\left[h_{i,t-1}-\mu_i\right]\right)^2\right)\right\}.\end{aligned} \tag{1.B.5}$$

Both algorithms EM-1 and EM-2 require some starting values. They are set in the same way for both alternatives. That is:

$$\begin{aligned}\hat{\alpha}^{(0)} &= \left(\left[(XX')^{-1}X\right] \otimes I_k\right)y^0, \\ \hat{B}^{(0)} &= (T^{-1}\hat{U}\hat{U}')^{\frac{1}{2}}Q, \quad \text{with } \hat{U} = Y^0 - \hat{A}X,\end{aligned}$$

where Q is a $K \times K$ orthogonal matrix uniformly drawn from the space of K-dimensional orthogonal matrices. Furthermore, we set the $r \times 1$ vectors:

$$\begin{aligned}\hat{\phi}^{(0)} &= [0.95, \ldots, 0.95]', \\ \hat{s}^{(0)} &= [0.02, \ldots, 0.02]',\end{aligned}$$

which correspond to persistent heteroskedasticity with initial kurtosis of about 3.7 for the estimated structural shocks $\hat{\varepsilon}_i, i = 1, \ldots, r$.
Note that in order to satisfy linear restriction (1.2.10) we set for $i = 1, \ldots, r$ and $l \geq 1$:

$$\hat{\mu}_i^{(l-1)} = -0.5\hat{s}_i^{(l-1)} \Big/ \left(1 - \left(\hat{\phi}_i^{(l-1)}\right)^2\right).$$

EM-1

Because of $\hat{\varepsilon}_t^{(l-1)} = \hat{B}^{(l-1)}(y_t - \hat{A}^{(l-1)}x_t)$, it is equivalent to condition the approximate smoothing densities π_G^c and their moments to $\left(\theta^{(l-1)}, \hat{\varepsilon}_i^{(l-1)}\right)$ or $\left(\theta^{(l-1)}, y\right)$, respectively. Based on starting values $\theta^{(0)} = \left[\left(\hat{\alpha}^{(0)}\right)', \mathrm{vec}\left(\hat{B}^{(0)}\right)', \left(\hat{\phi}^{(0)}\right)', \left(\hat{s}^{(0)}\right)'\right]'$, the EM algorithm iteratively cycles through the following steps for $l \geq 1$:

1. E-step: For $i = 1, \ldots, r$, evaluate the moments of the approximate smoothing densities, mean $\bar{\delta}_i^c$ and variance $\bar{Q}_i^{-1} - \bar{Q}_i^{-1} A_h' \left(A_h \bar{Q}_i^{-1} A_h'\right)^{-1} A_h \bar{Q}_i^{-1}$, as described in Appendix 1.B.1. Thereby, directly inverting $\bar{Q}_i$ is unnecessary costly since we only need its diagonal elements representing the marginal variances $\text{Var}(h_{it}|\theta^{(l-1)}, y)$ and the entries of the first off-diagonal corresponding to $\text{Cov}(h_{it}, h_{i,t-1}|\theta^{(l-1)}, y)$. Similar to the Kalman smoother recursions, they can be obtained without computing the whole inverse using sparse matrix routines based on Takahashi's equations (Rue et al., 2009). An efficient implementation in Matlab is available at the MathWorks File Exchange (see *sparseinv* by Tim Davis).

2. M-step: Conditional on the approximate smoothing density of log-variances h_i $(i = 1, \ldots, r)$, we update parameters of both state and measurement equation of the SV-SVAR model.

   (a) Update ϕ_i and s_i for $i = 1, \ldots, r$:
   Conditional on the moments of the approximate smoothing density we maximize the expected value of the complete data log-likelihood (1.B.5) with respect to the state equation parameters. Therefore, define:

$$\nabla G(\hat{\phi}, \hat{s}) = \text{E}\left[\frac{\partial \mathcal{L}_c}{\partial \phi'}, \frac{\partial \mathcal{L}_c}{\partial s'}\right]'_{\phi=\hat{\phi}, s=\hat{s}}, \qquad H(\hat{\phi}, \hat{s}) = \text{E}\begin{pmatrix} \frac{\partial^2 \mathcal{L}_c}{\partial \phi \partial \phi'} & \frac{\partial^2 \mathcal{L}_c}{\partial \phi \partial s'} \\ \frac{\partial^2 \mathcal{L}_c}{\partial s \partial \phi'} & \frac{\partial^2 \mathcal{L}_c}{\partial s \partial s'} \end{pmatrix}_{\phi=\hat{\phi}, s=\hat{s}}.$$

   The detailed expressions for first and second derivatives of the complete data log-likelihood are printed in 1.B.4. Then, set $\hat{\phi}_k = \hat{\phi}^{(l-1)}$ and $\hat{s}_k = \hat{s}^{(l-1)}$ and update parameters using Newton-Raphson, i.e. set:

$$\begin{pmatrix} \hat{\phi}_{k+1} \\ \hat{s}_{k+1} \end{pmatrix} = \begin{pmatrix} \hat{\phi}_k \\ \hat{s}_k \end{pmatrix} - \left(H\left(\hat{\phi}_k, \hat{s}_k\right)\right)^{-1} \nabla G\left(\hat{\phi}_k, \hat{s}_k\right),$$

   until $\left\| \begin{pmatrix} \hat{\phi}_{k+1} \\ \hat{s}_{k+1} \end{pmatrix} - \begin{pmatrix} \hat{\phi}_k \\ \hat{s}_k \end{pmatrix} \right\|$ is smaller than a specified threshold, e.g. 0.001. Then, set $\hat{\phi}^{(l)} = \hat{\phi}_{k+1}$ and $\hat{s}^{(l)} = \hat{s}_{k+1}$.

   (b) Update α. Let $Z = (X' \otimes I_K)$, then:

$$\hat{\alpha}^{(l)} = (Z' \tilde{\Sigma}_u^{-1} Z)^{-1} (Z' \tilde{\Sigma}_u^{-1} y^0),$$

with $\tilde{\Sigma}_u^{-1} = \left(\left[\left(\hat{B}^{(l-1)}\right)^{-1}\right]' \otimes I_T\right)\hat{\Sigma}_e^{-1}\left(\left(\hat{B}^{(l-1)}\right)^{-1} \otimes I_T\right)$ and $\hat{\Sigma}_e^{-1} = \text{diag}(\text{vec}(\hat{V}_{(-1)}))$. Furthermore, it is:

$$\begin{aligned}\hat{V}_{(-1)} =& \text{E}(V_{(-1)}|\theta^{(l-1)}, y) = (\hat{v}_1^{(-1)}, \ldots, \hat{v}_T^{(-1)}) \in \mathbb{R}^{K\times T}, \quad \text{with}\\ \hat{v}_t^{(-1)} =& \exp\left(-\text{E}(h_t|\theta^{(l-1)}, y) + \frac{1}{2}\text{Var}(h_t|\theta^{(l-1)}, y)\right).\end{aligned}$$

The latter is based on the properties of a log-normal distribution. Note that for $i = r+1, \ldots, K$, $\hat{v}_{it}^{(-1)} = 1$.

(c) Update B. Therefore, define $\hat{U} = Y^0 - \hat{A}^{(l)}X$, then:

$$\begin{aligned}\hat{B}^{(l)} =& \underset{B\in\mathbb{R}^{K\times K}}{\arg\max}\ \text{E}\left[\mathcal{L}_c(B)\middle|\hat{A}^{(l)}, \hat{\phi}^{(l)}, \hat{s}^{(l)}, y\right]\\ \propto& -T\ln|B| - \frac{1}{2}\text{vec}(B^{-1}\hat{U})'\hat{\Sigma}_e^{-1}\text{vec}(B^{-1}\hat{U}).\end{aligned}$$

3. Set $\theta^{(l)} = \left[\left(\hat{\alpha}^{(l)}\right)', \text{vec}\left(\hat{B}^{(l)}\right)', \left(\hat{\phi}^{(l)}\right)', \left(\hat{s}^{(l)}\right)'\right]'$, $l = l+1$ and return to step 1.

We iterate between steps 1.-3. until the relative change in the expected complete data log-likelihood becomes negligible. To be more precise, the algorithm is a Generalized EM algorithm since the M-step of impact matrix B depends on VAR coefficients α.

EM-2

In EM-2, the expectations in the E-step are approximated by MCMC integration. Based on starting values, $\theta^{(0)}$, the algorithm iterates between the following steps for $l \geq 1$:

1. E-Step: In order to compute the expectations necessary in the EM algorithm, we recur to Monte Carlo integration. In particular, for each of the heteroskedastic shocks ($i = 1, \ldots, r$), we simulate random draws of the mixture indicators $z_i^{(j)}$ for $j = 1, \ldots, R$ and compute:

$$Q(\theta, \theta^{(l-1)}) \approx \frac{1}{R}\sum_{j=1}^{R}\text{E}_{\theta^{(l-1)}}^{(j)}[\mathcal{L}(\theta)], \tag{1.B.6}$$

where the expectations are taken with respect to the tractable distribution $p(h|\theta^{(l-1)}, z^{(j)}, y)$. To generate random draws of z, we rely on the methodology of Kim et al. (1998). For each of the heteroskedastic shocks ($i = 1, \ldots, r$), this involves iteratively drawing from the following conditional distributions:

(a) $z_i^{(j)} \sim p\left(z_i \middle| \theta^{(l-1)}, h_i^{(j-1)}, y\right)$. The mixture indicators are drawn for each $t = 1, \ldots, T$ from the discrete conditional distribution $P\left(z_{it}^{(j)} = k\right) = q_{it,k}$ for $k = 1, \ldots, 7$ where:

$$q_{it,k} = \frac{p_k \phi\left(y_{it}^* - h_{it}; m_k, v_k^2\right)}{\sum_{k=1}^{7} p_k \phi\left(y_{it}^* - h_{it}; m_k, v_k^2\right)},$$

with $y_{it}^* = \log\left[\left(\hat{\varepsilon}_{it}^{(l-1)}\right)^2\right]$, $\hat{\varepsilon}_t^{(l-1)} = \left(\hat{B}^{(l-1)}\right)^{-1}\left(y_t - \hat{A}^{(l-1)} x_t\right)$ and $\phi(\cdot; m_k, v_k^2)$ indicating the pdf of a normal distribution with mean m_k and variance v_k^2. Mixture parameters p_k's, m_k's and v_k's are tabulated in Table 1.B.1.

(b) $h_i^{(j)} \sim p(h_i | \theta^{(l-1)}, z_i^{(j)}, y)$. To draw the log-variances, first a random sample from the unconstrained conditional distribution $\tilde{h}_i^{(j)} \sim \mathcal{N}(\bar{\delta}_{ij}, \Sigma_{ij})$ is generated using the precision sampler of Chan and Jeliazkov (2009). The unconstrained moments are given as:

$$\begin{aligned}
\Sigma_{ij}^{-1} &= H_i' \Sigma_{h_i}^{-1} H_i + G_{ij}, \\
\bar{\delta}_{ij} &= \Sigma_{ij}\left(H_i' \Sigma_{h_i}^{-1} H_i \delta_i + G_{ij}(y_i^* - m_{ij})\right),
\end{aligned}$$

and

$$\begin{aligned}
y_i^* &= \left(\log\left[\left(\hat{\varepsilon}_{i1}^{(l-1)}\right)^2\right], \ldots, \log\left[\left(\hat{\varepsilon}_{iT}^{(l-1)}\right)^2\right]\right)', \\
G_{ij} &= \operatorname{diag}\left(v^2\left(z_{i1}^{(j)}\right), \ldots, v^2\left(z_{iT}^{(j)}\right)\right)^{-1}, \\
m_{ij} &= \operatorname{diag}\left(m\left(z_{i1}^{(j)}\right), \ldots, m\left(z_{iT}^{(j)}\right)\right).
\end{aligned}$$

In a next step, the draw is corrected to account for the linear constraint. That is:

$$h_i^{(j)} = \tilde{h}_i^{(j)} - \Sigma_{ij} A_h' (A_h \Sigma_{ij} A_h')^{-1}\left(A_h \tilde{h}_i^{(j)} - \hat{\mu}_i^{(l-1)}\right),$$

which yields a draw from the correct distribution under the linear constraint. The moments of this distribution are:

$$\begin{aligned}
\bar{\delta}_{ij}^c &= \bar{\delta}_{ij} - \Sigma_{ij} A_h' (A_h \Sigma_{ij} A_h')^{-1}\left(A_h \bar{\delta}_{ij} - \hat{\mu}_i^{(l-1)}\right), \\
\operatorname{Cov}\left(h_i \middle| \theta^{(l-1)}, z_i^{(j)}, y, A_h h_i = \hat{\mu}_i^{(l-1)}\right) &= \Sigma_{ij} - \Sigma_{ij} A_h' (A_h \Sigma_{ij} A_h')^{-1} A_h \Sigma_{ij}.
\end{aligned}$$

Note that the corrected moments are those used to compute the Monte Carlo expected complete data log-likelihood from equation (1.B.6). As in EM-1, we only compute the diagonal and first off-diagonal of the covariance matrix Σ_{ij} using the same sparse matrix routines.

2. M-steps: Conditional on the mixture indicators $z_i^{(j)}$ $(i = 1, \ldots, r; j = 1, \ldots, R)$, first and second moments of h_i's are given. Thus, as in EM-1, we maximize the expected complete data log-likelihood using Newton-Raphson updates in state equation parameters, a closed-form update in VAR parameters and numerical optimization in the impact matrix.

   (a) Update ϕ_i and s_i for $i = 1, \ldots, r$: Conditional on the mixture indicators z, the expected value of the complete data log-likelihood (1.B.5) is maximized. To do so, define:

$$\nabla G_R(\hat{\phi}, \hat{s}) = \frac{1}{R} \sum_{j=1}^{R} \mathrm{E}\left(\frac{\partial \mathcal{L}_c}{\partial \phi'}, \frac{\partial \mathcal{L}_c}{\partial s'} \middle| z^{(j)} \right)'_{\phi=\hat{\phi}, s=\hat{s}}$$

$$H_R(\hat{\phi}, \hat{s}) = \frac{1}{R} \sum_{j=1}^{R} \mathrm{E}\left(\begin{matrix} \frac{\partial^2 \mathcal{L}_c}{\partial \phi \partial \phi'} & \frac{\partial^2 \mathcal{L}_c}{\partial \phi \partial s'} \\ \frac{\partial^2 \mathcal{L}_c}{\partial s \partial \phi'} & \frac{\partial^2 \mathcal{L}_c}{\partial s \partial s'} \end{matrix} \middle| z^{(j)} \right)_{\phi=\hat{\phi}, s=\hat{s}}.$$

   The detailed expressions are printed in Section 1.B.4. All expectations of functions of the log-variances are uniquely determined by the sampled mixture indicators. Then, set $\hat{\phi}_k = \hat{\phi}^{(l-1)}$ and $\hat{s}_k = \hat{s}^{(l-1)}$ and update parameters using Newton-Raphson, i.e. set

$$\begin{pmatrix} \hat{\phi}_{k+1} \\ \hat{s}_{k+1} \end{pmatrix} = \begin{pmatrix} \hat{\phi}_k \\ \hat{s}_k \end{pmatrix} - \left(H_R\left(\hat{\phi}_k, \hat{s}_k \right) \right)^{-1} \nabla G_R\left(\hat{\phi}_k, \hat{s}_k \right)$$

   until $\left\| \begin{pmatrix} \hat{\phi}_{k+1} \\ \hat{s}_{k+1} \end{pmatrix} - \begin{pmatrix} \hat{\phi}_k \\ \hat{s}_k \end{pmatrix} \right\|$ is smaller than a specified threshold, e.g. 0.001. Then, set $\hat{\phi}^{(l)} = \hat{\phi}_{k+1}$ and $\hat{s}^{(l)} = \hat{s}_{k+1}$.

   (b) Update α. Let $Z = (X' \otimes I_K)$, then:

$$\hat{\alpha}^{(l)} = (Z' \tilde{\Sigma}_u^{-1} Z)^{-1} (Z' \tilde{\Sigma}_u^{-1} y^0),$$

   where everything is as in EM-1 but:

$$\hat{v}_t^{(-1)} = R^{-1} \sum_{j=1}^{R} \exp\left(-\mathrm{E}\left(h_t | \theta^{(l-1)}, z_t^{(j)}, y \right) + \frac{1}{2} \mathrm{Var}\left(h_t | \theta^{(l-1)}, z_t^{(j)}, y \right) \right).$$

(c) Update B as in EM-1.

3. Set $\theta^{(l)} = \left[\left(\hat{\alpha}^{(l)}\right)', \text{vec}\left(\hat{B}^{(l)}\right)', \left(\hat{\phi}^{(l)}\right)', \left(\hat{s}^{(l)}\right)'\right]'$, $l = l + 1$ and return to step 1.

We recommend to set the starting values based on the results of EM-1, which are quickly available. We increase the number of MCMC replications deterministically over the EM iterations. This is necessary since automated strategies as the ascent-based MCEM algorithm (Caffo et al., 2005) fail to converge due to the substantial amount of parameters to be estimated in the VAR equation. That is, we first run a burn-in period of 300 EM steps using $R = 50$ and then proceed with another 100 EM iterations using $R = 500$. Subsequently, we increase R to 50,000 and iterate EM steps until the stopping criterion of Caffo et al. (2005) applies. This usually happens after a small number of additional EM steps using 50,000 MCMC replications.

Table 1.B.1: Mixture approximation log $\chi^2_{(1)}$-distribution (Kim et al., 1998)

k	$p_k = \Pr(z_{it} = k)$	m_k	v_k^2
1	0.00730	−10.12999	5.79596
2	0.10556	−3.97281	2.61369
3	0.00002	−8.56686	5.17950
4	0.04395	2.77786	0.16735
5	0.34001	0.61942	0.64009
6	0.24566	1.79518	0.34023
7	0.25750	−1.08819	1.26261

Note: Seven Normal Mixture components to approximate a $\log\left(\chi^2_{(1)}\right)$ distribution adjusted by its mean −1.2704.

1.B.4 Derivatives complete data log-likelihood

The respective derivatives of the complete data log-likelihood (1.B.5) are given in the following. Let $\bar{h}_{it} = h_{it} - \mu_i$ for $i = 1, \ldots, r$ and $t = 1, \ldots, T$. First, note that $\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial \phi_i \partial \phi_j} = \frac{\partial^2 \mathcal{L}_c(\theta)}{\partial s_i \partial s_j} = \frac{\partial^2 \mathcal{L}_c(\theta)}{\partial \phi_i \partial s_j} = 0$ for all $i \neq j$. The remaining first and second derivatives with respect to state equation parameters ϕ_i and s_i are given as follows:

$$
\begin{aligned}
\frac{\partial \mathcal{L}_c(\theta)}{\partial s_i} &= -\frac{1}{2s_i}\left(T - \frac{1-\phi_i^2}{s_i}\bar{h}_{i1}^2 + \bar{h}_{i1} + \sum_{t=2}^{T}\left[\frac{\bar{h}_{it}-\phi_i\bar{h}_{i,t-1}}{1+\phi_i} - \frac{\left(\bar{h}_{it}-\phi_i\bar{h}_{i,t-1}\right)^2}{s_i}\right]\right),\\
\frac{\partial \mathcal{L}_c(\theta)}{\partial \phi_i} &= -\frac{\phi_i}{1-\phi_i^2}\left(1+\bar{h}_{i1}\right) + \frac{\phi_i}{s_i}\bar{h}_{i1}^2\\
&\quad - \frac{1}{s_i}\sum_{t=2}^{T}\left[\left(\bar{h}_{it}-\phi_i\bar{h}_{i,t-1}\right)\left(\frac{s_i\phi_i(1-\phi_i)}{\left(1-\phi_i^2\right)^2} - \bar{h}_{i,t-1}\right)\right],\\
\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial \phi_i \partial s_i} &= -\frac{\phi_i}{2\left(1-\phi_i^2\right)^2} + \frac{\phi_i\bar{h}_{i1}}{\left(1-\phi_i^2\right)s_i} - \frac{\phi_i\bar{h}_{i1}^2}{s_i^2} + \frac{1}{s_i^2}\sum_{t=2}^{T}\left[\left(\bar{h}_{it}-\phi_i\bar{h}_{i,t-1}\right)\right.\\
&\quad \left.\left(\frac{s_i\phi_i(1-\phi_i)}{\left(1-\phi_i^2\right)^2} - \bar{h}_{i,t-1}\right)\right] - \frac{1}{s_i}\sum_{t=2}^{T}\left[\frac{1}{2\left(1+\phi_i\right)}\left(\frac{s_i\phi_i(1-\phi_i)}{\left(1-\phi_i^2\right)^2} - \bar{h}_{i,t-1}\right)\right.\\
&\quad \left.+\left(\bar{h}_{it}-\phi_i\bar{h}_{i,t-1}\right)\left(\frac{\phi_i(1-\phi_i)}{\left(1-\phi_i^2\right)^2} - \frac{1}{2\left(1-\phi_i^2\right)}\right)\right],\\
\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial s_i^2} &= \frac{1}{s_i}\left(\frac{T}{2s_i} + \frac{\bar{h}_{i1}}{s_i} - \frac{\bar{h}_{i1}^2\left(1-\phi_i^2\right)}{s_i^2} - \frac{1}{4\left(1-\phi_i^2\right)} - \frac{T-1}{4\left(1+\phi_i\right)^2}\right.\\
&\quad \left.+\frac{1}{s_i}\sum_{t=2}^{T}\left[\frac{\bar{h}_{it}-\phi_i\bar{h}_{i,t-1}}{1+\phi_i} - \frac{\left(\bar{h}_{it}-\phi_i\bar{h}_{i,t-1}\right)^2}{s_i}\right]\right),\\
\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial \phi_i^2} &= -\frac{1+\phi_i^2}{\left(1-\phi_i^2\right)^2}\left(1+\bar{h}_{i1}\right) - \frac{s_i\phi_i^2}{\left(1-\phi_i^2\right)^3} + \frac{\bar{h}_{i1}^2}{s_i} + \frac{2\phi_i^2\bar{h}_{i1}^2}{\left(1-\phi_i^2\right)^2}\\
&\quad - \frac{1}{s_i}\sum_{t=2}^{T}\left(\frac{s_i\phi_i(1-\phi_i)}{\left(1-\phi_i^2\right)^2} - \bar{h}_{i,t-1}\right)^2\\
&\quad - \left(\frac{1-3\phi_i}{\left(1-\phi_i^2\right)^2} + \frac{4\phi_i^2(1-\phi_i)}{\left(1-\phi_i^2\right)^3}\right)\sum_{t=2}^{T}\left(\bar{h}_{it}-\phi_i\bar{h}_{i,t-1}\right).
\end{aligned}
$$

Furthermore, let $\Sigma_t = BV_tB'$, $\beta = \text{vec}(B)$, $\alpha = \text{vec}(A)$, $\tilde{X}_t = (x_t' \otimes I_K)$, such that $\text{vec}(Ax_t) = \tilde{X}_t\alpha$ and $K^{(K,K)}$ be the $K^2 \times K^2$ commutation matrix. Then, the first and second derivatives of (1.B.5) with respect to α and β are given as:

$$
\begin{aligned}
\frac{\partial \mathcal{L}_c(\theta)}{\partial \alpha'} &= \left(\sum_{t=1}^{T} y_t' \Sigma_t^{-1} \tilde{X}_t\right) - \alpha' \left(\sum_{t=1}^{T} \tilde{X}_t' \Sigma_t^{-1} \tilde{X}_t\right), \\
\frac{\partial \mathcal{L}_c(\theta)}{\partial \beta'} &= -T \operatorname{vec}\left(\left[B^{-1}\right]'\right)' + \operatorname{vec}\left(\sum_{t=1}^{T} \left[B^{-1}\right]' V_t^{-1} B^{-1} u_t u_t' \left[B^{-1}\right]'\right)', \\
\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial \alpha' \partial \beta} &= -\sum_{t=1}^{T} \left[\left(\varepsilon_t' \otimes \tilde{X}_t' \left[B^{-1}\right]' V_t^{-1} B^{-1}\right) + \left(\varepsilon_t' V_t^{-1} B^{-1} \otimes \tilde{X}_t' \left[B^{-1}\right]'\right) K^{(K,K)}\right], \\
\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial \alpha \partial \alpha'} &= -\left(\sum_{t=1}^{T} \tilde{X}_t' \Sigma_t^{-1} \tilde{X}_t\right), \\
\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial \beta \partial \beta'} &= T\left(B^{-1} \otimes \left[B^{-1}\right]'\right) K^{(K,K)} \\
&\quad - \sum_{t=1}^{T} \left(I_K \otimes \left[B^{-1}\right]' V_t^{-1}\right) \left(K^{(K,K)} + I_{K^2}\right) \left(B^{-1} u_t u_t' \left[B^{-1}\right]' \otimes B^{-1}\right) \\
&\quad - \sum_{t=1}^{T} \left(B^{-1} u_t u_t' \left[B^{-1}\right]' V_t^{-1} B^{-1} \otimes \left[B^{-1}\right]'\right) K^{(K,K)}.
\end{aligned}
$$

Note that the cross derivatives $\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial \phi_i \partial \alpha}$, $\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial \phi_i \partial \beta}$, $\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial s_i \partial \alpha}$ and $\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial s_i \partial \beta}$ are equal to zero due to the structure of the complete data log-likelihood (1.B.5).

1.B.5 Inference on structural impulse response functions

Following Lütkepohl (2005), the IRFs are elements of the coefficient matrices $\Theta_j = \Phi_j B$ in the Vector Moving Average (VMA) representation of the model:

$$
y_t = \mu_y + \sum_{j=0}^{\infty} \Phi_j B \varepsilon_t,
$$

where $\varepsilon_t = V_t^{\frac{1}{2}} \eta_t$ are the structural shocks, $\mu_y = (I_K - A_1 - \ldots - A_p)^{-1} \nu$ is the unconditional mean of y_t and $\Phi_j \in \mathbb{R}^{K \times K}$ $(j = 0, 1, \ldots)$ is a sequence of exponentially decaying matrices given as: $\Phi_j = J \mathbf{A}^j J'$ with $J = [I_K, 0, \ldots, 0]$ and

$$
\mathbf{A} = \begin{pmatrix}
A_1 & A_2 & \ldots & A_{p-1} & A_p \\
I_K & 0 & \ldots & 0 & 0 \\
0 & I_K & & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & 0 \\
0 & 0 & \ldots & I_K & 0
\end{pmatrix}.
$$

The elements of Θ_i, $\theta_{jk,i}$'s are the impulse response functions in variable j to a structural innovation k after i periods.

We conduct inference on the estimated quantities $\hat{\Theta}_i$ based on their asymptotic distribution. Given that the IRFs are nonlinear functions of the model parameters, the distribution can be inferred based on the result that $T^{1/2}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}(\theta)^{-1})$. Let $\alpha = \text{vec}(A)$ with $A = (A_1, \ldots, A_p)$, $\beta = \text{vec}(B)$ and partition the asymptotic covariance matrix of $\hat{\theta}$ into:

$$\mathcal{I}(\theta)^{-1} = \Sigma_\theta = \begin{pmatrix} \Sigma_\nu & & & & \\ \Sigma_{\nu,\alpha} & \Sigma_\alpha & & & \\ \Sigma_{\nu,\beta} & \Sigma_{\alpha,\beta} & \Sigma_\beta & & \\ \Sigma_{\nu,\phi} & \Sigma_{\alpha,\phi} & \Sigma_{\beta,\phi} & \Sigma_\phi & \\ \Sigma_{\nu,s} & \Sigma_{\alpha,s} & \Sigma_{\beta,s} & \Sigma_{\phi,s} & \Sigma_s \end{pmatrix}.$$

As in Brüggemann et al. (2016), an application of the Delta method yields the asymptotic distribution of the structural impulse responses:

$$\sqrt{T}(\hat{\Theta}_i - \Theta_i) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\hat{\Theta}_i}), \qquad i = 0, 1, 2, \ldots,$$

where:

$$\Sigma_{\hat{\Theta}_i} = C_{i,\alpha}\Sigma_\alpha C_{i,\alpha}' + C_{i,\beta}\Sigma_\beta C_{i,\beta}' + C_{i,\alpha}\Sigma_{\alpha,\beta}' C_{i,\beta}' + C_{i,\beta}\Sigma_{\alpha,\beta} C_{i,\alpha}',$$

with $C_{0,\alpha} = 0$, $C_{i,\alpha} = \frac{\partial \text{vec}(\Theta_i)}{\partial \alpha'} = (B' \otimes I_K)G_i$ and $G_i = \frac{\partial \text{vec}(\Phi_i)}{\partial \alpha'} = \sum_{j=0}^{i-1} \left[J(\mathbf{A}')^{i-1-j} \right] \otimes \Phi_j$ for $i \geq 1$. Finally, $C_{i,\beta} = \frac{\partial \text{vec}(\Theta_i)}{\partial \beta'} = (I_K \otimes \Phi_i)$ for $i \geq 0$. Similarly, for the accumulated structural impulse responses $\Xi_n = \sum_{i=0}^{n} \Theta_i$, we get:

$$\sqrt{T}\left(\hat{\Xi}_n - \Xi_n\right) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\hat{\Xi}_n}), \qquad n = 0, 1, 2, \ldots,$$

where:

$$\Sigma_{\hat{\Xi}_n} = P_n \Sigma_\alpha P_n' + \bar{P}_n \Sigma_\beta \bar{P}_n' + P_n \Sigma_{\alpha,\beta}' \bar{P}_n' + \bar{P}_n \Sigma_{\alpha,\beta} P_n',$$

with $P_n = (B' \otimes I_K)F_n$, $F_0 = 0$, $F_n = G_1 + \cdots + G_n$, $\bar{P}_n = (I_K \otimes \Psi_n)$ and $\Psi_n = \sum_{i=0}^{n} \Phi_i$.

Appendix 1.C Data and complementary results

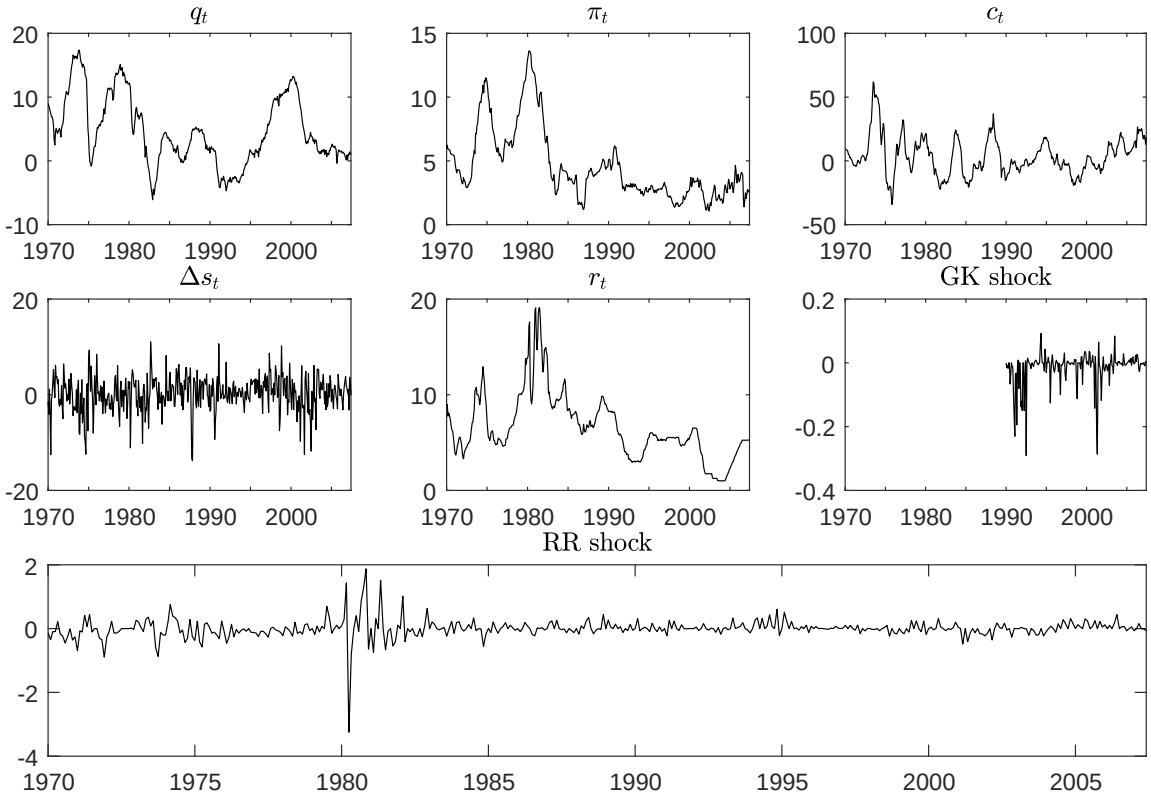
The time series data used in Section 1.5 is based on $y_t = (q_t, \pi_t, c_t, \Delta s_t, r_t)'$, where

- q_t is the logarithm of industrial production (linearly detrended),
- π_t is the growth rate of the consumer price index (in %),
- c_t denotes the annualized change in the logarithm of the World Bank commodity price index (in %),
- Δs_t is the first difference of the logarithm of the CPI deflated real S&P500 index,
- r_t is the Federal Funds rate.

As in Lütkepohl and Netšunajev (2017b) and Lütkepohl and Netšunajev (2017a), we use the updated sample period 1970M1-2007M6. Except for c_t, the data can be downloaded from the FRED. The commodity price index is provided by the World Bank. The transformed data set is readily available at `http://sfb649.wiwi.hu-berlin.de/fedc/discussionPapers_formular_content.php`.

The monetary policy instruments of Gertler and Karadi (2015) and Romer and Romer (2004) are obtained from the homepage of Valerie Ramey: `http://econweb.ucsd.edu/~vramey/research.html#data`. Note that the RR series used in our analysis is the one extended by Wieland and Yang (2016).

Figure 1.C.1: Time series plots in monetary SVAR



1.C.1 Complementary results

Table 1.C.1: Tests on standardized structural shocks

	Normality		Heteroskedasticity			
	MJB	p-value	Q_1	p-value	Q_2	p-value
Linear	12,911.0	0.000	52.34	0.000	1433.70	0.000
STVAR	49,785.0	0.000	40.25	0.000	1475.50	0.000
MS(2)	291.8	0.000	13.59	0.004	811.21	0.000
MS(3)	48.6	0.000	9.40	0.024	844.74	0.000
GARCH	555.0	0.000	8.22	0.042	627.45	0.904
SV	16.1	0.096	3.25	0.355	623.88	0.921

Note: Multivariate Jarque-Bera (MJB) test conducted as in Lütkepohl (2005, p. 181). Test statistics Q_1 and Q_2 as discussed in Section 1.3.4, applied to estimated standardized structural shocks $\hat{\varepsilon}_t \,/\, \exp\left(\hat{h}_t/2\right)$.

Table 1.C.2: Test for overidentifying restrictions (EM-2)

H_0	H_1	LR	dof	p-value	$p_{.025}$	$p_{.975}$
R1	UC	27.341	10	0.0023	0.0017	0.0032
R2	UC	23.693	7	0.0013	0.0009	0.0018
R3	UC	25.868	9	0.0021	0.0015	0.0030
R1	R3	1.543	1	0.2142	0.1390	0.3438
R4rr	UC	5.779	4	0.2163	0.1388	0.3294
R4gk	UC	256.590	4	0.0000	0.0000	0.0000

Note: For details about overidentifying restrictions see Section 1.5.1. Likelihood ratio test statistics are computed as $2\left(\ln L_{H_1} - \ln L_{H_0}\right)$ and are approximatively χ^2-distributed under H_0. Right columns report an approximate 95%-confidence interval for the p-value resulting from an application of the batch means method to the LR test statistic.

Chapter 2

Directed graphs and variable selection in large vector autoregressive models

2.1 Introduction

Vector autoregressive (VAR) models are popular tools in the analysis of multiple time series. Typical applications include forecasting or impulse response analysis. The popularity of the VAR model is at least partly due to the fact that it typically does not require strong economic theory assumptions. Often VAR models without any restrictions on the parameters are used to describe the joint dynamics of a set of economic time series. While the general VAR lag structure allows to uncover dynamic relations between the variables included in the system, the use of unrestricted VARs comes at a cost: The number of parameters to be estimated from the data increases with the square of the number of variables in the system. Even in moderately large VARs the degrees of freedom exhaust quickly. Thus applied researchers have to choose the number of variables to be included in the VAR wisely. On the one hand, a researcher would like to include all relevant variables to avoid omitted variable bias and to get a complete picture of the underlying dynamics. On the other hand, including too many variables makes parameter estimates unreliable and estimation uncertainty may lead to rather uninformative results such as estimated impulse responses with very wide confidence intervals.

Given variables of interest, our paper suggests to use a graphical modeling approach in order to select a 'minimal' VAR containing only variables that are relevant for predicting the variables of interest. This approach is helpful in selecting the relevant variables for VAR analysis in a data-driven way. We argue that this is a useful addition to the toolbox of time series econometricians as on the one hand, it exploits the information from large dimensional data sets but on the other hand eventually uses smaller VAR models for forecasting and structural analysis.

To fix ideas, suppose a researcher is interested in a set of variables denoted by y^{I}, including say GDP growth, the consumer price (CPI) inflation, and a key interest rate. She wants to conduct an impulse response analysis for the variables in y^{I}. For this purpose, typically a large cross-section of time series on e.g. output, income, consumption, the labor market, orders and inventories, money and credit, interest and exchange rates, financial market variables and various price measures, is available.[1]

In recent years, suggestions have been made on how to include the information from a large dimensional data set into VARs. Factor-augmented VARs (FAVARs) (see e.g. Bernanke et al. (2005) and Stock and Watson (2016)) condense the information from a large time series data set into a few factor time series, which are then included in a VAR model.

[1]Large data sets of this type have been used in various studies. See e.g. Stock and Watson (2003) or McCracken and Ng (2016) with references therein. They typically contain up to 130 variables.

Factor-augmented models have been used for forecasting and structural analysis.[2] Clearly, these models are only suitable if the underlying data has a factor structure, i.e. if the large number of time series are really driven by a small number of common factors (see e.g. Uhlig (2009) on this point). An alternative are large Bayesian VARs (BVARs) as suggested by Banbura et al. (2010).[3] In large dimensional settings, however, these models require to use a very tight prior. Consequently, using a large BVAR might impose more structure on the model than typical VAR users feel comfortable with. Other shrinkage methods have also been used for large VAR models, including the least absolute shrinkage and selection operator (LASSO) (see e.g. Kascha and Trenkler (2015)).[4] The LASSO approach can handle large dimensional VAR models by setting some VAR coefficients to zero and at the same time shrinking the remaining coefficients. While already frequently used in applied work, the theoretical underpinning is still developing and it is not entirely clear how to conduct inference based on LASSO-VAR models.

Consequently, using FAVARs or large BVARs may not be ideal in some situations faced by applied time series econometricians. Researchers may actually prefer to use smaller VARs also because they might be easier to interpret and resemble more closely small scale dynamic stochastic general equilibrium (DSGE) models used in macroeconomics. At the same time, the researcher would like to include additional 'relevant' variables that affect predictions and/or impulse responses for the variables of interest. Against the background of the large number of time series available today, this entails a variable selection procedure.

Regarding structural analysis, our paper is concerned with the question of how to choose variables for the smallest ('minimal') VAR that contains all variables that are 'relevant' for the impulse responses of variables of interest y^{I}. To answer this, we develop a variable selection strategy based on a graphical modeling approach. The first contribution of our paper is to use so-called strongly connected components (SCCs) and the relation among these components for variable selection. We first represent a sparse VAR structure as a directed graph with vertices and edges. From this graph we identify all SCCs in our VAR model by using a simple graphical modeling algorithm. The concept of SCCs is well-established in the graphical modeling literature but to the best of our knowledge, this concept has not been used in econometrics. We show how the SCCs and their connections to other SCCs are helpful in identifying the set of 'relevant' variables. Effectively, the set of relevant variables can be found from the graphical representation of the SCCs, known as a component graph. In a second theoretical contribution, we show the relation between the SCCs and the concept

[2]See e.g. Stock and Watson (2002), Ludvigson and Ng (2007), Eickmeier and Ziegler (2008), Ludvigson and Ng (2009), Stock and Watson (2012b), Clements (2016) and Cheng and Hansen (2015) for applications of factor-augmented regressions and FAVARs.

[3]See e.g. Carriero et al. (2015, 2009, 2012), Giannone et al. (2014), and Koop (2013) for applications of this method.

[4]See also Stock and Watson (2012b).

of multi-step causality as in Dufour and Renault (1998). In particular, given the variables of interest y^{I}, we show that a minimal VAR chosen by SCCs is identical to the VAR that contains y^{I} and all variables that are multi-step causal for y^{I}.

Methodologically, our paper is related to the literature on using graphical models in econometrics. Following the work on causal analysis of multivariate data (see e.g. Lauritzen (1996), Pearl (2000) and Edwards (2000)), graphical models have also been introduced for time series models. Brillinger (1996) and Dahlhaus (2000) are the first papers mentioning the use of graphical modeling for time series data and present concepts based on the partial correlation and partial spectral coherence.[5] Dahlhaus and Eichler (2003) introduce causality graphs based on the autoregressive representation. Our work is most closely related to the work of Eichler (2006, 2007, 2012), who shows the close relation of different causality concepts (Granger-causality and multi-step causality) to graphical representations in vector autoregressive models. We add to this work the link from causality structures to variable selection using the concept of strongly connected components.[6]

Our paper is also related to the work of Jarociński and Maćkowiak (2017), who investigate the same research question of variable choice for VAR analysis, albeit with a different econometric approach. Based on the concept of Granger-causal priority (see Sims (1982, 2015), Doan and Todd (2010)), their paper evaluates in a Bayesian setup the posterior probability of Granger-causal priority. For a given set of variables of interest y^{I}, Jarociński and Maćkowiak (2017) would drop a variable y_j, say, if the variables in y^{I} are likely to be Granger-causal prior to y_j. Thus, their method may also be used to choose variables for VAR analysis.

We illustrate the usefulness of our suggested variable selection in an application to US macroeconomic data. The variables of interest in y^{I} are US output, CPI inflation and the federal funds rate, three variables often used in stylized three-variable VARs for the US. Given y^{I}, we use our variable selection method based on SCCs to select a minimal VAR from 41 US time series for a period between 1975 and 2014. Starting point is a sparse VAR structure obtained from applying the LASSO to the large VAR. Regardless of the considered estimation period, six out of the 41 variables are always selected into the model and the selection is fairly stable over different samples before the financial crisis in 2008/09. Moreover, additional variables are selected into the model in a number of periods. Consequently, the 'minimal VAR' is still relatively large, indicating that the underlying relations are typically quite complex and may not be captured adequately in a three-variable

[5] See also Flamm et al. (2012) for an overview of different approaches.

[6] Graphical modelling has also been used for identifying the instantaneous relations. The first work in this area is the paper by Swanson and Granger (1997), followed by a number of studies that use graphical modeling for identifying structural VAR models (see e.g. Demiralp and Hoover (2003), Hoover et al. (2009) and Heinlein and Krolzig (2012)).

VAR. We also find that including the selected variables into the VAR leads to more reasonable responses to a monetary policy shock, indicating that the selection is useful.

The remainder of the paper is structured as follows. Section 2.2 shows how VARs can be represented as directed graphs. We also introduce the concept of strongly connected components and explain how this can be used for variable selection and for finding a 'minimal VAR'. Section 2.3 relates the graph-theoretical concepts to multi-step causality and shows how variable selection based on both concepts leads to the same set of relevant variables. In Section 2.4, we illustrate the usefulness of our method in an empirical application. Section 2.5 concludes. All proofs are deferred to the appendix.

2.2 Vector autoregressive models, directed graphs and strongly connected components

In this section, we explain how VAR models can be represented by directed graphs. We then review the concept of strongly connected components (SCCs) in directed graphs and explain how SCCs can be used for selecting relevant variables.

We denote the VAR model of order p, a VAR(p) for the K-dimensional time series vector $y_t = (y_{1,t}, y_{2,t}, \ldots, y_{K,t})'$ by

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \tag{2.2.1}$$

where $A_1, \ldots, A_p$ are $K \times K$ coefficient matrices and u_t is a zero mean white noise disturbance vector with non-singular covariance matrix Σ_u. We have not included deterministic terms (e.g. intercepts) into the VAR in (2.2.1) to simplify the notation. Adding deterministic terms would not change the results discussed below and they can be included in empirical work. The VAR(p) can be written in VAR(1) companion form as

$$Y_t = \mathbf{A} Y_{t-1} + U_t, \tag{2.2.2}$$

where $Y_t = (y_t', y_{t-1}', \ldots, y_{t-p+1}')'$ and $U_t = (u_t', 0, \ldots, 0)'$ are $Kp \times 1$ vectors and

$$\mathbf{A} := \begin{pmatrix} A_1 & A_2 & A_3 & \ldots & A_p \\ I_K & 0 & 0 & \ldots & 0 \\ 0 & I_K & 0 & \ldots & 0 \\ 0 & 0 & \ddots & & \vdots \\ 0 & 0 & \ldots & I_K & 0 \end{pmatrix}. \tag{2.2.3}$$

To make use of graphical modeling concepts, we represent the VAR in (2.2.2) as a directed graph. Following the standards in graphical modeling, a directed graph G is described by a set of vertices V and a set of edges E that are ordered pairs of vertices. In our application the vertices correspond to the Kp elements in vector Y_t and the edges are determined by the elements of the companion form matrix $\mathbf{A}$ as in the following definition:

Definition 2.1 (Directed VAR Graph)**.** *Given a VAR(p) model as in* (2.2.1)*, the associated directed graph is* $G = (V, E)$ *with* $V = \{1, 2, 3, \ldots, Kp\}$ *and*

$$(i, j) \in E \quad \Leftrightarrow \quad \mathbf{a}_{ij} \neq 0, \tag{2.2.4}$$

where $\mathbf{a}_{ij}$ *is the element in row* i *and column* j *of the companion matrix* $\mathbf{A}$*.*

Remark 2.1. In this graph, a directed edge (i, j) leads from vertex i to vertex j. This is standard in the graphical modeling literature. In our context, using this definition $(i, j) \in E$ implies that the i-th variable *depends on* the j-th component in Y_t, however, the arrow would point from vertex i to vertex j. Thus, the direction of the arrows is reversed compared to the type of arrows sometimes used to denote (Granger-)causality.

Remark 2.2. Different definitions of directed graphs for VAR models are possible. Eichler (2007), among others, uses a different definition for VAR(p) models. In particular, Eichler works directly on the matrices $A_s, s = 1, \ldots, p$ and defines the VAR graph $G^1 = \left(V^1, E^1\right)$ with $V^1 = \{1, \ldots, K\}$ and

$$(i, j) \in E^1 \quad \Leftrightarrow \quad \exists s \in \{1, \ldots, p\} : A_{ij,s} \neq 0,$$

where $A_{ij,s}$ denotes the element in row i and column j of A_s. Obviously, for $p > 1$ this definition would yield a different graph compared to that obtained from Definition 2.1. However, as discussed below, our central result does not depend on whether we use Definition 2.1 or that from Eichler (2007) (see Remark 2.11).

Given a sparse VAR structure, i.e. a VAR with zero restrictions on the VAR coefficients, we may use the associated directed graph to learn about the set of relevant variables. We do so by using the notion of the strongly connected components (SCCs) in a graph (Tarjan (1972)). In order to define these, one makes use of the concept of a path or a pathway. A path is defined to be a sequence of vertices to go from one vertex to another. More formally, a path P of length k leading from vertex u to u' in graph $G = (V, E)$ is a sequence $P = (v_0, v_1, \ldots, v_k)$ of vertices such that $u = v_0$ and $u' = v_k$ and $(v_{i-1}, v_i) \in E$ for $i = 1, 2, \ldots, k$. If there is a path from u to u', we say that u' is *reachable* from u, denoted as $u \overset{P}{\rightsquigarrow} u'$. We may now define the strongly connected components of a directed graph:

Definition 2.2 (Strongly Connected Components (Tarjan (1972)))**.** *Let $G = (V, E)$ be a directed graph and let two vertices u and v in G be equivalent if $u \overset{P}{\rightsquigarrow} v$ and $v \overset{P}{\rightsquigarrow} u$, i.e. u and v are mutually reachable. Call the corresponding equivalence classes of vertices V_i and let $C_i = (V_i, E_i)$ where $E_i = \{(u, v) \in E : u, v \in V_i\}$ for $i = 1, \ldots, k$. The subgraphs C_i are called the strongly connected components of G.*

Remark 2.3. Note that if we refer to SCC C_i in the following, we mean the corresponding set of vertices V_i that belongs to C_i in the sense of Definition 2.2. For example, if we write $j \in C_i$, we mean $j \in V_i$.

Remark 2.4. Note that each vertex of graph G belongs to exactly one strongly connected component. Consequently, the set of all equivalence classes $V_1, \ldots, V_k$ belonging to the strongly connected components $C_1, \ldots, C_k$ forms a partition of the set of vertices V such that $V = V_1 \cup \ldots \cup V_k$ (see e.g. Duff and Reid (1978)).

Remark 2.5. Following Tarjan (1972), a depth-first search algorithm may be used to compute the strongly connected components efficiently. We have implemented a variant of Tarjan's algorithm using Matlab according to the exposition in Cormen et al. (2009).

Remark 2.6. Duff and Reid (1978) suggest to order the SCCs such that there is no path from one strongly connected component to another later in the sequence, i.e. the SCCs $C_1, \ldots, C_k$ may be ordered such that there is no path from C_i to any C_j for $j > i$. The associated reordered matrix of the graph is then lower block-triangular. Each block on the diagonal corresponds to one of the SCCs. In the context of economic applications, the structure of the SCCs may give additional insights on the relevance of different variables.

In the next step, we condense the information of the graph G by moving from graph G to a graph of the SCCs. The resulting graph is called a component graph, which is defined next.

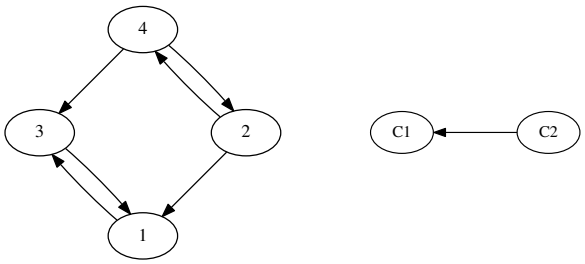
Definition 2.3 (Component Graph)**.** *A component graph is defined as $G^{SCC} = (V^{SCC}, E^{SCC})$, where $V^{SCC} = \{C_1, \ldots, C_k\}$ is the set of strongly connected components of graph G. There is an edge $(C_i, C_j) \in E^{SCC} \Leftrightarrow \exists x \in C_i : \exists y \in C_j : (x, y) \in E$.*

Remark 2.7. Definition 2.3 implies that there is only an edge (C_i, C_j) between two strongly connected components if the original graph G has a directed edge from one member of the SCC C_i to a member of the SCC C_j.

Remark 2.8. The component graph may be viewed as a condensed view of the original graph. Essentially, the component graph collapses all edges of the original graph whose vertices are contained in the same SCC.

Figure 2.1: VAR matrix, associated directed graph, and component graph

$$\begin{pmatrix} a_{11} & 0 & a_{13} & 0 \\ a_{21} & 0 & 0 & a_{24} \\ a_{31} & 0 & a_{33} & 0 \\ 0 & a_{42} & a_{43} & a_{44} \end{pmatrix}$$



Finally, for a given set of variables of interest, say $y^{\mathrm{I}} \subseteq y$, i.e. $\mathrm{I} \subseteq \{1, \ldots, K\}$, we would like to identify a 'minimal' set of variables which have to be taken into account when modeling y^{I}. We denote this set of relevant variables as $R(y^{\mathrm{I}})$. For that purpose, we define the set $R_{G^{SCC}}(C_i)$ as the set of all variables contained in SCCs that are *reachable* from C_i in G^{SCC} (including C_i). Thereby, note that *reachability* in G^{SCC} is defined analogous to G. That is, $R_{G^{SCC}}(C_i)$ can be interpreted as the set of variables on which C_i 'depends' and which have to be taken into consideration when forecasting variables in C_i. We specify the 'minimal' VAR system in the following definition.

Definition 2.4 (Relevant Variables)**.** *Given a subset of interest $y^{\mathrm{I}} \subseteq y$, the minimal VAR is a VAR composed of the series that are contained in the relevant SCCs given by:*

$$R(y^{\mathrm{I}}) := \bigcup_{\{C_i \,:\, y^{\mathrm{I}} \cap C_i \neq \emptyset\}} R_{G^{SCC}}(C_i).$$

Thereby, the set of variables that are reachable from SCC C_i ($i = 1, \ldots, k$) is defined as:

$$R_{G^{SCC}}(C_i) := \left\{ y_j \in y : (j \in C_i) \vee \left(\exists l \in \{1, \ldots, k\} : j \in C_l : C_i \overset{P}{\rightsquigarrow} C_l \right) \right\}.$$

Remark 2.9. Definition 2.4 states that the minimal VAR is the one composed of the series in the SCCs which contain elements of y^{I} and all SCCs that may be reached from these SCCs.

We illustrate the graph theoretical concepts using a simple example, starting with a four-dimensional VAR(1) with coefficient matrix as shown on the left side in Figure 2.1.[7] The associated directed graph indicates that this system has two strongly connected components. The set of vertices of the first SCC C_1 consists of variables 1 and 3, i.e. $V_1 = \{1, 3\}$, as vertex 1 may be reached from vertex 3 and vice versa. The set of vertices of the second SCC C_2 consists of variables 2 and 4, i.e. $V_2 = \{2, 4\}$. Note that the vertices within each strongly

[7] To avoid cluttering of the graph, we exclude self-loops from the graphical representation.

connected component are mutually reachable (see Definition 2.2) and each variable (vertex) is in exactly one SCC (see Remark 2.4). Moreover, SCC C_1 is reachable from SCC C_2 but not vice versa.

We may also illustrate Remark 2.6 as it is easy to see that we may reorder the variables such that a lower block-triangular matrix results. For this purpose, we order the SCC that has no leaving edges first. In our example, SCC C_1 containing the set of vertices $V_1 = \{1, 3\}$, has no leaving edges and is hence ordered first. The new ordering of variables is then (1, 3, 2, 4), which results in the following reordered VAR matrix:

$$A^* = \left(\begin{array}{cc|cc} a_{11} & a_{13} & 0 & 0 \\ a_{31} & a_{33} & 0 & 0 \\ \hline a_{21} & 0 & 0 & a_{24} \\ 0 & a_{43} & a_{42} & a_{44} \end{array}\right) = \left(\begin{array}{c|c} A_{11} & 0 \\ \hline A_{21} & A_{22} \end{array}\right),$$

where the coefficients within the matrix A^* still have their original names. Obviously, this matrix has the desired block-triangular form. After having grouped the variables according to the strongly connected components, we may now draw a corresponding component graph according to the partition of the matrix A^*. The resulting component graph is shown in the right panel of Figure 2.1. This component graph indicates that component C_1 may be reached from component C_2 but not vice versa. Consequently, if the variable(s) of interest are included in component C_1, then the minimal VAR only includes the variables contained in C_1 but not those contained in C_2. In contrast, if the variable(s) of interest are contained in C_2, a corresponding VAR needs to include the variables from C_2 and in addition, the variables from C_1 as C_1 may be reached from C_2. For instance, if the variable of interest is e.g. variable 3, then the VAR needed in forecasting or structural analysis needs to contain all variables that are in the corresponding component C_1. In our example, these are the variables 1 and 3. In contrast, if the variable of interest is variable 2, then we need to include all variables in C_2 and C_1, i.e. all four variables, in the VAR model. Similarly, we may find the minimal VAR from the component graph even if we have more than one variable of interest. In our example, if variables 1 and 3 are of interest, a VAR for just these two variables suffices as both variables form a strongly connected component (C_1) and no other strongly connected component can be reached from C_1. In contrast, if variables 2 and 4 are of interest, we need a VAR with all variables from C_1 and C_2 as C_1 may be reached from C_2. In other words, we need all four variables. Now assume that variables 1 and 2 are of interest. Then again, we need to consider all variables from the strongly connected components that include variable 1 and variable 2, which in our example boils down to again using all variables since there are only the two components C_1 and C_2.

2.3 Econometric causality concepts and graphs

The graph theoretical concepts discussed in Section 2.2 have a close relation to multi-step causality concepts in time series econometrics. In this section we explain how the two concepts are related and show that the set of variables selected for a minimal VAR by the SCC method as in Definition 2.4 of Section 2.2 coincides with the set of variables that are multi-step causal for at least one of the variables of interest.

The simple notion of Granger causality (see Granger (1969)) is known to neglect any indirect effects and influences of 'auxiliary' variables as it is based on 1-step ahead predictability. Consequently, the original definition of Granger non-causality is not helpful in the context of variable selection. A more general causality concept that takes into account all indirect effects of auxiliary variables is known as multi-step causality and has been formally introduced into the literature by Dufour and Renault (1998).[8] Informally, a subset y^B of the variables causes another subset y^A at a specific horizon h if the best linear forecast for y^A at horizon h can be improved by including the variables in y^B in the information set. Dufour and Renault (1998) discuss necessary and sufficient conditions for non-causality at different forecast horizons h. Dufour et al. (2006) focus on developing corresponding multi-step non-causality tests in the context of VAR models. For our purpose, it is convenient to note that multi-step non-causality at different horizons may be formulated as linear exclusion restrictions on the so-called direct VAR model. For $h \geq 1$, we write this direct VAR model as:

$$y_{t+h} = \Pi_1^{(h)} y_t + \cdots + \Pi_p^{(h)} y_{t-p+1} + u_{t+h}^{(h)}, \tag{2.3.1}$$

where this representation is obtained by successive substitution from the VAR in (2.2.1). Dufour and Renault (1998) show that $\Pi_1^{(0)} = I_K$, $\Pi_s^{(1)} = A_s$, $\Pi_s^{(h+1)} = A_{s+h} + \sum_{l=1}^{h} A_{h-l+1}\Pi_s^{(l)} = \Pi_{s+1}^{(h)} + \Pi_1^{(h)} A_s$ and the MA($h-1$) innovation term $u_{t+h}^{(h)} = \sum_{j=0}^{h-1} \Pi_1^{(h)} u_{t-j}$.

Given sets of indices A and B, let $\Pi_{AB,s}^{(h)}$ denote the submatrix of $\Pi_s^{(h)}$ consisting of the intersection of rows with indices in A and columns with indices in B. If A and B are singletons, say $A = \{k\}$, $B = \{l\}$, we simply write $\Pi_{kl,s}^{(h)}$. We reproduce Theorem 3.1 of Dufour and Renault (1998) tailored to the regular, finite VAR case.

Theorem 2.1 (Dufour-Renault (1998))**.** *Given $y^A, y^B \subseteq y$ and y is generated by a regular, finite-order VAR(p) as in (2.2.1), it is:*

$$y^B \nrightarrow_h y^A \quad \Leftrightarrow \quad \forall s = 1, \ldots, p : \Pi_{AB,s}^{(h)} = 0,$$

[8] The effect of intermediate variables have also been pointed out earlier by e.g. Sims (1980), Penm and Terrell (1986) and Lütkepohl (1993) but Dufour and Renault (1998) were the first who formalized the concept of multi-step causality in a general framework.

where 0 *indicates a zero matrix of appropriate dimension. That is,* y^B *does not cause* y^A *at horizon* h *if and only if all the relevant coefficients in the direct VAR model for horizon* h *are zero.*

By definition, y^B *causes* y^A *at lag* h if at least one of the parameter matrices in the above theorem is not zero. For some indices I, we denote by $C(y^{\mathrm{I}})$ the set consisting of the variables in y^{I} itself and all variables that cause y^{I} at any horizon in the above sense. When I is a singleton, say $\mathrm{I} = \{i\}$, we write $C(y_i)$ instead. Formally, we define the causal variables in the following definition.

Definition 2.5 (Causal variables)**.** *Given* $y^{\mathrm{I}} \subseteq y$*, the set of variables that cause* y^{I} *is given by:*

$$C\left(y^{\mathrm{I}}\right) := \left\{ y_j \in y : \left(y_j \in y^{\mathrm{I}}\right) \vee \left(\exists h \in \mathbb{N} : y_j \to_h y^{\mathrm{I}}\right) \right\}.$$

First, we investigate the case of a VAR with $p = 1$. For this case, we show that the coefficients of the direct VAR representation $\Pi_1^{(h)}$ are related to the set of paths in the directed graph representing the VAR model. To show this, note that the direct VAR representation for $p = 1$ is:

$$y_{t+h} = \Pi_1^{(h)} y_t + u_{t+h}^{(h)}, \quad \text{with} \quad \Pi_1^{(h)} = A_1^h \text{ for all } h \in \mathbb{N}. \tag{2.3.2}$$

By induction, each element $\Pi_{ij,1}^{(h)}$ in $\Pi_1^{(h)}$ can be linked to the set of all paths that lead from the associated vertices i to j in h steps. We state this result formally in Theorem 2.2.

Theorem 2.2. *Given two variables* $y_i, y_j \in y$ *following a regular VAR(1) as in (2.2.1), the entry at position* (i, j)*,* $\Pi_{ij}^{(h)}$*, corresponds to the set of paths:*

$$\mathbb{P}_{ij}^{(h)} = \{P : P = (e_1, \ldots, e_h) : \forall k = 1, \ldots, h : e_k = (v_{k-1}, v_k) \in E, v_0 = i, v_h = j\}, \tag{2.3.3}$$

leading from vertex i *to vertex* j *for all* $h \in \mathbb{N}$ *in that:*

$$\Pi_{ij}^{(h)} = \sum_{P \in \mathbb{P}_{ij}^{(h)}} \prod_{(l,m) \in P} a_{lm}. \tag{2.3.4}$$

Proof. See Appendix 2.A.1.

$\mathbb{P}_{ij}^{(h)}$ is the set of all paths of length h leading from vertex i to vertex j. Theorem 2.2 essentially states, that the coefficients of the direct VAR $\Pi_{ij}^{(h)}$ can be written in terms of sums of products of autoregressive coefficients in A_1, where the indices correspond to edges on different paths from i to j. To illustrate this, consider again our simple VAR(1) from Figure

2.1. In this example, there are two paths of length $h = 2$ from variable 2 to variable 3, thus the set of paths is:

$$\mathbb{P}_{23}^{(2)} = \{\langle(2,1),(1,3)\rangle, \langle(2,4),(4,3)\rangle\}.$$

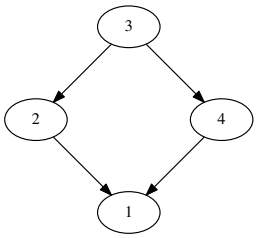
Using the result from Theorem 2.2, we find for $h = 2$, $i = 2$ and $j = 3$:

$$\Pi_{23,1}^{(2)} = a_{21}a_{13} + a_{24}a_{43}.$$

Note that the indices of the VAR coefficients match with the edges of the path set $\mathbb{P}_{23}^{(2)}$. Obviously, the same $\Pi_{23,1}^{(2)}$ would be obtained from the direct VAR coefficient definition.

At first sight, the result of Theorem 2.2 seems to imply that variable j would be multi-step causal for variable i, whenever there exists at least one path from i to j. However, $\mathbb{P}_{ij}^{(h)} \neq \emptyset$ does *not* imply that variable y_j causes y_i as the following example illustrates. Consider a VAR(1) with the associated directed graph:

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.5 & 1 & 0 & 0 \\ 0 & 0.5 & 1 & 0.5 \\ -0.5 & 0 & 0 & 1 \end{pmatrix}$$



In this example, we have $\mathbb{P}_{31}^{(1)} = \emptyset$ and $\mathbb{P}_{31}^{(2)} = \{\langle(3,2),(2,1)\rangle, \langle(3,4),(4,1)\rangle\}$ but variable y_3 is not caused by y_1 at neither horizon one nor horizon two since:

$$\Pi_{31}^{(2)} = a_{32}a_{21} + a_{34}a_{41} = 1/4 - 1/4 = 0.$$

Furthermore, one can easily verify $\Pi_{31}^{(h)} = 0$ and $\mathbb{P}_{31}^{(h)} \neq \emptyset$ for all $h \geq 2$. This 'canceling-out' effect of course happens very rarely with any real data set and most reasonable estimation methods. Therefore, one might exclude it by assumption.

Assumption 2.1. *Given a VAR(1) system with* $\Pi_{ij}^{(h_1)} = 0$ *for all* $h_1 > 1$*, then it is:*

$$\forall h_2 \in \mathbb{N} : \forall k \in \{1,\ldots,K\} : a_{ik}\Pi_{kj}^{(h_2)} = 0.$$

We basically assume that if variable j is multi-step non-causal for variable i, then this is because there is no path form i to j and not because there is a path from i to j with VAR coefficients such that there is 'canceling-out'. Assumption 2.1 thus ensures the correspondence between paths and causality as in Lemma 2.1.

Lemma 2.1. *Given a VAR(1) system and Assumption 2.1, for all $h \geq 1$:*

$$\mathbb{P}_{ij}^{(h)} \neq \emptyset \quad \Leftrightarrow \quad \Pi_{ij}^{(h)} \neq 0. \tag{2.3.5}$$

Proof. See Appendix 2.A.2.

This result states that variable j is multi-step causal for variable i if and only if there is at least one path from variable i to variable j. Under Assumption 2.1, the strongly connected components can be now be interpreted very easily.

Lemma 2.2. *Given a VAR(1) system and Assumption 2.1, the strongly connected components are sets of variables that are mutually causal.*

Proof. See Appendix 2.A.3.

This follows immediately from the definition of a SCC as for each pair i and j, there is a path from i to j and from j to i. The last step is to analyze multi-step causality across SCCs. Consider a strongly connected component C_i and, as in Section 2.2, denote the set of all SCCs that are reachable from C_i by $R_{G^{SCC}}(C_i)$. Then, all variables in $R_{G^{SCC}}(C_i)$ are multi-step causal for variables in C_i as there is a path from any variable in C_i to the variables in $R_{G^{SCC}}(C_i)$.

Finally, based on the foregoing discussion and for a given set of variables of interest y^{I}, we note that the variables, which are multi-step causal for y^{I} are all the variables in the SCCs that contain elements of y^{I} and all variables in all SCCs that may be reached from these SCCs. Note that under Assumption 2.1, this coincides with Definition 2.4 of the minimal set of relevant variables $R(y^{\mathrm{I}})$ from Section 2.2. Moreover, remember Definition 2.5 of causal variables $C(y^{\mathrm{I}})$. In case of 'canceling-out', the set of relevant variables $R(y^{\mathrm{I}})$ will be larger than the set of causal variables $C(y^{\mathrm{I}})$. We summarize this result more formally in Theorem 2.3.

Theorem 2.3. *Given a VAR(1) system, $C(y^{\mathrm{I}}) \subseteq R(y^{\mathrm{I}})$, that is, all variables that cause y^{I} are contained in the set of relevant variables. If Assumption 2.1 is true, then the set of relevant variables is identical to the set of causal variables, $C(y^{\mathrm{I}}) = R(y^{\mathrm{I}})$.*

Proof. See Appendix 2.A.4.

This establishes the relation between the set of relevant variables found from the graph of strongly connected components and the variables that are multi-step causal for the variables of interest.

To transfer this result to a general VAR(p) model, let Assumption 2.1 hold for the companion matrix of the corresponding VAR(1) representation. Then, we show that

it is equivalent to define the set of causal variables $C(y^{\mathrm{I}})$ based on either the VAR(p) representation or the VAR(1) companion form in Theorem 2.4.

Theorem 2.4. *Given a finite-order VAR(p) model* (2.2.1)*, Assumption 2.1 for the companion matrix* (2.2.3) *of the corresponding VAR(1) representation* (2.2.2) *and* $y^{\mathrm{I}} \subseteq y$*:*

$$\exists h_1 \in \mathbb{N} : y_j \rightarrow_{h_1} y^{\mathrm{I}} \quad \Leftrightarrow \quad \exists h_2 \in \mathbb{N} : \Pi_{\mathrm{I}j,1}^{(h_2)} \neq 0. \tag{2.3.6}$$

Proof. See Appendix 2.A.5.

Consequently, the set of causal variables is independent of the fact whether we use the VAR(p) or the corresponding VAR(1) companion form to define it given that Assumption 2.1 holds for the companion matrix. This is because for $i, j \leq K$, causality in the companion form representation only depends on the upper left $(K \times K)$ block of all $\mathbf{A}^h$ which is $\Pi_1^{(h)}$. Furthermore, the set of relevant variables $R\left(y^{\mathrm{I}}\right)$ as constructed in Definition 2.4 is based on the concept of *reachability* in G. This directed VAR graph G defined in Definition 2.1 relies on a general finite-order VAR(p) model written in VAR(1) companion form. Hence, the set of relevant variables is defined in the same way for $p = 1$ and $p > 1$, i.e. one first constructs the companion matrix $\mathbf{A}$ which determines graph G and finally $R\left(y^{\mathrm{I}}\right)$. Therefore, Theorem 2.3 also holds for a general finite-order VAR(p) model.

Remark 2.10. Given that Assumption 2.1 holds for the companion matrix of the VAR(1) representation of a general finite-order VAR(p) model, all results transfer to VAR(p) processes. We also note again that Assumption 2.1 is not restrictive at all as the 'canceling-out' effect will essentially never occur in practice.

Remark 2.11. One can show that *reachability* in our sense based on Definition 2.1 is equivalent to this imposed by the graph definition of Eichler (2007) (see Remark 2.2). This is shown formally in Appendix 2.B. Thus, it does not matter whether one defines the set of relevant variables $R\left(y^{\mathrm{I}}\right)$ based on our directed VAR graph or based on the graph defined by Eichler (2007).

2.4 Empirical illustration

To illustrate the usefulness of the graph theoretical approach for selecting the relevant information set, we apply the method to a large set of US economic time series. We focus on the selection of variables and on impulse response analysis of the selected models.

We start from a set of 41 quarterly economic time series that includes a large variety of macroeconomic and financial series over a period from 1975Q1-2014Q4. The collection of variables is similar to related studies as e.g. Jarociński and Maćkowiak (2017) and Kascha

and Trenkler (2015) and includes real GDP and its components, business cycle indicators, various price measures and interest rates, monetary aggregates and a number of labor market variables. In addition, the data includes exchange rate data together with three key variables for the Euro area (Euro area GDP, Euro area CPI and a Euro area interest rate). A detailed list with variables and data sources is provided in Appendix 2.C.

To apply the approach discussed in Section 2.2, we first transform the data to stationarity. This involves taking logarithms and/or differences depending on the property of the respective variable.[9] We describe the details of data preparation and document the transformations by reporting the transformation codes listed in Appendix 2.C.

In what follows, we apply the graph theoretical methods to a sparse VAR, i.e. a VAR with a number of zero coefficients in the autoregressive matrices. In our application, these sparse VARs are selected by applying the least absolute selection and shrinkage operator (LASSO) in the context of the VAR model. There is ample evidence in the literature that LASSO is a useful device and often leads to forecasts that are more precise than standard (unrestricted or subset) VARs (see e.g. Kascha and Trenkler (2015) and references therein). While in principle other methods for subset selection may be employed, we only use LASSO and point out that the subset selection is not the main focus of our paper. Instead, we start from a given subset structure and explore how this structure can be used to detect the smallest possible VAR system.

2.4.1 Variable selection

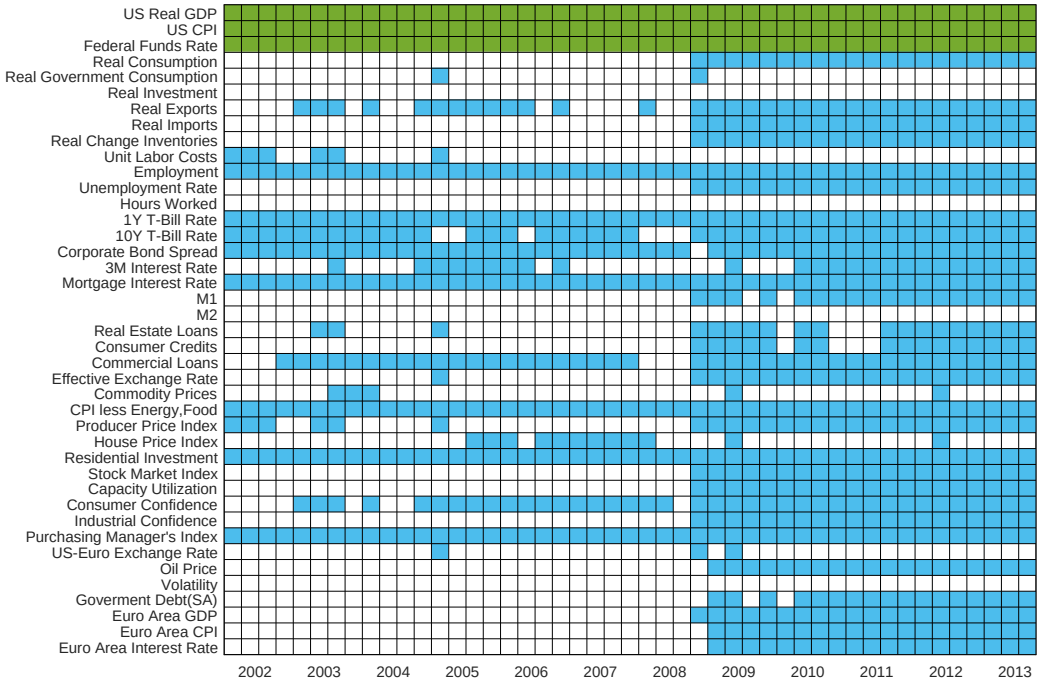
To illustrate the variable selection, we choose real US GDP, the US consumer prices (CPI), and the federal funds rate as variables of interest y^{I}. This includes three key economic variables often analyzed with VARs and also corresponds to the variables chosen by Jarociński and Maćkowiak (2017). Then, we estimate a LASSO-VAR(4) in all 41 variables such that the system contains lags up to one year. The shrinkage parameter in the LASSO approach is chosen by the Bayesian information criterion (BIC).[10]

To investigate which variables are selected into a VAR and how the selection changes over time, we use our method recursively in an expanding window setup, where initial estimation ends in 2002Q2. Consequently, we report selection results for several periods and show them in graphical form in Figure 2.2. The rows in the checkerboard graph correspond to the different economic variables, whereas the columns refer to different ends of estimation samples. The filled green squares correspond to the variables of interest (here: GDP, CPI

[9] Transforming the variables to stationarity cancels possible common trends and cointegration relations between the variables. Extending the graphical methods to models with common trends such as cointegrated VARs and vector error correction models is left for future research.

[10] As for the implementation of the VAR-LASSO we follow the paper by Kascha and Trenkler (2015) and refer to their paper for details.

Figure 2.2: Variable selection results



Variables of interest y^{I}: US Real GDP, CPI and FFR (green). Relevant variables as selected by graphical method (blue) and variables not selected (white). Sample period: 1975Q1-2014Q4.

and the Federal Funds Rate), a filled blue square in a specific row indicates that the variable in that row is selected into the minimal VAR in the period corresponding to the column. Accordingly, a white square indicates that the variable has not been selected into the minimal VAR in a particular period.

We find that six variables are always selected into the minimal VAR in each of the periods considered. This includes employment, the 1-year T-Bill rate, the mortgage interest rate, the CPI less food and energy price index, residential investment and the purchasing manager's index. A tentative interpretation of this result may be that these variables form the minimal set of additional variables that should be considered if a VAR in output, inflation and interest rate is of interest. We note that some of these variables have typically not been included in related empirical studies on the effects of monetary policy shocks. We also compared our set of selected variables with those from Jarociński and Maćkowiak (2017) and note that the variables of interest y^{I} show a very low posterior probability (< 0.1) of being Granger-causally prior to the variables (except residential investment) that have been always selected by our graphical methods. However, based on Table 1 of Jarociński and Maćkowiak (2017), one would select 22 series since y^{I} has a posterior probability of less than 0.1 to be Granger-causal prior to them. Hence, though the variables selected by our approach are

also chosen using the method of Jarociński and Maćkowiak (2017), the graphical technique applied in this paper seems to be more in favor of a small VAR system of relevant variables.

In a number of periods additional variables are selected into the minimal VAR. The number of selected variables is quite large (on average 21.5 out of the 41 are selected). This provides evidence that the dynamic relationship between economic variables is more complex than small scale VARs tend to suggest. While there are some changes as we increase the estimation sample, the overall selection of variables is relatively stable for the period before the 2008/2009 economic crisis. Interestingly, when the estimation sample extends beyond the crisis period, we observe that our method tends to select more variables, possibly suggesting that the linkages between variables have become more pronounced.

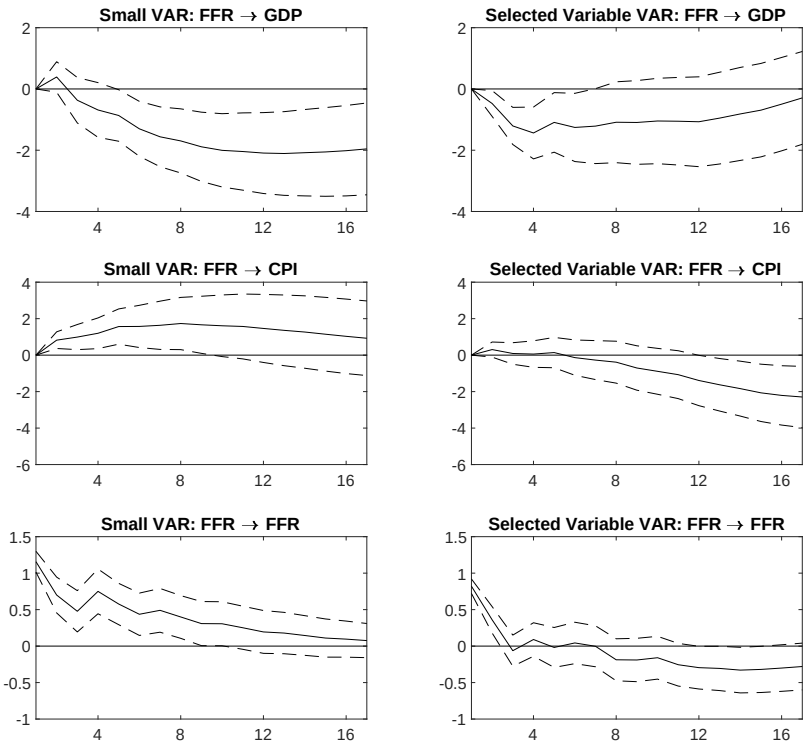
2.4.2 Impulse response analysis

We illustrate the effect of including the selected variables into a small VAR system on estimated impulse responses. As in Section 2.4.1, the small VAR consists of the variables of interest with real GDP, the CPI and the federal funds rates (FFR). These type of systems have also been used by Jarociński and Maćkowiak (2017) and Banbura et al. (2010).

Following standard specifications from the literature (see e.g. Christiano et al. (1999)), we have used VAR(4) models in the (log) levels of the variables of interest for the comparison of impulse responses. For the VAR with selected variables, we have added the six variables (again in (log) levels) that have been selected for all considered sample periods in Section 2.4.1. We use the standard ordering of variables and thus include employment, the CPI less food and energy price index, residential investment, and the purchasing manager's index in the group of 'slow moving' variables, i.e. they are ordered above the federal funds rate variable. In contrast, the 1-year T-Bill rate and the mortgage interest rate are in the group of 'fast moving' variables and are consequently ordered below the federal funds rate. Using a Cholesky decomposition, this ordering implies that a shock in the federal funds rate (typically labeled a monetary policy shock) may have immediate impact on the 'fast moving' variables, while the 'slow moving' variables may only react with a lag of one quarter. All VAR models are estimated by unrestricted multivariate LS (i.e. no shrinkage is applied) and the reported (pointwise) confidence intervals are asymptotic 95% intervals obtained using the 'delta method' (see e.g. Lütkepohl (2005, Section 3.7)).

In Figure 2.3, we report results for a sample that ends in 2008 to exclude the effects of the 2008/09 financial crisis and to take into account the break in the variable selection after 2008 (see Figure 2.3). The left panel of the figure shows the responses of GDP, CPI and the federal funds rate to a contractionary shock in the federal funds rate within the small VAR containing three variables. We find the typical pattern with a significant and persistent drop in output. We also find a significant increase in CPI, which is known as the 'price puzzle'

Figure 2.3: Impulse responses in small VAR and selected variable VAR



Left: Responses to a shock in FFR in 3-variable (small) VAR(4) including GDP, FFR, and CPI. Right: Responses to a shock in FFR in 9-variable VAR(4) with six selected additional variables. Sample period: 1975Q1-2008Q4.

because economic theory suggests a decrease rather than an increase in the price level after tightening monetary policy. In other words, the response of the price level in the small VAR is counter-intuitive. Adding the six selected variables changes the response patterns substantially. First, the drop in output is now much less persistent. In fact, two years after the shock the response of output is no longer significantly different from zero. Moreover, the price puzzle disappears: the reaction of consumer prices to a monetary policy shock is not significantly different from zero for more than three years. Thereafter, it shows the expected sign since in line with conventional economic theory it starts to be significantly negative. Thus, including the variables selected by our method leads to much more reasonable impulse response patterns and changes the interpretation of the results substantially. Using the right information set is of obvious importance for structural analysis.

We report additional results for different sample periods in Appendix 2.D. In the literature, it is more common to let the sample period end in 2007 when the researcher wants to exclude the global financial crisis. Therefore, Figure 2.D.1(a) uses data from 1975Q1 until 2007Q4 and shows responses that are very similar to those in Figure 2.3. Moreover, Figure 2.D.1(b) shows the responses for an extended sample that ends after 2014. Again, the addition of our

six selected variables clearly mitigates the prize puzzle, even though it does not impose the reaction of prices to a monetary policy shock to be significantly negative. Given our selected variables, the earliest possible starting date is 1972. Figure 2.D.2 shows the responses estimated using data from 1972Q1 up to the end of 2007 in (a) and up to the end of 2008 in (b), respectively. Besides the fact that the reaction of output seems to be quite persistent even after adding the six selected variables, the results are qualitatively and quantitatively very similar to these presented in the main analysis. Just note that for some of the samples considered, the response of consumer prices in the very first period after the monetary shock is still significantly positive. However, in these cases the lower confidence band is very close to zero.

It is interesting to note that the changes in the response pattern obtained by just adding six variables are to some extent similar to the changes obtained by Banbura et al. (2010) in their medium (20 variables) and large (131 variables) system for monthly data. In other words, it seems that our methodology provides a perspective that complements the 'large VAR' idea of Banbura et al. (2010).

2.5 Conclusion

This paper uses concepts from graph theory for variable selection in VAR models. To this end, we identify strongly connected components from the directed graph representing the dynamic relationships among the variables in a sparse VAR. We suggest to use relations among the strongly connected components in a so-called component graph to identify a minimal set of variables that we need to include in a VAR analysis for a small set of variables if forecasts or impulse response analysis are of interest. The paper adds to the existing literature by introducing a graphical method, which to the best of our knowledge has not been used for variable selection in econometrics.

We also show that there is a simple relation between the graph theoretical concept and multi-step causality and relate the paths in the graph to coefficients of a direct VAR system. It follows from the results in the paper that the set of relevant variables selected from the graphical approach coincides with the set of variables that are multi-step causal for the variables of interest.

We illustrate the usefulness of the variable selection method in a structural analysis of a small US monetary system (real GDP, CPI inflation and the federal funds rate) as the variables of interest. Given this set, we apply the graphical approach to select additional variables out of a large set of macroeconomic variables. The selected VAR typically includes some variables from the real sector (employment and residential investment), a forward looking indicator (purchasing manager's index), different interest rates (mortgage interest

rate and 1 year T-Bill rate) and a CPI related measure (CPI less food and energy). The selection of variables seems sensible from an economic point of view. Interestingly, we find that this list includes some variables that other researchers typically have not included in small monetary systems. Moreover, we find that including the selected variables for impulse response analysis is useful: In the small monetary system in output, inflation and interest rate, we find that including the selected variables avoids the so-called 'price puzzle'.

Overall our empirical results suggest that using graphical modeling for variable selection is a useful addition to the VAR econometricians' toolbox. The method complements existing methods for large data sets and is particularly useful if a researcher prefers to work with smaller scale models, e.g. for maintaining consistency with small scale theoretical models. Moreover, compared to alternative methods for large data sets, a graphical representation of the strongly connected components may give useful insights on the (causal) relationships among the VAR variables.

Extensions of the current paper could use graphical models for variable selection taking also the contemporaneous relationships among variables into account. Moreover, extending the approach to models with integrated and cointegrated variables would be of interest. We leave this for future research.

References

Banbura, M., D. Giannone, and L. Reichlin (2010). "Large Bayesian vector autoregressions". *Journal of Applied Econometrics* 25, 71–92.

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.

Brillinger, D. R. (1996). "Remarks concerning graphical models for time series and point processes". *Revista de Econometria* 16, 1–23.

Carriero, A., T. E. Clark, and M. Marcellino (2015). "Bayesian VARs: Specification choices and forecast accuracy". *Journal of Applied Econometrics* 30 (1), 46–73.

Carriero, A., G. Kapetanios, and M. Marcellino (2009). "Forecasting exchange rates with a large Bayesian VAR". *International Journal of Forecasting* 25 (2), 400–417.

– (2012). "Forecasting government bond yields with large Bayesian vector autoregressions". *Journal of Banking and Finance* 36 (7), 2026–2047.

Cheng, X. and B. E. Hansen (2015). "Forecasting with factor-augmented regression: A frequentist model averaging approach". *Journal of Econometrics* 186 (2), 280–293.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999). "Monetary policy shocks: What have we learned and to what end?" In: *Handbook of Macroeconomics*. Ed. by Taylor, J. B. and Woodford, M. Vol. 1. Handbook of Macroeconomics. Elsevier. Chap. 2, 65–148.

Clements, M. P. (2016). "Real-time factor model forecasting and the effects of instability". *Computational Statistics and Data Analysis* 100, 661–675.

Cormen, T. H., C. E. Leiserson, R. L. Rivest, and C. Stein (2009). "Introduction to Algorithms". In: The MIT Press. Chap. 22.

Dahlhaus, R. (2000). "Graphical interaction models for multivariate time series". *Metrika* 51 (2), 157–172.

Dahlhaus, R. and M. Eichler (2003). "Causality and graphical models for time series". In: *Highly structured stochastic systems*. Ed. by Green, P., Hjort, N., and Richardson, S. Oxford University Press, 115–137.

Demiralp, S. and K. D. Hoover (2003). "Searching for the causal structure of a vector autoregression". *Oxford Bulletin of Economics and Statistics* 65, 745–767.

Doan, T. A. and R. Todd (2010). "Causal ordering for multivariate linear systems". *Unpublished Work, Estima*.

Duff, I. S. and J. Reid (1978). "An implementation of Tarjan's algorithm for the block triangularization of a matrix". *ACM Transactions on Mathematical Software* 4 (2), 137–147.

Dufour, J. M., D. Pelletier, and E. Renault (2006). "Short run and long run causality in time series: Inference". *Journal of Econometrics* 132 (2), 337–362.

Dufour, J. M. and E. Renault (1998). "Short run and long run causality in time series: Theory". *Econometrica* 66 (5), 1099–1125.

Edwards, D. (2000). *Introduction to Graphical Modelling*. 2nd ed. Springer Texts in Statistics. New York: Springer-Verlag.

Eichler, M. (2006). "Graphical modeling of dynamic relationships in multivariate time series". In: *Handbook of Time Series Analysis: Recent Developments and Applications*. Ed. by Schelter, B., Winterhalder, M., and Timmer, J. Wiley.

Eichler, M. (2007). "Granger causality and path diagrams for multivariate time series". *Journal of Econometrics* 137 (2), 334–353.

– (2012). "Graphical modelling of multivariate time series". *Probability Theory and Related Fields* 153 (1-2), 233–268.

Eickmeier, S. and C. Ziegler (2008). "How successful are dynamic factor models at forecasting output and inflation? A meta-analytic approach". *Journal of Forecasting* 27 (3), 237–265.

Flamm, C., U. Kalliauer, M. Deistler, M. Waser, and A. Graef (2012). "Graphs for dependence and causality in multivariate time series". In: *System Identification, Environmental Modelling, and Control System Design*. Ed. by Wang, L. and Garnier, H. London: Springer London, 133–151.

Giannone, D., M. Lenza, D. Momferatou, and L. Onorante (2014). "Short-term inflation projections: A Bayesian vector autoregressive approach". *International Journal of Forecasting* 30 (3), 635–644.

Granger, C. W. J. (1969). "Investigating causal relations by econometric models and cross-spectral methods". *Econometrica* 37, 424–438.

Heinlein, R. and H. M. Krolzig (2012). "Effects of monetary policy on the US dollar/UK pound exchange rate. Is there a 'delayed overshooting puzzle'?" *Review of International Economics* 20 (3), 443–467.

Hoover, K., S. Demiralp, and S. J. Perez (2009). "Empirical identification of the vector autoregression: The causes and effects of U.S. M2". In: *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*. Ed. by Castle, J. and Shepard, N. Oxford: Oxford University Press, 37–58.

Jarociński, M and B. Maćkowiak (2017). "Granger causal priority and choice of variables in vector autoregressions". *Review of Economics and Statistics* 99 (2), 319–329.

Kascha, C. and C. Trenkler (2015). "Forecasting VARs, model selection, and shrinkage". *Working Paper ECON* (15-07).

Koop, G. M. (2013). "Forecasting with medium and large Bayesian VARs". *Journal of Applied Econometrics* 28 (2), 177–203.

Lauritzen, S. L. (1996). *Graphical Models*. Oxford Statistical Science Series. Oxford: Oxford University Press.

Ludvigson, S. C. and S. Ng (2007). "The empirical risk-return relation: A factor analysis approach". *Journal of Financial Economics* 83 (1), 171–222.

– (2009). "Macro factors in bond risk premia". *Review of Financial Studies* 22 (12), 5027–5067.

Lütkepohl, H. (1993). "Testing for causation between two variables in higher dimensional VAR models". In: *Studies in Applied Econometrics*. Ed. by Schneeweiß, H. and Zimmermann, K. F. Springer-Verlag, Heidelberg, 75–91.

– (2005). *New Introduction to Multiple Time Series Analysis*. Springer Science & Business Media.

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

Pearl, J. (2000). *Causality*. New York: Cambridge University Press.

Penm, J. and R. Terrell (1986). "The 'derived' moving-average model and its role in causality". In: *Essays in time series and allied processes : papers in honour of E. J. Hannan*. Ed. by Gani, J. and Priestly, M. Sheffield: Applied Probability Trust, 99–111.

Sims, C. A. (1980). "Macroeconomics and reality". *Econometrica* 48 (1), 1–48.
– (1982). "Policy analysis with econometric-models". *Brookings Papers on Economic Activity* (1), 107–164.
– (2015). "Causal orderings and exogeneity". *Lecture Notes*.
Stock, J. H. and M. W. Watson (2002). "Forecasting using principal components from a large number of predictors". *Journal of the American Statistical Association* 97 (460), 1167–1179.
– (2003). "Forecasting output and inflation: The role of asset prices". *Journal of Economic Literature* 41 (3), 788–829.
– (2012b). "Generalized shrinkage methods for forecasting using many predictors". *Journal of Business and Economic Statistics* 30 (4), 481–493.
– (2016). "Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics". In: *Handbook of Macroeconomics*. Ed. by Taylor, J. B. and Uhlig, H. Vol. 2. Elsevier. Chap. 8, 415–525.
Swanson, N. R. and C. W. J. Granger (1997). "Impulse response functions based on a causal approach to residual orthogonalization in vector autoregressions". *Journal of the American Statistical Association* 92 (437), 357–367.
Tarjan, R. E. (1972). "Depth-first search and linear graph algorithms". *SIAM Journal on Computing* 1 (2), 146–160.
Uhlig, H. (2009). "Comment on 'How has the Euro changed the monetary transmission mechanism?'" In: *NBER Macroeconomics Annual 2008*. Ed. by Acemoglu, D., Rogoff, K., and Woodford, M. Vol. 23. University of Chicago Press, 141–152.

Appendix 2.A Proofs

2.A.1 Proof of Theorem 2.2

Proof. We prove Theorem 2.2 by mathematical induction.

Base Case Let $h = 1$. Then, it is $\Pi_{ij}^{(h)} = \Pi_{ij}^{(1)} = a_{ij}$.

If $\mathbb{P}_{ij}^{(1)}$ is empty, $(i,j) \notin E$ what imposes $a_{ij} = 0$ because of Definition 2.1. Then, equality (2.3.4) holds trivially.

If $\mathbb{P}_{ij}^{(1)} \neq \emptyset$, it follows that $\mathbb{P}_{ij}^{(1)} = \{(i,j)\}$. Consequently, $\sum_{P \in \mathbb{P}_{ij}^{(1)}} \prod_{(l,m)\in P} a_{lm} = a_{ij} = \Pi_{ij}^{(1)}$, which proves (2.3.4) for $h = 1$.

Induction Step Let (2.3.4) hold for $h-1$, i.e.

$$\forall i,j \in \{1,\dots,K\} : \Pi_{ij}^{(h-1)} = \sum_{P \in \mathbb{P}_{ij}^{(h-1)}} \prod_{(l,m)\in P} a_{lm}. \tag{2.A.1}$$

Moreover, define:

$$E_i := \{k \in \{1,\dots,K\} : (i,k) \in E\} \quad \Rightarrow \quad \mathbb{P}_{ij}^{(h)} = \bigcup_{k \in E_i} \bigcup_{P \in \mathbb{P}_{kj}^{(h-1)}} [\{(i,k)\} \cup P]\,. \tag{2.A.2}$$

This just means that any path from vertex i to vertex j of length h can be decomposed into a tuple (i,k) and a path from vertex k to vertex j of length $(h-1)$ for all $k \in \{1,\dots,K\}$ with $a_{ik} \neq 0$ and $\mathbb{P}_{kj}^{(h-1)} \neq \emptyset$. Using this, we get:

$$\begin{aligned}
\Pi_{ij}^{(h)} &= \sum_{k=1}^{K} a_{ik} \Pi_{kj}^{(h-1)} \\
&\overset{(2.A.1)}{=} \sum_{k=1}^{K} a_{ik} \sum_{P \in \mathbb{P}_{kj}^{(h-1)}} \prod_{(l,m)\in P} a_{lm} \\
&= \sum_{k \in E_i} \sum_{P \in \mathbb{P}_{kj}^{(h-1)}} a_{ik} \prod_{(l,m)\in P} a_{lm} \\
&= \sum_{k \in E_i} \sum_{P \in \mathbb{P}_{kj}^{(h-1)}} \prod_{(l,m)\in [\{(i,k\}\cup P]} a_{lm} \\
&\overset{(2.A.2)}{=} \sum_{P \in \mathbb{P}_{ij}^{(h)}} \prod_{(l,m)\in P} a_{lm}.
\end{aligned}$$

□

2.A.2 Proof of Lemma 2.1

Proof. We prove both directions of the if-and-only-if statement (2.3.5):

"⇒" Let $\mathbb{P}_{ij}^{(h)} \neq \emptyset$. We prove $\Pi_{ij}^{(h)} \neq 0$ by mathematical induction.

Base Case Let $h = 1$. Then, it is $\mathbb{P}_{ij}^{(1)} = \{(i, j)\}$, i.e. $a_{ij} \neq 0$ due to Definition 2.1. Consequently, it is $\Pi_{ij}^{(1)} = a_{ij} \neq 0$.

Induction Step Let one direction of (2.3.5) hold for $(h-1)$, i.e.

$$\forall i, j \in \{1, \dots K\} \text{ with } \mathbb{P}_{ij}^{(h-1)} \neq 0, \text{ it is } \Pi_{ij}^{(h-1)} \neq 0. \tag{2.A.3}$$

Now, let $\mathbb{P}_{ij}^{(h)} \neq \emptyset$:

$$\begin{aligned} &\Rightarrow \exists k \in \{1, \dots, K\} : a_{ik} \neq 0 \wedge \mathbb{P}_{kj}^{(h-1)} \neq \emptyset \\ &\overset{(2.A.3)}{\Rightarrow} \Pi_{kj}^{(h-1)} \neq 0 \Rightarrow a_{ik}\Pi_{kj}^{(h-1)} \neq 0 \\ &\Rightarrow \Pi_{ij}^{(h)} = \sum_{l=1}^{K} a_{il}\Pi_{lk}^{(h-1)} \neq 0 \end{aligned}$$

because of Assumption 2.1 and $a_{ik}\Pi_{kj}^{(h-1)} \neq 0$.

"⇐" Let $\Pi_{ij}^{(h)} \neq 0$. Assume $\mathbb{P}_{ij}^{(h)} = \emptyset$. Consequently, by Theorem 2.2 it is $\Pi_{ij}^{(h)} = \sum_{P \in \mathbb{P}_{ij}^{(h)}} \prod_{(l,m) \in P} a_{lm} = 0$ because it is an empty sum. This is a contradiction. Thus, $\mathbb{P}_{ij}^{(h)} \neq \emptyset$ has to hold.

□

2.A.3 Proof of Lemma 2.2

Proof. Without loss of generality, let $i, j \in C_k$. By Definition 2.2, y_i and y_j are mutually reachable. Therefore, it is:

$$\begin{aligned} &\exists h_i \in \mathbb{N} : \mathbb{P}_{ij}^{(h_i)} \neq \emptyset \overset{\text{Lemma 2.1}}{\Longleftrightarrow} \Pi_{ij}^{(h_i)} \neq 0 \overset{\text{Theorem 2.1}}{\Longleftrightarrow} y_j \to_{h_i} y_i \Rightarrow y_j \in C(y_i), \\ &\exists h_j \in \mathbb{N} : \mathbb{P}_{ji}^{(h_j)} \neq \emptyset \overset{\text{Lemma 2.1}}{\Longleftrightarrow} \Pi_{ji}^{(h_j)} \neq 0 \overset{\text{Theorem 2.1}}{\Longleftrightarrow} y_i \to_{h_j} y_j \Rightarrow y_i \in C\left(y_j\right). \end{aligned}$$

Consequently, y_i and y_j are mutually causal. □

2.A.4 Proof of Theorem 2.3

Proof. First, we prove the general subset relation $C\left(y^{\mathrm{I}}\right) \subseteq R\left(y^{\mathrm{I}}\right)$.

"$\subseteq$" Let $y_j \in C\left(y^{\mathrm{I}}\right)$ and w.l.o.g. $j \in C_{j^*}$. Due to Definition 2.5, either

(a) $y_j \in y^{\mathrm{I}}$, or

(b) $\exists h \in \mathbb{N} : y_j \to_h y^{\mathrm{I}}$ has to hold.

In case (a), $y_j \in C_{j^*} \cap y^{\mathrm{I}}$ and therefore $C_{j^*} \cap y^{\mathrm{I}} \neq \emptyset$. By Definition 2.4, $y_j \in R_{G^{SCC}}\left(C_{j^*}\right)$ has to hold what directly imposes $y_j \in R\left(y^{\mathrm{I}}\right)$.
In case (b):

$$\exists h \in \mathbb{N} : y_j \to_h y^{\mathrm{I}} \overset{\text{Theorem 2.1}}{\iff} \Pi_{\mathrm{I}j}^{(h)} \neq 0 \Leftrightarrow \exists i \in \mathrm{I} : \Pi_{ij}^{(h)} \neq 0 \overset{\text{Theorem 2.2}}{\Rightarrow} \mathbb{P}_{ij}^{(h)} \neq \emptyset.$$

Let $i \in C_{i^*}$. Obviously, j is *reachable* from i in G. Hence, C_{j^*} is also *reachable* from C_{i^*} in G^{SCC}. Consequently, it is $y_j \in R_{G^{SCC}}\left(C_{i^*}\right)$. Because of $i \in C_{i^*}$ and $i \in \mathrm{I}$, it is $y_i \in y^{\mathrm{I}} \cap C_{i^*}$ and therefore $y^{\mathrm{I}} \cap C_{i^*} \neq \emptyset$. Thus, $R_{G^{SCC}}\left(C_{i^*}\right) \subseteq R\left(y^{\mathrm{I}}\right)$ and therefore $y_j \in R\left(y^{\mathrm{I}}\right)$.

"$\supseteq$" Let Assumption 2.1 hold, $y_j \in R\left(y^{\mathrm{I}}\right)$ and w.l.o.g. $j \in C_{j^*}$. Due to Definition 2.4, then:

$$\exists i^* \in \{1, \ldots, k\} : C_{i^*} \cap y^{\mathrm{I}} \neq \emptyset : y_j \in R_{G^{SCC}}\left(C_{i^*}\right).$$

Again by Definition 2.4, this means that either

(i) $j \in C_{i^*}$, or

(ii) $C_{i^*} \overset{P}{\rightsquigarrow} C_{j^*}$ has to hold.

In case (i):

$$C_{i^*} \cap y^{\mathrm{I}} \neq \emptyset \Rightarrow \exists i \in \mathrm{I} : i \in C_{i^*} \Rightarrow i, j \in C_{i^*} \overset{\text{Lemma 2.2}}{\Longrightarrow} y_j \to_h y_i \Rightarrow y_j \in C\left(y^{\mathrm{I}}\right).$$

In case (ii):

$$\exists i_1 \in C_{i^*} : \exists j_1 \in C_{j^*} : \exists h \in \mathbb{N} : \mathbb{P}_{i_1 j_1}^{(h)} \neq \emptyset. \tag{2.A.4}$$

As Assumption 2.1 holds, we can apply Lemma 2.2 to get:

$$\forall i_1 \in C_{i^*} : \exists h_1 \in \mathbb{N} : y_{i_1} \to_{h_1} y_i \overset{\text{Theorem 2.1}}{\iff} \Pi_{i,i_1}^{(h_1)} \neq 0 \overset{\text{Lemma 2.1}}{\iff} \mathbb{P}_{i,i_1}^{(h_1)} \neq \emptyset,$$
$$\forall j_1 \in C_{j^*} : \exists h_2 \in \mathbb{N} : y_j \to_{h_2} y_{j_1} \overset{\text{Theorem 2.1}}{\iff} \Pi_{j_1,j}^{(h_2)} \neq 0 \overset{\text{Lemma 2.1}}{\iff} \mathbb{P}_{j_1,j}^{(h_2)} \neq \emptyset.$$

As we have paths from vertex i to i_1, vertex i_1 to j_1 and vertex j_1 to j, there is also a path from vertex i to j as we can simply connect them, i.e.

$$\mathbb{P}_{ij}^{(h+h_1+h_2)} \neq \emptyset \overset{\text{Lemma 2.1}}{\iff} \Pi_{ij}^{(h+h_1+h_2)} \neq 0 \Rightarrow \Pi_{\mathrm{I}j}^{(h+h_1+h_2)} \neq 0 \overset{\text{Theorem 2.1}}{\iff} y_j \to_{(h+h_1+h_2)} y^{\mathrm{I}} \Rightarrow y_j \in C\left(y^{\mathrm{I}}\right).$$

□

2.A.5 Proof of Theorem 2.4

Proof. We prove both directions of the if-and-only-if statement (2.3.6) whereas the second implication is trivial:

"⇒" Let y_j be a variable that causes y^{I}, i.e. $\exists h_1 \in \mathbb{N} : y_j \to_{h_1} y^{\mathrm{I}}$. By Theorem 2.1 it follows:

$$\exists s_1 \in \{1,\dots,p\} : \Pi_{\mathrm{I}j,s_1}^{(h_1)} \neq 0 \Rightarrow \exists i \in \mathrm{I} : \Pi_{ij,s_1}^{(h_1)} \neq 0.$$

For $s_1 = 1$, $\Pi_{ij,1}^{(h_1)} \neq 0$ and thus $\Pi_{\mathrm{I}j,1}^{(h_1)} \neq 0$ follows immediately. So, let $s_1 \in \{2,\dots,p\}$. Note that one can easily show that $J\mathbf{A}^h = (\Pi_1^{(h)},\dots,\Pi_p^{(h)})$ with $J = (I_K, 0, \dots, 0)$ being a $(K \times Kp)$ selection matrix. Consequently, $\Pi_{ab,s}^{(h)} = \mathbf{A}_{a,(s-1)K+b}^{h}$ holds for all $a, b \in \{1,\dots,K\}$, $s \in \{1,\dots,p\}$ and $h \in \mathbb{N}$.
Therefore, $\mathbf{A}_{i,(s_1-1)K+j}^{h_1} = \Pi_{ij,s_1}^{(h_1)} \neq 0$ holds. Moreover, $\mathbf{A}_{l,l-K} = 1$ for $l = K+1,\dots,Kp$ by construction of the companion matrix (2.2.3). Applying this argument for $l = (s_1 - 1)K + j$ and using Assumption 2.1, it follows that $\Pi_{ij,s_1-1}^{(h_1+1)} = \mathbf{A}_{i,(s_1-2)K+j}^{h_1+1} \neq 0$. Continuing this argument, one shows that $\Pi_{ij,1}^{(h_1+s_1-1)} = \mathbf{A}_{ij}^{h_1+s_1-1} \neq 0$. So, $\Pi_{\mathrm{I}j,1}^{(h_1+s_1-1)} \neq 0$ is true.

"⇐" Assume that there is a $h \in \mathbb{N}$ such that $\Pi_{\mathrm{I}j,1}^{(h)} \neq 0$. Due to Theorem 2.1, this directly implies $y_j \to_h y^{\mathrm{I}}$.

□

Appendix 2.B Equivalence of graph definitions

In this part, we show that it is equivalent to define *reachability* for a VAR(p) in G or in G^1. Therefore, we first restate the definition of *reachibility* for both graphs.

Definition 2.6. *We write $u \overset{P}{\rightsquigarrow} u'$, if vertex u' is reachable from u in G, i.e. if there is a path from u to u' in E. Analogously, we write $u \overset{P^1}{\rightsquigarrow} u'$, if vertex u' is reachable from u in G^1, i.e. if there is a path from u to u' in E^1.*

Remark 2.12. Note that the following relations hold:

$$\text{Let } (i,j) \in E \text{ for } i = K+1,\dots,Kp \Leftrightarrow \mathbf{A}_{ij} \neq 0 \Leftrightarrow j = i - K. \tag{2.B.1}$$

$$\text{Let } (i,j) \in E^1 \Leftrightarrow \exists s \in \{1,\dots,p\} : A_{ij,s} = \mathbf{A}_{i,(s-1)K+j} \neq 0 \Leftrightarrow (i,[s-1]K+j) \in E. \tag{2.B.2}$$

Before stating the result of interest, we first prove two lemmas.

Lemma 2.3. *Let $(i,j) \in E^1$. Then, j is reachable from i in G, i.e. $i \overset{P}{\rightsquigarrow} j$.*

Proof. Let $(i,j) \in E^1$. Because of (2.B.2), $\exists s \in \{1,\dots,p\} : (i,[s-1]K+j) \in E$. For $s = 1$, $(i,j) \in E$ follows directly and therefore $i \overset{P}{\rightsquigarrow} j$. So, let $s \in \{2,\dots,p\}$. Due to (2.B.1), it is:

$$([s-1]K+j,[s-2]K+j),\dots,(j+K,j) \in E.$$

Consequently, there is a path of length s in E from i to j, i.e. $i \overset{P}{\rightsquigarrow} j$. □

Lemma 2.4. *Let $i,j \leq K$ and $i \overset{P}{\rightsquigarrow} j$ through path $P^* = (v_0,\dots,v_s)$ in E such that $v_l > K$ f.a. $l = 1,\dots,s-1$ with $s \in \{1,\dots,p\}$. Then, $(i,j) \in E^1$ holds.*

Proof. Because of $(v_{s-1},v_s) \in E$, $v_s = j$ and (2.B.1), it is $v_{s-1} = v_s + K = j + K$. Applying the same argument repeatedly, we get $v_1 = j + (s-1)K$. Moreover, $(v_0,v_1) = (i, j + [s-1]K) \in E$. Due to (2.B.2), (i,j) is contained in E^1. □

Using these results, we show the equivalence of *reachability* in G and G^1, respectively.

Corollary 2.1. *Let $i,j \leq K$. Then, it holds:*

$$i \overset{P^1}{\rightsquigarrow} j \quad \Leftrightarrow \quad i \overset{P}{\rightsquigarrow} j. \tag{2.B.3}$$

Proof. We prove both directions of the if-and-only-if statement:

"⇒" Let $i \overset{P^1}{\rightsquigarrow} j$. Then, there is a path $P^* = (v_0,\dots,v_n)$ of length $n \in \{1,\dots,K-1\}$ in E^1. So, for all $l \in \{1,\dots,n\} : (v_{l-1},v_l) \in E^1$. Lemma 2.3 yields $v_{l-1} \overset{P}{\rightsquigarrow} v_l$ for all $l \in \{1,\dots,n\}$. Moreover, we can simply connect these paths to find that there is a path from $v_0 = i$ to $v_n = j$, i.e. $i \overset{P}{\rightsquigarrow} j$.

"⇐" Let $i \overset{P}{\rightsquigarrow} j$. Then, there is a path $P^* = (v_0,\dots,v_s)$ of length $s \in \{1,\dots,p\}$ with $v_0 = i, v_s = j$ and $(v_{l-1},v_l) \in E$ for all $l = 1,\dots,s$. Let $L := \{l : v_l \leq K\}$ and its cardinality be equal to $n \in \{2,\dots,s+1\}$. Define $0 = l_1 < \dots < l_n = s$ such that $l_m \in L$ for $m = 1,\dots,n$. Moreover, define $P_m^* := \left(v_{l_m},\dots,v_{l_{m+1}}\right)$ for $m = 1,\dots,n-1$. This means that we have decomposed P^* into $n-1$ contiguous subpaths P_m^* that satisfy

the assumptions of Lemma 2.4.
Due to Lemma 2.4, $\left(v_{l_m}, v_{l_{m+1}}\right) \in E^1$ for all $m = 1, \ldots, n-1$. Consequently, there is a path $P^{**} = \left(v_{l_1}, \ldots, v_{l_n}\right)$ with $v_{l_1} = v_0 = i$ and $v_{l_n} = v_s = j$ in E^1, i.e. $i \overset{P^1}{\rightsquigarrow} j$.

□

Because we show in Corollary 2.1 that *reachability* in G and G^1 is equivalent, these two definitions determine the same SCCs. Moreover, they impose the same component graph why the resulting sets of relevant variables are also identical.

Appendix 2.C Data

We describe the data used in the empirical illustrations. Raw data for most series are obtained from the FRED database and Table 2.C.1 shows the corresponding FRED mnemonics. We construct some variables from splicing two series in order to obtain long time series: As a measure for the exchange rate, we use the US/DM exchange rate (`EXGEUS`) until 1998Q4. From 1999Q1 we use `EXUSEU` and splice both series accordingly. The resulting variable is called `EXCH`. We follow McCracken and Ng (2016) and use `OILPRICE` (Spot Oil Price) until 1985Q4 and `MCOILWTICO` (Crude Oil Price, Cushing) since 1986Q1, since the former series has been discontinued. The resulting series is labeled `POIL` in our data set. To obtain a crude measure of stock market volatility, we simply use the squared stock market returns, since the time series of volatility indices in FRED are rather short. This series is called `VOLA`. Seasonally adjusted series have been taken from FRED where necessary. The time series on Government Debt (`GFDEBTN`) has been seasonally adjusted by the authors using X-ARIMA-13. The resulting series is `GFDEBTNSA`. The Euro area time series have been added using the update 15 to the AWM database maintained at the ECB. The AWM mnemonics for the real GDP, CPI, and a short-term interest rates are `YER`, `HICP`, and `STN`. `HICP` has been seasonally adjusted by the authors using X-ARIMA-13. We use `EMUGDP`, `EMUHICPSA`, and `EMURS` to denote the three Euro area variables.

The last columns in Table 2.C.1 lists the transformation codes 1-6, corresponding to the following transformations of the series y_t: (1) no transformation, y_t, (2) Δy_t, (3) $\Delta^2 y_t$, (4) $400 \times \log(y_t)$, (5) $400 \times \Delta \log(y_t)$, (6) $400 \times \Delta^2 \log(y_t)$.

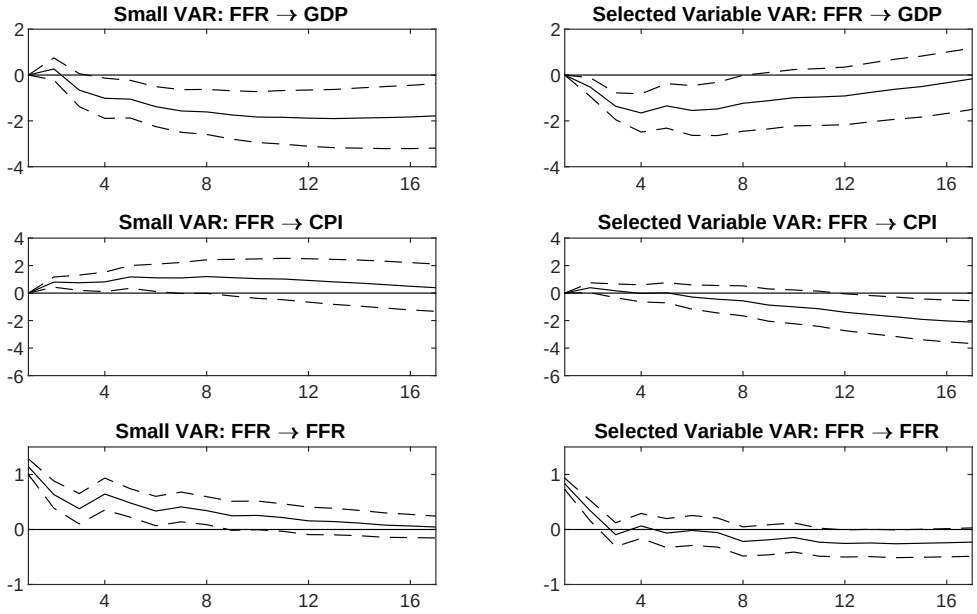
Table 2.C.1: Variables, data sources and transformations: Graph-VAR

Name	Mnemonic	Transf.Code
Real GDP	GDPC96	5
CPI	CPIAUCSL	6
Federal Funds Rate	FEDFUNDS	2
Real Consumption	PCECC96	5
Real Government Consumption	GCEC1	5
Real Investment	GPDIC1	5
Real Exports	EXPGSC1	5
Real Imports	IMPGSC1	5
Change in Real Inventories	CBIC96	1
Unit Labor Cost	ULCNFB	5
Employment	PAYEMS	5
Unemployment Rate	UNRATE	2
Hours worked	HOHWMN02USQ065S	1
1-year T-Bill Rate	GS1	2
10-year T-Bill Rate	GS10	2
Corporate Bond Spread	AAAFFM	1
Lending Rate to NFCs	TB3MS	3
Mortgage Rate	MORTG	2
M1	M1SL	6
M2	M2SL	6
Government Debt GFDEBTNSA	GFDEBTN. seas.adj: X-13	5
Real Estate Loans	REALLN	5
Consumer Credits	TOTALSL	5
Commercial Loans	BUSLOANS	5
Dollar/Euro Exchange Rate (EXCH)	spliced from EXGEUS and EXUSEU	5
Effective Exchange Rate	NNUSBIS	5
Oil Price (POIL)	spliced from OILPRICE and MCOILWTICO	5
Commodity Prices	CUSR0000SAC	6
Consumer Prices (excl. food, energy)	CPILFESL	6
Producer Price Index	PPIACO	5
House Prices	USSTHPI	6
Real Housing Investment	PRFI	5
Total Share Prices	SPASTT01USQ661N	5
Volatility Index	VIXCLS	5
Capacity Utilization	CUMFNS	2
Consumer Confidence	CSCICP03USM665S	2
Industrial Confidence	BSCICP03USM665S	2
Purchasing Manager's Index	NAPM	1
Real GDP (Euro Area)	AWM mnemonic: YER	5
CPI (Euro Area)	AWM mnemonic: HICP, seas.adj: X-13	6
Short term interest rate (Euro Area)	AWM mnemonic: STN	2

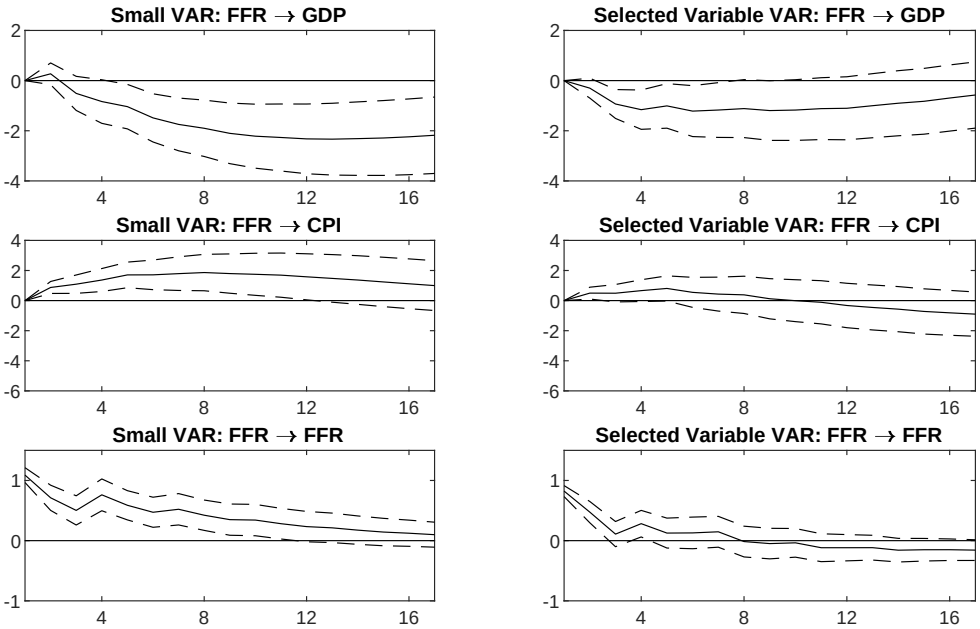
Note: The table shows FRED and AWM database names together with the transformation codes. See Appendix 2.C for a detailed description of the transformations.

Appendix 2.D Additional results

Figure 2.D.1: Impulse responses with selected variables: robustness analysis I



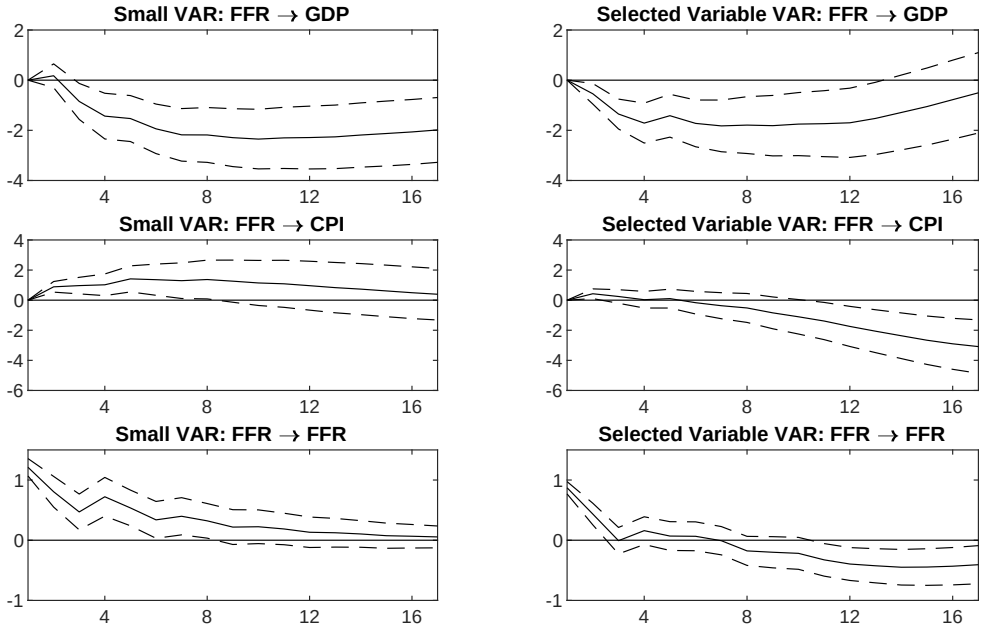
(a) Sample period: 1975Q1-2007Q4



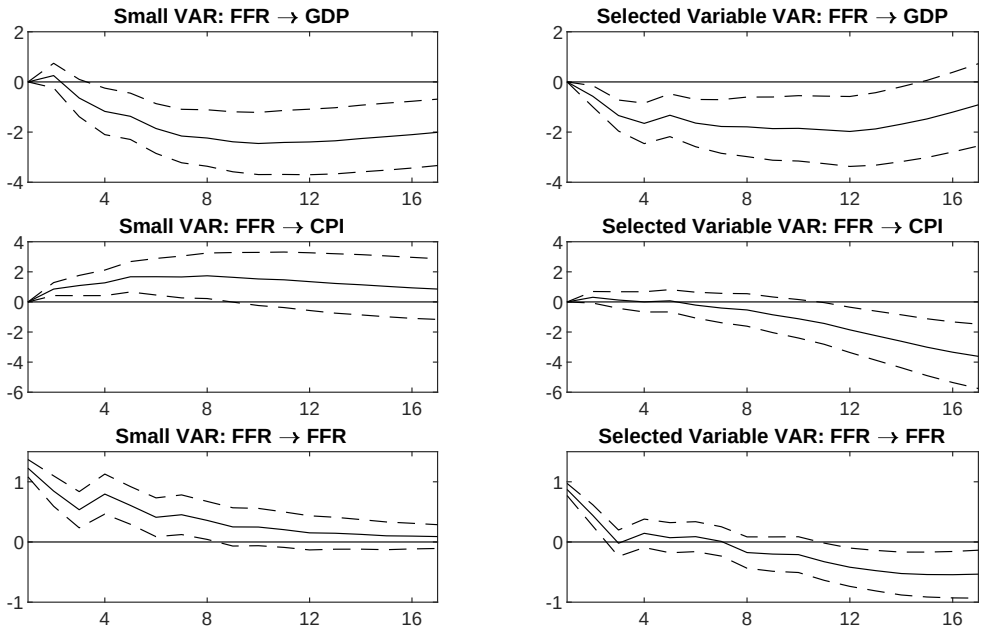
(b) Sample period: 1975Q1-2014Q4

Left: Responses to a shock in FFR in 3-variable (small) VAR(4) including GDP, FFR, and CPI. Right: Responses to a shock in FFR in 9-variable VAR(4) with six selected additional variables.

Figure 2.D.2: Impulse responses with selected variables: robustness analysis II



(a) Sample period: 1972Q1-2007Q4



(b) Sample period: 1972Q1-2008Q4

Left: Responses to a shock in FFR in 3-variable (small) VAR(4) including GDP, FFR, and CPI. Right: Responses to a shock in FFR in 9-variable VAR(4) with six selected additional variables.

Chapter 3

The effects of oil supply shocks on the macroeconomy: a Proxy-FAVAR approach

3.1 Introduction

Exogenous oil price movements have been found to be important drivers of US business cycles (Hamilton, 2009). Since Kilian (2009), the literature has further acknowledged the importance of decomposing these movements into supply and demand components, given that they are associated with very different implications for policy makers. In particular, supply shocks are of interest, as they are associated with a toxic combination of increasing consumer prices and decreasing economic activity. A sound understanding of the exact nature of supply shocks is essential for the design of effective policy responses.

In this paper, I develop novel time series methodology to simultaneously identify the effects of two oil supply shocks with a very distinct nature, labeled as a 'flow' supply shock and a 'news' supply shock. Flow shocks concern the amount of oil that comes out of the ground and therefore reveal a contemporaneous drop in global oil production and are caused e.g. by periods of political turmoil and wars. In turn, news shocks capture unexpected shifts in the expectation about future oil production. These shocks can arise if global oil producers announce production targets below or above the expectation of market participants. While both shocks have been studied separately in the literature, in this paper I focus on their joint identification within a single model. This ensures their unconfoundedness, and further allows me to assess their relative importance in explaining fluctuations in oil market variables and macroeconomic indicators. To achieve this goal in a credible way, I develop a structural factor-augmented vector autoregressive (FAVAR) model. In particular, I extend the standard oil market model of Kilian and Murphy (2014) by the large data set of Kerssenfischer (2019), which contains 132 US macro and financial variables. Within that model, I identify the supply shocks by instrumental variables (IV). Here, I use the exogenous supply shock for the Organization of the Petroleum Exporting Countries (OPEC) constructed by Kilian (2008a) as an instrument for flow shocks, and the high-frequency shock of Känzig (2019) as an instrument for news shocks.

The benefits of this modeling choice are threefold. First, a simultaneous identification of both shocks in one model ensures that the shocks are always orthogonal, which is not the case if they are identified in two separate models. Second, the incorporation of a large information set allows me to address concerns of informational deficiency and provides a consistent tool to assess the effects of supply shocks on a wide range of macroeconomic variables. Finally, identification via external instruments gives a credible identification scheme that avoids controversial restrictions typically imposed to identify oil market models. My main findings indicate that news shocks about future oil supply heavily contribute to the oil price movements, while the effects of flow surprises are rather low. On the other hand, an unexpected change in contemporaneous oil supply significantly impacts macroeconomic and

financial indicators like economic activity, stock markets, interest rates and employment. Here, the effects of the news shock on these variables remain neglectable. Hence, surges of the oil price stemming from supply depletions seem to affect the economy much more severe than news shocks which are estimated to be a more important driver of the real price of oil.

The present work makes two major contributions. The first is empirical, where the paper contributes to the oil market literature by assessing the relative importance of flow shocks and news shocks as drivers of the US business cycle. Both shocks have been studied extensively in the literature, however, always in a standalone manner.[1] Prominent studies that have assessed the importance of flow supply shocks are Kilian (2009), Kilian and Murphy (2012, 2014), Baumeister and Hamilton (2019) and Caldara et al. (2019). While these studies largely agree on the directional effects of these shocks on macroeconomic variables, their relative importance for driving oil prices remains highly controversial. The disagreement is mainly driven by identifying assumptions about the model implied (short-run) oil price elasticities.[2] A benefit of the modeling strategy adapted in this paper is that by using an external instrument approach, I can identify the shock without taking a side in this debate. The series provided by Kilian (2008a) serves as a proxy for the flow shock in this paper. Regarding oil supply news shocks, Känzig (2019) provides the pioneering work studying their effects on key US macroeconomic variables. Using a structural VAR identified by external instruments he identifies strong effects of news shocks on oil market variables and macroeconomic indicators. His novel identification strategy relies on exploiting high-frequency movements in oil price futures in short windows around OPEC announcements. The shock series of Känzig (2019) is used in this paper as an instrument for the news shocks.

Methodologically, this paper contributes by developing a FAVAR model that can be identified by multiple external instruments. Such a modeling framework is useful to mitigate two potential shortcomings in conventional SVAR analysis: informational insufficiency and a noncredible identification (Bruns, 2018). With respect to the first, there is a common perception that in order to reduce the number of parameters to estimate, researchers tend to choose the number of variables included in the system too low. As noted in Giannone et al. (2004), Bernanke et al. (2005) and Juvenal and Petrella (2015), this can harm the identification of shocks due to informational insufficiency and omitted variables. Instead, a FAVAR model is able to cope with the curse of dimensionality by summarizing the common time series dynamics of a large panel of variables in a distinctly lower number of latent factors,

[1] One exception is Känzig (2019) who also disentangles these two shocks in a short robustness analysis. While he computes impulse response functions, he does not further proceed with the assessment of their relative importance for driving macroeconomic variables. Furthermore, he focuses on a small scale VAR and hence does not benefit from the approach based on a rich information set used in this paper.

[2] In particular, models imposing a very low short run price elasticity of supply find very little effects of flow supply shocks, while models that allow for larger values find more pronounced effects (Herrera et al., 2019).

which in turn are assumed to follow VAR dynamics.[3] Moreover, a FAVAR model allows to consider the effect of the structural shocks of interest on a much broader set of variables than allowed by conventional SVAR models. Within SVAR models, econometricians tend to circumvent this problem by adding one variable at a time to a baseline specification (see e.g. Beaudry and Portier (2014) or Gertler and Karadi (2015)). Instead, a factor model secures that the responses are estimated internally consistent because it models the impact of the shocks on all variables within the same model. Recent studies that use this technique to assure informational sufficiency for identification are Juvenal and Petrella (2015) and Kerssenfischer (2019). The second potential shortcoming is credible identification of the shocks without relying on model-internal restrictions. As surveyed in Kilian and Lütkepohl (2017), most of the conventional identification approaches rely on such internal restrictions to identify the shock, e.g. by sign restrictions (Canova and De Nicoló, 2002; Faust, 1998; Uhlig, 2005) or exclusion restrictions (Bernanke, 1986; Blanchard and Quah, 1989) on the effects on endogenous variables. These approaches need to rely heavily on economic theory which often is not available or remains highly controversial. In this paper, I incorporate the identification strategy inherent in so-called 'Proxy-SVARs', proposed independently by Stock and Watson (2012a) and Mertens and Ravn (2013). Their identification strategy relies on external variables that are correlated with the shocks of interest, and uncorrelated with all other shocks in the model. Thereby, one can elegantly exploit external information to set identifying restrictions instead of imposing any internal restrictions on the structural parameters. To get reliable results, a researcher has to ensure that the employed instruments are fully exogenous.

The challenging task of inference is tackled using a Bayesian framework, a deliberate choice for various reasons. To begin with, Bayesian techniques are extremely useful when multiple structural shocks are to be identified by external instruments. In this case, in addition to the standard relevance and exogeneity conditions, more restrictions are required to disentangle the shocks of interest (Mertens and Ravn, 2012). Within classical inference, Lunsford (2015), Mertens and Montiel Olea (2018) and Lakdawala (2019) impose additional exclusion constraints for that purpose. However, these exclusion restrictions are chosen for convenience and can be economically questionable, particularly in the oil market application targeted by this paper. In turn, a Bayesian approach as proposed in Arias et al. (2019), Caldara and Herbst (2019) and Giacomini et al. (2019), allows for a consistent treatment of set

[3]This can be best illustrated by a simple counting example. Think of a data set with 100 variables with 300 observations each that is modeled using five factors and six autoregressive lags. Hence, slightly above 2,150 parameters have to be estimated. This is less than the number of unknowns in a classical VAR with six lags and 20 variables. In general, for factor models the number of parameters increases proportional to the number of variables N while VAR parameters rise with N^2.

identification, and hence allows to disentangle the multiple shocks e.g. by sign restrictions.[4] Another main point is that standard Bayesian inference algorithms are valid even if the instrument is weak (Poirier, 1998), while this does not hold under the frequentist paradigm (Montiel-Olea et al., 2016). Consequently, researchers can apply the algorithm presented in this paper without judging a priori whether the instruments used are weak or not. In contrast, if a researcher believes that an instrument is not weak, she can reflect this through the prior distribution as advocated by Arias et al. (2019) and Caldara and Herbst (2019). Next, a Bayesian framework is appropriate to handle highly parameterized models. Note that though FAVAR models shrink the dynamics in the economy to factors and therefore decrease the number of parameters remarkably, the inclusion of many lags can still be a challenge in terms of valid inference based on a rather low number of observations. For instance, in monthly oil VAR applications as in Hamilton and Herrera (2004) or Kilian (2008b) using 12 to 24 lags has become a standard, and conventional inference can quickly become unreliable. Finally, frequentist inference (e.g. by bootstrap) for factor-augmented models is non-trivial as illustrated by Yamamoto (2019). Hence, a Bayesian approach seems a more straightforward choice for the model studied in this paper.

The methodological contribution in this paper is related to several other papers. The closest is that of Bruns (2018), who extends the Proxy-SVAR specification of Caldara and Herbst (2019) to introduce a Bayesian Proxy-FAVAR (BP-FAVAR) model. The main difference of the present paper is that I adopt the Proxy-SVAR specification advocated by Arias et al. (2019), which I incorporate into the FAVAR Gibbs sampler using a Metropolis-Hasting steps. Extracting the information of more than one proxy to identify one structural shock in the way of Caldara and Herbst (2019) implies an overidentified $\mathcal{A}$-matrix that can only be handled using a non-conjugate prior on the structural parameters. However, Caldara and Herbst (2019) are unable to identify multiple shocks simultaneoulsy because their non-conjugate prior discriminates between observationally equivalent models such that the identified structural shocks cannot be disentangled in a reasonable way. Thus, employing the approach of Caldara and Herbst (2019) may violate the orthogonality assumption since only one structural shock can be identified at a time. Instead, the specification of Arias et al. (2019) works with conjugate prior distributions for the structural parameters of the model. Hence, their approach ensures that equal prior and posterior weight is assigned to observationally equivalent models, which is important if multiple shocks are to be identified by external instruments.

The remainder of this paper is organized as follows. Section 3.2 introduces the structural FAVAR model, revises identification by external instruments, carefully lists and justifies the

[4]See Piffer and Podstawski (2017) who identify an uncertainty shock and a news shock using proxy variables and sign restrictions, respectively, in a frequentist way. However, the way they explore the set of observationally equivalent models is effectively a Bayesian procedure.

choice of priors for unknown parameters and sketches the resulting posterior distributions. In Section 3.3, I use the model to simultaneously identify flow and news supply shocks and investigate their relative importance for driving both oil market specific and US macroeconomic variables. Section 3.4 concludes.

3.2 Proxy-FAVAR model

In this section I introduce the structural FAVAR model identified by proxy variables. To simplify notation, all variables are assumed to have zero mean. This allows to suppress an intercept in all equations. The core of the model will be a structural VAR model (SVAR) for a set of n_1 latent factors F_t and n_2 observed time series W_t. Using notation from Rubio-Ramírez et al. (2010), the structural model reads:

$$z_t' A_0 = \sum_{j=1}^{p} z_{t-j}' A_j + \varepsilon_t', \qquad t = p+1, \ldots, T, \tag{3.2.1}$$

where $z_t' = (F_t', W_t')$ is of dimension $n = n_1 + n_2$ and consists of unobserved factors as well as observed time series. Furthermore, A_j are $n \times n$ matrices for $j = 0, \ldots, p$ with A_0 invertible and structural shocks are denoted by $\varepsilon_t \sim \mathcal{N}(0, I_n)$.

The model will be identified by external instruments, and hence, vector z_t will be modeled within a Proxy-SVAR. A coherent treatment of Proxy-SVARs can be achieved using the framework of Arias et al. (2019), which involves augmenting equation (3.2.1) by a $k \times 1$ vector of proxy variables m_t. Let $\tilde{z}_t' = (z_t', m_t')$ be an augmented vector of dimension $n + k$. Then, equation (3.2.1) is extended to the following Proxy-SVAR:

$$\tilde{z}_t' \tilde{A}_0 = \sum_{j=1}^{p} \tilde{z}_{t-j}' \tilde{A}_j + \tilde{\varepsilon}_t', \qquad t = p+1, \ldots, T, \tag{3.2.2}$$

where $\tilde{A}_j$ are $(n + k) \times (n + k)$ matrices for $j = 0, \ldots, p$ with $\operatorname{rank}(\tilde{A}_0) = n + k$ and $\tilde{\varepsilon}_t' = (\varepsilon_t', v_t') \sim \mathcal{N}(0, I_{n+k})$ with v_t being a $k \times 1$ innovation vector for the proxies that may have the interpretation of a measurement error (Mertens and Ravn, 2012).

The purpose of adding equations for m_t to the model is only to identify k structural shocks. Note that in this framework exactly one proxy variable identifies exactly one structural shock. Therefore, $n \geq k$ has to hold to ensure that there are not more proxies than shocks contained in the model. Equation (3.2.2) is a standard structural VAR in simultaneous equation form with p autoregressive lags in $\tilde{z}_t$. To ensure that the time series vector z_t still satisfies the

SVAR dynamics of equation (3.2.1), a set of restrictions must be imposed such that m_t does not affect the dynamics of z_t. Following Arias et al. (2019), these restrictions read:

$$\tilde{A}_j = \begin{pmatrix} A_j & \Gamma_{j,1} \\ 0_{k\times n} & \Gamma_{j,2} \end{pmatrix}, \tag{3.2.3}$$

where $\Gamma_{j,1}$ is a $n \times k$ matrix and $\Gamma_{j,2}$ is a $k \times k$ matrix. Note that with these restrictions, the dynamics implied by the first n equations of (3.2.2) are identical with those implied by the classical SVAR of equation (3.2.1). To have a more compact notation, define $x_t' = (\tilde{z}_{t-1}', \ldots, \tilde{z}_{t-p}')$ and $\tilde{A}_+' = (\tilde{A}_1', \ldots, \tilde{A}_p')$. Then, equation (3.2.2) can be written more compactly as:

$$\tilde{z}_t' \tilde{A}_0 = x_t' \tilde{A}_+ + \tilde{\varepsilon}_t', \qquad t = p+1, \ldots, T. \tag{3.2.4}$$

Finally, it is worth mentioning that equation (3.2.4) gives a structural equation representation at the time series process $\tilde{z}_t$. Corresponding reduced form dynamics are given by postmultiplying with $\tilde{A}_0^{-1}$, yielding:

$$\tilde{z}_t' = x_t' B + u_t', \qquad t = p+1, \ldots, T, \tag{3.2.5}$$

where $B = \tilde{A}_+ \tilde{A}_0^{-1}$ and $u_t \sim \mathcal{N}(0, \Sigma_u)$ with a full covariance matrix $\Sigma_u = (\tilde{A}_0^{-1})' \tilde{A}_0^{-1}$.

Note that F_t are unobserved factors, and therefore an additional measurement equation is necessary to complete the specification of the proxy-augmented FAVAR model. In the following, denote a large panel of time series data by Y_t and let its dimension be $N \times 1$. Following Bernanke et al. (2005), I assume that these informational time series summarized in Y_t can be modeled by the latent factors F_t and the observed time series W_t using the following observation equation:

$$Y_t = \Lambda_F F_t + \Lambda_W W_t + e_t, \qquad t = 1, \ldots, T, \tag{3.2.6}$$

where Λ_F is a $N \times n_1$ matrix of factor loadings, Λ_W is a $N \times n_2$ matrix and $e_t \sim \mathcal{N}(0, R)$ is an idiosyncratic error term that is assumed to be serially and cross-sectionally uncorrelated. Hence, the covariance matrix R is diagonal with elements r_i $(i = 1, \ldots, N)$. However, since the factors F_t are unobserved, factor model (3.2.6) is not identified without additional constraints (Bai et al., 2016). Sufficient identifying restrictions are to set the upper $n_1 \times n_1$ block of Λ_F to identity and the upper $n_1 \times n_2$ block of Λ_W to zero (Bernanke et al., 2005), i.e. $\Lambda_F' = (I_{n_1}, \tilde{\Lambda}_F')$ and $\Lambda_W' = (0_{n_2 \times n_1}, \tilde{\Lambda}_W')$ where $0_{n_2 \times n_1}$ is a $n_2 \times n_1$ matrix of zeros. Note that the identification condition in Λ_F corresponds to the named factor normalization of Stock and Watson (2016). Moreover, the benefit of this identification strategy is that it allows

to leave the VAR dynamics implied for factors and observed time series (equation (3.2.5)) completely unrestricted, which allows to proceed with the identification analysis of the structural shocks in a standard manner. Moreover, note that the measurement equation (3.2.6) has a dynamic factor model representation (Stock and Watson, 2016) with $\Lambda = (\Lambda_F, \Lambda_W)$:

$$Y_t = \Lambda z_t + e_t, \qquad t = 1, \ldots, T. \tag{3.2.7}$$

Note that all variables in Y_t are assumed to be stationary. However, in case of data sets with potential cointegration relationships, it can be reasonable to include non-stationary variables in the model and consider how they are impacted by structural shocks. Therefore, this framework allows to include non-stationary time series in the observed part W_t. This practice is standard in the factor literature (see e.g. Bernanke et al. (2005) or Kerssenfischer (2019)). In addition, to ensure a common scaling of all variables and to avoid the inclusion of an intercept, Y_t, W_t and m_t are standardized to have zero mean and unit variance, another common procedure in factor analysis. In the following, I will refer to (3.2.4) and (3.2.6) as the Proxy-FAVAR model.

3.2.1 Identification

Incorporating the Proxy-SVAR approach of Arias et al. (2019) implies that several structural shocks can be simultaneously identified by proxy variables. Hence, their orthogonality is ensured by construction. The benefit of this for practitioners in a factor model framework can be illustrated based on the empirical application of Stock and Watson (2012a): they use various instruments to identify different structural shocks - one at a time - in a dynamic factor model.[5] The pairwise empirical correlation coefficients between the resulting shock series suggest that most innovations identified in this work are far away from being orthogonal to each other, suggesting that the results are confounded to some extent. In my framework, the structural shocks of interest are orthogonal by construction.

Intuitively, each of the k external instruments in m_t identifies one structural shock in ε_t. Without loss of generality, let the k structural innovations ordered last, i.e. $\varepsilon_{n-k+1,t}, \ldots, \varepsilon_{n,t}$ be the shocks identified. Following Mertens and Ravn (2012), the IV identifying restrictions are given by:

$$\mathrm{E}\left[\varepsilon_{1:n-k,t} m_t'\right] = 0_{(n-k)\times k}, \tag{3.2.8}$$

$$\mathrm{E}\left[\varepsilon_{n-k+1:n,t} m_t'\right] = V, \tag{3.2.9}$$

[5] They identify an oil shock, a monetary policy shock, a productivity shock, an uncertainty shock, a liquidity shock and a fiscal policy shock using different instruments for each.

where $\varepsilon_{a:b,t}$ are columns a to b of the vector of structural shocks ε_t and V is a nonsingular $k \times k$ matrix. Equation (3.2.9) is known as relevance condition and requires that the instruments carry some information for the structural shocks they were designed for. Equation (3.2.8) states that the proxies are uncorrelated with all other shocks considered, also known as exogeneity constraint.

Arias et al. (2019) show how exogeneity constraint (3.2.8) and relevance condition (3.2.9) can be translated to restrictions for the Proxy-SVAR parameters. First, due to the structure of partition matrix $\tilde{A}_0$ stated in (3.2.3), one can easily determine its inverse following Magnus and Neudecker (2019):

$$\tilde{A}_0^{-1} = \begin{pmatrix} A_0^{-1} & -A_0^{-1}\Gamma_{0,1}\Gamma_{0,2}^{-1} \\ 0_{k\times n} & \Gamma_{0,2}^{-1} \end{pmatrix}. \tag{3.2.10}$$

As outlined in (3.2.5), postmultiplication of (3.2.4) with $\tilde{A}_0^{-1}$ yields the reduced form dynamics for vector $\tilde{z}_t$:

$$\tilde{z}_t' = x_t'\tilde{A}_+\tilde{A}_0^{-1} + \tilde{\varepsilon}_t'\tilde{A}_0^{-1}, \qquad t = p+1,\dots,T. \tag{3.2.11}$$

Since $\tilde{z}_t' = (F_t', W_t', m_t')$ and $\tilde{\varepsilon}_t' = (\varepsilon_t', v_t')$, the last k equations of (3.2.11) combined with the explicit exposition of $\tilde{A}_0^{-1}$ in (3.2.10) yield the reduced form dynamics for the instrument vector:

$$m_t' = x_t'\tilde{A}_+ \begin{pmatrix} -A_0^{-1}\Gamma_{0,1}\Gamma_{0,2}^{-1} \\ \Gamma_{0,2}^{-1} \end{pmatrix} - \varepsilon_t' A_0^{-1}\Gamma_{0,1}\Gamma_{0,2}^{-1} + v_t'\Gamma_{0,2}^{-1}, \qquad t = p+1,\dots,T. \tag{3.2.12}$$

Moreover, the structural shocks ε_t are orthogonal to x_t and v_t. Hence, combining proxy equation (3.2.12) with exogeneity restriction (3.2.8) and relevance condition (3.2.9) directly implies:

$$\mathrm{E}(\varepsilon_t m_t') = -A_0^{-1}\Gamma_{0,1}\Gamma_{0,2}^{-1} = \begin{pmatrix} 0_{(n-k)\times k} \\ V \end{pmatrix}. \tag{3.2.13}$$

Note that $-A_0^{-1}\Gamma_{0,1}\Gamma_{0,2}^{-1}$ represents the upper right block of $\tilde{A}_0^{-1}$ (see (3.2.10)). Hence, because of (3.2.13), the upper right $(n-k)\times k$ block of $\tilde{A}_0^{-1}$ is restricted to zero by exogeneity constraint (3.2.8). This complements the zero restriction on the lower left $k \times n$ block of $\tilde{A}_0$ specified in (3.2.3).

Generally, the IV condition does yield two sets of shocks, one which is correlated with the instruments and one that is uncorrelated. If the size of one group is greater than one, additional constraints are necessary to disentangle shocks within that category. Hence, if

$k > 1$ holds, one has to further think about additional restrictions to disentangle the shocks of interest. In the empirical application considered in this paper, I impose sign restrictions on corresponding columns of A_0^{-1}.

Finally, if a researcher is confident about the relevance of the instrument(s), she might impose this as prior information (Caldara and Herbst, 2019). A convenient measure for instruments strength is the reliability matrix $\left(\left[\Gamma_{0,2}^{-1}\right]' \Gamma_{0,2}^{-1} + VV'\right)^{-1} VV'$ derived in Mertens and Ravn (2012). Intuitively, its eigenvalues can be interpreted as the share of the variance in m_t that can be explained by the structural shocks $\varepsilon_{n-k+1:n,t}$. In the empirical application, I follow Arias et al. (2019) and impose a lower bound on the eigenvalues, thereby ruling out irrelevance of the instruments a priori.

3.2.2 Priors measurement equation

Although Bernanke et al. (2005) also consider a Bayesian approach, due to its tremendous computational burden the literature about FAVAR models is highly dominated by frequentists. However, taking into account the points made in Section 3.1, I use a Bayesian approach for inference of the Proxy-FAVAR model. Alternative Bayesian approaches rely on Kalman filtering techniques to sample the latent factors (Belviso and Milani, 2006; Bruns, 2018). Instead, I make use of methods from the Gaussian Markov random fields literature (Chan, 2017; Rue et al., 2009) to increase the efficiency in sampling the factors considerably. Before outlining a Markov Chain Monte Carlo (MCMC) algorithm to summarize the posterior distribution, I start with the exposition of the exact prior distributions used for the parameters.

First, consider the prior distributions of the unknown parameters in the measurement equation (3.2.6). These include the unrestricted elements of the factor loadings Λ_F, the free entries of Λ_W and the diagonal entries r_i $(i = 1, \ldots, N)$ of covariance matrix R. Furthermore, the dynamic model structure implies implicit prior distributions for the unobserved factors F_t $(t = 1, \ldots, T)$. If there are any missing data points in data matrix $Y = (Y_1, \ldots, Y_T)$ or the matrix of observed time series $W = (W_1, \ldots, W_T)$, priors for these missings are also implied by the model.[6] For details on these missing data points, I refer to Appendix 3.A.

[6]Note that the proxy variables m_t are incorporated into the state equation in order to identify structural shocks and not to be modeled using an autoregressive structure. Hence, imputing potential missing values in the proxies would just mean to sample noise and would not add any value. Therefore, it is easier to set missing values to zero.

Conditional on the factors and the observed time series, observation equation (3.2.7) is a normal linear regression model. Since the covariance matrix R is assumed to be diagonal, I formulate a conjugate normal-inverse-gamma prior for the parameters:

$$r_i \sim i\mathcal{G}\left(\underline{a}, \underline{b}\right), \qquad i = 1, \ldots, N, \tag{3.2.14}$$

$$\tilde{\lambda}_i | r_i \sim \mathcal{N}\left(\underline{\mu}_{\tilde{\lambda}_i}, r_i \cdot \underline{V}_{\tilde{\lambda}_i}\right), \qquad i = n_1 + 1, \ldots, N, \tag{3.2.15}$$

where $i\mathcal{G}(\cdot, \cdot)$ is the inverse-gamma distribution, $\underline{a}, \underline{b} > 0$ are scalars, the $1 \times n$ vector $\tilde{\lambda}_i'$ is the $(i - n_1)$-th row of $\tilde{\Lambda} = (\tilde{\Lambda}_F, \tilde{\Lambda}_W)$, $\underline{\mu}_{\tilde{\lambda}_i}$ is a $n \times 1$ vector and $\underline{V}_{\tilde{\lambda}_i}$ is a symmetric and positive semi-definite $n \times n$ matrix. As will become evident later, this choice of prior is convenient as it leads to a straightforward updating step in the MCMC algorithm.

The prior of the unobserved factors F_t is implicitly given by the underlying dynamics of the FAVAR model. First, the latent vectors for $t = 1, \ldots, p$ are treated as presample values and given a simple normal prior:

$$F_t \sim \mathcal{N}\left(0, \Sigma_0\right),$$

with Σ_0 being a $n_1 \times n_1$ symmetric and positive semi-definite matrix. Second, the prior for the remaining factors is provided by state equation (3.2.5) that can be rewritten by:

$$\tilde{z}_t = B_1 \tilde{z}_{t-1} + \ldots + B_p \tilde{z}_{t-p} + u_t, \qquad t = p + 1, \ldots, T,$$

with $B_j = \left(\tilde{A}_j \tilde{A}_0^{-1}\right)'$ and $u_t \sim \mathcal{N}\left(0, \Sigma_u\right)$. Similar to Chan and Jeliazkov (2009), I define $(n + k)T \times 1$ vectors $\bar{z} = \left(\tilde{z}_1', \ldots, \tilde{z}_T'\right)'$ and $\bar{u} = \left(u_1', \ldots, u_T'\right)'$ to further manipulate the model to matrix notation:

$$H_{\bar{z}} \bar{z} = \bar{u}, \qquad \bar{u} \sim \mathcal{N}\left(0, \Sigma_{\bar{z}}\right),$$

whereas $(n+k)T \times (n+k)T$ matrices $H_{\bar{z}}$ and $\Sigma_{\bar{z}}$ are defined by:

$$H_{\bar{z}} = \begin{pmatrix} I_{n+k} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & I_{n+k} & \ddots & & & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & & & \vdots \\ 0 & \cdots & 0 & I_{n+k} & \ddots & & & \vdots \\ -B_p & \cdots & \cdots & -B_1 & I_{n+k} & \ddots & & \vdots \\ 0 & -B_p & \cdots & \cdots & -B_1 & I_{n+k} & \ddots & \vdots \\ \vdots & \ddots & \ddots & & & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -B_p & \cdots & \cdots & -B_1 & I_{n+k} \end{pmatrix}, \quad \Sigma_{\bar{z}} = \begin{pmatrix} I_p \otimes \Sigma_0 & 0 \\ 0 & I_{T-p} \otimes \Sigma_u \end{pmatrix}.$$

However, the goal of this manipulation is to derive the distribution of the latent factors F_t conditional on the observed time series W_t and the proxies m_t. Therefore, an alternative reordering of vector $\bar{z}$ is required. Define $n_1 \times T$ matrix $F = (F_1, \ldots, F_T)$ and $k \times T$ matrix $M = (m_1, \ldots, m_T)$ to write $(n+k)T \times 1$ vector $\hat{z}' = [\text{vec}(F')', \text{vec}(W')', \text{vec}(M')'] = (f', w', m')$. Then, similar to $H_{\bar{z}}$ and $\Sigma_{\bar{z}}$ before, I can define $H_{\hat{z}}$ and $\Sigma_{\hat{z}}$ such that the state equation has the representation:

$$H_{\hat{z}}\hat{z} = u, \qquad u \sim \mathcal{N}(0, \Sigma_{\hat{z}}),$$

with $u = \text{vec}(U')$ and $U = (u_1, \ldots, u_T)$. Although $H_{\hat{z}}$ and $\Sigma_{\hat{z}}$ cannot be exposed as simple as $H_{\bar{z}}$ and $\Sigma_{\bar{z}}$ above, they are also sparse band matrices and can easily be implemented in the same fashion. Consequently, vector $\hat{z}$ has the following unconditional distribution:

$$\hat{z} = \begin{pmatrix} f \\ w \\ m \end{pmatrix} \sim \mathcal{N}\left(0, \left[H_{\hat{z}}'\Sigma_{\hat{z}}^{-1}H_{\hat{z}}\right]^{-1}\right), \qquad H_{\hat{z}}'\Sigma_{\hat{z}}^{-1}H_{\hat{z}} = \begin{pmatrix} Q_f & Q_{f,w,m} \\ Q_{f,w,m}' & Q_{w,m} \end{pmatrix},$$

with $n_1T \times n_1T$ matrix Q_f, $n_1T \times (n_2+k)T$ matrix $Q_{f,w,m}$ and $(n_2+k)T \times (n_2+k)T$ matrix $Q_{w,m}$ being the corresponding blocks of the inverse covariance matrix that is determined by the previously defined matrices $H_{\hat{z}}$ and $\Sigma_{\hat{z}}$. Since w and m are observed, the prior distribution of f conditional on it can finally be derived using a standard result for multivariate Gaussian distributions:

$$f|w, m, \tilde{A}_0, \tilde{A}_+, \Sigma_0 \sim \mathcal{N}\left(-Q_f^{-1}Q_{f,w,m}(w', m')', Q_f^{-1}\right). \tag{3.2.16}$$

3.2.3 Priors SVAR parameters

With respect to the priors for the SVAR parameters $\tilde{A}_0$ and $\tilde{A}_+$, I follow Arias et al. (2019) using a normal-generalized-normal distribution, which is a conjugate prior for the simultaneous equation representation. The density function is given by:

$$\mathcal{NGN}_{(\nu,\Phi,\Psi,\Omega)}\left(\tilde{A}_0,\tilde{A}_+\right) \propto \left|\det(\tilde{A}_0)\right|^{\nu-n} \exp\left(-\frac{1}{2}\text{vec}\left(\tilde{A}_0\right)'\Phi\text{vec}\left(\tilde{A}_0\right)\right) \exp\left(-\frac{1}{2}\left[\text{vec}\left(\tilde{A}_+\right)-\Psi\text{vec}\left(\tilde{A}_0\right)\right]'\Omega^{-1}\left[\text{vec}\left(\tilde{A}_+\right)-\Psi\text{vec}\left(\tilde{A}_0\right)\right]\right),$$

where $\nu \geq n+k$ is a scalar, Φ is a $(n+k)^2 \times (n+k)^2$ symmetric and positive definite matrix, Ψ is a $p(n+k)^2 \times (n+k)^2$ matrix and Ω is a $p(n+k)^2 \times p(n+k)^2$ symmetric and positive definite matrix. This distribution is constrained such that the structural parameters satisfy the restrictions discussed in (3.2.3) and (3.2.13). Unlike in Caldara and Herbst (2019), this prior is conjugate and furthermore, ensures that observationally equivalent models get the same prior and hence, posterior weight. Particularly, for any orthogonal $(n+k)\times(n+k)$ matrix Q that rotates the structural parameters subject to the Proxy-SVAR restrictions, the prior obtains the same density value for $(\tilde{A}_0,\tilde{A}_+)$ and $(\tilde{A}_0Q,\tilde{A}_+Q)$. This feature is particularly appealing if multiple shocks are to be identified by IV, which requires additional restrictions. If these restrictions yield a set-identified model, the Bayesian approach adopted in this paper ensures that the prior is not informative about the set of structural parameters besides the identifying restrictions imposed by the econometrician.

As suggested by Arias et al. (2019) and Caldara and Herbst (2019), I also introduce a relevance prior that sets a lower bound γ for the percentage of the instrument's variance that is related to the corresponding structural shock. Therefore, I restrict the minimum eigenvalue of the so-called reliability matrix $\left(\left[\Gamma_{0,2}^{-1}\right]'\Gamma_{0,2}^{-1}+VV'\right)^{-1}VV'$ to be greater equal to the specified γ within each iteration.

3.2.4 Posterior inference

In the following, I outline how to conduct posterior inference. Define the vector of all parameters and latent variables as $\theta' = \left(f',\text{vec}(\tilde{A}_0,\tilde{A}_+)',r',\text{vec}(\tilde{\Lambda})'\right)$, and let θ_{-S} be the parameter vector without quantity S. Note again that if there are missing values in data matrix Y or observed time series W, they can be treated as unknown parameters and incorporated in θ. In order to simplify notation, details about that are moved to Appendix 3.A. The posterior distribution $p(\theta|Y,W)$ is proportional to the product of likelihood and prior, i.e. $p\left(\theta|Y,W\right) \propto p\left(Y,W|\theta\right)p(\theta)$. However, since this posterior is of unknown form, I design a MCMC algorithm that cycles through three major blocks of conditional distributions:

1. Draw latent factors f using Gaussian Markov random fields methods (Chan, 2017; Rue et al., 2009).

2. Draw the structural parameters $(\tilde{A}_0, \tilde{A}_+)$ following Arias et al. (2019).

3. Draw $\tilde{\Lambda}$, containing the factor loadings $\tilde{\Lambda}_F$ and the loadings for the observed time series $\tilde{\Lambda}_W$, and covariance matrix R.

Note that all choices for prior parameters, the initializations required and the detailed algorithm are provided in Appendix 3.A. This section only sketches the resulting posterior distributions of the unknown parameters.

First, sample the latent factors summarized in f. Since for each of the n_1 unobserved factors the full dynamic process for $t = 1, \dots, T$ has to be generated, they could be treated as time-varying parameters (Belviso and Milani, 2006; Bruns, 2018). Consequently, Kalman Filter techniques as the algorithm of Carter and Kohn (1994) are employed to draw F_t for $t = 1, \dots, T$. As an alternative, I make use of the fact that equation (3.2.6) in matrix notation can easily be manipulated such that it reads as a normal linear regression model. Conditional on θ_{-f} everything except the regression coefficients f in this normal linear regression model is known. Hence, combined with normal prior (3.2.16) that crucially depends on current draws of the structural parameters $(\tilde{A}_0, \tilde{A}_+)$, a classical normal posterior for f conditional on θ_{-f}, Y and W can be derived. According to the dimension of f, its posterior covariance matrix is of size $n_1 T \times n_1 T$.[7] Sampling such highly dimensional vectors usually requires a lot of computational time. However, the posterior covariance matrix is a sparse band matrix because the prior covariance Q_f^{-1} is. This allows to apply methods from the Gaussian Markov random fields literature (Chan, 2017; Rue et al., 2009). These techniques decrease the computational burden considerably. Therefore, this procedure to draw the latent factors f is a huge gain in efficiency compared to commonly applied filtering approaches.

The most difficult block is to sample the structural parameters $(\tilde{A}_0, \tilde{A}_+)$, as these parameters might be only set-identified. Conditional on f, equation (3.2.4) represents a Proxy-SVAR. Since the prior is conjugate, the posterior of the structural parameters is also a normal-generalized-normal distribution subject to constraints (3.2.3) and (3.2.13) potentially amended by sign restrictions to further disentangle multiple shocks. Arias et al. (2019) derive an importance sampler to draw from the restricted normal-generalized-normal posterior of the structural parameters. In their method, they make use of an alternative parameterization of equation (3.2.4) which can be easier to draw from. In a second step, draws from the alternative parameterization are mapped back to the parameters of interest and importance

[7] Remember that the number of latent factors n_1 typically varies between four and nine and a common time series length T in macroeconomics is around 400 data points. Consequently, the dimension of f can easily be above 2000.

weights are calculated using the change of variables theorem. I embed the importance sampler of Arias et al. (2019) into the MCMC algorithm by an accept-reject Metropolis-Hastings (ARMH) step (Chib and Greenberg, 1995; Tierney, 1994). This is necessary due to the fact that the proposal distribution of *Algorithm 2* of Arias et al. (2019) does not dominate the restricted normal-generalized-normal posterior of the structural parameters. Hence, I correct for that in a Metropolis-Hastings step. An alternative way to embed an importance sampling step in a Gibbs sampler is to use sampling-importance-resampling (Koch, 2007) as suggested by Antolín-Díaz et al. (2018). The machinery of Arias et al. (2019) gives all tools to do the ARMH step: a proposal distribution to draw independently from and a change of variables theorem that allows to compute the density of the proposal. For details on this, see step 3 in Appendix 3.A.2.

The remaining parameters of the measurement equation (3.2.6), factor loadings Λ_F, the loadings of the observed time series Λ_W and the error term variances r_i ($i = 1, \dots, N$), can be drawn relying on basic results from the Bayesian regression models. Conditional on latent factors and observed time series, (3.2.6) can be written equation by equation as univariate normal linear regression models with coefficients being the i-th row of $\Lambda = (\Lambda_F, \Lambda_W)$ and innovation variances r_i ($i = 1, \dots, N$). Since prior (3.2.14) and (3.2.15) is conjugate, this yields a (conditional) normal-inverse-gamma distribution. Thus, first variances r_i are drawn from its inverse-gamma posterior with moments conditional on the most recent draw of Λ. Then, conditional on r_i, the rows of $\tilde{\Lambda} = (\tilde{\Lambda}_F, \tilde{\Lambda}_W)$ are sampled from their normal posterior.

Based on the posterior draws of θ, visualization of the implied results with impulse response functions or forecast error variance decompositions in textbook form is straightforward. Furthermore, I investigate the contribution of the identified structural shocks to the dynamics of specific time series via historical decompositions. Although Gaussian Markov random fields techniques are applied to gain efficiency in sampling the latent factors, due to the time-consuming ARMH step for the structural parameters the computational burden is still quite high (~1,000 posterior draws per hour with $n_1 = 5$, $n_2 = 4$, $k = 2$, $N = 132$ and $p = 24$). Convergence diagnostics are conducted using the concept of relative numerical efficiency proposed by Geweke (1992).

3.3 The macroeconomic effects of oil supply shocks

In the remainder of this paper, I will use the Proxy-FAVAR model to compare the effects of two distinct oil supply shocks on the US macroeconomy: supply flow shocks and supply news shocks. Both shocks are of particular interest to policy makers and a sound understanding of the exact magnitudes as well as their relative importance is vital to design effective policy responses and to construct meaningful counterfactual scenario analysis.

To the best of my knowledge, such a direct comparison is novel to the literature. Instead, there exists a large amount of literature that has gathered evidence for the effects of these shocks in separately. With respect to a classical supply flow shock that causes an immediate drop in oil production, prominent papers are Kilian and Murphy (2012, 2014), Antolín-Díaz and Rubio-Ramírez (2018), Baumeister and Hamilton (2019), Caldara et al. (2019) and Herrera et al. (2019). It is commonly accepted that flow shocks are associated with negative effects on economic activity, while prices tend to rise. However, there is remarkable disagreement concerning the relative importance and the size of the effect of an oil flow shock. In turn, the work on oil news shocks is more recent and pioneered by Känzig (2019). He finds that an oil news shock implies a gradual fall in oil production and is a very important driver for the price of oil. All of the aforementioned papers study the corresponding shocks within small scale SVAR models. Here, instead, I fully take advantage of a large panel of US data in order to identify and estimate their effects using a rich set of information. Thereby, the paper follows the idea used in Juvenal and Petrella (2015) and Stock and Watson (2016), however, with the novelty of identifying multiple shocks by instrumental variables.

To identify both shocks simultaneously, I use a combination of instrumental variable restrictions and sign restrictions. The (flow) supply shock series of Kilian (2008a) and the news shock of Känzig (2019) are used to disentangle two shocks in the structural FAVAR that are correlated with these instruments, and uncorrelated with all other shocks. To further distinguish between flow and news shocks, I impose a sign restriction on the effects on inventories. In response to a negative supply flow shock, oil inventories are assumed to decrease instantly in order to smooth out some of the impact. Instead, in response to a (negative) news shock, oil inventories are build up in order to prepare for the future decrease in production.[8]

The empirical analysis is structured as follows. I first introduce the data set, modeling choices within the FAVAR as well as the identification strategy. Then, I present the main results including impulse response functions, forecast error variance decompositions and historical decompositions.

3.3.1 Model specification and identification strategy

The FAVAR model is specified at the monthly frequency and is estimated based on oil market and macroeconomic variables. The core of the model is a block of observed factors that closely resembles the design of other oil market models specified in the literature (Baumeister and Hamilton, 2019; Känzig, 2019; Kilian and Murphy, 2014). Particularly, I include as observed factors W_t the log of global oil production, the log of world industrial production (Baumeister and Hamilton, 2019), the log of the real price of oil and the log of crude oil

[8]Negative shocks are normalized to decrease production and hence, increase oil prices.

inventories.[9] Note that as in Känzig (2019), I include these variables in levels, as to allow for possible cointegration amongst the oil market block.

I augment this model with a set of unobserved factors that broadly capture the macroeconomic conditions in the United States. These factors are extracted from a panel of 132 time series borrowed from Kerssenfischer (2019). The panel closely follows the FRED-MD database with some minor differences in variables and transformations. All of these variables are first transformed to stationarity and then standardized to have zero mean and unit variance, a standard procedure in factor analysis. A detailed overview of all economic indicators including the transformations used is provided in Table 3.B.1 in Appendix 3.B.1. Due to availability of the oil market variables and the data set of Kerssenfischer (2019), the sample covers 1974M1 to 2016M9.

In order to identify the two oil supply shocks, I augment the set of variables with two external instruments. As outlined in Section 3.2, they are used to augment the SVAR in latent factors and observed time series by proxies (see equation (3.2.2)). The first instrument is designed to capture classical exogenous disruptions in oil supply (flow shocks). Here, I use the series constructed in Kilian (2008a) who computes country-specific counterfactual oil productions by extrapolating the production level of a certain country, following certain exogenous disruptive events.[10] The extrapolation is based on the average growth rate of oil production in countries with similar macroeconomic conditions as the country of interest that are not directly affected by the extraordinary political event. The difference of this counterfactual and the actual production series at each point in time is the country-specific exogenous shortfall series. The OPEC series is obtained by summing the individual series for the member countries and is usually expressed as a share of world oil production. Finally, the change over time in this percentage acts as the proxy for exogenous oil supply flow shocks. Note that Montiel-Olea et al. (2016) find that this instrument may be weak, which would require using non-standard inference in a classical setting. However, within the Bayesian approach adopted in this paper this issue is less severe, and I impose instrument relevance through the prior distribution. In practice, this is implemented by imposing a lower bound on the eigenvalues of the reliability matrix as discussed Arias et al. (2019). Note that the results presented in the following are robust with respect to the choice of this lower bound. Given that the monthly series of Kilian (2008a) is no longer available on his website, I use an extended version constructed in Braun and Brüggemann (2019).

[9]The oil price is based on the Refiners Acquisition Costs (RAC). RAC, oil production and oil inventories raw data is available at the Energy Information Administration. A proxy of global crude oil inventories is constructed as in Kilian and Murphy (2014).

[10]Such events are the Arab oil embargo 1973/74, the Iranian Revolution 1978/79, the Iran-Iraq War 1980-88, the Gulf War 1990/91, the Venezuela Civil Unrest 2002 and the Iraq War 2003.

The second instrument is designed to capture news shocks about oil supply and was recently proposed by Känzig (2019). The idea is to exploit high-frequency movements in oil price futures around announcements that affect market participants expectations on future oil supply. The construction of the series is based on two inputs: the announcements about future production quotas of the OPEC and data on oil futures. The OPEC unites 14 member countries that account for about 44% of global oil production. Member countries are currently Algeria, Angola, Congo, Ecuador, Equatorial Guinea, Gabon, Iran, Iraq, Libya, Nigeria, Saudi Arabia, United Arab Emirates and Venezuela. The oil ministers and their delegations of these member countries meet on a regular basis, twice a year at the OPEC conference. In addition, there can be extraordinary meetings if necessary. Following these conferences, the OPEC publishes announcements about future production ceilings that on average become effective after one month. This practice came into force in 1982 while before the OPEC focused on controlling the price of oil directly. Due to the big power of the OPEC as the most important player on the oil market, these announcements crucially impact agents' oil price expectations. Since there are very liquid markets for oil futures, these expectations can be measured by the price of e.g. West Texas Intermediate (WTI) crude oil futures. Hence, Känzig (2019) constructs a surprise component of future oil supply by measuring the shift in these oil price expectations caused by the OPEC announcement. For that purpose, he takes the settlement price of the oil future on the day of the press release minus the one on the preceding trading day based on futures with a maturity of six months. This daily series is then aggregated to monthly frequency. This instrument is suitable to proxy an oil supply news shock because the announced production quotas only become effective with a lag. Consequently, in the month of the OPEC conference, oil supply should not be affected. Instead, these releases are news about future oil supply. Since the OPEC announcements may also be driven by the current state of the economy, one could argue against the exogeneity of this instrument. However, the changes in future prices are measured in tight windows around the OPEC announcements to isolate the news component. Känzig (2019) finds his proxy to be not autocorrelated, not forecastable by macroeconomic variables and uncorrelated with other shock measures such that exogeneity of the instrument seems to be a reasonable assumption. Moreover, its F-statistic of 24.2 is clearly above the threshold value proposed by Montiel-Olea et al. (2016). Hence, also the relevance condition does not seem to be violated.

With two instruments and two shocks, one additional restriction is necessary to disentangle the classical supply from the news shock. Here, I exploit common sense sign restrictions that allow me to disentangle these shocks. Particularly, I follow the argument in Känzig (2019) that a distinguished feature of oil supply news shocks is that they lead to an increase of inventories on impact, reflecting agents' precautionary build-up of their stocks in oil. In turn,

a contemporaneous flow shock is associated with a sudden decrease of inventories, as they are used to mitigate the impact of the sudden disruption in supply. The sign restriction might lead to a set-identified model. However, the Bayesian approach adopted in this framework allows for a coherent treatment of this feature, in that it assigns equal prior (and hence posterior) weight to observationally equivalent structural parameters (Arias et al., 2019). In a robustness check, Känzig (2019) also uses the same instruments to compare the news shock with a flow shock in a conventional SVAR. Here, he relies on the assumption that news shocks do not affect production within one month. However, this might not be reasonable given that a good share of global oil producers are not obliged to follow any production targets and hence, can instantly adjust their production as a reaction to increasing oil prices.

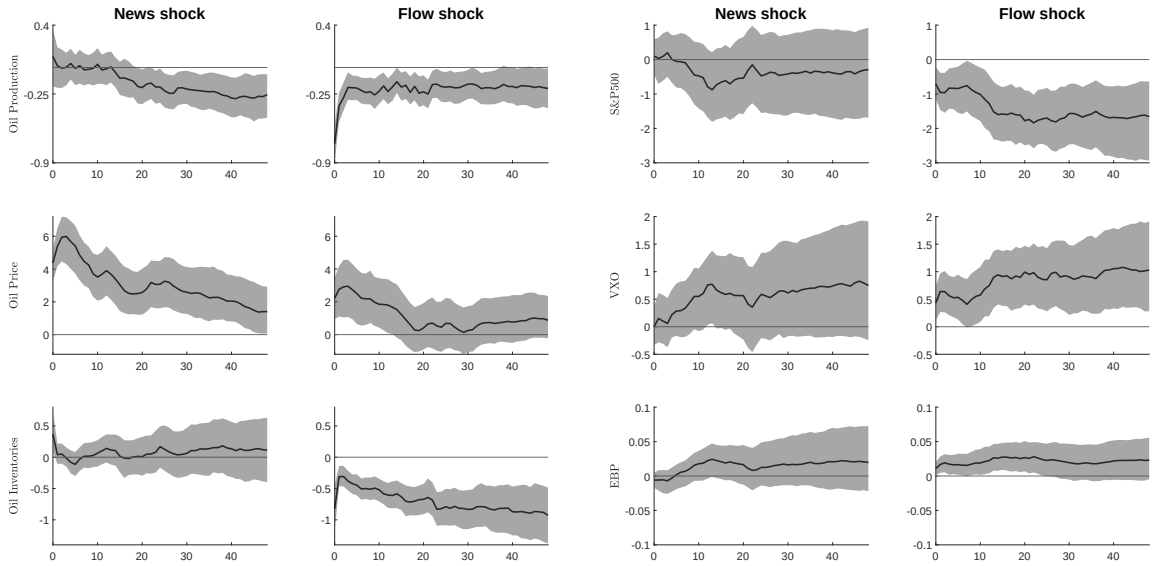
Two more model choices are required for the FAVAR: selecting the number of lags and the number of latent factors that summarize the dynamics of the US macroeconomy. With respect to the lag length, I ensure to capture all autoregressive structure by selecting $p = 24$. Hereby I implicitly intend to follow the convention put forward in the oil market models of Kilian (2009), Kilian and Murphy (2012, 2014) and Herrera et al. (2019). With respect to the number of latent factors, I choose five supported by the information criterion of Onatski (2010).[11] Due to the structure of model (3.2.4) and (3.2.6), the first five variables in data matrix Y are the so-called naming variables (Stock and Watson, 2016). It is desirable that the naming variables are not too correlated and should be representative for the other series in the data matrix. Therefore, I select them to be main indicators from five different variable categories: real personal consumption expenditures, industrial production, the unemployment rate, the S&P 500 index and the consumer price index in their stationary transformations. The largest absolute correlation between these variables is 0.15 such that they clearly satisfy the requirement to be sufficiently different from each other. The named factor normalization makes the interpretation of the factors straightforward: I consider a consumption factor, an activity factor, a labor market factor, a financial factor and a price factor.

3.3.2 Dynamic effects of oil supply shocks

I present responses of a selection of oil market variables and key US macroeconomic indicators to oil flow and news shocks. For the variables transformed in logs (see Table 3.B.1 in Appendix 3.B.1), responses can be interpreted in percentages. These variables are the oil variables, S&P500, the activity variables, consumer prices, income, the foreign exchange rate and M2 money. For the remaining indicators, the responses are simply effects

[11]The most common way in the literature to determine the number of factors is to use the criterion of Bai and Ng (2002). However, it often heavily overestimates the number of factors. For the data set of interest in this paper, it always selects the maximum number of unobserved factors. This is not applicable since the number of unknown parameters would be too large.

Figure 3.1: FAVAR impulse responses: oil and financial variables



(a) Oil variables (b) Financial variables

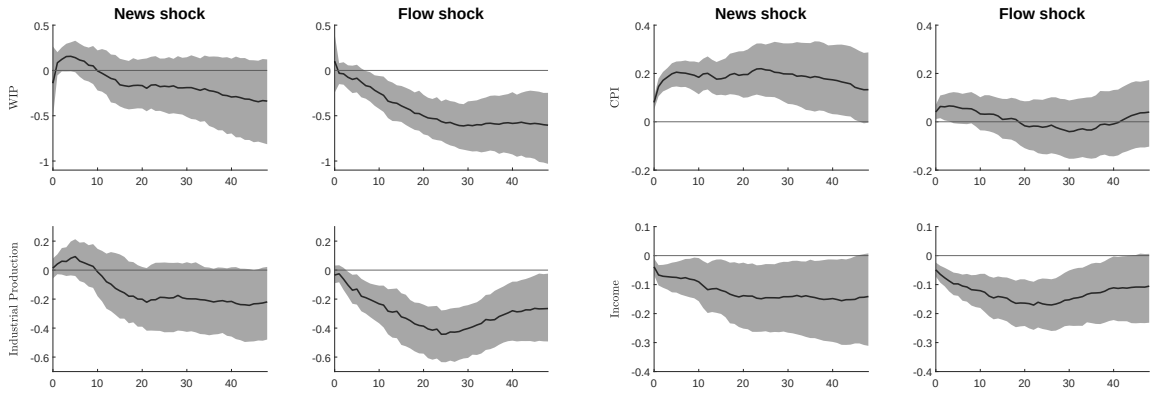
Responses of oil variables in (a) and financial variables in (b) to an oil supply news shock (1st and 3rd column) and an oil supply flow shock (2nd and 4th column). Shaded areas are 68% posterior confidence bands.

on levels. These variables are the VXO volatility index, the excess bond premium (EBP), the unemployment rate and the federal funds rate.

Figure 3.1(a) shows the responses of the oil variables to a shortfall in oil supply. With respect to the oil market, both shocks impose an immediate jump in the real price of oil. Here, a news shock causes a larger rise of the price and is much more persistent than the unanticipated flow shock. In line with Känzig (2019), a news shock is followed by a gradual decrease in oil production that becomes significant only after 1.5 years, and a short-term increase in inventories. The sluggish fall in production may be observed because non-OPEC countries increase their production for a few months following to the ceiling due to the increased price. Consequently, positive and negative effects on production could cancel in the short run before the setting of quotas for OPEC members dominates significantly. Regarding a flow supply shock, I observe an instant downturn of production that remains significant for almost all horizons, and a very persistent decrease in oil inventories. These results are in line with estimates observed in Kilian and Murphy (2014) and Caldara et al. (2019).

The IRFs for the financial variables are given in Figure 3.1(b) and I find remarkable differences between flow and news shocks. In general, news shocks seem to have less pronounced effects than traditional (flow) supply shocks. With respect to the S&P500, no

Figure 3.2: FAVAR impulse responses: activity and price/income variables



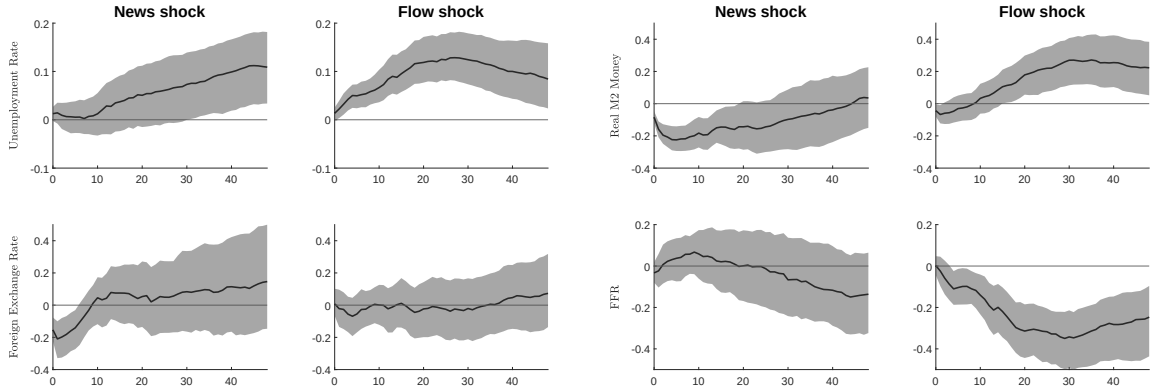
(a) Activity variables (b) Income and Prices

Responses of world industrial production and US industrial production in (a) and consumer prices and real personal income in (b) to a news shock in the 1st and 3rd column and a flow shock in the 2nd and 4th column. Shaded areas are 68% posterior confidence bands.

significant response is found for the news shock, while the flow shock induces a severe downturn. This is in line with Baumeister and Hamilton (2019) who also find that a flow shock harms the fundamentals steeper than other oil shocks. The insignificant response to a news shock in stock markets is less strong than in Känzig (2019), who finds that stock markets decline significantly but only after about one year. In line with the effects on stock markets, a flow shock also substantially raises financial uncertainty measured by the VXO index. Moreover, this increased uncertainty on the financial markets imposes that investors claim a higher risk premium. Therefore, credit market conditions indicated by the excess bond premium of Gilchrist and Zakrajšek (2012) do also worsen significantly at least for several months. The reactions of VXO and EBP to a news shock have the same sign as for a flow shock, but are insignificant for almost all periods. Since the shock under consideration does not increase uncertainty significantly, it can be interpreted as a news shock rather than an uncertainty shock about future oil supply. However, VXO and EBP may not be the perfect indicators to measure uncertainty on the oil market. Nevertheless, this is worth mentioning, because Alquist and Kilian (2010) show that in general news and uncertainty shocks on the oil market can have very similar effects.

Figure 3.2(a) provides IRFs for variables capturing economic activity. A supply (flow) shock strongly impacts economic activity in a negative and persistent way, while a news shock has less pronounced effects with considerably higher uncertainty in the estimates. This finding does hold for both, global and US industrial production, and significantly differs from those found in the robustness check of Känzig (2019) studying similar IRFs in a SVAR

Figure 3.3: FAVAR impulse responses: unemployment, exchange rate and Fed variables



(a) Unemployment and Foreign Exchange Rate (b) Federal Reserve variables

Responses of the unemployment rate and trade weighted foreign exchange rate against major currencies in (a) and real M2 money and the federal funds rate in (b) to a news shock in the 1st and 3rd column and a flow shock in the 2nd and 4th column. Shaded areas are 68% posterior confidence bands.

environment. My differences result from a larger information set what is discussed in more detail at the end of this subsection. Finally, the results suggest that the effects of a news shock might in fact closely resemble a so-called inventory demand shock studied e.g. in Kilian and Murphy (2014), Juvenal and Petrella (2015) or Baumeister and Hamilton (2019), who find similar results on economic activity than I find for the news shock.

Turning to prices in Figure 3.2(b), I observe that consumer prices increase strongly for almost four years following a news shock. This is in line with the relatively strong response I find in oil prices, which passes through to consumer prices almost immediately. As in Lippi and Nobili (2012), CPI's response to a traditional exogenous supply shock is barely significant. The adverse reaction of real personal income is very similar for a news shock and a flow shock in terms of magnitude and significance: it falls for some years before it returns to its original level slowly.

The last set of IRFs studies the effect of the supply shocks on unemployment, exchange rates, the monetary base and interest rates. Consistent with the persistent slowdown in economic activity and stock markets, the unemployment rate in Figure 3.3(a) exhibits a long-lasting and significant rise due to an unexpected change in oil supply. The fact that the US as net oil importer might be hit harder than other economies by the oil price increase caused by a news shock, justifies that the dollar depreciates for about six months. Since a flow shock seems to affect the US and the global economy by similar strength, the exchange rates remain unchanged.

Bernanke et al. (1997) argue that the Federal Reserve increases their interest rate due to the inflationary pressure caused by oil price shocks. However, Figure 3.3(b) shows that the federal funds rate lowers in response to an exogenous shock in oil supply. This confirms the result of Kilian and Lewis (2011) who justify this by the Fed's interpretation of an increased oil price stemming from a flow shock as an adverse aggregate demand shock. This decrease of the policy rate is enforced by increasing the money supply. Similar to Känzig (2019), the federal funds rate does not seem to react significantly to a news shock. Since this news shock causes a rise in consumer prices, real money naturally falls for two years. Yet the Fed reacts to that to balance these inflationary effects on real money supply.

The results presented are robust towards replacing the sign restriction on inventories with a zero constraint on the contemporaneous reaction of oil production to a news shock, as exploited in Känzig (2019). Corresponding impulse response functions closely resemble those in Figures 3.1, 3.2 and 3.3 (see Figure 3.C.1 in Appendix 3.C.1).

To illustrate the benefits from adopting a FAVAR model to jointly study the effects of oil supply shocks, Appendix 3.D provides a comparison with estimates obtained from conventional approaches. One popular strategy is to use small scale SVAR models augmenting the core oil market variables by one macroeconomic indicator at a time. Here, one can either identify a flow shock and a news shock separately using the corresponding instruments, or identify them jointly employing the sign restriction on oil inventories used in this section. Figures 3.D.1 and 3.D.2 provide corresponding impulse responses obtained by both approaches. The differences in the estimates are remarkable. All small scale models yield puzzling responses of the flow shock on macroeconomic variables, casting serious doubt on the credibility of the results. In particular, disruptive oil flow shocks are estimated to decrease consumer prices and increase economic activity, thereby contradicting any conventional economic theory. The fact that flow shocks are estimated poorly also yields much stronger and significant effects of a news shock on macroeconomic variables, compared to the FAVAR model adopted in this paper. These findings clearly motivate the need for a larger information set to study the effects of oil market shocks on the macroeconomy.

In Figure 3.D.3, I further illustrate the benefits from jointly identifying both shocks in one model. Therefore, I plot the impulse response functions obtained by two separate FAVAR models using either the proxy variable for the flow shock or the instrument for the news shock. If the two instruments were truly unrelated to each other, there would be little need to further disentangle the effects using the methodology proposed in this paper. First, note that by specifying a FAVAR model one does no longer obtain puzzling effects of the flow supply shock, and the estimates are more in line with the findings of this paper. However, there are some differences regarding the effects of news shocks on the macroeconomy. If oil supply shocks are not identified jointly, the effects of a news shock on key macroeconomic variables

Table 3.1: Forecast error variance decompositions

	News shock			Flow shock		
	0	24	48	0	24	48
Oil production	0.03 [0.00,0.11]	0.13 [0.01,0.34]	0.12 [0.02,0.33]	0.35 [0.20,0.55]	0.12 [0.01,0.32]	0.08 [0.01,0.23]
Oil price	0.46 [0.25,0.65]	0.29 [0.12,0.54]	0.14 [0.02,0.39]	0.12 [0.02,0.28]	0.04 [0.00,0.16]	0.08 [0.01,0.26]
Oil inventories	0.02 [0.00,0.09]	0.03 [0.00,0.12]	0.03 [0.00,0.12]	0.12 [0.04,0.25]	0.31 [0.15,0.50]	0.21 [0.06,0.41]
S&P500	0.02 [0.00,0.10]	0.05 [0.00,0.20]	0.05 [0.00,0.19]	0.08 [0.01,0.21]	0.20 [0.05,0.40]	0.16 [0.03,0.38]
VXO	0.02 [0.00,0.09]	0.10 [0.01,0.32]	0.10 [0.01,0.33]	0.07 [0.01,0.19]	0.14 [0.03,0.35]	0.13 [0.02,0.34]
EBP	0.04 [0.00,0.16]	0.10 [0.01,0.36]	0.14 [0.02,0.40]	0.07 [0.01,0.21]	0.13 [0.02,0.36]	0.09 [0.01,0.29]
WIP	0.03 [0.00,0.11]	0.05 [0.01,0.18]	0.05 [0.01,0.19]	0.02 [0.00,0.06]	0.27 [0.11,0.50]	0.12 [0.02,0.32]
IP	0.02 [0.00,0.10]	0.06 [0.01,0.23]	0.09 [0.01,0.28]	0.03 [0.00,0.10]	0.26 [0.11,0.45]	0.10 [0.01,0.28]
CPI	0.29 [0.11,0.47]	0.20 [0.06,0.43]	0.08 [0.01,0.26]	0.08 [0.01,0.22]	0.02 [0.00,0.11]	0.03 [0.00,0.14]
RPI	0.11 [0.01,0.31]	0.15 [0.02,0.39]	0.12 [0.01,0.33]	0.18 [0.04,0.34]	0.19 [0.05,0.39]	0.07 [0.01,0.24]
Unemployment	0.05 [0.00,0.18]	0.08 [0.01,0.27]	0.19 [0.03,0.42]	0.05 [0.01,0.18]	0.28 [0.12,0.47]	0.11 [0.01,0.30]
FX	0.17 [0.05,0.36]	0.09 [0.01,0.32]	0.15 [0.02,0.43]	0.02 [0.00,0.10]	0.07 [0.01,0.22]	0.06 [0.01,0.22]
M2	0.20 [0.06,0.39]	0.07 [0.01,0.26]	0.05 [0.00,0.19]	0.06 [0.01,0.18]	0.13 [0.02,0.31]	0.13 [0.02,0.35]
FFR	0.04 [0.00,0.15]	0.04 [0.00,0.16]	0.09 [0.01,0.29]	0.02 [0.00,0.09]	0.30 [0.11,0.48]	0.20 [0.04,0.44]

Entries show fraction of the forecast error variance in variables in the first column that is explained by a news shock and a flow shock, respectively, at horizon $h = 0, 24, 48$. Posterior 68% credible intervals are displayed in parantheses.

are more pronounced and significant. For example, the magnitudes on industrial production and income are almost twice as large and significant than responses in the model that further orthogonalizes these shocks. This points towards a slight risk of confounding the effects of supply news and flow shocks, which is avoided in the strategy adopted in this paper.

3.3.3 Relative importance of flow and news shocks

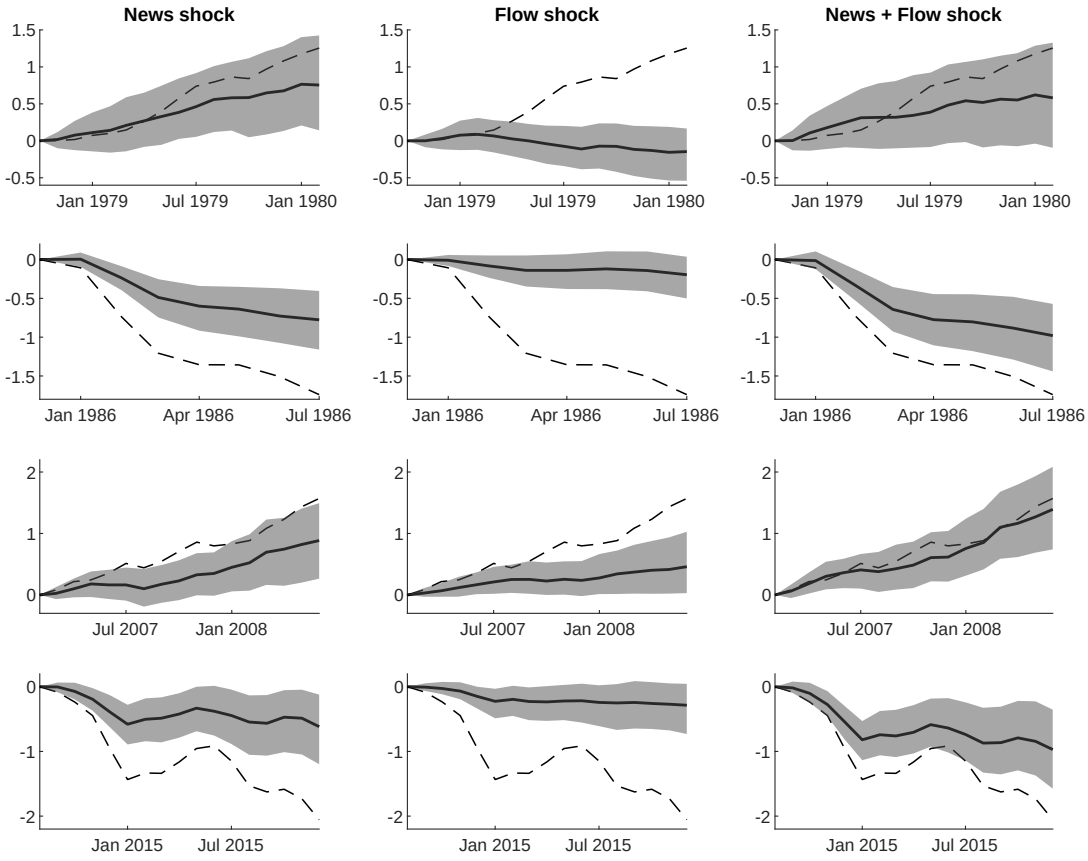
In this section I evaluate the relative importance of the two oil shocks for driving variations in the oil market and the US business cycle. For both shocks, Table 3.1 provides corresponding posterior credibility regions for forecast error variance decompositions (FEVD) at various horizons.

First of all, news shocks seem to be major drivers of the oil price, especially in the short run where it accounts for about half of the variability. Note that these estimates are slightly lower than values obtained by Känzig (2019) who focuses on the news shock only. As in his paper, these large shares in the FEVD of the oil price directly translate to the results obtained in consumer prices, where the news shock is estimated to account for up to one third of the price variation. For both oil prices and consumer prices the relative importance of a news shock peaks on impact and decreases gradually in the following. Additionally, it accounts for a substantial amount in the impact variation of exchange rates. Interestingly, the news shock drives very little of oil inventories. The fact that oil inventories seem to be almost unrelated to the news shock speaks against putting it on a level with an oil inventory demand shock studied extensively in the literature by Kilian and Murphy (2014), Juvenal and Petrella (2015) and Baumeister and Hamilton (2019).

The exogenous oil supply shock attributes greatly to the volatility in oil production, particularly in the short run. Furthermore, it also explains a large fraction of the variation in oil inventories at all horizons, a finding in line with Juvenal and Petrella (2015). With shares up to 25%, flow shocks also explain a considerable fraction of industrial production, the S&P500 stock market and unemployment. These values are somewhat larger than estimated in the SVAR models by Antolín-Díaz and Rubio-Ramírez (2018) or Caldara et al. (2019), and are rather in line with those obtained in Baumeister and Peersman (2013). Finally, flow shocks contribute for a non-negligible part of the medium-term and long-term variance in monetary policy variables. Overall, the FAVAR estimates of the FEVDs suggest that news shocks are able to explain a large amount of variation in oil and consumer prices, especially on impact. In turn, more conventional flow shocks are more important drivers of real and financial variables. These effects can also be observed in the medium and the long run.

There is a large debate in the literature about the importance of oil supply shocks for driving oil prices. A considerable set of papers argues that oil prices are largely driven by demand components and less by supply (Antolín-Díaz and Rubio-Ramírez, 2018; Kilian, 2009; Kilian and Murphy, 2012, 2014). In turn, recent papers find much larger contributions of supply shocks (Baumeister and Hamilton, 2019; Caldara et al., 2019). My estimates suggest that the traditional supply shock as studied in their papers contributes to a negligible part of oil prices, a result that would support the first set of papers. However, studying the news component of supply completely turns this result, and both supply shocks together can account up to half of the variation. These results are robust to using a smaller information set or to identifying each structural separately as outlined in Appendix 3.D. Hence, there is evidence for the fact that a news shock about future oil supply is a very important driver of the oil price while an oil flow shock is of minor interest here.

Figure 3.4: Historical decompositions



Historical decompositions of the real price of oil around key historical events for a news shock and a flow shock. Shaded areas are 68% posterior error bands. Dashed line is actual price of oil.

Finally, I study the contributions of the two types of oil supply shocks to variations in the oil price using historical decompositions. In general, historical decompositions $\mathrm{HD}_{ij}^{(t)}$ calculate the cumulative effect of the j-th structural shock ($j = 1, \ldots, n$) on the i-th variable ($i = 1, \ldots, N$) at time $t = 1, \ldots, T$:

$$\mathrm{HD}_{ij}^{(t)} = \sum_{l=0}^{t-1} e_{i,N}' \Lambda \Theta_l e_{j,n} e_{j,n}' \varepsilon_{t-l}, \qquad \text{with}$$

$$\Theta_k = \begin{cases} (A_0^{-1})', & \text{for } k = 0, \\ \sum_{l=0}^{k} (A_l A_0^{-1})' \Theta_{k-l} & \text{for } 1 \leq k \leq p, \\ \sum_{l=0}^{p} (A_l A_0^{-1})' \Theta_{k-l} & \text{for } k > p, \end{cases}$$

where $e_{i,N}$ and $e_{j,n}$ are the i-th and the j-th column of I_N and I_n, respectively. Following Antolín-Díaz and Rubio-Ramírez (2018), Figure 3.4 displays the historical decompositions in certain historical episodes that were associated with either a large rise or drop of oil prices starting from the first month in the period of interest, i.e. normalized to start from zero.

First, consider the rise in oil prices during the Iranian Revolution in 1979. Barsky and Kilian (2002) and Kilian and Murphy (2014) doubt that an exogenous oil supply disruption caused the tremendous surge and instead, they advocate that speculative components did so. This view is supported by the FAVAR estimates finding that the oil price was mainly driven by a news shock and not by a supply flow shock. To be more specific, OPEC countries except for Saudi Arabia agreed for a higher target price twice in 1979.[12] Thereby, the local authorities reacted to the freeze of Iranian dollar accounts by the US during the oil crisis. Obviously, these news shocks about future oil supply explained the major part of the price upturn.

At the end of 1985 Saudi Arabia announced to flow the market with oil. Their goal was to put pressure on other OPEC members to make them comply to a new production ceiling. Obviously, this severely crashed down the price of oil. Kilian and Murphy (2014) interpret this as an exogenous oil supply surprise. However, the second row shows that it was rather a news shock about future oil supply than a traditional supply shortfall that accounts for this decrease in the price. Note that this fall in the price cannot be completely explained by the unexpected changes under consideration. Thus, demand shocks stemming from sources other than information about future oil supply like a drop in speculative demand were also responsible for the decrease in the price in 1986 (Kilian and Murphy, 2014).

With respect to the sharp rise in the oil price observed in 2007-08, the Proxy-FAVAR finds that both a news shock and a supply flow shock contributed moderately. Though the news shock seemed to be slightly more important than the flow surprise, this difference is not significant. Apparently, the sum of both shocks explained the price surge very well. Hamilton (2009) states that there were declines in oil supply in the North Sea, Mexico and Indonesia at that time. Moreover, he argues that the oil producing countries failed to increase supply between 2005 and 2007 what can be interpreted as a news shock.

Finally, Caldara et al. (2019) and also Braun (2019) find the 2014-15 oil slump to be jointly caused by a shortfall in oil supply and a negative oil-specific demand shock. However, Caldara et al. (2019) view the unwillingness of the OPEC to cut production to yield an oil-specific demand shock. Following Känzig (2019), this should rather be interpreted as an oil supply news shock. The last row illustrates that in my analysis both the news shock and the exogenous supply shock accounted for the fall in the oil price to some extent, whereas the news component adapts the shape of the decrease slightly better. The sum of surprises could

[12]Note that before 1982 OPEC countries decided about a target price directly. Of course the prices were controlled by reducing/extending production.

not fully explain the price decline. According to Baumeister and Kilian (2016) I can justify this by the fact that the general slowdown in global economic activity was also responsible for the downward trend in the price.

Again, I find these results to be robust to using a different identification strategy that disentangles supply flow from news shocks. Table 3.C.1 and Figure 3.C.2 in Appendix 3.C.2 show that there is little difference when instead of sign restrictions on inventories, a zero restriction is placed on the contemporaneous reaction of news shocks on oil production.

3.4 Conclusion

The ultimate goal of this paper is to credibly disentangle two types of oil supply shocks simultaneously relying on a rich information set: conventional oil flow shocks that shift oil production contemporaneously, and news shocks about future oil supply that rather cause a decrease in supply in some months. So far, the effects of these shocks are investigated separately in small scale VAR models.

For this purpose, I develop novel econometric methodology which allows me to identify multiple shocks by instrumental variables within a rich information framework. My approach involves augmenting a structural FAVAR model by equations for the instruments, and imposing a certain set of exclusion restrictions such that these variables do not affect the dynamics of the FAVAR model. On the methodological side, I show how the recently proposed Bayesian inference framework for Proxy-SVARs by Arias et al. (2019) translates into a FAVAR environment. Specifically, I show how one can efficiently incorporate their algorithms into a standard Gibbs sampler for FAVAR models, using an accept-reject Metropolis-Hastings step. Adopting a Bayesian framework for the Proxy-FAVAR has two major benefits. First, it allows to easily incorporate prior information in order to regularize a highly parametrized VAR model for the factors. Second, one has the possibility to treat potential set identification in a coherent way, assigning equal prior and posterior weights to observationally equivalent parameters. Set identification becomes an important feature if multiple shocks are to be identified by instrumental variables, and sign restrictions are used to further disentangle these shocks. The possibility to identify multiple shocks by IV contrasts another recently proposed Bayesian Proxy-FAVAR by Bruns (2018), which is designed to identify the effects of only one shock.

My empirical results suggest that both shocks seem to have very similar directional effects but their magnitude might differ remarkably. With respect to the news shocks, I find that in line with Känzig (2019), they seem to be very important drivers of oil prices and hence, consumer prices in the US. However, when looking at financial and real variables, the traditional oil supply (flow) shock seems to have more fundamental effects explaining

considerably larger fractions of the forecast error variance than the news shock. Furthermore, a rich information set is crucial to reasonably identify an oil flow shock because small scale models result in puzzling effects like decreasing prices and increasing economic activity. Identifying two types of oil supply shocks simultaneously and thereby ensuring that they are unconfounded, allows to figure out which shock harms the economy more severe. In terms of economic activity, an oil flow shock has a stronger impact on industrial production than a news shock. FAVAR models that only identify one type of oil supply shock at a time are unable to make this point. These findings are important for policy makers, as precise estimates of their effects are vital to conduct meaningful scenario analysis, assess the risks that are associated with distinct supply shocks, and to design effective policy measures including financial regulation or monetary policy.

References

Alquist, R. and L. Kilian (2010). "What do we learn from the price of crude oil futures?" *Journal of Applied Econometrics* 25 (4), 539–573.

Antolín-Díaz, J., I. Petrella, and J. F. Rubio-Ramírez (2018). "Structural scenario analysis with SVARs". *CEPR Discussion Paper No. DP12579*.

Antolín-Díaz, J. and J. F. Rubio-Ramírez (2018). "Narrative sign restrictions for SVARs". *American Economic Review* 108 (10), 2802–29.

Arias, J. E., J. F. Rubio-Ramírez, and D. F. Waggoner (2018). "Inference based on structural vector autoregressions identified with sign and zero restrictions: Theory and applications". *Econometrica* 86 (2), 685–720.

– (2019). "Inference in Bayesian Proxy-SVARs". *FRB Atlanta Working Paper*.

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

Barsky, R. B. and L. Kilian (2002). "Do we really know that oil caused the great stagflation? A monetary alternative". *NBER Macroeconomics Annual* 16, 137–183.

Baumeister, C. and J. D. Hamilton (2019). "Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks". *American Economic Review* 109 (5), 1873–1910.

Baumeister, C. and L. Kilian (2016). "Understanding the decline in the price of oil since June 2014". *Journal of the Association of Environmental and Resource Economists* 3 (1), 131–158.

Baumeister, C. and G. Peersman (2013). "The role of time-varying price elasticities in accounting for volatility changes in the crude oil market". *Journal of Applied Econometrics* 28 (7), 1087–1109.

Beaudry, P. and F. Portier (2014). "News-driven business cycles: Insights and challenges". *Journal of Economic Literature* 52 (4), 993–1074.

Belviso, F. and F. Milani (2006). "Structural factor-augmented VARs (SFAVARs) and the effects of monetary policy". *Topics in Macroeconomics* 6 (3).

Bernanke, B. S. (1986). "Alternative explanations of the money-income correlation". In: *Carnegie-Rochester Conference Series on Public Policy*. Vol. 25. Elsevier, 49–99.

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.

Bernanke, B. S., M. Gertler, and M. Watson (1997). "Systematic monetary policy and the effects of oil price shocks". *Brookings Papers on Economic Activity* 1, 91–157.

Blanchard, O. J. and D. Quah (1989). "The dynamic effects of aggregate demand and supply disturbances". *The American Economic Review* 79 (4), 655–673.

Braun, R. (2019). "Three essays on identification in structural vector autoregressive models". PhD thesis, University of Konstanz.

Braun, R. and R. Brüggemann (2019). "Identification of SVAR models by combining sign restrictions with external instruments". *Working Paper, University of Konstanz.*

Bruns, M. (2018). "Combining factors models and external instruments to identify uncertainty shocks". *Unpublished working paper*.

Caldara, D., M. Cavallo, and M. Iacoviello (2019). "Oil price elasticities and oil price fluctuations". *Journal of Monetary Economics* 103, 1–20.

Caldara, D. and E. Herbst (2019). "Monetary policy, real activity, and credit spreads: Evidence from Bayesian Proxy SVARs". *American Economic Journal: Macroeconomics* 11 (1), 157–92.

Canova, F. and G. De Nicoló (2002). "Monetary disturbances matter for business fluctuations in the G-7". *Journal of Monetary Economics* 49, 1131–1159.

Carter, C. K. and R. Kohn (1994). "On Gibbs sampling for state space models". *Biometrika* 81 (3), 541–553.

Chan, J. C. C. (2017). "The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling". *Journal of Business & Economic Statistics* 35 (1), 17–28.

Chan, J. C. C. and I. Jeliazkov (2009). "Efficient simulation and integrated likelihood estimation in state space models". *International Journal of Mathematical Modelling and Numerical Optimisation* 1 (1-2), 101–120.

Chib, S. and E. Greenberg (1995). "Understanding the Metropolis–Hastings algorithm". *The American Statistician* 49 (4), 327–335.

Faust, J. (1998). "The robustness of identified VAR conclusions about money". *Carnegie-Rochester Conference Series on Public Policy* 49, 207–244.

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.

Gertler, M. and P. Karadi (2015). "Monetary policy surprises, credit costs, and economic activity". *American Economic Journal-Macroeconomics* 7 (1), 44–76.

Geweke, J. (1992). "Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments". *Bayesian statistics* 4, 641–649.

Giacomini, R., T. Kitagawa, and M. Read (2019). "Robust Bayesian inference in Proxy SVARs". *Cemmap Working Paper 23*.

Giannone, D., L. Reichlin, and L. Sala (2004). "Monetary policy in real time". *NBER Macroeconomics Annual* 19, 161–200.

Gilchrist, S. and E. Zakrajšek (2012). "Credit spreads and business cycle fluctuations". *American Economic Review* 102 (4), 1692–1720.

Hamilton, J. D. (2009). "Causes and consequences of the oil shock of 2007-08". *Brookings Papers on Economic Activity* 40 (1), 215–283.

Hamilton, J. D. and A. Herrera (2004). "Oil shocks and aggregate macroeconomic behavior: The role of monetary policy." *Journal of Money, Credit, and Banking* 36 (2), 265.

Herrera, A. M., M. B. Karaki, and S. K. Rangaraju (2019). "Oil price shocks and US economic activity". *Energy Policy* 129, 89–99.

Juvenal, L. and I. Petrella (2015). "Speculation in the oil market". *Journal of Applied Econometrics* 30 (4), 621–649.

Känzig, D. R. (2019). "The macroeconomic effects of oil supply shocks: New evidence from OPEC announcements". *Working Paper*.

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. (2008a). "Exogenous oil supply shocks: How big are they and how much do they matter for the U.S. economy?" *The Review of Economics and Statistics* 90 (2), 216–240.
– (2008b). "The economic effects of energy price shocks". *Journal of Economic Literature* 46 (4), 871–909.
– (2009). "Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market". *American Economic Review* 99 (3), 1053–1069.
Kilian, L. and L. T. Lewis (2011). "Does the Fed respond to oil price shocks?" *The Economic Journal* 121 (555), 1047–1072.
Kilian, L. and H. Lütkepohl (2017). *Structural Vector Autoregressive Analysis*. Themes in Modern Econometrics. Cambridge University Press.
Kilian, L. and D. P. Murphy (2012). "Why agnostic sign restrictions are not enough: Understanding the dynamics of oil market VAR models". *Journal of the European Economic Association* 10 (5), 1166–1188.
– (2014). "The role of inventories and speculative trading in the global market for crude oil". *Journal of Applied Econometrics* 29 (3), 454–478.
Koch, K. (2007). "Gibbs sampler by sampling-importance-resampling". *Journal of Geodesy* 81 (9), 581–591.
Lakdawala, A. (2019). "Decomposing the effects of monetary policy using an external instruments SVAR". *Journal of Applied Econometrics* 34 (6), 934–950.
Lippi, F. and A. Nobili (2012). "Oil and the macroeconomy: A quantitative structural analysis". *Journal of the European Economic Association* 10 (5), 1059–1083.
Lunsford, K. G. (2015). "Identifying structural VARs with a proxy variable and a test for a weak proxy". *FRB of Cleveland Working Paper 15-28*.
Magnus, J. R. and H. Neudecker (2019). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons.
McCracken, M. W. and S. Ng (2016). "FRED-MD: A monthly database for macroeconomic research". *Journal of Business & Economic Statistics* 34 (4), 574–589.
Mertens, K. and M. O. Ravn (2012). "Empirical evidence on the aggregate effects of anticipated and unanticipated US tax policy shocks". *American Economic Journal-Economic Policy* 4 (2), 145–181.
– (2013). "The dynamic effects of personal and corporate income tax changes in the United States". *American Economic Review* 103 (4), 1212–1247.
Mertens, K. and J. L. Montiel Olea (2018). "Marginal tax rates and income: New time series evidence". *The Quarterly Journal of Economics* 133 (4), 1803–1884.
Montiel-Olea José, L, J. H. Stock, and M. W. Watson (2016). "Uniform inference in SVARs identified with external instruments". *Harvard Manuscript*.
Onatski, A. (2010). "Determining the number of factors from empirical distribution of eigenvalues". *The Review of Economics and Statistics* 92 (4), 1004–1016.
Piffer, M. and M. Podstawski (2017). "Identifying uncertainty shocks using the price of gold". *The Economic Journal* 128 (616), 3266–3284.
Poirier, D. J. (1998). "Revising beliefs in nonidentified models". *Econometric Theory* 14 (4), 483–509.
Rubio-Ramírez, J. F., D. F. Waggoner, and T. Zha (2010). "Structural vector autoregressions: Theory of identification and algorithms for inference". *The Review of Economic Studies* 77 (2), 665–696.

Rue, H., S. Martino, and N. Chopin (2009). "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations". *Journal of the Royal Statistical Society Series B-Statistical Methodology* 71 (2), 319–392.

Stock, J. H. and M. W. Watson (2012a). "Disentangling the channels of the 2007-09 recession". *Brookings Papers on Economic Activity* 43 (1), 81–156.

– (2016). "Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics". In: *Handbook of Macroeconomics*. Ed. by Taylor, J. B. and Uhlig, H. Vol. 2. Elsevier. Chap. 8, 415–525.

Tierney, L. (1994). "Markov chains for exploring posterior distributions". *The Annals of Statistics* 22 (4), 1701–1728.

Uhlig, H. (2005). "What are the effects of monetary policy on output? Results from an agnostic identification procedure". *Journal of Monetary Economics* 52, 381–419.

Waggoner, D. F. and T. Zha (2003). "A Gibbs sampler for structural vector autoregressions". *Journal of Economic Dynamics and Control* 28 (2), 349–366.

Yamamoto, Y. (2019). "Bootstrap inference for impulse response functions in factor-augmented vector autoregressions". *Journal of Applied Econometrics* 34 (2), 247–267.

Appendix 3.A Algorithm

Before the algorithm is presented in detail, quickly restate the Proxy-FAVAR model:

$$Y_t = \Lambda_F F_t + \Lambda_W W_t + e_t = \Lambda z_t + e_t, \qquad e_t \sim \mathcal{N}(0, R), \tag{3.A.1}$$

$$\tilde{z}_t' \tilde{A}_0 = \sum_{j=1}^{p} \tilde{z}_{t-j}' \tilde{A}_j + \tilde{\varepsilon}_t' = x_t' \tilde{A}_+ + \tilde{\varepsilon}_t', \qquad \varepsilon_t \sim \mathcal{N}(0, I_{n+k}), \tag{3.A.2}$$

where Y_t is a $N \times 1$ vector of time series, F_t is a $n_1 \times 1$ vector of latent factors, W_t is a $n_2 \times 1$ vector of observed time series and e_t is a $N \times 1$ vector of error terms. Furthermore, it is $\Lambda = (\Lambda_F, \Lambda_W)$ with $N \times n_1$ factor loadings Λ_F and $N \times n_2$ loadings for observed time series Λ_W, and $z_t' = (F_t', W_t')$. Imposed constraints on observation equation (3.A.1) are identifying restrictions $\Lambda_F' = (I_{n_1}, \tilde{\Lambda}_F')$ and $\Lambda_W' = (0_{n_2 \times n_1}, \tilde{\Lambda}_W')$ as well as a diagonal covariance $R = \text{diag}(r_1, \ldots, r_N)$. State equation (3.A.2) models $\tilde{z}_t' = (z_t', m_t')$ with m_t being a $k \times 1$ vector of instruments, and $\tilde{\varepsilon}_t' = (\varepsilon_t', v_t')$ where ε_t is a $n \times 1$ vector of structural shocks and v_t is a $k \times 1$ innovation vector for the proxies. Matrices $\tilde{A}_j$ are $(n + k) \times (n + k)$ for $j = 0, \ldots, p$, $\tilde{A}_+ = (\tilde{A}_1', \ldots, \tilde{A}_p')'$ and $x_t' = (\tilde{z}_{t-1}', \ldots, \tilde{z}_{t-p}')$ for $t = p + 1, \ldots, T$. The structural parameters are subject to block and exogeneity constraints:

$$J \tilde{A}_j L_1' = 0_{k \times n}, \qquad \text{for } j = 0, \ldots, p, \tag{3.A.3}$$

$$J(\tilde{A}_0^{-1})' L_2' = 0_{k \times (n-k)}, \tag{3.A.4}$$

where $J = (0_{k \times n}, I_k)$, $L_1 = (I_n, 0_{n \times k})$ and $L_2 = (I_{n-k}, 0_{(n-k) \times 2k})$.

To simplify notation, define $f = \text{vec}(F')$ with $F = (F_1, \ldots, F_T)$, $w = \text{vec}(W')$ with $W = (W_1, \ldots, W_T)$ and $m = \text{vec}(M')$ with $M = (m_1, \ldots, m_T)$. If there is an unobserved data point in data matrix $Y = (Y_1, \ldots, Y_T)$, I mark this with a star, i.e. $Y_{i,t}^*$ represents a missing data point in the (i, t)-th entry of Y. Furthermore, let y^* be the vector of all missing observations in Y and Y_{obs} be the data matrix without missing observations. Analogously, I define w^* and W_{obs} for the observed time series factors.

Orthogonal triangular-block parameterization Due to practical reasons that will be clarified in the following, I consider an alternative parameterization of transition equation (3.A.2) (see Arias et al. (2019)). Consider an upper-triangular $(n + k) \times (n + k)$ matrix $\tilde{\Pi}_0$ that has strictly positive diagonal entries and a $p(n + k) \times (n + k)$ matrix $\tilde{\Pi}_+$ that satisfies the same constraints as $\tilde{A}_+ = (\tilde{A}_1', \ldots, \tilde{A}_p')'$ specified in (3.A.3). Furthermore, let Q_1 and Q_2

be a $n \times n$ and a $k \times k$ orthogonal matrix, respectively. Moreover, let function $f(\cdot,\cdot)$ map $(\tilde{\Pi}_0, \tilde{\Pi}_+, Q_1, Q_2)$ to $(\tilde{A}_0, \tilde{A}_+)$ using $Q = \text{diag}(Q_1, Q_2)$:

$$f(\tilde{\Pi}_0, \tilde{\Pi}_+, Q_1, Q_2) \mapsto (\underbrace{\tilde{\Pi}_0 Q}_{=\tilde{A}_0}, \underbrace{\tilde{\Pi}_+ Q}_{=\tilde{A}_+}).$$

Obviously, parameters $\tilde{A}_0$ and $\tilde{A}_+$ resulting from this mapping fulfill restrictions (3.A.3) by construction. Additionally, Arias et al. (2019) show that exogeneity restrictions (3.A.4) can be translated into linear constraints on the columns of Q_1 that are determined by entries of $\tilde{\Pi}_0$. Based on that, they find a way to formulate a unique mapping g from $(\tilde{\Pi}_0, \tilde{\Pi}_+)$ towards $(\tilde{\Pi}_0, \tilde{\Pi}_+, Q_1, Q_2)$. Hence, Arias et al. (2019) have found an unique function $(f \circ g)$ that maps $(\tilde{\Pi}_0, \tilde{\Pi}_+)$ to $(\tilde{A}_0, \tilde{A}_+)$. Thus, I consider an observationally equivalent state equation that is obtained multiplying (3.A.2) with Q' from the right:

$$\tilde{z}_t' \tilde{\Pi}_0 = x_t' \tilde{\Pi}_+ + \tilde{\eta}_t', \qquad t = p+1, \ldots, T, \tag{3.A.5}$$

where $\tilde{\eta}_t$'s are standard normally distributed and $x_t' = (\tilde{z}_{t-1}', \ldots, \tilde{z}_{t-p}')$.

3.A.1 Prior distributions

Given that there are missing values in the matrix of observed time series W, state equation (3.A.2) serves as prior distribution for the missing observations (Chan and Jeliazkov, 2009). This works analogously to the prior for the latent factors specified in Section 3.2.2. Let vector w separate into the vector of actually observed time series $w_{\text{obs}} = \text{vec}(W_{\text{obs}}')$ and the vector of missing values w^*. Then, rearrange $\hat{z}' = (f', w', m')$ to obtain $\mathring{z} = (\underbrace{f', w_{\text{obs}}', m'}_{=\breve{z}'}, w^{*'})$.

If we rearrange $H_{\hat{z}}$ and $\Sigma_{\hat{z}}$ that were specified in Section 3.2.2 accordingly, we get $H_{\mathring{z}}$ and $\Sigma_{\mathring{z}}$ such that:

$$H_{\mathring{z}} \mathring{z} = \mathring{u}, \qquad \mathring{u} \sim \mathcal{N}(0, \Sigma_{\mathring{z}}).$$

Hence, vector $\mathring{z}$ has the unconditional distribution:

$$\mathring{z} = \begin{pmatrix} \breve{z} \\ w^* \end{pmatrix} \sim \mathcal{N}\left(0, \left[H_{\mathring{z}}' \Sigma_{\mathring{z}}^{-1} H_{\mathring{z}}\right]^{-1}\right), \qquad H_{\mathring{z}}' \Sigma_{\mathring{z}}^{-1} H_{\mathring{z}} = \begin{pmatrix} Q_{\breve{z}} & Q_{\breve{z},w^*} \\ Q_{\breve{z},w^*}' & Q_{w^*} \end{pmatrix}.$$

Consequently, the resulting prior for the missing data points in the observed time series conditional on latent factors, actually observed data points and proxies reads:

$$w^*|\breve{z}, \tilde{A}_0, \tilde{A}_+, \Sigma_0 \sim \mathcal{N}\left(-Q_{w^*}^{-1} Q'_{\breve{z},w^*} \breve{z}, Q_{w^*}^{-1}\right). \tag{3.A.6}$$

Again, this distribution is conditional on the structural parameters of the state equation.

Next, set a prior distribution for the potentially unavailable data points in time series matrix Y, denoted by $Y_{i,t}^*$ ($i = 1, \ldots, N$; $t = 1, \ldots, T$). Here, measurement equation (3.A.1) serves as the conditional prior:

$$Y_{i,t}^*|\lambda_i, z_t, r_i \sim \mathcal{N}\left(\lambda_i' z_t, r_i\right), \tag{3.A.7}$$

where λ_i' is the i-th row of loading matrix Λ.

The remaining prior distributions for model (3.A.1) and (3.A.2) are specified in the main body and read:

$$f|w, m, \tilde{A}_0, \tilde{A}_+, \Sigma_0 \sim \mathcal{N}\left(-Q_f^{-1} Q_{f,w,m}(w', m')', Q_f^{-1}\right), \tag{3.A.8}$$

$$(\tilde{A}_0, \tilde{A}_+) \propto \mathcal{NGN}_r(\nu, \Phi, \Psi, \Omega), \tag{3.A.9}$$

$$r_i \sim i\mathcal{G}\left(\underline{a}, \underline{b}\right), \quad i = 1, \ldots, N, \tag{3.A.10}$$

$$\tilde{\lambda}_i|r_i \sim \mathcal{N}\left(\underline{\mu}_{\tilde{\lambda}_i}, r_i \cdot \underline{V}_{\tilde{\lambda}_i}\right), \quad i = n_1 + 1, \ldots, N, \tag{3.A.11}$$

where matrices Q_f and $Q_{f,w,m}$ are defined as pointed out in detail in Section 3.2.2, $\tilde{\lambda}_i$ is the $(i - n_1)$-th row of $\tilde{\Lambda} = (\tilde{\Lambda}_F, \tilde{\Lambda}_W)$ and $\mathcal{NGN}_r(\nu, \Phi, \Psi, \Omega)$ denotes the normal-generalized-normal distribution subject to restrictions (3.A.3), (3.A.4) and potentially additional sign restrictions that can be required to distinguish between different structural shocks under consideration.[13] Furthermore, I restrict the minimum eigenvalue of the reliability matrix that is implied by the structural parameters as outlined in Section 3.2.1 to be at least equal to γ.

Prior moments The prior covariance matrix of the presample latent factors F_t ($t = 1, \ldots, p$) is set to $\Sigma_0 = I_{n_1}$. I set the prior moments for the structural parameters $\nu = n + k$, $\Phi = I_{(n+k)^2}$, $\Psi = 0_{p(n+k)^2 \times (n+k)^2}$ and $\Omega = I_{p(n+k)^2}$. Next, the minimum share of the proxy variance that is explained by the corresponding structural shocks is lower bounded by $\gamma = 0.05$. For the remaining prior parameters I use standard quantities: $\underline{a} = (3 - n_2 - k)/2$, $\underline{b} = 1.5$, $\underline{\mu}_{\tilde{\lambda}_i} = 0_{n \times 1}$ and $\underline{V}_{\tilde{\lambda}_i} = I_n$.

[13] The density of the normal-generalized-normal is presented in Section 3.2.3.

3.A.2 Posterior algorithm

Before I report the individual steps of a MCMC sampler, I briefly sketch the initializations of the parameters. First, set potential missings in the data matrix to zero, i.e. $Y^*_{i,t} = 0$ $(i = 1,\dots,N; t = 1,\dots,T)$, and obtain eventual missing data points in the matrix of observed time series W by linear interpolation and extrapolation, respectively. Then, initialize the latent factors F_t with the first n_1 time series of data matrix Y. Given these initializations for F and W, estimate state equation (3.A.2) in reduced form and set $\hat{\Pi}_0^{-1}$ equal to the upper-triangular Cholesky factor of the estimated least squares covariance matrix. Departing from that $\hat{\Pi}_0$, draw an initialization of the structural parameters $(\tilde{A}_0, \tilde{A}_+)$ by generating a candidate like in the ARMH step 3 described below. The unrestricted part of the factor loadings is initialized by least squares accordingly: $\tilde{\Lambda} = (Y_{n_1+1:N}Z')(ZZ')^{-1}$ with $Y_{n_1+1:N}$ being rows $n_1 + 1$ to N of data matrix Y and $Z = (z_1,\dots,z_T)$. In addition, set $r_i = 0.01$ for $i = 1,\dots,n_1$ and $r_i = (1/T)\sum_{t=1}^{T} \hat{e}^2_{i,t}$ for $i = n_1 + 1,\dots,N$ using $\hat{e}_{i,t} = Y_{i,t} - \tilde{\lambda}_i' z_t$.

Let θ be the vector of all unknown parameters and generally θ_{-S} the parameter vector without quantity S, i.e. $\theta \setminus S$:[14]

$$\theta = \left(\tilde{f}', w^{*'}, \mathrm{vec}(\tilde{A}_0, \tilde{A}_+)', r', \mathrm{vec}(\tilde{\Lambda})', y^{*'}\right)',$$
$$\theta_{-f} = \left(w^{*'}, \mathrm{vec}(\tilde{A}_0, \tilde{A}_+)', r', \mathrm{vec}(\tilde{\Lambda})', y^{*'}\right)'.$$

Then, the algorithm generates draws $\theta^{(l)}$ for $l = 1,\dots,M$ from its posterior by iterating through the following blocks.[15]

1. First, sample the latent factors f. Considering measurement equation (3.A.1) in matrix notation and vectorizing it yields:

$$\begin{aligned} Y &= \Lambda_F F + \Lambda_W W + E, && \text{with } E = (e_1,\dots,e_T), \\ Y_W' &= Y' - W'\Lambda_W' = F'\Lambda_F' + E', \\ y_W &= \mathrm{vec}(Y_W') = (\Lambda_F \otimes I_T) f + e, && \text{with } e = \mathrm{vec}(E') \sim \mathcal{N}(0, R \otimes I_T). \end{aligned}$$

[14] Note that θ does not contain $\hat{\Pi}_0$ because it is not a parameter of model (3.A.1) and (3.A.2). However, $\hat{\Pi}_0$ is required to make inference on the structural parameters $(\tilde{A}_0, \tilde{A}_+)$. Therefore, it is also updated through the algorithm.

[15] I use 1,000 burn-in draws and 4,000 draws for inference. To break autocorrelation, only every 10-th draw is stored.

Combined with prior (3.A.8), this yields a classical normal posterior:

$$
\begin{aligned}
p(f|Y_{\text{obs}}, W_{\text{obs}}, \theta_{-f}) &\sim \mathcal{N}(\bar{\mu}_f, \bar{V}_f), \\
\text{with } \bar{V}_f &= \left(Q_f + (\Lambda_F \otimes I_T)'(R^{-1} \otimes I_T)(\Lambda_F \otimes I_T)\right)^{-1}, \\
\bar{\mu}_f &= \bar{V}_f \left(-Q_{f,w,m}(w', m')' + (\Lambda_F \otimes I_T)'(R^{-1} \otimes I_T) y_W\right).
\end{aligned}
\tag{3.A.12}
$$

Thereby, using techniques from the Gaussian Markov random fields literature (Chan, 2017; Rue et al., 2009), one can sample from this posterior in a very efficient way. The intuition behind is that the highly dimensional $n_1 T \times n_1 T$ matrix $\bar{V}_f$ is sparse. Therefore, it can be manipulated to be a sparse band matrix what yields a remarkable decrease of the computational burden.

2. Analogously, I can sample potential missing data points in the observed time series following the same procedure as in step 1:

$$
\begin{aligned}
y_F &= \text{vec}(Y' - F'\Lambda_F') = (\Lambda_W \otimes I_T)w + e, \\
y_F^* &= y_F - \Lambda_W^{\text{obs}} w_{\text{obs}} = \Lambda_W^* w^* + e,
\end{aligned}
$$

where Λ_W^{obs} and Λ_W^* are the columns of $(\Lambda_W \otimes I_T)$ that are related to w_{obs} and w^*, respectively. I combine this likelihood with prior (3.A.6) to get the normal posterior:

$$
\begin{aligned}
p(w^*|Y_{\text{obs}}, W_{\text{obs}}, \theta_{-w^*}) &\sim \mathcal{N}(\bar{\mu}_{w^*}, \bar{V}_{w^*}), \\
\text{with } \bar{V}_{w^*} &= \left(Q_{w^*} + \Lambda_W^{*'}(R^{-1} \otimes I_T)\Lambda_W^*\right)^{-1}, \\
\bar{\mu}_{w^*} &= \bar{V}_{w^*} \left(-Q'_{\breve{z},w^*}\breve{z} + \Lambda_W^{*'}(R^{-1} \otimes I_T) y_F^*\right).
\end{aligned}
\tag{3.A.13}
$$

As in step 1, Gaussian Markov random fields methods to sample from posterior (3.A.13) efficiently are applied.

3. The most challenging and computationally most demanding block is to sample the structural parameters of state equation (3.A.2):

$$
\tilde{Z}\tilde{A}_0 = X\tilde{A}_+ + \tilde{\varepsilon},
$$

where $\tilde{Z} = (\tilde{z}_{p+1}, \ldots, \tilde{z}_T)'$, $X = (x_{p+1}, \ldots, x_T)'$ and $\tilde{\varepsilon} = (\tilde{\varepsilon}_{p+1}, \ldots, \tilde{\varepsilon}_T)'$. The posterior distribution of the structural parameters conditional on the factors is proportional to a

normal-generalized-normal with respect to restrictions (3.A.3), (3.A.4) and potential sign restrictions:

$$p(\tilde{A}_0, \tilde{A}_+ | Y_{\text{obs}}, W_{\text{obs}}, \theta_{-(\tilde{A}_0, \tilde{A}_+)}) \propto \mathcal{NGN}_r \left(\tilde{\nu}, \tilde{\Phi}, \tilde{\Psi}, \tilde{\Omega}\right), \tag{3.A.14}$$

where $\tilde{\nu} = T + \nu$, $\tilde{\Omega} = (I_{n+k} \otimes X'X + \Omega^{-1})^{-1}$, $\tilde{\Psi} = \tilde{\Omega}(I_{n+k} \otimes X'\tilde{Z} + \Omega^{-1}\Psi)$ and $\tilde{\Phi} = I_{n+k} \otimes \tilde{Z}'\tilde{Z} + \Phi + \Psi'\Omega^{-1}\Psi + \tilde{\Psi}'\tilde{\Omega}^{-1}\tilde{\Psi}$ (Arias et al., 2019). In general, Arias et al. (2018) show how to sample independently from any posterior distribution of the structural parameters for point-identified and set-identified models. However, for $p > 0$, Proxy-SVAR model (3.A.2) with restrictions (3.A.3) and (3.A.4) is overidentified.[16] Therefore, Arias et al. (2019) adjust their procedure to be able to use it in a Proxy-SVAR setting. Particularly, they sample $(\tilde{\Pi}_0, \tilde{\Pi}_+)$ from the orthogonal triangular-block parameterization (3.A.5) using the algorithm of Waggoner and Zha (2003) and map these draws to the structural parameters. Note that the resulting draws $(\tilde{A}_0, \tilde{A}_+)$ are not from the restricted normal-generalized-normal distribution. Nevertheless, Arias et al. (2019) show how to use an importance sampler to generate draws from the desired distribution. I embed this importance sampler in an accept-reject Metropolis-Hastings algorithm following the ideas of Tierney (1994) and Chib and Greenberg (1995). This is required since the restricted normal-generalized-normal posterior of the structural parameters is not dominated by the proposal distribution presented in *Algorithm 2* of Arias et al. (2019). Thus, I correct for that in a Metropolis-Hastings step. An alternative way to use an importance sampling step within a Gibbs sampler would be to apply sampling-importance-resampling (Koch, 2007) as proposed by Antolín-Díaz et al. (2018).

(a) *Accept-reject step*: Generate candidate draws for the structural parameters $(\tilde{A}_0, \tilde{A}_+)$:

  (i) First, starting from $\hat{\Pi}_0^{(l-1)}$, sample candidates for the parameters of orthogonal triangular-block parameterization (3.A.5) from its normal-generalized-normal posterior $\mathcal{NGN}_R(\hat{\nu}, \hat{\Phi}, \hat{\Psi}, \hat{\Omega})$ that is subject to the corresponding constraints using the algorithm of Waggoner and Zha (2003). I set the posterior moments of this normal-generalized-normal equal to that of the structural parameters specified above, i.e. $\hat{\nu} = \tilde{\nu}$, $\hat{\Omega} = \tilde{\Omega}$, $\hat{\Psi} = \tilde{\Psi}$ and

[16] In particular, (3.A.3) imposes $(p+1)k$ constraints on each of the first n equations. The algorithm of Arias et al. (2018) allows at most $n + k - j$ restrictions in equation j, i.e. for $j = n$ it allows for at most k restrictions. Hence, in case of $p > 0$, (3.A.3) implies too many restrictions.

$\hat{\Phi} = \tilde{\Phi}$.[17] For $i = 1, \ldots, n+k$, let $\hat{\pi}_{0,i}$, $\tilde{\pi}_{0,i}$ and $\tilde{\pi}_{+,i}$ be the column vectors of the free parameters in column i of matrices $\hat{\Pi}_0$, $\tilde{\Pi}_0$ and $\tilde{\Pi}_+$, respectively:

$$\hat{\Pi}_{0,i} = U_i \hat{\pi}_{0,i}, \quad \tilde{\Pi}_{0,i} = U_i \tilde{\pi}_{0,i}, \quad \tilde{\Pi}_{+,i} = V_i \tilde{\pi}_{+,i}, \tag{3.A.15}$$

where U_i are the first i columns of I_{n+k}, V_i equals $I_{p(n+k)}$ except columns $n+1, \ldots, n+k, 2n+k+1, \ldots, 2(n+k), \ldots, pn+(p-1)k+1, \ldots, p(n+k)$ for $i = 1, \ldots, n$ and $V_i = I_{p(n+k)}$ for $i = n+1, \ldots, n+k$. Like Arias et al. (2019), set for $i = 1, \ldots, n+k$:

$$\begin{aligned}
\nu_i &= \hat{\nu} - n - k, \\
H_i &= \left(V_i' \hat{\Omega}_i^{-1} V_i \right)^{-1}, \\
P_i &= H_i V_i' \hat{\Omega}_i^{-1} \hat{\Psi}_i U_i, \\
S_i &= \left(1/\nu_i \left(U_i' \hat{\Phi}_i U_i + U_i' \hat{\Psi}_i' \hat{\Omega}_i^{-1} \hat{\Psi}_i U_i - P_i' H_i^{-1} P_i \right) \right)^{-1},
\end{aligned}$$

where $\hat{\Phi}_i$, $\hat{\Psi}_i$ and $\hat{\Omega}_i$ are the i-th blocks of the block diagonal matrices $\hat{\Phi}$, $\hat{\Psi}$ and $\hat{\Omega}$, respectively. To draw candidates for the triangular-block parameters, iterate for $i = 1, \ldots, n+k$:

- Select $w_{(i)}$ as a vector to be orthogonal to all columns except the i-th column of $\hat{\Pi}_0^{(l-1)}$. Then, set:

  $$w_{1,i} = T_i' U_i' w_{(i)} / ||T_i' U_i' w_{(i)}||,$$

  with T_i being the Cholesky decomposition of S_i. Choose $w_{2,i}, \ldots, w_{i,i}$ such that together with $w_{1,i}$ they build an orthonormal basis of $\mathbb{R}^i$.
- Sample $\zeta_1, \ldots, \zeta_{T+1} \sim \mathcal{N}\left(0, \frac{1}{T}\right)$, set $r_\zeta = \sum_{t=1}^{T+1} \zeta_t^2$ and $\beta_1 = \pm\sqrt{r_\zeta}$ with equal probabilities. Furthermore, draw $\beta_2, \ldots, \beta_i \sim \mathcal{N}\left(0, \frac{1}{T}\right)$.
- Set quantities:

  $$\hat{\pi}_{0,i,c} = T_i \sum_{l=1}^{i} w_{l,i} \beta_l, \quad \tilde{\pi}_{0,i,c} = \hat{\pi}_{0,i,c} / \text{sign}\left(\hat{\pi}_{0,i,c}(i)\right),$$

[17] Arias et al. (2019) argue that this choice can cause low effective sample sizes and suggest a more tailored choice. However, this does not seem to be a problem for this model.

where $\hat{\pi}_{0,i,c}(i)$ is the i-th element of column vector $\hat{\pi}_{0,i,c}$. Conditional on that, sample $\hat{\pi}_{+,i,c}|\hat{\pi}_{0,i,c} \sim \mathcal{N}\left(P_i\hat{\pi}_{0,i,c}, H_i\right)$ and set:

$$\tilde{\pi}_{+,i,c} = \hat{\pi}_{+,i,c}/\text{sign}\left(\hat{\pi}_{0,i,c}(i)\right).$$

Finally, using equations (3.A.15), map the drawn vectors to candidates $\hat{\Pi}_{0,c}$, $\tilde{\Pi}_{0,c}$ and $\tilde{\Pi}_{+,c}$.

(ii) Draw Q_1 and Q_2 to map $(\tilde{\Pi}_{0,c}, \tilde{\Pi}_{+,c})$ to $(\tilde{A}_{0,c}, \tilde{A}_{+,c})$ as described in Arias et al. (2019):

- For $i = 1, \ldots, n$, sample $\xi_{1,i} \in \mathbb{R}^{n+1-i-a_i}$ from the standard normal distribution with $a_i = k \cdot \mathbb{1}_{\{i \leq n-k\}}$ and set $\rho_{1,i} = \xi_{1,i}/||\xi_{1,i}||$. Then, set $Q_1 = \left(q_{1,1}, \ldots, q_{1,n}\right)$ by recursively defining $q_{1,i} = K_{1,i}\rho_{1,i}$. Thereby, the columns of $K_{1,i}$ form an orthonormal basis of the null space of matrix $M_{1,i}$:

$$M_{1,i} = \begin{cases} \left(q_{1,1}, \ldots, q_{1,i-1}, L_1\tilde{\Pi}_{0,c}^{-1}J', D_{1,i}\right), & \text{if } 1 \leq i \leq n-k, \\ \left(q_{1,1}, \ldots, q_{1,i-1}, D_{1,i}\right), & \text{if } n-k+1 \leq i \leq n, \end{cases}$$

where reference matrix $D_{1,i}$ is equal to the first $n+1-i-a_i$ columns of I_n.

- For $i = 1, \ldots, k$ draw $\xi_{2,i} \in \mathbb{R}^{k+1-i}$ from the standard normal distribution and set $\rho_{2,i} = \xi_{2,i}/||\xi_{2,i}||$. Then, set $Q_2 = \left(q_{2,1}, \ldots, q_{2,k}\right)$ by recursively defining $q_{2,i} = K_{2,i}\rho_{2,i}$. Here, the columns of $K_{2,i}$ form an orthonormal basis of the nullspace of $M_{2,i} = \left(q_{2,1}, \ldots, q_{2,i-1}, D_{2,i}\right)$ with $D_{2,i}$ being equal to the first $k+1-i$ columns of I_k.

(iii) Repeat steps (i) and (ii) until $(\tilde{A}_{0,c}, \tilde{A}_{+,c}) = (\tilde{\Pi}_{0,c}Q, \tilde{\Pi}_{+,c}Q)$ with $Q = \text{diag}(Q_1, Q_2)$ fulfill possible sign restrictions as well as the corresponding reliability matrix satisfies the high relevance prior, i.e. until the following inequality holds:

$$\left(\left[\Gamma_{0,2,c}^{-1}\right]'\Gamma_{0,2,c}^{-1} + V_cV_c'\right)^{-1}V_cV_c' \geq \gamma,$$

whereas $\Gamma_{0,2,c}$ and V_c are directly defined through candidate draw $\tilde{A}_{0,c}$ and equations (3.2.3) and (3.2.13).

(iv) Define the SVAR-parameter vectors $\tilde{a} = \text{vec}\left(\left[\tilde{A}_0', \tilde{A}_+'\right]'\right)$, $\tilde{a}_c = \text{vec}\left(\left[\tilde{A}_{0,c}', \tilde{A}_{+,c}'\right]'\right)$ and $\tilde{a}^{(l-1)} = \text{vec}\left(\left[\tilde{A}_0^{(l-1)'}, \tilde{A}_+^{(l-1)'}\right]'\right)$. Accept draw $\tilde{a}_c$ with probability equal to the importance weight of candidate $\tilde{a}_c$ (see *Algorithm 2* of Arias et al. (2019):[18]

$$\alpha_{\text{AR}}(\tilde{a}_c) = \min\left\{1, \left[c \cdot v_{(f\circ g)^{-1}}(\tilde{a}_c)\right]^{-1}\right\},$$

where $v_{(f\circ g)^{-1}}$ is the 'volume element' of the inverse mapping $(f \circ g)^{-1}$ that maps the vectorized structural parameters $\tilde{a}$ back to the vectorized orthogonal triangular-block parameterization. Following Theorem 3 of Arias et al. (2018), it is:

$$v_{(f\circ g)^{-1}}(\tilde{a}_c) = |N_{\tilde{a}_c}' \mathcal{J}_{(f\circ g)^{-1}}(\tilde{a}_c)' \mathcal{J}_{(f\circ g)^{-1}}(\tilde{a}_c) N_{\tilde{a}_c}|^{\frac{1}{2}},$$

with $\mathcal{J}_{(f\circ g)^{-1}}(\tilde{a}_c) = \left.\frac{\partial (f\circ g)^{-1}(\tilde{a})}{\partial \tilde{a}'}\right|_{\tilde{a}=\tilde{a}_c}$ being the Jacobian of the inverse mapping evaluated at $\tilde{a}_c$ and $N_{\tilde{a}_c}$ such that its columns form an orthonormal basis for the null space of $\left.\frac{\partial \text{vec}\left(J\left[\tilde{A}_0^{-1}\right]' L_2'\right)}{\partial \tilde{a}'}\right|_{\tilde{a}=\tilde{a}_c}$. Repeat steps (i)-(iii) until the draw is accepted.[19]

(b) *Metropolis-Hastings step*: Let $\mathcal{D} = \left\{\tilde{a} : c \cdot v_{(f\circ g)^{-1}}(\tilde{a}) \geq 1\right\}$ and $\mathcal{D}^C$ its complement. Then, accept proposal $\tilde{a}_c$ with probability $\alpha_{\text{MH}}\left(\tilde{a}_c | \tilde{a}^{(l-1)}\right)$ that is defined as:

$$\alpha_{\text{MH}}\left(\tilde{a}_c | \tilde{a}^{(l-1)}\right) = \begin{cases} 1, & \text{if } \tilde{a}^{(l-1)} \in \mathcal{D}, \\ c \cdot v_{(f\circ g)^{-1}}\left(\tilde{a}^{(l-1)}\right), & \text{if } \tilde{a}^{(l-1)} \in \mathcal{D}^C, \tilde{a}_c \in \mathcal{D}, \\ \frac{v_{(f\circ g)^{-1}}\left(\tilde{a}^{(l-1)}\right)}{v_{(f\circ g)^{-1}}(\tilde{a}_c)}, & \text{if } \tilde{a}^{(l-1)}, \tilde{a}_c \in \mathcal{D}^C. \end{cases}$$

In case of acceptance, do not only set $\tilde{A}_0^{(l)} = \tilde{A}_{0,c}$ and $\tilde{A}_+^{(l)} = \tilde{A}_{+,c}$ but also $\hat{\Pi}_0^{(l)} = \hat{\Pi}_{0,c}$. Otherwise, set $\tilde{A}_0^{(l)} = \tilde{A}_0^{(l-1)}$, $\tilde{A}_+^{(l)} = \tilde{A}_+^{(l-1)}$ and $\hat{\Pi}_0^{(l)} = \hat{\Pi}_0^{(l-1)}$.

[18] This acceptance probability is simplified because due to the specific choice $\tilde{\nu} = \hat{\nu}$, $\tilde{\Phi} = \hat{\Phi}$, $\tilde{\Psi} = \hat{\Psi}$ and $\tilde{\Omega} = \hat{\Omega}$, density values of $\mathcal{NGN}_{r,(\tilde{\nu},\tilde{\Phi},\tilde{\Psi},\tilde{\Omega})}(\tilde{A}_0, \tilde{A}_+)$ and $\mathcal{NGN}_{R,(\hat{\nu},\hat{\Phi},\hat{\Psi},\hat{\Omega})}(\tilde{\Pi}_0, \tilde{\Pi}_+)$ coincide.

[19] Tuning parameter c is iteratively updated during the burn-in period to two times the average volume element. This ensures a reasonable trade-off between accept-reject acceptance and Metropolis-Hastings acceptance.

4. Draw the loadings $\tilde{\Lambda}$ and the diagonal elements of covariance matrix R from its normal-inverse-gamma posterior. First, sample the error term variances based on prior (3.A.10) and observation equation $e_{i,t} = Y_{i,t} - \lambda_i' z_t$:

$$p(r_i|Y_{\text{obs}}, W_{\text{obs}}, \theta_{-r}) \sim i\mathcal{G}\left(\underline{a} + T/2, \underline{b} + \sum_{t=1}^{T} e_{i,t}^2/2\right), \qquad i = 1, \ldots, N.$$

Conditional on that, draw the free elements of the factor loadings equation by equation. For that purpose, consider equation $i = n_1 + 1, \ldots, N$ for the measurement equation:

$$Y_{(i)} = \tilde{\lambda}_i' Z + E_{(i)}, \qquad E_{(i)}' \sim \mathcal{N}(0, r_i I_T), \tag{3.A.16}$$

where $Y_{(i)}$ and $E_{(i)}$ are the i-th row of Y and E, respectively. Combined with prior (3.A.11), the posterior for the loadings reads:

$$\begin{aligned} p\left(\tilde{\lambda}_i|Y_{\text{obs}}, W_{\text{obs}}, \theta_{-\tilde{\Lambda}}\right) &\sim \mathcal{N}\left(\bar{\mu}_{\tilde{\lambda}_i}, r_i \cdot \bar{V}_{\tilde{\lambda}_i}\right), \\ \text{with } \bar{V}_{\tilde{\lambda}_i} &= \left(\underline{V}_{\tilde{\lambda}_i}^{-1} + ZZ'\right)^{-1}, \\ \bar{\mu}_{\tilde{\lambda}_i} &= \bar{V}_{\tilde{\lambda}_i}\left(\underline{V}_{\tilde{\lambda}_i}^{-1}\underline{\mu}_{\tilde{\lambda}_i} + ZY_{(i)}'\right). \end{aligned} \tag{3.A.17}$$

5. Missing observations in the data matrix $Y_{i,t}^*$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$ are simply drawn from their conditional distribution given the data and the previous blocks:

$$p(Y_{i,t}^*|Y_{\text{obs}}, W_{\text{obs}}, \theta_{-y^*}) \sim \mathcal{N}\left(\lambda_i' z_t, r_i\right), \tag{3.A.18}$$

with λ_i' being the i-th row of Λ.

Appendix 3.B Data

3.B.1 Data and transformations

Table 3.B.1: Variables, data sources and transformations: Proxy-FAVAR

#	Description	tcode	Source
	Observed factors		
	Global crude oil production	4	EIA
	World industrial production index	4	Baumeister and Hamilton (2019)
	Real price of oil	4	EIA
	Crude oil inventories	4	EIA

	Proxy variables		
	Future price variation around OPEC announcements	1	Känzig (2019)
	Exogenous oil supply shock series of Kilian (2008a)	1	Braun and Brüggemann (2019)
	Output and income		
2	IP: Index	5	FRED
6	Real Personal Income	5	FRED
7	Real personal income excluding current transfer receipts	5	FRED
10	IP: Final Products and Nonindustrial Supplies	5	FRED
11	IP: Final Products (Market Group)	5	FRED
12	IP: Consumer Goods	5	FRED
13	IP: Durable Consumer Goods	5	FRED
14	IP: Nondurable Consumer Goods	5	FRED
15	IP: Business Equipment	5	FRED
16	IP: Materials	5	FRED
17	IP: Durable Materials	5	FRED
18	IP: Nondurable Materials	5	FRED
19	IP: Manufacturing (SIC)	5	FRED
20	IP: Residential utilities	5	FRED
21	IP: Fuels	5	FRED
22	Capacity Utilization: Manufacturing (SIC)	2	FRED
117	ISM Purchasing Manager Index: Production	5	Datastream
	Labor market		
3	Unemployment Rate	2	FRED
23	Civilian Labor Force Level	5	FRED
24	Employment Level	5	FRED
25	Average Weeks Unemployed	2	FRED
26	Number Unemployed for Less Than 5 Weeks	5	FRED
27	Number Unemployed for 5-14 Weeks	5	FRED
28	Number Unemployed for 15 Weeks & Over	5	FRED
29	Number Unemployed for 15-26 Weeks	5	FRED
30	Number Unemployed for 27 Weeks & Over	5	FRED
31*	Initial Claims	5	FRED
32	All Employees, Total Nonfarm	6	FRED
33	All Employees, Goods-Producing	6	FRED
34	All Employees: Mining and Logging: Mining	6	FRED
35	All Employees, Construction	6	FRED
36	All Employees, Manufacturing	6	FRED
37	All Employees, Durable Goods	6	FRED
38	All Employees, Nondurable Goods	6	FRED
39	All Employees, Service-Providing	6	FRED
40	All Employees, Trade, Transportation, and Utilities	6	FRED
41	All Employees, Wholesale Trade	6	FRED
42	All Employees, Retail Trade	6	FRED
43	All Employees, Financial Activities	6	FRED
44	All Employees, Government	6	FRED
45	Average Weekly Hours: Goods-Producing	2	FRED
46	Average Weekly Overtime Hours: Manufacturing	2	FRED
47	Average Weekly Hours: Manufacturing	2	FRED
108	Average Hourly Earnings: Goods-Producing	6	FRED
109	Average Hourly Earnings: Construction	6	FRED
110	Average Hourly Earnings: Manufacturing	6	FRED
118	ISM Manufacturers survey: Employment Index	5	Datastream
	Housing		
48	Housing Starts: Total: New Privately Owned Housing	5	FRED
49	Housing Starts in Northeast Census Region	5	FRED
50	Housing Starts in Midwest Census Region	5	FRED

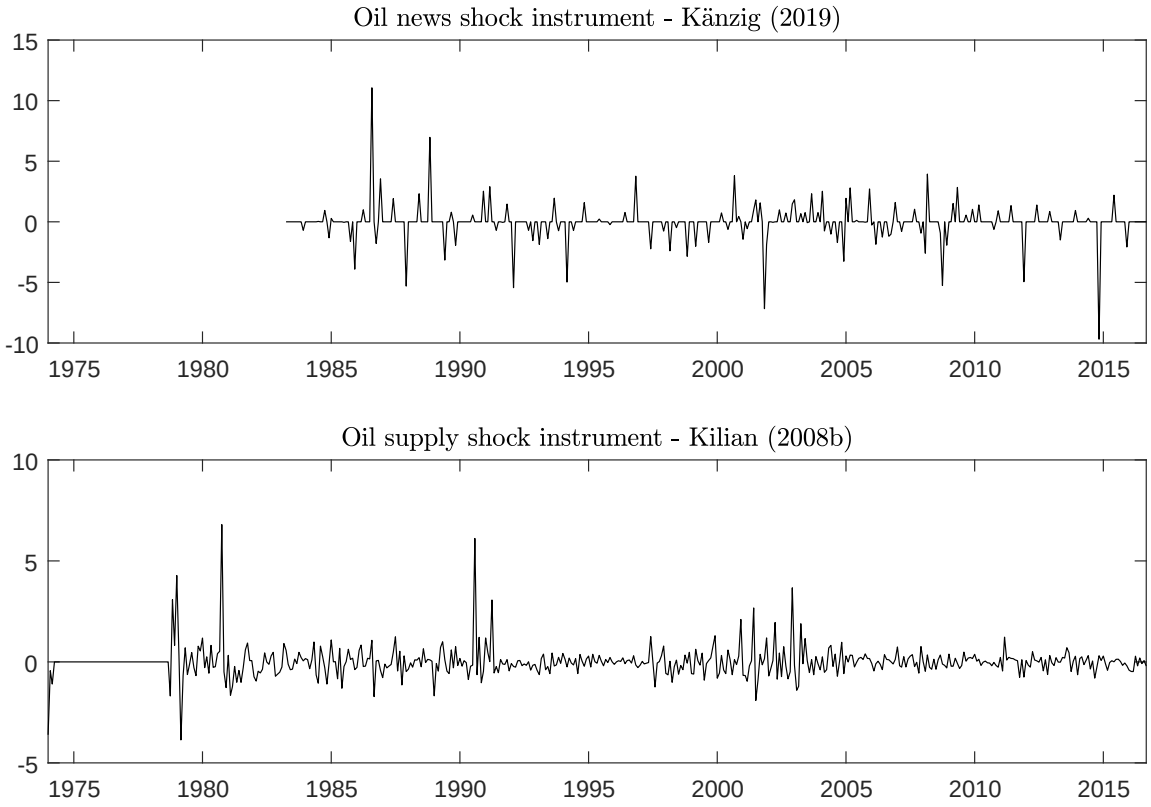
51	Housing Starts in South Census Region	5	FRED
52	Housing Starts in West Census Region	5	FRED
53	New Private Housing	5	FRED
54	New Private Housing: Northeast	5	FRED
55	New Private Housing: Midwest	5	FRED
56	New Private Housing: South	5	FRED
57	New Private Housing: West	5	FRED
	Consumption, orders and inventories		
1	Real personal consumption expenditures	5	FRED
8*	Real Manufacturing and Trade Industries Sales	5	FRED
9*	Retail Sales: Retail and Food Services, Total	5	FRED
58*	Manufacturers' New Orders: Durable Goods	5	FRED
59*	Manufacturers' New Orders: Nondefense Capital Goods	5	FRED
60*	Manufacturers' Unfilled Orders for Durable Goods	5	FRED
61*	Total Business Inventories	5	FRED
62*	Total Business: Inventories to Sales Ratio	2	FRED
119	ISM Purchasing Managers Index	5	Datastream
120	ISM Manufacturers Survey: New Orders Index	5	Datastream
121	ISM Manufacturers Survey: Employment Index	5	Datastream
122	ISM Manufacturers Survey: Inventories Index	5	Datastream
	Prices		
5	CPI: All Items	5	FRED
89	PPI: Finished Goods	6	FRED
90	PPI: Personal Consumption Goods	6	FRED
91	PPI: Processed Goods for Intermediate Demand	6	FRED
92	PPI: Unprocessed Goods for Intermediate Demand	6	FRED
93*	Spot Crude Oil Price: West Texas Intermediate	6	FRED
94	PPI: Primary nonferrous metals	6	FRED
95	CPI: Apparel	5	FRED
96	CPI: Transportation	5	FRED
97	CPI: Medical Care	5	FRED
98	CPI: Commodities	5	FRED
99	CPI: Durables	5	FRED
100	CPI: Services	5	FRED
101	CPI: All Items Less Food	5	FRED
102	CPI: All Items Less Shelter	5	FRED
103	CPI: All Items Less Medical Care	5	FRED
104	Real PCE: Energy goods and services	5	FRED
105	Real PCE: Durables	5	FRED
106	Real PCE: Non-Durables	5	FRED
107	Real PCE: Services	5	FRED
123	ISM Manufacturers Survey: Prices Paid Index	5	Datastream
	Money and credit		
63	M1 Money Stock	6	FRED
64	M2 Money Stock	6	FRED
65	Real M2 Money Stock	5	FRED
66	St. Louis Adjusted Monetary Base	6	FRED
67	Total Reserves of Depository Institutions	6	FRED
68	Commercial and Industrial Loans, All Commercial Banks	6	FRED
69	Real Estate Loans, All Commercial Banks	6	FRED
70	Total Nonrevolving Credit Owned/Securitized, Outstanding	6	FRED
71*	Nonrevolving Consumer Credit to Personal Income	2	FRED
111	MZM Money Stock	6	FRED
112	Consumer Motor Vehicle Loans Outstanding	6	FRED
113	Total Consumer Loans and Leases Outstanding	6	FRED
114	Securities In Bank Credit, All Commercial Banks	6	FRED
	Stock markets		

4*	S&P 500	5	FRED
72*	S&P 500: Industrials	5	FRED
73*	S&P 500: Dividend Yields	2	FRED
74*	S&P 500: Prices-Earnings ratio	5	FRED
115*	VXO	2	FRED
	Interest and exchange rates		
75	Effective Federal Funds Rate	2	FRED
76*	3-Month Commercial Paper Rate	2	FRED
77	3-Month Treasury Bill: Secondary Market Rate	2	FRED
78	6-Month Treasury Bill: Secondary Market Rate	2	FRED
79	1-Year Treasury Constant Maturity Rate	2	FRED
80	5-Year Treasury Constant Maturity Rate	2	FRED
81	10-Year Treasury Constant Maturity Rate	2	FRED
82	Moody's Seasoned AAA Corporate Bond Yield	2	FRED
83	Moody's Seasoned BAA Corporate Bond Yield	2	FRED
84	Trade Weighted U.S. Dollar Index: Major Currencies, Goods	5	FRED
85*	Switzerland / U.S. Foreign Exchange Rate	5	FRED
86*	Japan / U.S. Foreign Exchange Rate	5	FRED
87*	U.K. / U.S. Foreign Exchange Rate	5	FRED
88*	Canada / U.S. Foreign Exchange Rate	5	FRED
116	30-Year Fixed Rate Mortgage Average	2	FRED
124	Excess Bond Premium	2	Gilchrist and Zakrajšek (2012)
125	Switzerland / U.S. short-term interest rate spread	2	Forni and Gambetti (2010)
126	Japan / U.S. short-term interest rate spread	2	Forni and Gambetti (2010)
127	U.K. / U.S. short-term interest rate spread	2	Forni and Gambetti (2010)
128	Canada / U.S. short-term interest rate spread	2	Forni and Gambetti (2010)
129	Switzerland / U.S. Real Exchange Rate	5	Forni and Gambetti (2010)
130	Japan / U.S. Real Exchange Rate	5	Forni and Gambetti (2010)
131	U.K. / U.S. Real Exchange Rate	5	Forni and Gambetti (2010)
132	Canada / U.S. Real Exchange Rate	5	Forni and Gambetti (2010)

* Data not directly from FRED but manipulated or extended as described in McCracken and Ng (2016).
Transformation codes are: 1 - levels, 2 - 1st differences, 4 - log levels, 5 - log differences, 6 - 2nd log differences.

3.B.2 Time series plots

Figure 3.B.1: Proxy variables for oil supply shocks

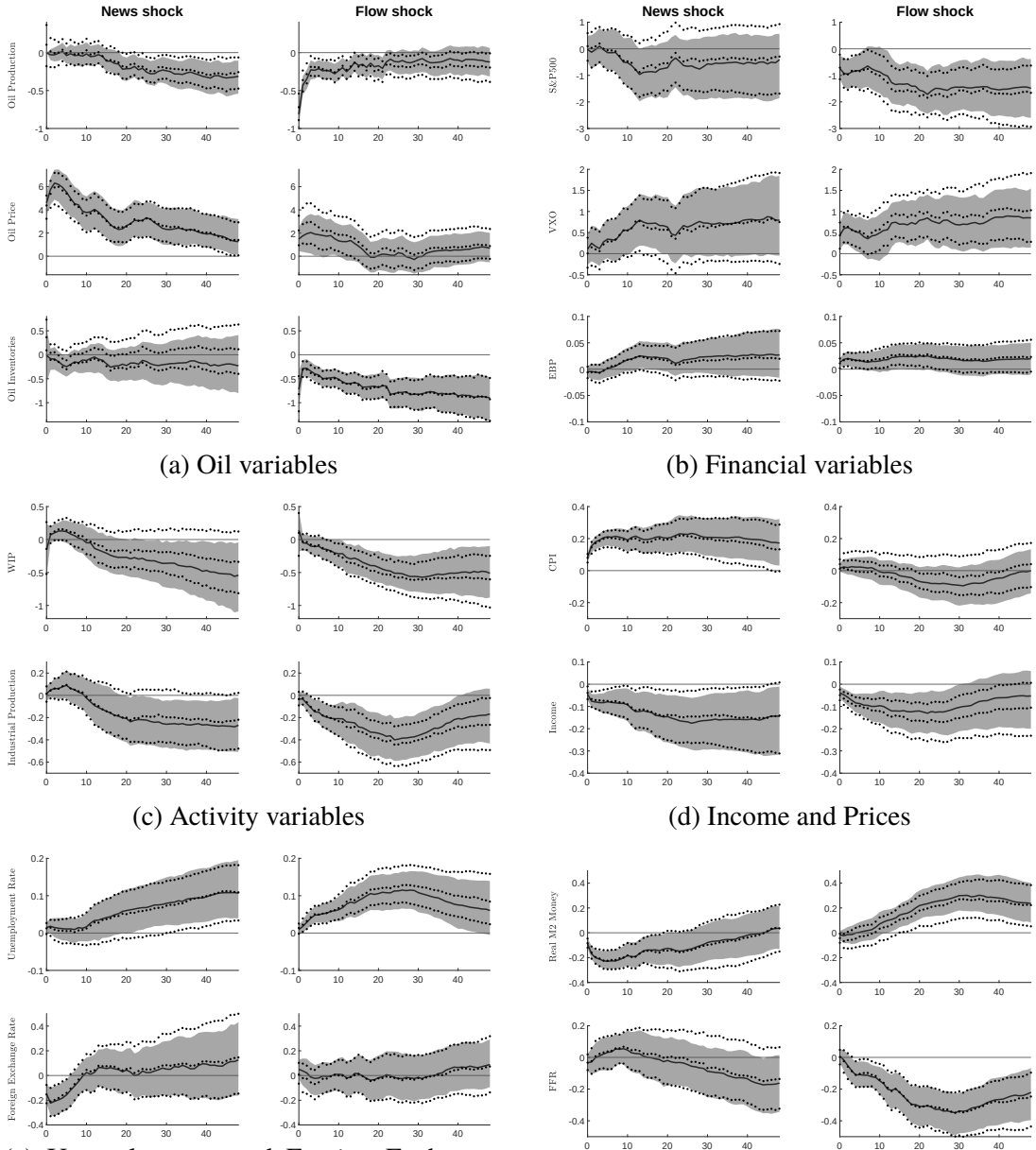


External instruments of Känzig (2019) and Kilian (2008a). Monthly version of Kilian (2008a) is taken from Braun and Brüggemann (2019).

Appendix 3.C Sensitivity analysis

3.C.1 Impulse response analysis

Figure 3.C.1: FAVAR impulse responses: sensitivity analysis



(a) Oil variables

(b) Financial variables

(c) Activity variables

(d) Income and Prices

(e) Unemployment and Foreign Exchange Rate

(f) Federal Reserve variables

Responses to an oil supply news shock (1st and 3rd column) and an oil supply flow shock (2nd and 4th column). Shocks are disentangled by a zero impact constraint on production for the news shock. Shaded areas are 68% posterior error bands. Dotted lines illustrate impulse response estimates from a Proxy-FAVAR model that identifies a news and a flow shock simultaneously (see Section 3.3).

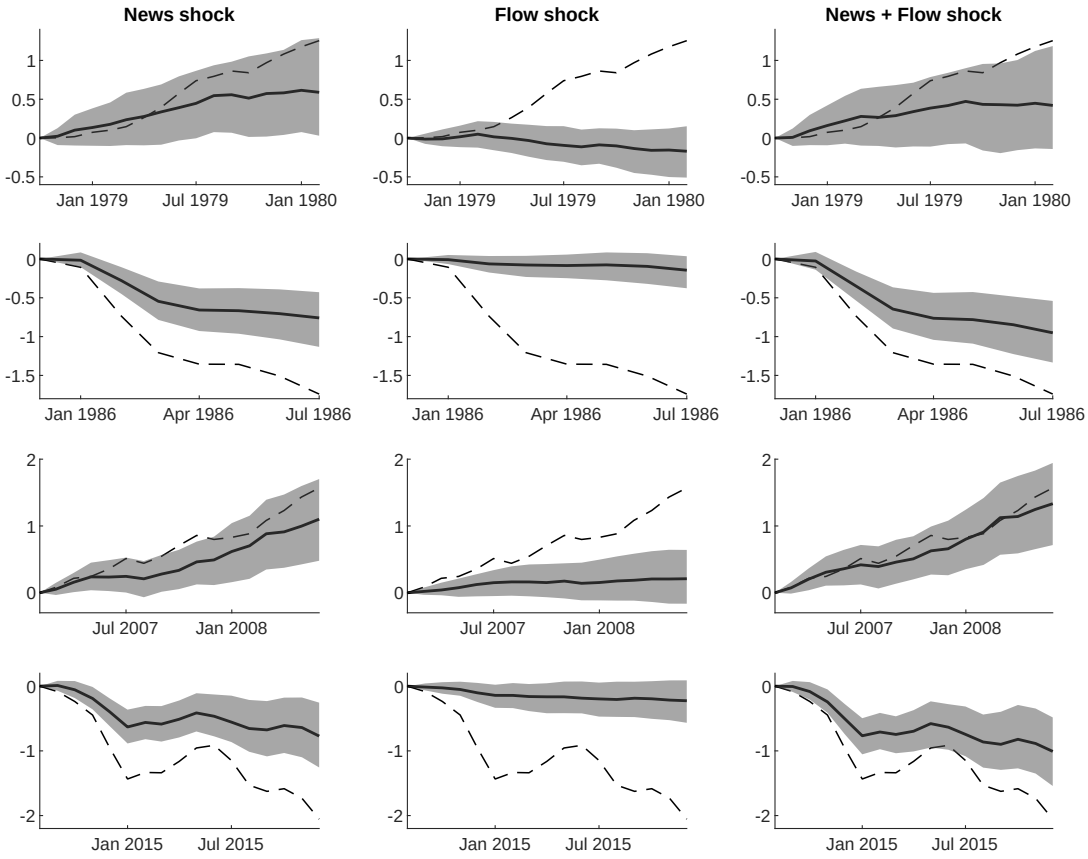
3.C.2 Variance and historical decompositions

Table 3.C.1: Forecast error variance decompositions: sensitivity analysis

	News shock			Flow shock		
	0	24	48	0	24	48
Oil production	0.00 [0.00,0.00]	0.15 [0.03,0.36]	0.18 [0.03,0.43]	0.45 [0.29,0.61]	0.06 [0.01,0.23]	0.05 [0.00,0.16]
Oil price	0.51 [0.31,0.70]	0.32 [0.10,0.54]	0.14 [0.01,0.41]	0.05 [0.01,0.14]	0.03 [0.00,0.13]	0.07 [0.01,0.24]
Oil inventories	0.03 [0.00,0.11]	0.04 [0.00,0.13]	0.04 [0.01,0.17]	0.09 [0.02,0.22]	0.30 [0.13,0.52]	0.19 [0.05,0.41]
S&P500	0.02 [0.00,0.08]	0.06 [0.00,0.21]	0.04 [0.00,0.20]	0.09 [0.01,0.22]	0.17 [0.03,0.34]	0.14 [0.02,0.33]
VXO	0.02 [0.00,0.09]	0.10 [0.01,0.33]	0.10 [0.01,0.32]	0.07 [0.01,0.20]	0.10 [0.01,0.28]	0.10 [0.01,0.27]
EBP	0.03 [0.00,0.12]	0.12 [0.01,0.37]	0.15 [0.02,0.42]	0.09 [0.01,0.23]	0.12 [0.01,0.34]	0.07 [0.01,0.27]
WIP	0.03 [0.00,0.11]	0.08 [0.01,0.27]	0.10 [0.01,0.32]	0.02 [0.00,0.07]	0.21 [0.05,0.41]	0.09 [0.01,0.24]
IP	0.03 [0.00,0.11]	0.08 [0.01,0.27]	0.12 [0.01,0.30]	0.02 [0.00,0.10]	0.22 [0.06,0.42]	0.06 [0.01,0.21]
CPI	0.32 [0.14,0.53]	0.23 [0.06,0.43]	0.13 [0.01,0.33]	0.02 [0.00,0.12]	0.04 [0.00,0.15]	0.03 [0.00,0.13]
RPI	0.15 [0.02,0.32]	0.17 [0.03,0.43]	0.14 [0.01,0.35]	0.10 [0.01,0.26]	0.11 [0.01,0.29]	0.05 [0.00,0.18]
Unemployment	0.06 [0.01,0.22]	0.08 [0.01,0.26]	0.17 [0.03,0.44]	0.03 [0.00,0.11]	0.23 [0.08,0.44]	0.06 [0.00,0.24]
FX	0.16 [0.04,0.37]	0.07 [0.01,0.27]	0.13 [0.01,0.38]	0.03 [0.00,0.12]	0.06 [0.01,0.24]	0.05 [0.00,0.21]
M2	0.22 [0.06,0.43]	0.06 [0.00,0.21]	0.04 [0.00,0.16]	0.02 [0.00,0.11]	0.18 [0.05,0.39]	0.16 [0.03,0.39]
FFR	0.04 [0.00,0.15]	0.03 [0.00,0.12]	0.09 [0.01,0.33]	0.02 [0.00,0.07]	0.29 [0.11,0.50]	0.16 [0.02,0.40]

Entries show fraction of the forecast error variance in variables in the first column that is explained by a news shock and a flow shock, respectively, at horizon $h = 0, 24, 48$. Shocks are disentangled by a zero impact constraint on production for the news shock. Posterior 68% credible intervals are displayed in parantheses.

Figure 3.C.2: Historical decompositions: sensitivity analysis

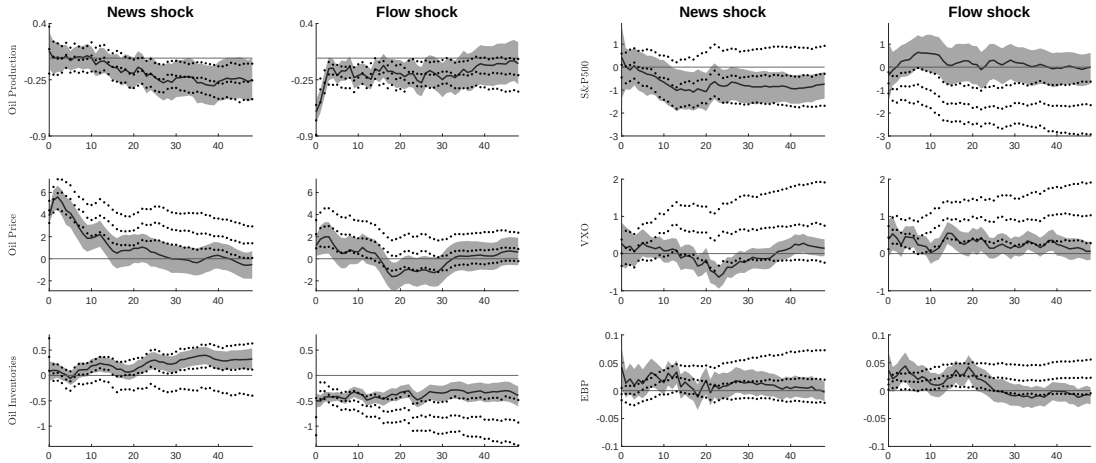


Historical decompositions of the real price of oil around key historical events for a news shock and a flow shock. Shaded areas are 68% posterior error bands. Dashed line is actual price of oil. Shocks are disentangled by a zero impact constraint on production for the news shock.

Appendix 3.D Additional results

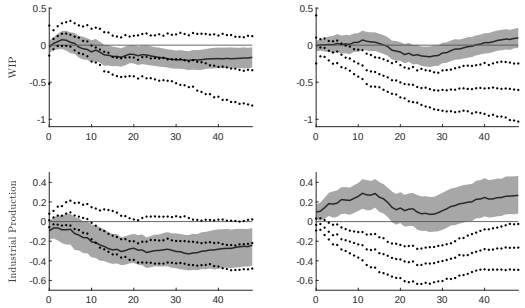
3.D.1 SVARs with one proxy

Figure 3.D.1: Impulse responses in SVARs with one proxy each

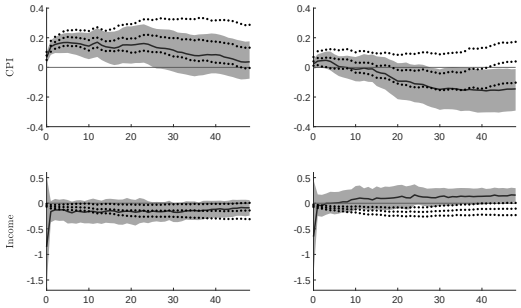


(a) Oil variables

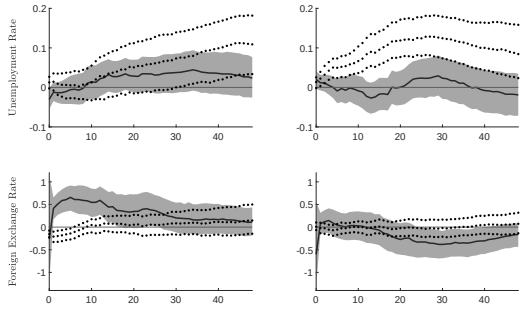
(b) Financial variables



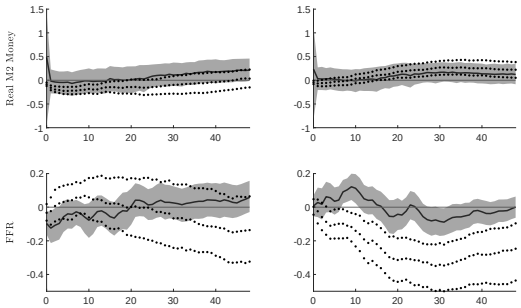
(c) Activity variables



(d) Income and Prices



(e) Unemployment and Foreign Exchange Rate



(f) Federal Reserve variables

Responses to an oil supply news shock (1st and 3rd column) and an oil supply flow shock (2nd and 4th column). Shocks are identified separately in SVAR models with only one proxy variable. Shaded areas are 68% posterior error bands. Dotted lines illustrate impulse response estimates from a Proxy-FAVAR model that identifies a news and a flow shock simultaneously (see Section 3.3).

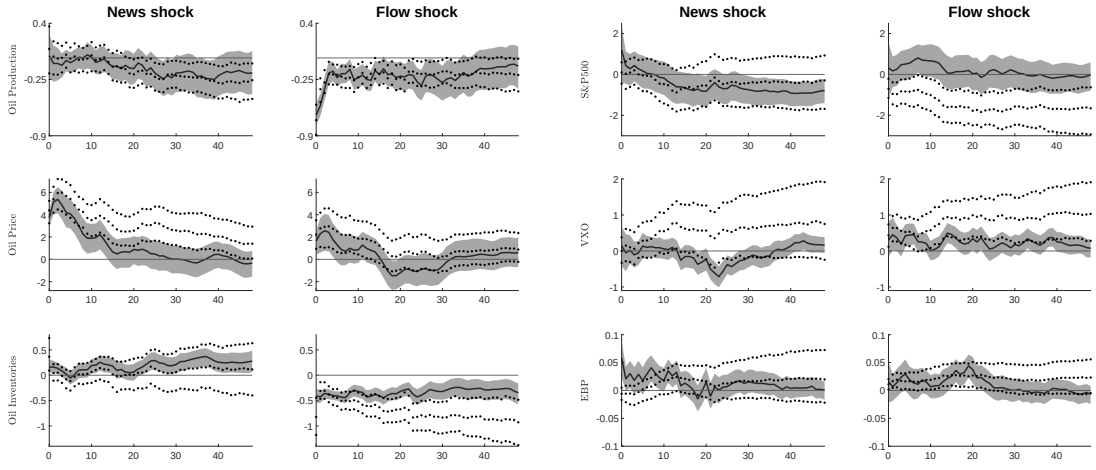
Table 3.D.1: Forecast error variance decompositions in SVARs with one proxy each

	News shock			Flow shock		
	0	24	48	0	24	48
Oil production	0.03 [0.00,0.11]	0.08 [0.01,0.28]	0.08 [0.01,0.28]	0.38 [0.23,0.53]	0.06 [0.01,0.21]	0.03 [0.00,0.14]
Oil price	0.69 [0.53,0.83]	0.25 [0.06,0.54]	0.11 [0.01,0.37]	0.05 [0.00,0.15]	0.07 [0.01,0.26]	0.05 [0.01,0.20]
Oil inventories	0.02 [0.00,0.10]	0.12 [0.01,0.35]	0.30 [0.08,0.55]	0.32 [0.18,0.48]	0.42 [0.19,0.65]	0.21 [0.05,0.43]
S&P500	0.04 [0.00,0.12]	0.12 [0.01,0.35]	0.14 [0.02,0.40]	0.02 [0.00,0.08]	0.06 [0.01,0.20]	0.05 [0.00,0.23]
VXO	0.02 [0.00,0.09]	0.28 [0.06,0.55]	0.09 [0.01,0.34]	0.02 [0.00,0.09]	0.04 [0.00,0.18]	0.08 [0.01,0.27]
EBP	0.05 [0.01,0.16]	0.07 [0.01,0.24]	0.08 [0.01,0.34]	0.04 [0.01,0.13]	0.05 [0.01,0.21]	0.09 [0.01,0.33]
WIP	0.02 [0.00,0.09]	0.09 [0.01,0.32]	0.09 [0.01,0.35]	0.02 [0.00,0.06]	0.05 [0.00,0.17]	0.07 [0.01,0.27]
IP	0.03 [0.00,0.12]	0.14 [0.02,0.39]	0.09 [0.01,0.32]	0.07 [0.01,0.18]	0.09 [0.01,0.29]	0.12 [0.01,0.36]
CPI	0.12 [0.01,0.31]	0.08 [0.01,0.26]	0.05 [0.00,0.19]	0.02 [0.00,0.08]	0.04 [0.01,0.16]	0.09 [0.01,0.29]
RPI	0.05 [0.01,0.13]	0.10 [0.01,0.37]	0.10 [0.01,0.36]	0.03 [0.00,0.10]	0.09 [0.01,0.32]	0.18 [0.02,0.45]
Unemployment	0.05 [0.01,0.17]	0.05 [0.00,0.18]	0.04 [0.00,0.17]	0.03 [0.00,0.09]	0.04 [0.00,0.15]	0.04 [0.00,0.18]
FX	0.04 [0.01,0.11]	0.11 [0.01,0.33]	0.07 [0.01,0.27]	0.03 [0.00,0.09]	0.09 [0.01,0.28]	0.06 [0.01,0.24]
M2	0.04 [0.01,0.11]	0.05 [0.01,0.23]	0.08 [0.01,0.27]	0.02 [0.00,0.08]	0.05 [0.00,0.21]	0.04 [0.00,0.17]
FFR	0.08 [0.01,0.21]	0.05 [0.00,0.22]	0.17 [0.02,0.48]	0.02 [0.00,0.06]	0.05 [0.00,0.19]	0.05 [0.01,0.21]

Entries show fraction of the forecast error variance in variables in the first column that is explained by a news shock and a flow shock, respectively, at horizon $h = 0, 24, 48$. Posterior 68% credible intervals are displayed in parantheses.

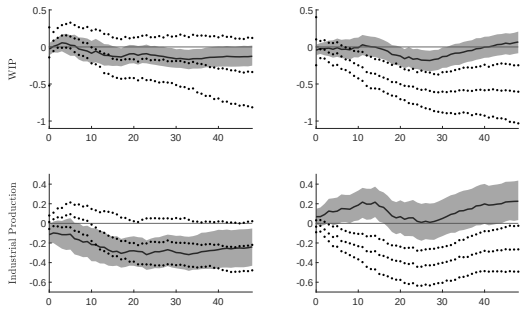
3.D.2 SVAR with two proxies

Figure 3.D.2: Impulse responses in SVAR with two proxies

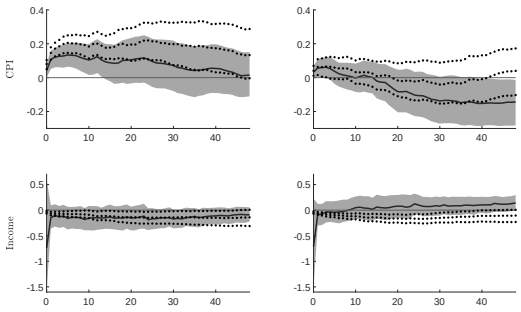


(a) Oil variables

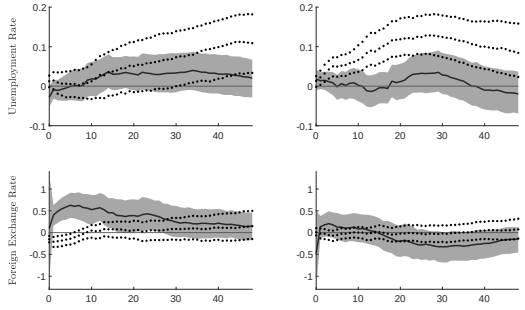
(b) Financial variables



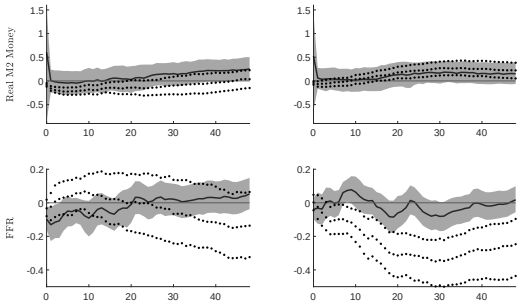
(c) Activity variables



(d) Income and Prices



(e) Unemployment and Foreign Exchange Rate



(f) Federal Reserve variables

Responses to an oil supply news shock (1st and 3rd column) and an oil supply flow shock (2nd and 4th column). Shocks are identified jointly in SVAR model with two proxy variables. Shaded areas are 68% posterior error bands. Dotted lines illustrate impulse response estimates from a Proxy-FAVAR model that identifies a news and a flow shock simultaneously (see Section 3.3).

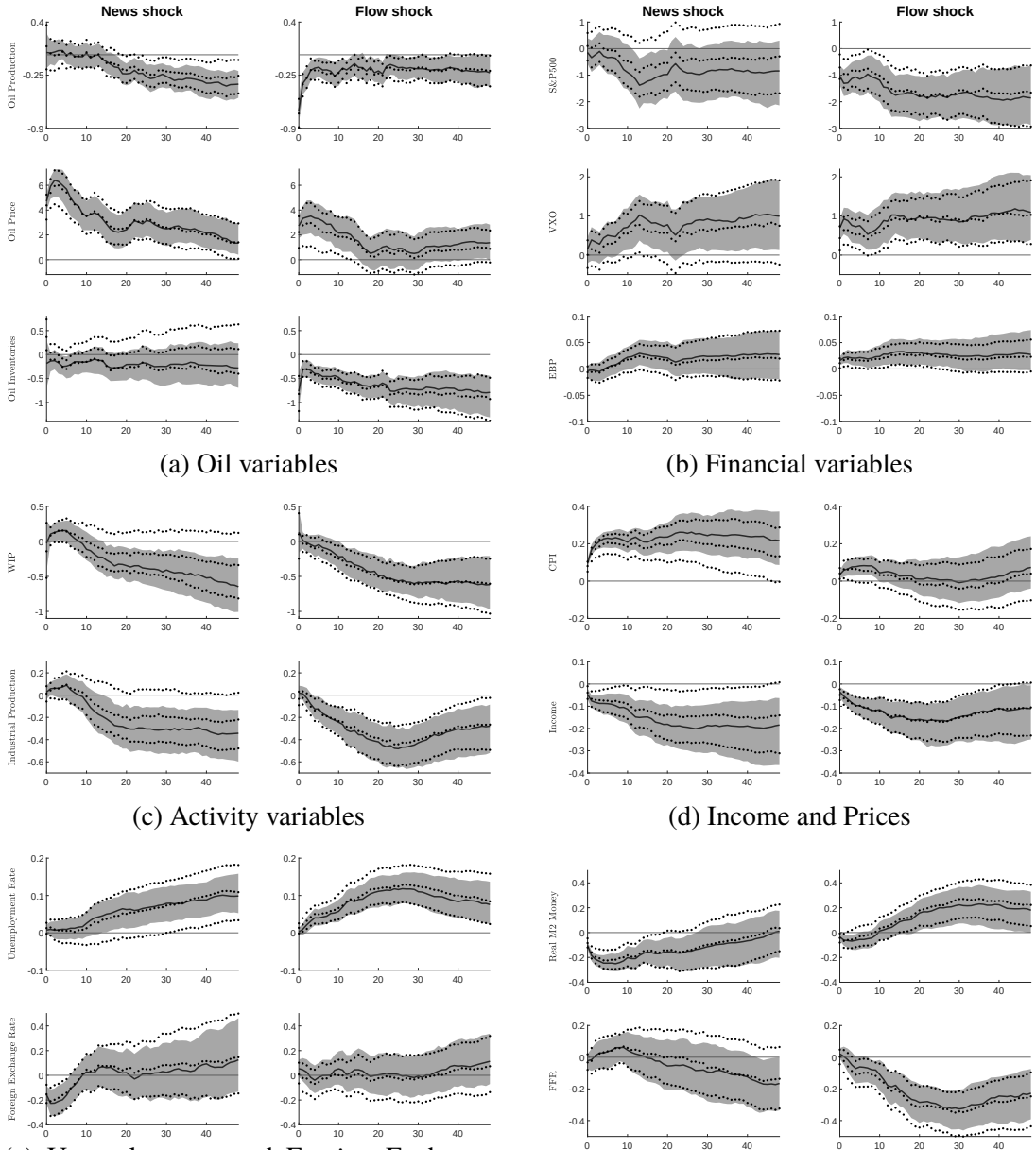
Table 3.D.2: Forecast error variance decompositions in SVAR with two proxies

	News shock			Flow shock		
	0	24	48	0	24	48
Oil production	0.03 [0.00,0.11]	0.04 [0.00,0.19]	0.05 [0.01,0.20]	0.41 [0.25,0.58]	0.15 [0.03,0.35]	0.04 [0.00,0.16]
Oil price	0.51 [0.33,0.70]	0.05 [0.00,0.18]	0.04 [0.00,0.19]	0.10 [0.01,0.25]	0.08 [0.01,0.27]	0.06 [0.01,0.21]
Oil inventories	0.03 [0.00,0.11]	0.16 [0.02,0.38]	0.12 [0.02,0.34]	0.19 [0.07,0.33]	0.38 [0.17,0.60]	0.20 [0.04,0.47]
S&P500	0.05 [0.01,0.16]	0.09 [0.01,0.28]	0.18 [0.02,0.45]	0.03 [0.00,0.08]	0.05 [0.00,0.20]	0.06 [0.00,0.21]
VXO	0.02 [0.00,0.07]	0.30 [0.08,0.53]	0.10 [0.01,0.33]	0.02 [0.00,0.08]	0.05 [0.00,0.20]	0.09 [0.01,0.30]
EBP	0.09 [0.02,0.21]	0.05 [0.01,0.23]	0.09 [0.01,0.27]	0.02 [0.00,0.08]	0.07 [0.01,0.26]	0.09 [0.01,0.36]
WIP	0.02 [0.00,0.08]	0.08 [0.01,0.26]	0.09 [0.01,0.29]	0.02 [0.00,0.07]	0.13 [0.02,0.38]	0.05 [0.01,0.22]
IP	0.06 [0.01,0.18]	0.23 [0.05,0.45]	0.11 [0.02,0.34]	0.03 [0.00,0.10]	0.04 [0.00,0.15]	0.11 [0.01,0.33]
CPI	0.06 [0.01,0.21]	0.05 [0.00,0.19]	0.04 [0.00,0.17]	0.04 [0.00,0.13]	0.04 [0.00,0.17]	0.13 [0.01,0.37]
RPI	0.04 [0.01,0.12]	0.12 [0.01,0.35]	0.09 [0.01,0.33]	0.03 [0.00,0.11]	0.07 [0.01,0.28]	0.15 [0.02,0.46]
Unemployment	0.05 [0.01,0.16]	0.04 [0.00,0.17]	0.04 [0.00,0.16]	0.03 [0.00,0.11]	0.04 [0.00,0.18]	0.04 [0.01,0.17]
FX	0.04 [0.00,0.12]	0.12 [0.02,0.36]	0.07 [0.01,0.26]	0.02 [0.00,0.09]	0.07 [0.01,0.24]	0.06 [0.01,0.25]
M2	0.04 [0.01,0.12]	0.06 [0.01,0.22]	0.10 [0.01,0.33]	0.03 [0.00,0.10]	0.07 [0.01,0.25]	0.05 [0.01,0.21]
FFR	0.08 [0.01,0.21]	0.06 [0.01,0.23]	0.15 [0.02,0.46]	0.02 [0.00,0.09]	0.05 [0.00,0.18]	0.07 [0.01,0.28]

Entries show fraction of the forecast error variance in variables in the first column that is explained by a news shock and a flow shock, respectively, at horizon $h = 0, 24, 48$. Posterior 68% credible intervals are displayed in parantheses.

3.D.3 FAVARs with one proxy

Figure 3.D.3: Impulse responses in FAVARs with one proxy each



(a) Oil variables

(b) Financial variables

(c) Activity variables

(d) Income and Prices

(e) Unemployment and Foreign Exchange Rate

(f) Federal Reserve variables

Responses to an oil supply news shock (1st and 3rd column) and an oil supply flow shock (2nd and 4th column). Shocks are identified separately in FAVAR models with only one proxy variable. Shaded areas are 68% posterior error bands. Dotted lines illustrate impulse response estimates from a Proxy-FAVAR model that identifies a news and a flow shock simultaneously (see Section 3.3).

Table 3.D.3: Forecast error variance decompositions in FAVARs with one proxy each

	News shock			Flow shock		
	0	24	48	0	24	48
Oil production	0.01 [0.00,0.06]	0.19 [0.02,0.39]	0.23 [0.07,0.43]	0.31 [0.17,0.48]	0.06 [0.01,0.25]	0.09 [0.01,0.26]
Oil price	0.53 [0.38,0.68]	0.33 [0.15,0.57]	0.18 [0.04,0.38]	0.18 [0.07,0.30]	0.05 [0.00,0.18]	0.14 [0.02,0.38]
Oil inventories	0.02 [0.00,0.09]	0.03 [0.00,0.14]	0.04 [0.00,0.13]	0.10 [0.02,0.23]	0.27 [0.11,0.44]	0.17 [0.05,0.36]
S&P500	0.02 [0.00,0.07]	0.07 [0.01,0.25]	0.07 [0.01,0.23]	0.17 [0.07,0.32]	0.22 [0.08,0.44]	0.19 [0.04,0.41]
VXO	0.03 [0.00,0.10]	0.13 [0.01,0.36]	0.16 [0.02,0.36]	0.18 [0.06,0.30]	0.19 [0.03,0.41]	0.17 [0.03,0.42]
EBP	0.03 [0.00,0.16]	0.14 [0.02,0.43]	0.18 [0.03,0.43]	0.13 [0.03,0.30]	0.19 [0.03,0.40]	0.12 [0.01,0.37]
WIP	0.02 [0.00,0.09]	0.12 [0.02,0.32]	0.14 [0.03,0.30]	0.02 [0.00,0.08]	0.25 [0.08,0.44]	0.13 [0.02,0.29]
IP	0.03 [0.00,0.11]	0.13 [0.02,0.31]	0.17 [0.04,0.41]	0.02 [0.00,0.08]	0.27 [0.12,0.49]	0.11 [0.01,0.29]
CPI	0.42 [0.23,0.60]	0.30 [0.12,0.51]	0.16 [0.04,0.42]	0.05 [0.01,0.15]	0.02 [0.00,0.07]	0.04 [0.00,0.18]
RPI	0.16 [0.03,0.35]	0.28 [0.07,0.52]	0.20 [0.04,0.46]	0.10 [0.02,0.24]	0.19 [0.06,0.42]	0.09 [0.01,0.26]
Unemployment	0.03 [0.00,0.13]	0.13 [0.02,0.32]	0.22 [0.06,0.49]	0.01 [0.00,0.06]	0.30 [0.13,0.48]	0.14 [0.02,0.33]
FX	0.17 [0.06,0.34]	0.06 [0.01,0.26]	0.10 [0.01,0.36]	0.02 [0.00,0.11]	0.05 [0.00,0.17]	0.06 [0.01,0.26]
M2	0.30 [0.16,0.48]	0.08 [0.01,0.26]	0.05 [0.00,0.21]	0.04 [0.00,0.12]	0.09 [0.01,0.24]	0.10 [0.01,0.24]
FFR	0.03 [0.00,0.13]	0.04 [0.00,0.14]	0.09 [0.01,0.33]	0.02 [0.00,0.09]	0.24 [0.11,0.46]	0.18 [0.03,0.41]

Entries show fraction of the forecast error variance in variables in the first column that is explained by a news shock and a flow shock, respectively, at horizon $h = 0, 24, 48$. Posterior 68% credible intervals are displayed in parantheses.

Complete References

Alquist, R. and L. Kilian (2010). "What do we learn from the price of crude oil futures?" *Journal of Applied Econometrics* 25 (4), 539–573.

Antolín-Díaz, J., I. Petrella, and J. F. Rubio-Ramírez (2018). "Structural scenario analysis with SVARs". *CEPR Discussion Paper No. DP12579*.

Antolín-Díaz, J. and J. F. Rubio-Ramírez (2018). "Narrative sign restrictions for SVARs". *American Economic Review* 108 (10), 2802–29.

Arias, J. E., J. F. Rubio-Ramírez, and D. F. Waggoner (2018). "Inference based on structural vector autoregressions identified with sign and zero restrictions: Theory and applications". *Econometrica* 86 (2), 685–720.

– (2019). "Inference in Bayesian Proxy-SVARs". *FRB Atlanta Working Paper*.

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

Banbura, M., D. Giannone, and L. Reichlin (2010). "Large Bayesian vector autoregressions". *Journal of Applied Econometrics* 25, 71–92.

Barsky, R. B. and L. Kilian (2002). "Do we really know that oil caused the great stagflation? A monetary alternative". *NBER Macroeconomics Annual* 16, 137–183.

Baumeister, C. and J. D. Hamilton (2019). "Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks". *American Economic Review* 109 (5), 1873–1910.

Baumeister, C. and L. Kilian (2016). "Understanding the decline in the price of oil since June 2014". *Journal of the Association of Environmental and Resource Economists* 3 (1), 131–158.

Baumeister, C. and G. Peersman (2013). "The role of time-varying price elasticities in accounting for volatility changes in the crude oil market". *Journal of Applied Econometrics* 28 (7), 1087–1109.

Beaudry, P. and F. Portier (2014). "News-driven business cycles: Insights and challenges". *Journal of Economic Literature* 52 (4), 993–1074.

Belviso, F. and F. Milani (2006). "Structural factor-augmented VARs (SFAVARs) and the effects of monetary policy". *Topics in Macroeconomics* 6 (3).

Bernanke, B. S. and I. Mihov (1998). "Measuring monetary policy". *Quarterly Journal of Economics* 113 (3), 869–902.

Bernanke, B. S. (1986). "Alternative explanations of the money-income correlation". In: *Carnegie-Rochester Conference Series on Public Policy*. Vol. 25. Elsevier, 49–99.

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.

Bernanke, B. S., M. Gertler, and M. Watson (1997). "Systematic monetary policy and the effects of oil price shocks". *Brookings Papers on Economic Activity* 1, 91–157.

Bjørnland, H. C. and K. Leitemo (2009). "Identifying the interdependence between US monetary policy and the stock market". *Journal of Monetary Economics* 56 (2), 275–282.

Blanchard, O. J. and D. Quah (1989). "The dynamic effects of aggregate demand and supply disturbances". *The American Economic Review* 79 (4), 655–673.

Braun, R. (2019). "Three essays on identification in structural vector autoregressive models". PhD thesis, University of Konstanz.

Braun, R. and R. Brüggemann (2019). "Identification of SVAR models by combining sign restrictions with external instruments". *Working Paper, University of Konstanz*.

Brillinger, D. R. (1996). "Remarks concerning graphical models for time series and point processes". *Revista de Econometria* 16, 1–23.

Brüggemann, R., C. Jentsch, and C. Trenkler (2016). "Inference in VARs with conditional heteroskedasticity of unknown form". *Journal of Econometrics* 191 (1), 69–85.

Bruns, M. (2018). "Combining factors models and external instruments to identify uncertainty shocks". *Unpublished working paper*.

Caffo, B. S., W. Jank, and G. L. Jones (2005). "Ascent-based Monte Carlo expectation-maximization". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67 (2), 235–251.

Caldara, D., M. Cavallo, and M. Iacoviello (2019). "Oil price elasticities and oil price fluctuations". *Journal of Monetary Economics* 103, 1–20.

Caldara, D. and E. Herbst (2019). "Monetary policy, real activity, and credit spreads: Evidence from Bayesian Proxy SVARs". *American Economic Journal: Macroeconomics* 11 (1), 157–92.

Canova, F. and G. De Nicoló (2002). "Monetary disturbances matter for business fluctuations in the G-7". *Journal of Monetary Economics* 49, 1131–1159.

Cappé, O., E. Moulines, and T. Ryden (2005). *Inference in Hidden Markov Models*. Springer.

Carriero, A., T. E. Clark, and M. Marcellino (2015). "Bayesian VARs: Specification choices and forecast accuracy". *Journal of Applied Econometrics* 30 (1), 46–73.

– (2019). "The identifying information in vector autoregressions with time-varying volatilities: An application to endogenous uncertainty". *Working Paper*.

Carriero, A., G. Kapetanios, and M. Marcellino (2009). "Forecasting exchange rates with a large Bayesian VAR". *International Journal of Forecasting* 25 (2), 400–417.

– (2012). "Forecasting government bond yields with large Bayesian vector autoregressions". *Journal of Banking and Finance* 36 (7), 2026–2047.

Carter, C. K. and R. Kohn (1994). "On Gibbs sampling for state space models". *Biometrika* 81 (3), 541–553.

Chan, J. C. C. (2017). "The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling". *Journal of Business & Economic Statistics* 35 (1), 17–28.

Chan, J. C. C. and A. L. Grant (2016). "On the observed-data deviance information criterion for volatility modeling". *Journal of Financial Econometrics* 14 (4), 772–802.

Chan, J. C. C. and I. Jeliazkov (2009). "Efficient simulation and integrated likelihood estimation in state space models". *International Journal of Mathematical Modelling and Numerical Optimisation* 1 (1-2), 101–120.

Cheng, X. and B. E. Hansen (2015). "Forecasting with factor-augmented regression: A frequentist model averaging approach". *Journal of Econometrics* 186 (2), 280–293.

Chib, S. and E. Greenberg (1995). "Understanding the Metropolis–Hastings algorithm". *The American Statistician* 49 (4), 327–335.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999). "Monetary policy shocks: What have we learned and to what end?" In: *Handbook of Macroeconomics*. Ed. by Taylor, J. B. and Woodford, M. Vol. 1. Handbook of Macroeconomics. Elsevier. Chap. 2, 65–148.

– (2005). "Nominal rigidities and the dynamic effects of a shock to monetary policy". *Journal of Political Economy* 113 (1), 1–45.

Clements, M. P. (2016). "Real-time factor model forecasting and the effects of instability". *Computational Statistics and Data Analysis* 100, 661–675.

Coibion, O. (2012). "Are the effects of monetary policy shocks big or small?" *American Economic Journal: Macroeconomics* 4 (2), 1–32.

Cormen, T. H., C. E. Leiserson, R. L. Rivest, and C. Stein (2009). "Introduction to Algorithms". In: The MIT Press. Chap. 22.

Dahlhaus, R. (2000). "Graphical interaction models for multivariate time series". *Metrika* 51 (2), 157–172.

Dahlhaus, R. and M. Eichler (2003). "Causality and graphical models for time series". In: *Highly structured stochastic systems*. Ed. by Green, P., Hjort, N., and Richardson, S. Oxford University Press, 115–137.

Danielsson, J. and J. F. Richard (1993). "Accelerated Gaussian importance sampler with application to dynamic latent variable models". *Journal of Applied Econometrics* 8, 153–173.

Demiralp, S. and K. D. Hoover (2003). "Searching for the causal structure of a vector autoregression". *Oxford Bulletin of Economics and Statistics* 65, 745–767.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). "Maximum likelihood from incomplete data via the EM algorithm". *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 39, 1–38.

Doan, T. A. and R. Todd (2010). "Causal ordering for multivariate linear systems". *Unpublished Work, Estima*.

Duff, I. S. and J. Reid (1978). "An implementation of Tarjan's algorithm for the block triangularization of a matrix". *ACM Transactions on Mathematical Software* 4 (2), 137–147.

Dufour, J. M., D. Pelletier, and E. Renault (2006). "Short run and long run causality in time series: Inference". *Journal of Econometrics* 132 (2), 337–362.

Dufour, J. M. and E. Renault (1998). "Short run and long run causality in time series: Theory". *Econometrica* 66 (5), 1099–1125.

Durbin, J. and S. J. Koopman (1997). "Monte Carlo maximum likelihood estimation for non-Gaussian state space models". *Biometrika* 84 (3), 669–684.

– (2000). "Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 62 (1), 3–56.

Edwards, D. (2000). *Introduction to Graphical Modelling*. 2nd ed. Springer Texts in Statistics. New York: Springer-Verlag.

Eichler, M. (2006). "Graphical modeling of dynamic relationships in multivariate time series". In: *Handbook of Time Series Analysis: Recent Developments and Applications*. Ed. by Schelter, B., Winterhalder, M., and Timmer, J. Wiley.

– (2007). "Granger causality and path diagrams for multivariate time series". *Journal of Econometrics* 137 (2), 334–353.

– (2012). "Graphical modelling of multivariate time series". *Probability Theory and Related Fields* 153 (1-2), 233–268.

Eickmeier, S. and C. Ziegler (2008). "How successful are dynamic factor models at forecasting output and inflation? A meta-analytic approach". *Journal of Forecasting* 27 (3), 237–265.

Faust, J. (1998). "The robustness of identified VAR conclusions about money". *Carnegie-Rochester Conference Series on Public Policy* 49, 207–244.

Faust, J., E. T. Swanson, and J. H. Wright (2004). "Identifying VARs based on high frequency futures data". *Journal of Monetary Economics* 51 (6), 1107–1131.

Fernández-Villaverde, J. and J. F. Rubio-Ramírez (2007). "Estimating macroeconomic models: A likelihood approach". *Review of Economic Studies* 74 (4), 1059–1087.

Flamm, C., U. Kalliauer, M. Deistler, M. Waser, and A. Graef (2012). "Graphs for dependence and causality in multivariate time series". In: *System Identification, Environmental Modelling, and Control System Design*. Ed. by Wang, L. and Garnier, H. London: Springer London, 133–151.

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.

Gertler, M. and P. Karadi (2015). "Monetary policy surprises, credit costs, and economic activity". *American Economic Journal-Macroeconomics* 7 (1), 44–76.

Geweke, J. (1992). "Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments". *Bayesian statistics* 4, 641–649.

Giacomini, R., T. Kitagawa, and M. Read (2019). "Robust Bayesian inference in Proxy SVARs". *Cemmap Working Paper 23*.

Giannone, D., M. Lenza, D. Momferatou, and L. Onorante (2014). "Short-term inflation projections: A Bayesian vector autoregressive approach". *International Journal of Forecasting* 30 (3), 635–644.

Giannone, D., L. Reichlin, and L. Sala (2004). "Monetary policy in real time". *NBER Macroeconomics Annual* 19, 161–200.

Gilchrist, S. and E. Zakrajšek (2012). "Credit spreads and business cycle fluctuations". *American Economic Review* 102 (4), 1692–1720.

Gourieroux, C., A. Monfort, and J. P. Renne (2017). "Statistical inference for independent component analysis: Application to structural VAR models". *Journal of Econometrics* 196 (1), 111–126.

Granger, C. W. J. (1969). "Investigating causal relations by econometric models and cross-spectral methods". *Econometrica* 37, 424–438.

Hamilton, J. D. (2009). "Causes and consequences of the oil shock of 2007-08". *Brookings Papers on Economic Activity* 40 (1), 215–283.

Hamilton, J. D. and A. Herrera (2004). "Oil shocks and aggregate macroeconomic behavior: The role of monetary policy." *Journal of Money, Credit, and Banking* 36 (2), 265.

Harvey, A., E. Ruiz, and N. Shephard (1994). "Multivariate stochastic variance models". *Review of Economic Studies* 61 (2), 247–264.

Heinlein, R. and H. M. Krolzig (2012). "Effects of monetary policy on the US dollar/UK pound exchange rate. Is there a 'delayed overshooting puzzle'?" *Review of International Economics* 20 (3), 443–467.

Herrera, A. M., M. B. Karaki, and S. K. Rangaraju (2019). "Oil price shocks and US economic activity". *Energy Policy* 129, 89–99.

Herwartz, H. and H. Lütkepohl (2014). "Structural vector autoregressions with Markov switching: Combining conventional with statistical identification of shocks". *Journal of Econometrics* 183 (1), 104–116.

Hoover, K., S. Demiralp, and S. J. Perez (2009). "Empirical identification of the vector autoregression: The causes and effects of U.S. M2". In: *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*. Ed. by Castle, J. and Shepard, N. Oxford: Oxford University Press, 37–58.

Jacquier, E., N. G. Polson, and P. E. Rossi (1994). "Bayesian analysis of stochastic volatility models". *Journal of Business & Economic Statistics* 12 (4), 371–389.

Jarociński, M and B. Maćkowiak (2017). "Granger causal priority and choice of variables in vector autoregressions". *Review of Economics and Statistics* 99 (2), 319–329.

Justiniano, A. and G. E. Primiceri (2008). "The time-varying volatility of macroeconomic fluctuations". *American Economic Review* 98 (3), 604–641.

Juvenal, L. and I. Petrella (2015). "Speculation in the oil market". *Journal of Applied Econometrics* 30 (4), 621–649.

Känzig, D. R. (2019). "The macroeconomic effects of oil supply shocks: New evidence from OPEC announcements". *Working Paper*.

Kascha, C. and C. Trenkler (2015). "Forecasting VARs, model selection, and shrinkage". *Working Paper ECON* (15-07).

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. (2008a). "Exogenous oil supply shocks: How big are they and how much do they matter for the U.S. economy?" *The Review of Economics and Statistics* 90 (2), 216–240.

– (2008b). "The economic effects of energy price shocks". *Journal of Economic Literature* 46 (4), 871–909.

– (2009). "Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market". *American Economic Review* 99 (3), 1053–1069.

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

Kilian, L. and H. Lütkepohl (2017). *Structural Vector Autoregressive Analysis*. Themes in Modern Econometrics. Cambridge University Press.

Kilian, L. and D. P. Murphy (2012). "Why agnostic sign restrictions are not enough: Understanding the dynamics of oil market VAR models". *Journal of the European Economic Association* 10 (5), 1166–1188.

– (2014). "The role of inventories and speculative trading in the global market for crude oil". *Journal of Applied Econometrics* 29 (3), 454–478.

Kim, S., N. Shephard, and S. Chib (1998). "Stochastic volatility: Likelihood inference and comparison with ARCH models". *Review of Economic Studies* 65 (3), 361–393.

Koch, K. (2007). "Gibbs sampler by sampling-importance-resampling". *Journal of Geodesy* 81 (9), 581–591.

Koop, G. M. (2013). "Forecasting with medium and large Bayesian VARs". *Journal of Applied Econometrics* 28 (2), 177–203.

Koop, G. and D. Korobilis (2010). "Bayesian multivariate time series methods for empirical macroeconomics". *Foundations and Trends in Econometrics* 3 (4), 267–358.

Koopman, S. J., N. Shephard, and D. Creal (2009). "Testing the assumptions behind importance sampling". *Journal of Econometrics* 149 (1), 2–11.

Lakdawala, A. (2019). "Decomposing the effects of monetary policy using an external instruments SVAR". *Journal of Applied Econometrics* 34 (6), 934–950.

Lanne, M., H. Lütkepohl, and K. Maciejowska (2010). "Structural vector autoregressions with Markov switching". *Journal of Economic Dynamics & Control* 34 (2), 121–131.

Lanne, M., M. Meitz, and P. Saikkonen (2017). "Identification and estimation of non-Gaussian structural vector autoregressions". *Journal of Econometrics* 196 (2), 288–304.

Lanne, M. and P. Saikkonen (2007). "A multivariate generalized orthogonal factor GARCH model". *Journal of Business & Economic Statistics* 25 (1), 61–75.

Lauritzen, S. L. (1996). *Graphical Models*. Oxford Statistical Science Series. Oxford: Oxford University Press.

Lewis, D. J. (2018). "Identifying shocks via time-varying volatility". *Federal Reserve Bank of New York Staff Report* (871).

Lippi, F. and A. Nobili (2012). "Oil and the macroeconomy: A quantitative structural analysis". *Journal of the European Economic Association* 10 (5), 1059–1083.

Louis, T. A. (1982). "Finding the observed information matrix when using the EM algorithm". *Journal of the Royal Statistical Society Series B-Methodological* 44 (2), 226–233.

Ludvigson, S. C. and S. Ng (2007). "The empirical risk-return relation: A factor analysis approach". *Journal of Financial Economics* 83 (1), 171–222.

– (2009). "Macro factors in bond risk premia". *Review of Financial Studies* 22 (12), 5027–5067.

Lunsford, K. G. (2015). "Identifying structural VARs with a proxy variable and a test for a weak proxy". *FRB of Cleveland Working Paper 15-28*.

Lütkepohl, H. and A. Netšunajev (2017a). "Structural vector autoregressions with smooth transition in variances". *Journal of Economic Dynamics & Control* 84, 43–57.

Lütkepohl, H. (1993). "Testing for causation between two variables in higher dimensional VAR models". In: *Studies in Applied Econometrics*. Ed. by Schneeweiß, H. and Zimmermann, K. F. Springer-Verlag, Heidelberg, 75–91.

– (2005). *New Introduction to Multiple Time Series Analysis*. Springer Science & Business Media.

Lütkepohl, H. and G. Milunovich (2016). "Testing for identification in SVAR-GARCH models". *Journal of Economic Dynamics and Control* 73, 241–258.

Lütkepohl, H. and A. Netšunajev (2017b). "Structural vector autoregressions with heteroskedasticity: A review of different volatility models". *Econometrics and Statistics* 1, 2–18.

Lütkepohl, H. and T. Schlaak (2018). "Choosing between different time-varying volatility models for structural vector autoregressive analysis". *Oxford Bulletin of Economics and Statistics* (forthcoming).

Magnus, J. R. and H. Neudecker (2019). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons.

Mahieu, R. J. and P. C. Schotman (1998). "An empirical application of stochastic volatility models". *Journal of Applied Econometrics* 13 (4), 333–359.

Maxand, S. (2018). "Identification of independent structural shocks in the presence of multiple Gaussian components". *Econometrics and Statistics*.

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

McLachlan, G. and T. Krishnan (2007). *The EM Algorithm and Extensions*. Vol. 382. John Wiley & Sons.

Melino, A. and S. M. Turnbull (1990). "Pricing foreign-currency options with stochastic volatility". *Journal of Econometrics* 45 (1-2), 239–265.

Mertens, K. and M. O. Ravn (2012). "Empirical evidence on the aggregate effects of anticipated and unanticipated US tax policy shocks". *American Economic Journal-Economic Policy* 4 (2), 145–181.

– (2013). "The dynamic effects of personal and corporate income tax changes in the United States". *American Economic Review* 103 (4), 1212–1247.

Mertens, K. and J. L. Montiel Olea (2018). "Marginal tax rates and income: New time series evidence". *The Quarterly Journal of Economics* 133 (4), 1803–1884.

Montiel-Olea José, L, J. H. Stock, and M. W. Watson (2016). "Uniform inference in SVARs identified with external instruments". *Harvard Manuscript*.

Nakamura, E. and J. Steinsson (2018). "High frequency identification of monetary non-neutrality: The information effect". *Quarterly Journal of Economics* (forthcoming).

Neal, R. M. and G. E. Hinton (1998). "A view of the EM algorithm that justifies incremental, sparse, and other variants". In: *Learning in graphical models*. Springer, 355–368.

Normandin, M. and L. Phaneuf (2004). "Monetary policy shocks: Testing identification conditions under time-varying conditional volatility". *Journal of Monetary Economics* 51 (6), 1217–1243.

Onatski, A. (2010). "Determining the number of factors from empirical distribution of eigenvalues". *The Review of Economics and Statistics* 92 (4), 1004–1016.

Pearl, J. (2000). *Causality*. New York: Cambridge University Press.

Penm, J. and R. Terrell (1986). "The 'derived' moving-average model and its role in causality". In: *Essays in time series and allied processes : papers in honour of E. J. Hannan*. Ed. by Gani, J. and Priestly, M. Sheffield: Applied Probability Trust, 99–111.

Piffer, M. and M. Podstawski (2017). "Identifying uncertainty shocks using the price of gold". *The Economic Journal* 128 (616), 3266–3284.

Poirier, D. J. (1998). "Revising beliefs in nonidentified models". *Econometric Theory* 14 (4), 483–509.

Primiceri, G. E. (2005). "Time varying structural vector autoregressions and monetary policy". *Review of Economic Studies* 72 (3), 821–852.

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

Rigobon, R. (2003). "Identification through heteroskedasticity". *Review of Economics and Statistics* 85 (4), 777–792.

Romer, C. D. and D. H. Romer (2004). "A new measure of monetary shocks: Derivation and implications". *American Economic Review* 94 (4), 1055–1084.

Roweis, S. and Z. Ghahramani (2001). "Learning nonlinear dynamical systems using the Expectation–Maximization algorithm". *Kalman filtering and neural networks* 6, 175–220.

Rubio-Ramírez, J. F., D. F. Waggoner, and T. Zha (2010). "Structural vector autoregressions: Theory of identification and algorithms for inference". *The Review of Economic Studies* 77 (2), 665–696.

Rue, H., S. Martino, and N. Chopin (2009). "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations". *Journal of the Royal Statistical Society Series B-Statistical Methodology* 71 (2), 319–392.

Ruiz, E. (1994). "Quasi-maximum likelihood estimation of stochastic volatility models". *Journal of Econometrics* 63 (1), 289–306.

Sims, C. A. (1980). "Macroeconomics and reality". *Econometrica* 48 (1), 1–48.

– (1982). "Policy analysis with econometric-models". *Brookings Papers on Economic Activity* (1), 107–164.

– (2015). "Causal orderings and exogeneity". *Lecture Notes*.

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

– (2003). "Forecasting output and inflation: The role of asset prices". *Journal of Economic Literature* 41 (3), 788–829.

– (2012a). "Disentangling the channels of the 2007-09 recession". *Brookings Papers on Economic Activity* 43 (1), 81–156.

– (2012b). "Generalized shrinkage methods for forecasting using many predictors". *Journal of Business and Economic Statistics* 30 (4), 481–493.

– (2016). "Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics". In: *Handbook of Macroeconomics*. Ed. by Taylor, J. B. and Uhlig, H. Vol. 2. Elsevier. Chap. 8, 415–525.

Swanson, N. R. and C. W. J. Granger (1997). "Impulse response functions based on a causal approach to residual orthogonalization in vector autoregressions". *Journal of the American Statistical Association* 92 (437), 357–367.

Tarjan, R. E. (1972). "Depth-first search and linear graph algorithms". *SIAM Journal on Computing* 1 (2), 146–160.

Tierney, L. (1994). "Markov chains for exploring posterior distributions". *The Annals of Statistics* 22 (4), 1701–1728.

Uhlig, H. (2005). "What are the effects of monetary policy on output? Results from an agnostic identification procedure". *Journal of Monetary Economics* 52, 381–419.

– (2009). "Comment on 'How has the Euro changed the monetary transmission mechanism?'" In: *NBER Macroeconomics Annual 2008*. Ed. by Acemoglu, D., Rogoff, K., and Woodford, M. Vol. 23. University of Chicago Press, 141–152.

Van der Weide, R. (2002). "GO-GARCH: A multivariate generalized orthogonal GARCH model". *Journal of Applied Econometrics* 17 (5), 549–564.

Waggoner, D. F. and T. Zha (2003). "A Gibbs sampler for structural vector autoregressions". *Journal of Economic Dynamics and Control* 28 (2), 349–366.

Wieland, J. F. and M.-J. Yang (2016). "Financial dampening". *NBER Working Paper 22141*.

Wright, J. H. (2012). "What does monetary policy do to long-term interest rates at the zero lower bound?" *The Economic Journal* 122 (564), 447–466.

Yamamoto, Y. (2019). "Bootstrap inference for impulse response functions in factor-augmented vector autoregressions". *Journal of Applied Econometrics* 34 (2), 247–267.

Eigenabgrenzung

Das erste Kapitel, *Identification of structural vector autoregressions by stochastic volatility*, ist in Zusammenarbeit mit Robin Braun, der zu dieser Zeit ebenfalls Promotionsstudent an der Universität Konstanz war, entstanden. Meine individuelle Leistung bei der Erstellung dieses Kapitels beträgt 50%.

Das zweite Kapitel, *Directed graphs and variable selection in large vector autoregressive models*, habe ich gemeinsam mit Prof. Dr. Ralf Brüggemann (Universität Konstanz) und Dr. Christian Kascha (damals Universität Zürich, heute S Rating und Risikosysteme GmbH) verfasst. Meine individuelle Leistung bei der Erstellung dieses Kapitels beträgt 40%.

Das dritte Kapitel, *The effects of oil supply shocks on the macroeconomy: a Proxy-FAVAR approach*, habe ich ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel erstellt.