Topics in Structural Vector Autoregressive Modelling

Dissertation

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

vorgelegt von
Tilmann Richard Härtl

Sektion Politik – Recht – Wirtschaft
Fachbereich Wirtschaftswissenschaften
Konstanz, 2024

Dissertation der Universität Konstanz
Tag der mündlichen Prüfung: 22. November 2024

1. Referent: Prof. Dr. Ralf Brüggemann
2. Referent: Prof. Dr. Carsten Trenkler
3. Referent: Prof. Dr. Stefan Niemann

Danksagung

Zunächst möchte ich mich bei Prof. Dr. Ralf Brüggemann für die Betreuung meiner Promotion bedanken. Mein besonderer Dank gilt dem entgegengebrachten Vertrauen, meine Forschungsgebiete selbst wählen und meine Ideen eigenständig entwickeln zu dürfen. Gleichzeitig bedanke ich mich für die zahlreichen hilfreichen Kommentare zu meiner Arbeit, die mir oft die Richtung gewiesen haben, in die ich die Projekte weiterentwickeln kann. Auch möchte ich mich für die finanzielle Unterstützung bedanken. Dank gilt in diesem Zusammenhang auch der Deutschen Forschungsgemeinschaft, deren Mittel einen Teil meiner Promotion finanziert haben.

Des Weiteren bedanke ich mich bei Prof. Dr. Carsten Trenkler und Prof. Dr. Stefan Niemann für die Übernahme des Zweit- und Drittgutachtens. Ebenso gilt mein Dank Prof. Dr. Winfried Pohlmeier für die interessanten Kommentare in verschiedenen Seminaren. Bedanken möchte ich mich ebenfalls bei meinem Ko-Autor Sascha Keweloh. Besonderer Dank gilt auch Verena Kretz, die mir nie eine Frage nicht beantworten konnte wenn ich Hilfe brauchte. Gleiches gilt für Jutta Obenland von der GSDS.

Außerdem gibt es viele Kolleg:innen und Freund:innen an der Universität, die dafür gesorgt haben, dass ich meine Zeit an der Universität Konstanz - und darüber hinaus - in guter Erinnerung behalten werde. Mein Dank gilt Julie und Mark für die schönen gemeinsamen Erlebnisse. Julie danke ich auch insbesondere dafür, dass sie mir am Anfang gezeigt hat, wie so eine Promotion läuft, was meinen Start sehr vereinfacht hat. Auch bei den anderen direkten Lehrstuhlkollegen, die immer eine sehr angenehme Arbeitsatmosphäre geschaffen haben, möchte ich mich bedanken: bei Maurizio für die Abende auf verschiedenen Konferenzen und bei Linus und Benedikt für die gemeinsamen Aktivitäten in Konstanz. Darüber hinaus gibt es viele andere Kolleg:innen/Freund:innen von F3 oder anderen Fluren der Universität, mit denen ich eine gute Zeit hatte. Um der Gefahr aus dem Weg zu gehen jemanden vergessen

zu haben, nenne ich keine weiteren Namen aber jeder, der sich jetzt angesprochen fühlt, tut dies zu Recht.

Zudem möchte ich Dominik danken, mit dem ich nur eine kurze gemeinsame Zeit an der Universität verbracht habe. Er hat mich aber in den Tennisclub eingeführt, in dem ich ebenfalls Freunde gefunden habe, mit denen ich während meiner Promotion sehr viel Freude hatte. Das gleiche gilt für alle meine Mitbewohner in Heidi's guter Stube. Ich hätte mir keinen schöneren Ort zum Wohnen vorstellen können.

Außerdem möchte ich meiner Familie danken, die mich immer unterstützt hat, meinen Weg zu gehen und immer ein toller Rückhalt war. Ich bedanke mich auch bei meinen Freunden aus der alten Heimat, mit denen ich schöne Urlaube verbracht habe, die eine schöne Abwechslung vom Arbeitsalltag waren. Der letzte Dank geht an Eva, die mich seit der zweiten Hälfte meiner Promotion begleitet. Die vielen schönen gemeinsamen Stunden und Erlebnisse haben es mir sehr leicht gemacht, abzuschalten und den Stress der Promotion hinter mir zu lassen. Auch Eva war ein toller Rückhalt und ich bin ihr sehr dankbar für ihre Unterstützung.

Table of Contents

Summary **1**

Zusammenfassung **5**

1 Identifying Proxy VARs with Restrictions on the FEV **10**

1.1 Introduction 11

1.2 Econometric Framework 14

1.2.1 The Structural VAR 14

1.2.2 The Proxy VAR 17

1.3 Proxy VAR with Bounds on the FEV 20

1.3.1 Bounding the Contribution to the FEV 20

1.3.2 Nonemptiness of the Set 21

1.3.3 Estimation and Inference 23

1.4 Proxy VAR and Max-Share 26

1.4.1 Ruling out Confounders via the Proxy VAR 27

1.4.2 Tackling the Max-Share Bias 28

1.4.3 Estimation and Inference 33

1.5 Empirical Illustrations 33

1.5.1 Max-Share 37

1.5.2 Bounds on the FEV 42

1.6 Conclusion 45

References 47

Appendix 1.A Proofs and Remarks 50

Appendix 1.B Simulation Results 59
Appendix 1.C Additional Empirical Results 74

2 Uncertainty Shocks: Insights from Synthetic Proxy Variables 78

2.1 Introduction 79
2.2 The Proxy VAR 83
2.3 The Synthetic Proxy VAR Estimator 85
2.4 Finite Sample Performance 88
2.5 Identifying Uncertainty Shocks 91
2.5.1 Data and Model Specification 93
2.5.2 The Proxy for Uncertainty Shocks 94
2.5.3 Labelling the Shocks 97
2.5.4 Impulse Responses 101
2.5.5 Forecast Error Variance Decompositions 105
2.6 Conclusion 106
References 108
Appendix 2.A Proofs 110
Appendix 2.B Additional Empirical Results 112

3 Learning Sets Accounting for Model and Estimation Uncertainty 127

3.1 Introduction 128
3.2 Econometric Framework 133
3.2.1 Learning the Union of Parameter Regions 133
3.2.2 Set-identified Structural VARs 137
3.2.3 Interpretation of the Learning Procedure 143
3.3 Simulation Study 149
3.3.1 Impulse Response Functions 152
3.3.2 Forecast Error Variance Decompositions 158

3.4 Empirical Illustrations 162
3.4.1 Comparison with Conventional Inference 163
3.4.2 Identifying an Uncertainty Shock 167
3.5 Conclusion 170
References 171
Appendix 3.A Proofs and Remarks 173
Appendix 3.B Additional Simulation and Empirical Results 176

Complete References **180**

Eigenabgrenzung **185**

List of Tables

1.B.1 Two Shocks Simulated with Scenario A and B 60
1.B.2 Max-Share with an Additional Inequality Constraint - Scenario B . . 63
1.B.3 Sufficient Condition for Bias Reduction - Scenario B 67
1.B.4 Bias of the IRFs, $T = 1000$ - Scenario B 68
1.B.5 Bias of the IRFs to the Technology Shock, $T = 250$ - Scenario B . . 69

2.1 Correlation between the Proxy and Uncertainty Measurements . . . 96
2.2 Testing Relations of the Proxies to Specific Variables 98
2.3 FEV Explained by the Financial Uncertainty Shock 105
2.4 FEV Explained by the Real Uncertainty Shock 106

3.1 Joint Coverages of the Augmented Learning Framework - IRFs . . . 157
3.2 Joint Coverages of the Augmented Learning Framework - FEVs . . . 162

List of Figures

1.1 Impulse Responses to the two MP Shocks - Max-Share 41
1.2 Impulse Responses to the two MP Shocks - Bound=0.05 44
1.A.1 Constrained Maximization of $\Omega_{1,1}(h)$ 54
1.A.2 Linear Constraint on q_{11} 55
1.A.3 Linear Constraint on q_{11} without Monotonicity 57
1.B.4 95% Point Estimate Bands, $\epsilon = 0.39$ - Scenario B 64
1.B.5 95% Point Estimate Bands, $\epsilon = 0$ - Scenario B 65
1.B.6 95% Point Estimate Bands, $\epsilon = 0.2$ - Scenario B 66
1.B.7 Bias of the IRFs to the Technology Shock, $T = 1000$ - Scenario B . . 70
1.B.8 Bias of the IRFs to the Monetary Shock, $T = 1000$ - Scenario B . . . 71
1.B.9 Bias of the IRFs to the Technology Shock, $T = 250$ - Scenario B . . 72
1.B.10 Bias of the IRFs to the Monetary Shock, $T = 250$ - Scenario B . . . 73
1.C.11 Data used in Section 1.5 74
1.C.12 Impulse Responses with the 10-year Treasury Yield 75
1.C.13 Impulse Responses with the 5-year Treasury Yield and $H = 4$ 76
1.C.14 Impulse Responses to the two MP Shocks - Bound=0.1 77

2.1 Finite Sample Performance using z_t^2 and z_t^3 90
2.2 Finite Sample Performance using only z_t^2 91
2.3 Different Measurements of Uncertainty 95
2.4 Labelling of the Uncertainty Shocks 99
2.5 Impulse Responses to the Financial Uncertainty Shock 102
2.6 Impulse Responses to the Real Uncertainty Shock 103
2.B.1 Data Used in the Baseline Analysis 112
2.B.2 Impulse Responses to the Financial Uncertainty Shock using only z_t^2 as a Synthetic Proxy 113
2.B.3 Impulse Responses to the Real Uncertainty Shock using only z_t^2 as a Synthetic Proxy 114

2.B.4 Impulse Responses to the Financial Uncertainty Shock using only z_t^3 as a Synthetic Proxy . . . 115
2.B.5 Impulse Responses to the Real Uncertainty Shock using only z_t^3 as a Synthetic Proxy . . . 116
2.B.6 Impulse Responses to the Financial Uncertainty Shock using the Lower Quartile Uncertainty Shock by Piffer and Podstawski (2018) . . . 117
2.B.7 Impulse Responses to the Real Uncertainty Shock using the Lower Quartile Uncertainty Shock by Piffer and Podstawski (2018) . . . 118
2.B.8 Impulse Responses to the Financial Uncertainty Shock using the Upper Quartile Uncertainty Shock by Piffer and Podstawski (2018) . . . 119
2.B.9 Impulse Responses to the Real Uncertainty Shock using the Upper Quartile Uncertainty Shock by Piffer and Podstawski (2018) . . . 120
2.B.10 Impulse Responses to the Financial Uncertainty Shock using Employment instead of Hours Worked . . . 121
2.B.11 Impulse Responses to the Real Uncertainty Shock using Employment instead of Hours Worked . . . 122
2.B.12 Impulse Responses to the Financial Uncertainty Shock with $p = 6$ . . . 123
2.B.13 Impulse Responses to the Real Uncertainty Shock with $p = 6$ . . . 124
2.B.14 Impulse Responses to the Financial Uncertainty Shock with $p = 18$ . . . 125
2.B.15 Impulse Responses to the Real Uncertainty Shock with $p = 18$ . . . 126

3.1 Learning the Tightest Band . . . 134
3.2 Learning Scenarios in the Limit . . . 145
3.3 Iso-Draw Curves for Different Numbers of M . . . 147
3.4 Coverages for the IRFs for H=25 . . . 153
3.5 Coverages for the IRFs for H=13 . . . 154
3.6 Coverages for the Baseline Learning Algorithm . . . 155
3.7 Coverages for the FEV decomposition for H=25 . . . 159
3.8 Coverages for the FEV decomposition for H=13 . . . 160
3.9 Comparing Learned to Frequentist and Bayesian Quantities . . . 164
3.10 Comparing Learned Bands to Joint Inference . . . 165
3.11 ECDFs for the FFR Responses . . . 166
3.12 Impulse Responses to an Uncertainty and a News Shock . . . 168
3.B.1 Coverages with 500 Residual-Based Moving Block Bootstrap Iterations 177
3.B.2 ECDFs for the GDP Responses . . . 178
3.B.3 ECDFs for the CPI Responses . . . 179

Summary

This dissertation concerns with different topics in structural vector autoregressions (VARs). These structural VARs, introduced by Sims (1980), aim to identify dynamic causal effects and are most often used to identify the effects of macroeconomic shocks in order to analyse their role in business cycle fluctuations. An important application is, for example, to assess how monetary policy shocks affect the economy. This knowledge is essential for policymakers such that they can design policies to steer the economy complaisantly. Thus, policymakers are the primary practitioners of structural VARs in order to analyse and keep track of economic developments.

The quantities of interest are, firstly, the structural impulse response functions (IRFs), which give the dynamic response of the included variables to the identified structural shocks. Secondly, the forecast error variance (FEV) decomposition tells the practitioner how much a certain structural shock contributes to the FEV of the variables of interest. Hence, the FEV decomposition uncovers how much a certain shock helps to explain the variation in the included variables. In a similar fashion, the historical decomposition quantifies how much a certain shock contributed to the fluctuation of the variable during certain time periods.

The key challenge in structural VARs is the identification of these structural parameters. This requires to impose restrictions on the model. Restrictions can either be imposed directly on the structural parameters of interest, yet, these restrictions need to be backed by economic theory or strong economic intuition. Depending on the rigidity of such imposed restrictions, one may not obtain unique identification of the parameters of interest but instead obtain a set of admissible structural models. This is referred to as set-identification. Another approach is an agnostic identification of the model based on its statistical properties. In doing so, one avoids the need to impose restrictions on the parameters one wants to learn something about, yet this agnostic identification requires ex post labelling of the structural shocks, which makes

the interpretation of the results less clear. Finally, also exogenous information can be used to identify the structural VAR through proxy variables, which are similar to instruments in the microeconometric literature.

The first two chapters of this dissertation focus on the identification of the structural parameters of interest. Both chapters innovate how to combine two identification strategies in order to complement their strengths while mitigating their respective weaknesses. The third chapter deviates and investigates the downstream problem after the identification: the inference on the identified parameters. In this chapter, I provide a new and practitioner-friendly tool for uncertainty quantification in set-identified structural VARs. The three chapters are independent research papers that resulted from my doctoral studies. The first and the third chapter were written entirely by myself, while the second chapter is joint work with Sascha Keweloh from the TU Dortmund. Hence, the author is referred to as "we" in the second chapter.

The first two chapters of this dissertation address identification via proxy variables. Proxies are a powerful tool as they allow the identification of the structural parameters of interest without the need to impose any structure on the quantities the researcher wants to learn something about. However, the proxy VAR has a significant limitation. As shown by Mertens and Ravn (2013), the identification of multiple shocks with multiple proxies requires additional restrictions, in addition to the exogenous information contained in the proxies. Once more, candidates are restrictions on the structural parameters themselves as well as identification by statistical means. In the first chapter, I propose to combine the proxy VAR with restrictions on the FEV decomposition that allow the identification of the structural IRFs, which are often the key parameters of interest, without the need for strict, or even any restrictions at all, on these response parameters. The second chapter investigates how proxy VARs can be identified statistically via non-Gaussian shocks and independence criteria. This preserves the agnostic nature of the proxy VAR, while at the same time, the information from the proxies softens the assumptions on the statistical properties of the underlying model that are necessary for the statistical identification.

In the first chapter, I propose to employ restrictions on the FEV decomposition in the proxy VAR. Two types of restrictions are considered. Firstly, building on the work of Volpicella (2022), less restrictive bounds are placed on the contributions to the FEV. This leads to set-identification of the structural parameters or helps to sharpen the identification when used in conjunction with other set-identifying restrictions, e.g. sign restrictions on the IRFs. Secondly, imposing more strict assumptions on the FEV decomposition, namely assuming that one shock maximizes the contribution to

the FEV of a target variable, leads to point-identification of the structural parameters. Yet, this so-called Max-Share framework is prone to bias concerns and I show how the proxy VAR helps to address these concerns. I propose an augmentation to the Max-Share framework with an inequality constraint and I show theoretically under which conditions this constraint removes or reduces the bias in the proxy Max-Share framework. Yet, this analysis is limited to the case of two shocks being identified with two proxies, which is a highly relevant case in practice. As a result, point-identification is achieved without the need for strict equality restrictions on structural IRFs. An empirical illustration identifies an unconventional and a conventional monetary policy shock for the United States in the pre-zero lower bound era by employing both types of FEV restrictions separately. The findings suggest that both types of monetary policy shocks have detrimental effects on the economy.

The second chapter proposes a novel identification strategy for structural VARs that blends proxy variables with statistical identification based on non-Gaussian shocks. The combination makes use of the information in the proxy variables in order to weaken the assumptions on the dependence structure of the structural shocks. Furthermore, our methodology allows us to identify multiple shocks with a single proxy variable through the use of synthetic proxies, which are nonlinear transformations of the single proxy variable. With our methodology we are, to the best of our knowledge, the first ones to identify two different uncertainty shocks for the United States in the proxy VAR, namely a financial and a real uncertainty shock. The combination of the proxy VAR and agnostic statistical identification addresses two main challenges in the identification of uncertainty shocks. The proxy VAR addresses reverse causality issues through the exogenous variation in the proxy variable, while the statistical identification avoids the need to impose restrictions on a priori unclear structural parameters. We find that financial uncertainty is largely driven by the financial uncertainty shock, which evokes only minor effects outside of financial markets. Real uncertainty is to a large part driven by non-uncertainty shocks, but at the same time, real uncertainty shocks cause an immediate and significant decrease of real economic activity.

In the third chapter, I address a different topic in structural VARs: the inference on identified structural parameters. Inference is especially challenging if the structural parameters are not identified sharply in the first place. In set-identified models, inference is difficult because one does not have a single point to construct e.g. confidence intervals around. The existing frequentist or Bayesian methods in the literature are often limited to certain identification schemes or computationally

costly. The third chapter proposes a general and computationally feasible alternative based on learning guarantees from the statistical learning literature. These learning guarantees quantify the uncertainty involved in the identification of the parameters and take the place of the typical frequentist or Bayesian interpretation. Based on the work of Montiel Olea and Nesbit (2021), I show how one can use these learning guarantees to achieve joint uncertainty quantification for the structural parameters in set-identified structural VARs. I augment their learning algorithm with a bootstrap and I show that the resulting learned bands have a clear asymptotic interpretation under certain assumptions. Furthermore, I investigate the small sample properties of the augmented learning algorithm in a simulation study. In comparison to existing methods of inference in the literature, the augmented learning algorithm is advantageous due to its low computational costs, its general applicability and its joint uncertainty quantification. At the same time, the learned bands achieve nominal coverage levels comparable to those of their frequentist or Bayesian competitors. Yet, the resulting learned bands do not admit the usual frequentist or Bayesian interpretation, which makes the interpretation more difficult. Finally, I illustrate the learned bands in two empirical applications. A stylized setting compares the bands to existing methodologies to conduct inference. The results suggest that the width of the learned bands is largely comparable to other frequentist intervals or Bayesian credible regions. The replication of a proper empirical study shows that the bands are also informative in a real empirical setting.

References

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), 1212–47.

Montiel Olea, J. L. and J. Nesbit (2021). "(Machine) Learning Parameter Regions". *Journal of Econometrics* 222. (1), 716–744.

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

Volpicella, A. (2022). "SVARs Identification Through Bounds on the Forecast Error Variance". *Journal of Business & Economic Statistics* 40. (3), 1291–1301.

Zusammenfassung

Politische Entscheidungsträger benötigen Kenntnisse über Faktoren, welche die Konjunkturzyklen antreiben oder beeinflussen, um fundierte Entscheidungen treffen zu können. In der empirischen Makroökonomie werden diese Faktoren häufig mit Hilfe von strukturellen vektorautoregressiven (VAR) Modellen identifiziert, die auf Sims (1980) zurückgehen. Diese Modelle ermöglichen die Schätzung dynamischer kausaler Effekte sogenannter struktureller Schocks. Diese beschreiben, wie die Wirtschaft auf bestimmte Umstände reagiert bzw. wie sie mittels politischer Instrumente gesteuert werden kann.

Strukturelle VARs liefern einerseits Impuls-Antworten, die die dynamische Reaktion der relevanten ökonomischen Variablen auf strukturelle Schocks messen. Zum anderen ermöglichen strukturelle VARs die Quantifizierung des Beitrags eines bestimmten strukturellen Schocks zur Variation ökonomischer Variablen, indem die Prognosefehlervarianz (PFV) zerlegt wird. Diese strukturellen Parameter ermöglichen es politischen Entscheidungsträgern zu prognostizieren, wie die Wirtschaft auf neue Impulse reagiert. Gleichzeitig helfen diese Informationen bei der Stabilisierung der wirtschaftlichen Entwicklungen durch politische Entscheidungen. Als Beispiel kann angeführt werden, dass Zentralbanken mittels struktureller VARs in der Lage sind, die Auswirkungen einer Leitzinsänderung auf relevante ökonomische Größen abzuschätzen.

Die Herausforderung bei der Modellierung struktureller VARs besteht in der Identifikation der strukturellen Schocks und der dazugehörigen Parameter. Die Identifikation erfordert die Vorgabe einer Struktur für das zugrundeliegende VAR-Modell, wobei diese Struktur unterschiedlichen Ursprungs sein kann. Zum einen kann auf die ökonomische Theorie zurückgegriffen werden. Auf Basis der Theorie werden einzelne strukturelle Parameter, wie z.B. Impuls-Antworten, gezielt mit Restriktionen belegt, um die restlichen Parameter zu identifizieren. Dies erfordert jedoch Vorwis-

sen über diese strukturellen Größen, wobei jedoch Spielraum besteht, wie strikt man diese Restriktionen auferlegt. Sind die Restriktionen nicht strikt genug, werden die relevanten Parameter nicht eindeutig identifiziert, sondern lediglich als Set. D.h. es gibt verschiedene valide strukturelle Modelle. Dieser Fall wird als Set-Identifikation bezeichnet.

Die Vorgabe einer Struktur basierend auf Vorwissen kann vermieden werden, indem man sich statistische Eigenschaften zunutze macht. Bestimmte statistische Annahmen über die stochastischen Prozesse des Modells erlauben die Identifikation der strukturellen Parameter. Allerdings ist zu beachten, dass die so identifizierten Schocks keine klar zugeordnete Bedeutung haben und ihre Effekte ex-post ökonomischen Einflüssen, beispielsweise der Geldpolitik, zugewiesen werden müssen. Dies erschwert die Interpretation der Ergebnisse.

Zudem kann externe Information von ausserhalb des VAR-Modells genutzt werden, um die strukturellen Parameter zu identifizieren. Diese Identifikation erfolgt insbesondere mittels sogenannter „Proxy-Variablen" oder „Proxies". Diese Proxy-Variablen ähneln den Instrumenten aus der mikroökonometrischen Literatur.

Die ersten beiden Kapitel dieser Dissertation befassen sich mit Methoden zur Identifikation der strukturellen Parameter. In beiden Kapiteln wird die Kombination von zwei bereits bekannten Identifikationsstrategien beschrieben, mit dem Ziel, deren Stärken zu bündeln und Schwächen auszugleichen. Das dritte Kapitel behandelt das nachgelagerte Problem nach der Identifikation der strukturellen Parameter, die Inferenz. Im dritten Kapitel wird ein neuer Ansatz vorgestellt, mit dem die Unsicherheit in den geschätzten Parametern quantifiziert werden kann. Die drei Kapitel sind eigenständige Forschungspapiere, die während meiner Zeit als Doktorand an der Universität Konstanz entstanden sind. Das erste und dritte Kapitel wurde eigenständig erarbeitet, während das zweite Kapitel in Zusammenarbeit mit Sascha Keweloh von der Technischen Universität Dortmund entstanden ist.

Die ersten beiden Kapitel befassen sich mit der Identifikation von Schocks mithilfe von Proxy-Variablen, welche prinzipiell die Identifikation ohne Vorgabe weiterer Struktur ermöglichen. Allerdings zeigen Mertens und Ravn (2013), dass die Information der Proxy-Variablen nicht ausreicht, wenn mehrere Schocks mit mehreren Proxies identifiziert werden. In diesem Fall ist es erforderlich, zusätzliche Struktur zu definieren, die entweder auf Vorwissen oder den statistischen Eigenschaften des Modells basiert. Im ersten Kapitel wird beschrieben, wie Restriktionen auf die PFV-Zerlegung genutzt werden, um die zusätzliche Struktur beizusteuern. Der Vorteil dieser Vorgehensweise besteht darin, dass die Impuls-Antworten, welche oft die

Hauptzielgröße der strukturellen VARs darstellen, identifiziert werden können, ohne dass Restriktionen auf die Impuls-Antworten angewandt werden müssen. Im zweiten Kapitel werden die strukturellen Schocks mithilfe statistischer Eigenschaften des Modells identifiziert. Die Struktur wird zusätzlich zu den Proxy-Variablen durch die Annahme eingeführt, dass die strukturellen Schocks teilweise nicht-Gaussianisch und unabhängig voneinander sind. Die Information der Proxies reduziert dabei die Anzahl der nötigen Annahmen. Beide Kombinationen in den ersten zwei Kapiteln versuchen also, die agnostische Natur der Proxy-Variablen so gut wie möglich zu bewahren.

Das erste Kapitel befasst sich mit der Kombination von Proxies mit Restriktionen auf die PFV-Zerlegung. Hierbei lassen sich zwei Möglichkeiten unterscheiden, die PFV-Zerlegung zu beschränken. Erstens kann der Beitrag der Schocks zur PFV von bestimmten Variablen begrenzt werden, um die strukturellen Parameter zu identifizieren. Diese Strategie wurde zuerst von Volpicella (2022) ohne Proxy-Variablen vorgeschlagen und arbeitet mit flexiblen Restriktionen. Der hier beschriebene Ansatz führt zur Set-Identifikation der strukturellen Parameter und kann ebenfalls mit anderen flexiblen Restriktionen kombiniert werden. Unter der Annahme, dass ein Schock den Beitrag zur PFV einer bestimmten Variable maximiert, können die strukturellen Effekte auch eindeutig identifiziert werden. Diese Strategie wird als „Max-Share" bezeichnet, führt jedoch häufig zu verzerrten Ergebnissen. Im ersten Kapitel wird dargelegt, wie die Kombination mit den Proxy-Variablen dazu beiträgt, Verzerrungen zu beseitigen oder zu reduzieren. Dafür schlage ich vor, dem Max-Share-Ansatz eine Ungleichheitsrestriktion hinzuzufügen und ich zeige theoretisch, unter welchen Bedingungen diese Erweiterung die Verzerrung behebt. Der Max-Share-Ansatz ist auf den Fall mit zwei Proxy-Variablen beschränkt, welcher jedoch in der Praxis hochrelevant ist. In einer empirischen Studie werden zwei geldpolitische Schocks in den Vereinigten Staaten identifiziert, die jeweils auf konventionelle bzw. unkonventionelle Politikinstrumente zurückzuführen sind. Die Ergebnisse zeigen, dass sich beide Schocks nachteilig auf die Konjunktur auswirken.

Im zweiten Kapitel erfolgt eine Kombination der Proxy-Variablen mit der Identifikation mittels nicht-Gaussianischer und unabhängiger Schocks. Die Anforderungen an die Annahmen über den stochastischen Prozess des Modells werden dabei durch die Informationen der Proxies reduziert. Gleichzeitig kann mit dieser Kombination die Identifikation mehrerer Schocks mit nur einer Proxy-Variable erfolgen. Dazu werden synthetische Proxies generiert, die nichtlineare Transformationen der einzigen zugrundeliegenden Proxy-Variable darstellen. Die Methode wird verwendet, um zwei verschiedene Unsicherheitsschocks in den Vereinigten Staaten zu identifizieren.

Ein Schock erfasst Unsicherheit auf Finanzmärkten, der andere Unsicherheit in der Realwirtschaft. Nach unserem Kenntnisstand sind wir die ersten, die dies mithilfe von Proxy-Variablen tun. Die aus exogener Variation von Unsicherheit bestehende Proxy-Variable löst das Problem der umgekehrten Kausalität. Die agnostische Identifikation vermeidet im zweiten Schritt das Auferlegen von Struktur auf das Modell, welche a priori nicht eindeutig von der ökonomischen Theorie prognostiziert werden kann. Unsere Ergebnisse zeigen, dass die beiden Unsicherheitsschocks verschiedene Effekte auf die Konjunktur haben. Während der Schock, der die Finanzmarktseite widerspiegelt, nur wenige Reaktionen außerhalb der Finanzmärkte hervorruft, zeigt sich beim Unsicherheitsschock der Realwirtschaft ein deutlicher Konjunktureinbruch. Unsicherheit auf den Finanzmärkten wird dabei hauptsächlich von den beiden Unsicherheitsschocks bedingt, während Unsicherheit in der Realwirtschaft zu großen Teilen von andersartigen Schocks abhängt. Somit lässt sich festhalten, dass Unsicherheit in der Realwirtschaft auch eine endogene Folge wirtschaftlicher Schwankungen ist.

Das dritte Kapitel präsentiert einen neuen Ansatz zur Quantifizierung der Unsicherheit in der Identifikation der strukturellen Parameter. Der Ansatz wurde für den Fall der Set-Identifikation entwickelt, in dem es keinen alleinigen geschätzten Parameter gibt, um den man beispielsweise Konfidenzintervalle konstruieren kann. Bayesianische oder frequentistische Methoden für die Inferenz set-identifizierter Modelle sind oft auf bestimmte Identifikationsstrategien beschränkt oder rechenintensiv. Das dritte Kapitel präsentiert eine Alternative, deren Aussage über die Unsicherheit auf so genannten „learning guarantees “aus dem Bereich des statistischen Lernens beruht. Basierend auf den „learning guarantees “zeige ich, wie die Unsicherheit für mehrere Parameter gleichzeitig quantifiziert werden kann. Dafür erweitere ich den Algorithmus von Montiel Olea und Nesbit (2021) mit einem Bootstrap-Verfahren und zeige, dass der erweiterte Algorithmus unter bestimmenten Annahmen eine klare asymptotische Interpretation hat. Die Eigenschaften in endlichen Stichproben untersuche ich mithilfe einer Simulationsstudie. Im Vergleich zu den bereits existierenden Inferezmethoden bietet der Lern-Algorithmus einen flexiblen Ansatz, der für die meisten Identifikationsstrategien anwendbar ist. Zusätzlich ist er nicht rechenintensiv und bietet die Möglichkeit zur gleichzeitigen Quantifizierung der Unsicherheit mehrerer struktureller Parameter. Die Interpretation, die auf den „learning guarantees “beruht, ist nicht so eindeutig und geläufig wie die Interpretation der konventionellen Inferenz. Abschließend wird der Nutzen des Ansatzes anhand empirischer Anwendungen veranschaulicht. In einer stilisierten Anwendung werden die gelernten

Größen mit konventioneller Inferenz verglichen. Die Ergebnisse legen nahe, dass die gelernten Größen einen ähnlichen Informationsgehalt aufweisen. Dieses Ergebnis wird auch durch die Replikation einer wirklichen empirischen Studie bestätigt.

Literatur

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), 1212–47.

Montiel Olea, J. L. and J. Nesbit (2021). "(Machine) Learning Parameter Regions". *Journal of Econometrics* 222. (1), 716–744.

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

Volpicella, A. (2022). "SVARs Identification Through Bounds on the Forecast Error Variance". *Journal of Business & Economic Statistics* 40. (3), 1291–1301.

Chapter 1

Identifying Proxy VARs with Restrictions on the Forecast Error Variance

1.1. Introduction

External instruments, also called proxies, are a powerful tool for the identification of structural vector autoregressions (VARs). Since the proxy VAR framework was introduced by Stock and Watson (2012) and Mertens and Ravn (2013), proxies are highly prevalent in the structural VAR literature, e.g. Gertler and Karadi (2015), Piffer and Podstawski (2018) or Lakdawala (2019). The proxies contain external information regarding the shocks that one wants to identify. Reminiscent to the classical instrumental variables framework the proxies need to satisfy two conditions. They have to be related to the target shocks of interest (relevance) while being unrelated to the remaining structural shocks that are not identified (exogeneity). However, if multiple shocks are identified with multiple instruments, Mertens and Ravn (2013) show that the information of the proxies alone is insufficient to disentangle the structural shocks and additional identifying restrictions are needed.

Obvious candidates for these additional restrictions are e.g. restrictions on the structural impulse response functions (IRFs) such as strict equality restrictions or less strict sign restrictions. Yet, this chapter focuses on the identification via the forecast error variance (FEV) decomposition. The key merit of restrictions on the FEV decomposition is that they do not impose restrictions on the structural impulse response functions which are often the key element of interest in the structural VARs. In an ideal scenario, the restrictions on the FEVs have the potential to identify the IRFs without the need to impose any structure on the response parameters. However, the FEV restrictions can also be used in conjunction with restrictions on the IRFs in particular and with other restrictions in general. This can be done either to weaken the requirements on the other restrictions or to simply sharpen the identification of existing identification schemes. This paper discusses two different approaches that use restrictions on the FEV for identification, which leads to either set- or point-identification, depending on which approach is employed.

Firstly, bounds on the contribution of the target shocks to the FEV of specified variables can be used as identifying restrictions. These bounds were introduced by Volpicella (2022) and this chapter describes how to use them in the proxy VAR framework. The bounds serve as inequality restrictions and, similar to other identification schemes that rely on inequality restrictions (e.g. sign restrictions), the structural parameters are set-identified. Consequently, the bound restrictions are subject to the trade-off between less strict assumptions and less sharp identification which is common to set-identification schemes. The narrower the bound constraints are set, the

sharper the identification. Yet, narrower bounds represent more strict assumptions on the underlying structural parameters.

The bound restrictions are useful in two distinct ways. On their own, they can identify the structural IRFs without the necessity of imposing any restrictions on them. Yet, in combination with other restrictions they help to sharpen the set-identification. To illustrate, consider a scenario in which the structural shocks are identified with relatively weak sign restrictions. Rather than imposing more strict sign restrictions on impulse responses, adding bounds on the FEVs can help to sharpen the identification. Building on the work of Volpicella (2022) this chapter describes how these bound restrictions can be used in the proxy VAR and, building on Giacomini et al. (2022), it provides robust Bayesian inference for the structural parameters of interest.

Secondly, one can also achieve sharp point-identification via restrictions on the FEV. In doing so, the key assumption is that one shock maximizes the contribution to the FEV of a specific variable, i.e. a technology shock is identified as the shock with maximum contribution to the FEV of a total factor productivity (TFP) measurement, see e.g. Francis et al. (2014). This strategy dates back to Faust (1998) and Uhlig (2004a) and was originally an alternative to the bias-prone long-run restrictions. Francis et al. (2014) coined the term 'Max-Share' for this identification strategy and in the following I use this expression to refer to it. In the literature it has, for instance, been used by Barsky and Sims (2011) to identify technology news shocks and by Ben Zeev and Pappa (2017) to identify defence spending news shocks.

The Max-Share portion of this chapter only discusses the case of two shocks being identified with two proxies. In practice many applications with multiple proxies are limited to the use of two instruments, e.g. Mertens and Ravn (2013), Piffer and Podstawski (2018) or Lakdawala (2019). Hence, the two instrument case is highly relevant for empirical work and it has the nice advantage that the identification of one shock with the Max-Share approach also yields identification of the second shock at the same time. The key advantage of fusing proxy VARs and the Max-Share approach is that sharp point-identification can be achieved without the need for strict equality restrictions on the structural parameters.

However, Dieppe et al. (2019) highlight that the Max-Share framework has a problem with confounding shocks leading to bias in the structural parameters. Shocks, other than the target shock of the Max-Share approach, are confounders if they also contribute to the FEV of the target variable. The proxy VAR partially addresses this concern as shocks that are not related to the proxies are ruled out as confounders.

I show that if one of the two shocks, related to the proxies, exclusively contributes to the FEV of the target variable, the Max-Share approach correctly identifies the true structural parameters. Other potential confounding shocks, not related to the proxies, do not bias the results.

However, in practice this exclusive contribution is not always a suitable assumption and in this case I propose to augment the Max-Share framework with an inequality constraint to eliminate the Max-Share bias. A similar strategy is proposed by Francis and Kindberg-Hanlon (2022) for the Max-Share without proxies. In contrast to Francis and Kindberg-Hanlon (2022) I provide sufficient conditions under which this augmentation is actually guaranteed to remove the Max-Share bias, yet, only for the fusion with the proxy VAR. Moreover, these conditions are applicable to any potential inequality constraint and I do not merely consider a simple sign restriction as the added constraint. One of the sufficient conditions simply depends on the functional form of the quantity that is constrained, which can be easily checked in practice. However, a second condition requires the economic intuition behind the inequality constraint to be correct. In order to assess the plausibility of this condition one needs to provide convincing evidence or reasoning for the imposed constraint. As with other identifying restrictions this reasoning can be based on economic theory, previous results or simply on economic intuition. The proofs for the respective conditions can be found in Appendix 1.A.

Two different types of inequality constraints serve as an example. Firstly, a constraint on a single impulse response at a single horizon can be sufficient to remove the Max-Share bias and I show how to assess whether the imposed constraint is suitable or not. Secondly, I describe the circumstances under which a constraint on the difference in impulse responses of the two identified shocks on a single variable, and again at a single horizon, helps to identify the underlying structural parameters. A simulation study presented in Appendix 1.B illustrates the theoretical results and shows how to assess the sufficient conditions in a potential empirical application. Furthermore, I show under which circumstances the Max-Share bias is guaranteed to decrease in the presence of an inequality constraint. The simulation study also illustrates this. Finally, I describe how to conduct inference for the structural parameters in the proxy Max-Share framework.

In summary, I present two different approaches to impose restrictions on the FEV decomposition. On the one hand, one can employ the less restrictive approach via bounds on the FEV, which can also be easily used in combination with other set-identifying restrictions in order to improve upon existing set-identification schemes.

On the other hand, the Max-Share framework sharply point-identifies the structural model but requires stronger assumptions and a good economic intuition for a potential inequality constraint. Hence, depending on the prior knowledge about the structure in the FEV decomposition a suitable identification scheme can be chosen.

These two approaches are employed to estimate the effects of a conventional and an unconventional monetary policy shock for the US in the pre-zero lower bound period during the Volcker and Greenspan era. Using the Romer and Romer (2004) monetary policy surprises and the monetary policy shock from Sims and Zha (2006) as instruments, I once disentangle the two shocks with the augmented Max-Share approach and once with the more agnostic bounds on the FEVs. The Max-Share approach builds on the assumption that the conventional monetary policy shock contributes more to the FEV of the federal funds rate than the unconventional shock. Potential bias concerns are addressed with an inequality constraint which is based on estimates by Swanson (2021). The details are described in the respective section. The results suggest that both types of monetary policy shocks have a detrimental effect on real economic activity.

The bounds on the FEV are imposed in conjunction with an agnostic set of sign restrictions on impulse responses. The bounds impose a modest immediate contribution of the unconventional monetary policy shock to the FEV of the federal funds rate. Imposing these bounds in addition to the sign restrictions on impulse responses sharpens the identification of the structural parameters, i.e. reduces the width of the identified sets. Qualitatively, the results indicate the same detrimental effects on the economy which were also derived with the Max-Share framework.

Section 1.2 commences with the introduction of the baseline structural VAR framework and the proxy VAR. Section 1.3 describes how the bounds constraints can be used in the proxy VAR, while section 1.4 outlines the fusion of the Max-Share framework with the proxy VAR. Section 1.5 presents the results of the empirical application.

1.2. Econometric Framework

1.2.1. The Structural VAR

The starting point is the k-dimensional stationary structural VAR(p) model:

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + B w_t, \quad t = 1, ...T, \tag{1.2.1}$$

where the $k \times 1$ vector w_t depicts the economically meaningful structural shocks, see e.g. Kilian and Lütkepohl (2017). The $k \times k$ impact matrix B maps the structural shocks w_t into the reduced form innovations, $u_t = Bw_t$. The elements of the $k \times 1$ white noise vector u_t are the reduced form innovations. To shorten the notation one can rewrite the structural VAR in (1.2.1) as:

$$y_t = Ax_t + Bw_t, \quad t = 1, ...T, \tag{1.2.2}$$

with $x_t = (y'_{t-1}, ..., y'_{t-p})'$ and $A = (A_1, ..., A_p)$. A constant is omitted for the brevity of the notation but it can be included in a straightforward manner.

Common to most identification strategies is the normalization $\mathbb{E}(w_t w'_t) \equiv \Sigma_w = I_K$, which yields the set of covariance restrictions $\mathbb{E}(u_t u'_t) \equiv \Sigma = BB'$. Without further assumptions these restrictions do not suffice to pin down the structural parameters as the resulting system of equations has many possible solutions. For example, the Cholesky decomposition of Σ, denoted by Σ_c, satisfies these covariance restrictions. Yet, they will also hold for every rotation with a $k \times k$ orthonormal matrix Q, $\Sigma = \Sigma_c \Sigma'_c = \Sigma_c Q Q' \Sigma'_c$. Giacomini et al. (2022) refer to this representation of the structural VAR as the 'orthogonal reduced form'.

Assuming stability of the VAR polynomial the moving average representation of the structural VAR is given by:

$$y_t = \sum_{m=0}^{\infty} C_m \Sigma_c Q w_{t-m}, \quad t = 1, ...T,$$

where the $k \times k$ matrices C_m contain the moving average coefficients which give the response of the system to the reduced form innovations m periods ago. The impulse response of variable i to shock j at horizon h is given by:

$$\eta_{i,j,h} = e'_i C_h \Sigma_c q_j,$$

where e_i is the ith column of I_k and q_j is the jth column of Q.

Apart from the impulse response functions the FEV decompositions are an element of interest in the structural VARs. In this chapter the FEV decomposition is of particular importance as identifying restrictions are placed on it. To formalise the FEV decomposition let the h-step-ahead forecast of y_t be:

$$y_{t+h|t} = \sum_{m=0}^{\infty} C_{h+m} \Sigma_c Q w_{t-m}.$$

The h-step ahead forecast error is then given by:

$$y_{t+h} - y_{t+h|t} = \sum_{m=0}^{h-1} C_m \Sigma_c Q w_{t+h-m},$$

and the h-step ahead forecast error covariance matrix is represented by:

$$\Omega(h) = \sum_{m=0}^{h-1} C_m \Sigma_c Q Q' \Sigma_c' C_m' = \sum_{m=0}^{h-1} C_m \Sigma C_m'.$$

The contribution from shock j to the total forecast error variance of variable i at horizon h is then:

$$\Omega_{i,j}(h) = \frac{e_i'(\sum_{m=0}^{h-1} C_m \Sigma_c q_j q_j' \Sigma_c' C_m') e_i}{e_i'(\sum_{m=0}^{h-1} C_m \Sigma C_m') e_i}.$$

In order to identify the structural parameters of interest, restrictions have to be placed on the model. As previously mentioned, two different identification schemes exist. Set-identification amounts to finding all the rotation matrices Q that satisfy the identification restrictions. In turn, the rotation matrices define the identified set for which e.g. the impulse response functions or the FEV decomposition can be computed. Common set-identification restrictions are, for instance, inequality restrictions on the structural IRFs. The stricter the identifying restrictions, the smaller the identified set. In point-identification schemes the restrictions are such that only one admissible rotation matrix Q exists. Typical point-identification restrictions are equality restrictions on the elements of the impact matrix B. For example, $k(k-1)/2$ independent equality restrictions on B are sufficient to point-identify the structural parameters.

The proxy VAR framework that is introduced in the next subsection allows for both, point- and set-identification. If one shock is identified using one proxy variable the parameters are point-identified up to sign and scale. Yet, identifying multiple shocks with multiple proxies requires additional restrictions in order to disentangle the shocks. In the latter case the aforementioned sign or equality restrictions are one potential candidate. Depending on the nature of the imposed restrictions the structural parameters are again either point- or set-identified.

1.2.2. The Proxy VAR

The notation of the proxy VAR framework is loosely based on the general framework proposed by Giacomini et al. (2022) as the inference for the set-identification part will be based on their work. As in the standard instrumental variable (IV) framework, the proxy variables - also called instruments interchangeably - have to satisfy two key assumptions. Without loss of generality, let z_t be a $l \times 1$ vector of instruments which are related to the first l structural shocks in w_t. The following two conditions have to be satisfied:

$$\mathbb{E}(z_t w_{(1:l),t}) = \Psi \quad \text{and} \quad \mathbb{E}(z_t w_{(l+1:k),t}) = 0, \tag{1.2.3}$$

where Ψ is an $l \times l$ matrix of full rank. These two conditions resemble the relevance and exogeneity conditions of the standard IV approach. The instruments have to be related to the target shocks and unrelated to the remaining structural shocks w_t. Assume that the proxies follow:

$$\Gamma_0 z_t = \Lambda w_t + \sum_{m=1}^{p_z} \Gamma_m z_{t-m} + \nu_t, \quad t = 1, ..., T. \tag{1.2.4}$$

The process in (1.2.4) indicates that the proxies are related to the structural shocks. Giacomini et al. (2022) assume that $(w_t', \nu_t')' | \mathcal{F}_{t-1} \sim N(0_{(k+l)\times 1}, I_{k+l})$, where $\mathcal{F}_{t-1}$ is the information set at time $t-1$.

The assumptions in (1.2.3) together with process (1.2.4) yield:

$$\mathbb{E}(z_t w_t') = \Gamma_0^{-1}\Lambda = [\Psi, 0_{l\times(k-l)}]. \tag{1.2.5}$$

Plugging model (1.2.2) into the process in (1.2.4) and left-multiplying by Γ_0^{-1} yields:

$$z_t = D y_t + G x_t + \sum_{m=1}^{p_z} H_m z_{t-m} + v_t, \quad t = 1, ..., T, \tag{1.2.6}$$

where $D = \Gamma_0^{-1}\Lambda B^{-1}, G = -\Gamma_0^{-1}\Lambda A$ and $H_m = \Gamma_0^{-1}\Gamma_m$ for each $m = 1, ..., p_z$.

Giacomini et al. (2022) show that (1.2.5) can also be represented by:

$$\mathbb{E}(z_t w_t') = D\Sigma_c Q = [\Psi, 0_{l\times(k-l)}], \tag{1.2.7}$$

implying that the relevance assumption $\text{rank}(\Psi) = l$ is fulfilled if and only if $\text{rank}(D) = l$. The exogeneity and relevance assumption regarding the proxies are fulfilled if the

rotation matrices Q are such that equality (1.2.7) holds, i.e. that the resulting matrix has the zero block. In this manner, the proxy VAR shrinks the identified set.

In the following I deviate from the notation of Giacomini et al. (2022) and use the proxy VAR framework by Piffer and Podstawski (2018). This allows me to handle both the bounds on the FEV decomposition and the combination with Max-Share approach in the same proxy VAR framework. The next section describes how the robust Bayesian inference algorithm of Giacomini et al. (2022) is adapted to the following representation of the proxy VAR below.

Following Piffer and Podstawski (2018), I decompose the reduced form errors into two components:

$$u_t = B_z w_{(1:l),t} + B_{-z} w_{(l+1:k),t}, \quad t = 1, ..., T, \tag{1.2.8}$$

where B_z is the $k \times l$ block of the impact matrix B that contains the first l columns and B_{-z} is the corresponding remaining part of B. B_z contains the structural parameters of the shocks related to the proxies whose identification is the goal of the proxy VAR. I refer to this matrix as the 'proxy impact matrix'. Considering (1.2.8) in conjunction with the assumption regarding the proxies in (1.2.3) yields $\mathbb{E}(u_t z_t') = B_z \Psi' = Z$. Equation (1.2.7) allows this expected value to be rewritten as follows:

$$\mathbb{E}(u_t z_t') = B\mathbb{E}(w_t z_t') = B\mathbb{E}(z_t w_t')' = \Sigma D',$$

with $B = \Sigma_c Q$ and $\mathbb{E}(z_t w_t') = D\Sigma_c Q$. Hence, $\Sigma D' = Z = B_z \Psi'$. Partitioning the matrices $\Sigma D'$ and B_z yields:

$$\Sigma D' = Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \text{ and } B_z = \begin{pmatrix} B_{11} \\ B_{21} \end{pmatrix}$$

and thus

$$B_{21} = Z_2 Z_1^{-1} B_{11} = Z_l B_{11},$$

where Z_1 is the upper $l \times l$ block of the matrix Z and B_{11} is the upper $l \times l$ block of B_z. Hence,

$$B_z = \begin{pmatrix} B_{11} \\ Z_l B_{11} \end{pmatrix}, \tag{1.2.9}$$

meaning that if the upper $l \times l$ block B_{11} is identified the remaining block of the proxy impact matrix is identified as well. In order to identify the upper block of B_z, partition the matrices of the standard covariance restrictions $\mathbb{E}(u_t u_t') \equiv \Sigma = BB'$ such that:

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \begin{pmatrix} B_{11} & B_{21} \\ B_{12} & B_{22} \end{pmatrix},$$

where Σ_{11} is the upper left $l \times l$ block of Σ. B_{11} is again the upper $l \times l$ block of B_z and therefore the upper left block of B. The remaining blocks of the two matrices have the according dimensions. It can be shown that $B_{11}B_{11}' = \Sigma_{11} - B_{12}B_{12}'$ (see Mertens and Ravn, 2013) with:

$$\begin{aligned} B_{12}B_{12}' &= (\Sigma_{21} - Z\Sigma_{11})'\Pi^{-1}(\Sigma_{21} - Z\Sigma_{11}), \\ \Pi &= \Sigma_{22} + Z'\Sigma_{11}Z' - \Sigma_{21}Z' - Z\Sigma_{21}'. \end{aligned}$$

Similar to the covariance restrictions $\mathbb{E}(u_t u_t') \equiv \Sigma = BB'$, the equation $B_{11}B_{11}' = \Sigma_{11} - B_{12}B_{12}'$ does not pin down the parameters of B_{11} uniquely. Let B_{11}^c be the Cholesky decomposition of $\Sigma_{11} - B_{12}B_{12}'$, then every rotation of B_{11}^c with an $l \times l$ orthonormal matrix Q will also satisfy $B_{11}^c B_{11}^{c'} = B_{11}^c QQ' B_{11}^{c'} = \Sigma_{11} - B_{12}B_{12}'$.

The exogeneity and relevance condition regarding the proxies is satisfied for B_z by construction. Hence, the identification boils down to finding the set of $l \times l$ orthonormal matrices Q that satisfy the additional identifying restrictions, e.g. inequality restrictions on structural parameters. In section 1.3, I describe how inequality restrictions on the FEV like in Volpicella (2022) fit into this framework.

It is also possible to point-identify the structural parameters, which amounts to finding the one $l \times l$ rotation matrix for which the restrictions are satisfied. If the resulting recursive structure of B_{11}^c for the contemporaneous impacts of the identified shocks is economically justifiable, the Cholesky decomposition immediately point-identifies the structural shocks, i.e. the corresponding rotation matrix is just $Q = I_l$. Another way to point-identify the structural parameters is the Max-Share framework and section 1.4 describes how it is used to obtain the single rotation matrix Q in the proxy VAR.

1.3. Proxy VAR with Bounds on the FEV

The bounds on the contributions to the FEV were introduced by Volpicella (2022) and this section applies them to the proxy VAR framework. In doing so, I loosely follow the notation of Volpicella (2022). Such bounds on the FEV are inequality restrictions in the spirit of the well-known sign restrictions on impulse response parameters, and thus the structural parameters are set-identified. Naturally, the challenges the set-identification literature deals with also apply to the identification via FEV bounds.

1.3.1. Bounding the Contribution to the FEV

In the context of the proxy VAR framework the contribution of shock j to the FEV of variable i at horizon h is:

$$\Omega_{ij}^{z}(h) = \frac{e_i'(\sum_{m=0}^{h-1} C_m B_z q_j q_j' B_z' C_m')e_i}{e_i'(\sum_{m=0}^{h-1} C_m \Sigma C_m')e_i}. \tag{1.3.1}$$

Uhlig (2004b) shows that equation (1.3.1) can also be written as:

$$\Omega_{i,j}^{z}(h) = q_j' R_{i,h} q_j, \tag{1.3.2}$$

where

$$R_{i,h} = \frac{\sum_{m=0}^{h-1} c_{i,m}' c_{i,m}}{e_i'(\sum_{m=0}^{h-1} C_m \Sigma C_m')e_i},$$

with $c_{i,m} = e_i C_m B_z$ being the ith row vector of $C_m B_z$. $R_{i,h}$ is a positive semidefinite, symmetric and real $l \times l$ matrix.

Given equation (1.3.2) the bounds on the contribution to the FEV of variable i by shock j at horizon h can be represented by:

$$\underline{\tau}_{i,j,h} \leq q_j' R_{i,h} q_j \leq \overline{\tau}_{i,j,h},$$

where $\underline{\tau}_{i,j,h}$ and $\overline{\tau}_{i,j,h}$ depict the lower and upper bound, respectively, and $0 \leq \underline{\tau}_{i,j,h} \leq \overline{\tau}_{i,j,h} \leq 1$. In accordance with the notation of Volpicella (2022), let $\mathcal{I}_{ij}$ represent a set of indices that denote whether the FEV of variable i is bounded for shock j and

$\mathcal{H}_{ij}$ collects the horizons h for which these bounds are imposed. The complete set of bound constraints is then expressed by

$$\underline{\tau}_{i,j,h} \leq q_j' R_{i,h} q_j \leq \overline{\tau}_{i,j,h}, \text{ for } i \in \mathcal{I}_{ij} \text{ and } h \in \mathcal{H}_{ij}.$$

These bounds on the contributions to the FEV can also be applied in conjunction with already existing set-identifying inequality restrictions, e.g. sign restrictions like introduced by Uhlig (2005). Furthermore, restrictions on the correlations of the proxies with the identified shocks are possible. These types of restrictions constrain the elements of Ψ and they can be checked employing the routine used by Piffer and Podstawski (2018). The identified set is then characterized by all the rotation matrices Q for which these FEV bounds and other potential restrictions are satisfied. As pointed out by Volpicella (2022) such bounds on the FEV contributions can be derived either through economic theory or simply through a strong economic intuition.

1.3.2. Nonemptiness of the Set

The bound restrictions, together with potential additional restrictions, are subject to set-identification specific considerations. On the one hand, if the bounds are not restrictive enough one may obtains potentially large identified sets which yield a fuzzy identification of the underlying structural effects. If, on the other hand, the bounds are too restrictive the identified set might be empty because no structural representation of the model satisfies them.

Unfortunately, no formal guidance helps to assess the restrictions in this regard, what in turn highlights the importance of the economic theory or intuition behind them. Yet, if the identified set is empty, this might be a sign that the imposed restrictions are not reasonable.

Furthermore, it is important to know whether the set is empty for the estimation procedure. Volpicella (2022) provides sufficient conditions for the nonemptiness of the identified set when only one shock is restricted. These sufficient conditions also apply in the same fashion to the proxy VAR framework. Recall from (1.3.2) that the contribution of the target shock j to the FEV of variable i is:

$$\Omega_{ij}^{z}(h) = q_j' R_{i,h} q_j,$$

Let λ_m^{ih} be the real eigenvalues of $R_{i,h}$ with $i \in \mathcal{I}_{ij}$, $h \in \mathcal{H}_{ij}$ and $m = 1, ..., l$. Uhlig (2004b) shows that finding the maximum (minimum) of (1.3.2) with respect to q_j amounts to finding the largest (smallest) eigenvalue λ_m^{ih} of $R_{i,h}$ and the maximum

(minimum) is achieved by using the corresponding eigenvector q_m as a rotation vector q_j. Hence, the eigenvalues λ_m^{ih} correspond to the contributions to the FEV of variable i at horizon h.

Proposition 1.1 adapts Proposition 3.1 of Volpicella (2022) to the proxy VAR framework and provides sufficient conditions for the nonemptiness of the identified set when a single target shock j is restricted. The proof is relegated to Appendix 1.A.

Proposition 1.1 (Nonemptiness)**.** *If the following conditions hold:*

(a) $\exists i \in \mathcal{I}_j, \exists h \in \mathcal{H}_{ij} \mid \underline{\tau}_{i,j,h} \leq \lambda_m^{ih} \leq \overline{\tau}_{i,j,h},\ R_{i,h}q_m = \lambda_m^{ih}q_m$ *for some* $m = 1, ..., l$,

(b) given q_m *from (a),* $\underline{\tau}_{i,j,h} \leq q_m' R_{i,h} q_m \leq \overline{\tau}_{i,j,h}\ \forall i \in \mathcal{I}_j$ *and* $\forall h \in \mathcal{H}_{ij}$*, and all other additional restrictions are satisfied,*

then the identified set is non-empty and bounded.

The sufficient condition requires to find an eigenvalue λ_m^{ih} which lies within the imposed bounds on $\Omega_{ij}^z(h)$. If for $q_j = q_m$ all other identifying restrictions are satisfied the identified set is non-empty.

A difference to Volpicella (2022) is the number of eigenvalues that are available for the assessment of the nonemptiness. In the proxy VAR only l eigenvalues are available, compared to the k eigenvalues in Volpicella (2022). In practice, the sufficient condition is more likely to be fulfilled with k eigenvalues compared to the case with only l eigenvalues as the chance is higher that one of the eigenvalues falls into the imposed bounds.

Yet, the fewer eigenvalues can nevertheless be useful. Suppose $l = 2$, then the two eigenvalues represent the maximum and minimum contribution to the FEV at the specific horizon. Hence, if $l = 2$ only the maximum and minimum contribution can be used to check the nonemptiness. If the bounds do not encompass the extreme values the sufficient conditions are not satisfied. In this case one can potentially see whether the identified set is empty or not by looking at these extreme values. If the upper bound $\overline{\tau}_{i,j,h}$ is smaller than the minimum eigenvalue of $R_{i,h}$ for one of the imposed bounds, or if the lower bound $\underline{\tau}_{i,j,h}$ is larger than the maximum eigenvalue the imposed bounds are outside the interval of possible contributions. Hence, the set is empty.

If the sufficient conditions of Proposition 1.1 are not fulfilled and the emptiness of the identified set is not determined by the reasoning described in the previous paragraph, a different approach helps to approximate the nonemptiness of the identified set. As often done in the literature, one can draw a specified number of matrices Q

from the orthonormal space. If none of these draws satisfy the restrictions, one can conclude that the set is empty.

Finally, a special case arises when the number of instruments is $l = 2$. Then Proposition 1.1 also helps to detect nonemptiness when both shocks are subject to identification restrictions. Knowing q_m amounts to knowing the whole 2×2 rotation matrix Q_m due to the properties of orthonormal matrices. Hence, in step (b) of Proposition 1.1 also the restrictions on the second shock can be checked.

1.3.3. Estimation and Inference

This subsection describes how the robust Bayesian inference framework by Giacomini et al. (2022) is employed to conduct inference in the presence of the FEV bound restrictions. The advantage of this approach is that it avoids specifying a prior over the rotation matrices Q that is not updated by the data. Baumeister and Hamilton (2015) show that when the prior for the rotation matrices Q is a uniform distribution over the space of orthonormal matrices, the common approach in the literature, the structural parameters are influenced by the prior distribution even asymptotically.

The remedy is to employ a distribution-free approach. In robust Bayesian inference the endpoints of the identified set are calculated numerically, or analytically if possible. The endpoints, or boundaries, of the identified set are the maximum and minimum values of the structural parameters of interest given all admissible rotation matrices Q. Giacomini et al. (2022) show that this procedure yields prior robust inference over the structural parameters.

As previously stated, I adapt the algorithm used in Giacomini et al. (2022) to the specification of the proxy VAR framework outlined in the previous section. I incorporate the exogeneity and relevance restrictions regarding the proxies via the proxy impact matrix B_z and only rotate the upper block B_{11} with an $l \times l$ orthonormal matrix Q. Hence, unlike in the algorithm by Giacomini et al. (2022) one does not have to deal with an $k \times k$ orthonormal matrix which needs to fulfil the exogeneity condition as depicted in (1.2.7). Incorporating the exogeneity via the proxy impact matrix in (1.2.9) reduces the computational demands of the Bayesian inference by reducing the dimension of the orthonormal matrix. Furthermore, this makes additional constraints in the optimization, which should ensure exogeneity of the instruments, obsolete. Another benefit of the approach via the proxy impact matrix is that it does not require a specific ordering of the variables in the VAR. Yet, the computational benefits come at the cost of being only able to identify the structural shocks that are related to the proxies, while the general framework of Giacomini et al. (2022) allows for

the identification of shocks beyond those targeted by the instruments. If this is an objective in practice the original algorithm by Giacomini et al. (2022) needs to be employed.

To describe the Bayesian algorithm let $\phi \in \Phi$ collect all the reduced form parameters in (1.2.2) and (1.2.6). For the following algorithm it is not important which prior for ϕ is used as long as one is capable of drawing from the posterior distribution of the reduced form parameters. For the results derived in section 1.5 I follow Giacomini et al. (2022) and use an (improper) Jeffrey's prior.

Suppose that the impulse responses $\eta_{i,j,h} = e_i' C_m B_z q_j$ are the structural parameters of interest. The upper and lower boundary of the identified set with respect to the imposed restrictions are depicted by $u_{i,h}(\phi)$ and $l_{i,h}(\phi)$. Algorithm 1.1 describes how to conduct robust Bayesian inference for the identified set of $\eta_{i,j,h}$.

Algorithm 1.1.

Step 1: Obtain draws ϕ from its posterior distribution and compute B_{11}^c.

Step 2: Check whether the identified set is empty. If the set is empty go back to Step 1. If the set is non-empty proceed with Step 3.

Step 3: Compute the boundaries of the identified set:

$$
\begin{aligned}
& l_{i,h}(\phi) = \min_{Q} e_i' C_m B_z q_j \\
& s.t. \quad \underline{\tau}_{i,j,h} \leq q_j' R_{i,h}(\phi) q_j \leq \overline{\tau}_{i,j,h}, \quad \forall i \in \mathcal{I}_j \text{ and } \forall h \in \mathcal{H}_{ij}, \\
& \qquad QQ' = I_l, \\
& \qquad \textit{potential sign restrictions and/or restrictions on } \Psi.
\end{aligned}
$$

The upper boundary $u_{i,h}(\phi)$ is obtained analogously.

Step 4: Repeat Steps 1 to 3 N times.

Step 5: Approximate the set of posterior means and the robust credible region as described in Giacomini et al. (2022).

For step 2, Proposition 1.1 helps to gauge whether the identified set is empty or not. If the sufficient conditions of Proposition 1.1 are not fulfilled a specified number of rotation matrices Q are drawn in order to approximate the set as being empty if none of the draws satisfy the identification restrictions. In this case, step 2 differs from Giacomini et al. (2022) as the $l \times l$ rotation matrices Q do not need to be drawn

considering the exogeneity conditions for the proxies. These conditions are already incorporated in the construction of B_z. Step 3 differs in the specification of the proxy VAR, and thus the dimension of the rotation matrix Q. Secondly, the added constraint in the maximization problem that represents the bounds on the FEV are a distinction to the algorithm in Giacomini et al. (2022).

Step 3 poses a potentially nonconvex optimization problem, and thus the typical approaches to handle this with gradient-based optimization techniques are necessary. The simple remedy is to use different initial values and to compute the maximum or minimum over the set of solutions which are derived with the different initial values. Giacomini et al. (2022) also provide an algorithm to approximate the boundaries of the identified set in order to check the convergence of the numerical optimization or as a standalone alternative.

Algorithm 1.2. *Replace step 3 of Algorithm 1.1 with:*

> *Step 3: Draw Q until X draws which satisfy the identification restrictions are reached. For each Q_x, $x = 1, ..., X$ compute $\eta_{x,i,j,h} = e_i' C_m B_z q_{x,j}$ and approximate $u_{i,h}(\phi)$ and $l_{i,h}(\phi)$ by the maximum and minimum of $\eta_{x,i,j,h}$ over all X draws.*

Montiel Olea and Nesbit (2021) show that the random sampling approximation of Algorithm 1.2 can be represented as a supervised learning problem. They provide the number of admissible draws of Q required to learn the set with a certain precision. In general, the approximated set will be smaller than the true set, yet with a sufficient number of draws the approximation error will be small. The theoretical results of Montiel Olea and Nesbit (2021) can be used to assess the precision of the approximation at a specific amount of draws X. Hence, this approximation introduces an additional layer of uncertainty into the Bayesian algorithm which needs to be acknowledged in the interpretation of the credible regions.

Giacomini et al. (2022) argue that this approximation might be advantageous under certain circumstances. Firstly, if the VAR system is large and one is interested in the impulse responses for many variables at many horizons, drawing many rotation matrices Q is computationally less costly than optimizing for every variable at every horizon. This is especially true with the representation of the proxy VAR used in this paper as it is relatively straightforward to draw simple $l \times l$ rotation matrices, given that l is small in most empirical applications. Secondly, if not only the impulse responses but also e.g. the FEV decomposition is of interest, the approximation has an advantage. Each draw of Q can be used to compute the impulse responses and

FEV decomposition of each variable at every horizon, whereas the optimization has to be carried out for each parameter, variable and horizon individually.

1.4. Proxy VAR and Max-Share

This section describes how the shocks in the proxy VAR can be disentangled using the Max-Share framework which was introduced by Faust (1998) and Uhlig (2004b). The key assumption behind the Max-Share approach is that the shock of interest j is the one that maximizes the contribution to the FEV of a target variable i over the considered horizon $\underline{h}$ up to $\overline{h}$. Identification of the target shock amounts to finding the rotation vector q_j for which this maximum is achieved:

$$q_j = \text{argmax} \sum_{h=\underline{h}}^{\overline{h}} \Omega_{i,j}^z(h) \quad \text{s.t.} \quad q_j' q_j = 1. \tag{1.4.1}$$

The closed-form solution of the baseline Max-Share maximization problem via the eigenvalues presented by Uhlig (2004b) also applies to the proxy VAR.

A key advantage of the Max-Share approach in the proxy VAR framework is that it does not require the assumption that the shock of interest is the one which maximizes the contribution to the FEV of a target variable among all possible shocks. The assumption reduces to the shock being the one which maximizes the contribution among the shocks related to the instruments. In practice the latter assumption is generally easier to defend depending on the empirical application. Yet, the circumstances under which the Max-Share approach correctly identifies the shocks of interests are discussed below.

In the following, I only consider the case when two shocks are identified with two instruments. In practice the case of two instruments is highly relevant, as finding multiple convincing proxy variables is difficult and finding two of them is already a challenging task. If two shocks are identified using two instruments, the rotation matrix Q has dimension 2×2. This means that identifying the first column of a 2×2 orthonormal matrix also identifies the second column up to a sign normalization due to the properties of orthonormal matrices:

$$Q = \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix}, \qquad 1 = q_{21}^2 + q_{22}^2 \quad \text{and} \quad 1 = q_{11}^2 + q_{12}^2.$$

The conditions of the 2×2 orthonormal matrices imply that each element of Q is between $-1 \leq q_{ij} \leq 1$. Furthermore, these conditions imply that $q_{21} = \pm\sqrt{1 - q_{11}^2}$, $q_{12} = \pm q_{21}$ and $q_{22} = \pm q_{11}$. Hence, every element of Q can be written in terms of just q_{11} up to sign normalizations. It follows that the identification of one shock via the Max-Share framework, i.e. the first column of the rotation matrix, also gives the structural parameters of the second shock up to a sign normalization.

Note, that if just a single shock is identified with a single proxy the structural parameters are already identified up to scale, and thus no additional restrictions like e.g. the Max-Share approach are needed. If more than two shocks are identified with more than two instruments the $l \times l$ rotation matrix Q will be of dimension larger than two. Hence, if one column is identified with the Max-Share approach the other columns of the rotation matrix, and thus the remaining structural shocks, are not identified. The generalization of the Max-Share approach proposed by Carriero and Volpicella (2024) could be an interesting extension for the identification of more than two shocks in the proxy VAR.

1.4.1. Ruling out Confounders via the Proxy VAR

Rather recently Dieppe et al. (2019) highlighted that the key identification assumption of the Max-Share framework is violated if another shock also contributes to the FEV of the target variable. To illustrate this point, assume that the technology shock is accountable for the majority of the FEV of a TFP measurement. If the technology shock is identified as the shock that maximizes the contribution to the TFP measurement, the results will be biased if other shocks also contribute to the FEV of the TFP measurement. Dieppe et al. (2019) present strategies how to circumvent this problem of confounding shocks in the baseline Max-Share identification without proxies. In this section I focus on the remedies to the bias concerns that the proxy VAR framework offers.

First and foremost, the proxy VAR helps with the confounding shocks as it rules out confounding shocks that are not related to the proxies. Proposition 1.2 formalizes this property of the Max-Share approach in the proxy VAR. The proof is relegated to Appendix 1.A.

Proposition 1.2. *Suppose two valid proxy variables z_1 and z_2 are available which are related to shocks w_1 and w_2. Without loss of generality, if $\Omega_{i,2}(h) = 0$ holds for $\underline{h} \leq h \leq \overline{h}$. Then*

$$q_1 = argmax \sum_{h=\underline{h}}^{\overline{h}} \Omega_{i,1}^{z}(h) \quad s.t. \quad q_1' q_1 = 1,$$

point-identifies the shocks w_1 and w_2 up to a sign normalization.

Proposition 1.2 establishes that if out of the shocks that are related to the instruments, one exclusively contributes to the FEV of the target variable, both shocks are correctly identified by the Max-Share framework. Proposition 1.2 also establishes that shocks which are not related to the proxies do not act as confounding shocks in the proxy VAR Max-Share approach. Given the exogeneity condition of the proxies, the proxy VAR separates the shocks identified with the proxy VAR from the remaining shocks and identifies them up to a linear combination. This linear combination is then disentangled with the Max-Share framework in the second step.

To illustrate, consider a scenario in which two proxies are used to identify two shocks. One of the shocks contributes the most to the FEV of a specified variable and out of the two shocks that are related to the proxies, it contributes exclusively to the FEV of this variable. If another shock that is not related to the proxies also contributes to the FEV of this target variable, it is ruled out as a confounding shock, as it is not related to the proxies. As out of the two shocks that are related to the proxies only one exclusively contributes to the FEV of the target variable, the Max-Share framework correctly disentangles and identifies both shocks of interest. The simulation study presented in Appendix 1.B highlights and confirms this property of the Max-Share approach in the proxy VAR.

The case of exclusive contribution is a rather strict assumption which would also allow the shocks to be identified through imposing a recursive structure on the proxy impact matrix. Yet, the next subsection describes how the Max-Share approach can be employed without the assumption of exclusive contribution.

1.4.2. Tackling the Max-Share Bias

If the assumption of exclusive contribution for one of the two identified shocks is not reasonable, an additional inequality constraint which augments the Max-Share framework can help to nevertheless correctly identify the true structural parameters. Let $IQ_{i,j,h}(q_{11})$ denote the function on which the inequality constraint is imposed

and below two examples for potential functions $IQ_{i,j,h}(q_{11})$ are given. Proposition 1.3 states sufficient conditions under which the structural parameters are identified when the additional inequality constraint is included in the Max-Share framework. The rotation matrix that corresponds to the true structural parameters is denoted by Q^* with its columns q_j^*.

In the following, I restrict the first element q_{11} of Q, to be positive, and thus $q_{11} \in [0, 1]$ by the properties of 2×2 orthogonal matrices. This simply helps with writing down the conditions for successful identification. In practice this means the shock is normalized to be an expansionary or contractionary shock - depending on the application - during the identification. Yet, after successful identification one can transform the shock into its contractionary/expansionary counterpart as usual by multiplying its respective column of the impact matrix or rotation matrix with -1.

Proposition 1.3. *Let $q_{11} \in [0, 1]$. Suppose two valid proxy variables z_1 and z_2 are available which are related to shocks w_1 and w_2. Without loss of generality, suppose out of w_1 and w_2, w_1 contributes the most, but not exclusively, to the FEV of variable i for $\underline{h} \leq h \leq \overline{h}$. The augmented Max-Share approach*

$$
\begin{aligned}
q_1 = \; & argmax \sum_{h=\underline{h}}^{\overline{h}} \Omega_{i,1}^{z}(h) \\
& s.t. \\
& q_1' q_1 = 1, \\
& q_1' q_2 = 0, \\
& q_{11} \geq 0, \\
& IQ_{i,1,h}(q_{11}) \lesseqgtr \epsilon,
\end{aligned}
\tag{1.4.2}
$$

identifies the true structural parameters if

(a) $IQ_{i,1,h}(q_{11}) \lesseqgtr \epsilon$ implies a single binding linear restriction on q_{11} and

(b) the restriction is set such that it implies $q_{11} \lesseqgtr q_{11}^*$.

The proof of Proposition 1.3 is relegated to Appendix 1.A. The sign of ϵ and therefore the direction of the inequality constraint depends on the specific application. Whether condition (a) of Proposition 1.3 is satisfied depends on the type of inequality restriction that is placed while condition (b) requires the economic intuition behind the chosen restriction to be correct. In practice, one can check whether condition

(a) is satisfied and in the following I provide an example of two different types of inequality constraints. The implications of condition (b) will also be discussed below.

Proposition 1.4 provides a sufficient condition under which condition (a) of Proposition 1.3 is satisfied.

Proposition 1.4. *Let $q_{11} \in [0,1]$. Suppose a single linear restriction is placed on $IQ_{i,1,h}(q_{11})$. If $IQ_{i,1,h}(q_{11})$ is monotonic in $q_{11} \in [0,1]$, $IQ_{i,1,h}(q_{11}) \lesseqgtr \epsilon$ implies a single linear restriction on q_{11} for $q_{11} \in [0,1]$.*

The proof is again relegated to Appendix 1.A. This chapter considers two types of inequality restrictions $IQ_{i,1,h}(q_{11})$ in more detail. Firstly, a standard sign restriction on the impulse response of the first shock on a specified variable for a single horizon h, namely $\eta_{i,1,h}(q_{11}) \lesseqgtr \epsilon$. The second constraint restricts the difference of the impulse responses of the two identified shocks on a single variable i for a single horizon h, i.e. $D_{i,h}(q_{11}) = \eta_{i,1,h} - \eta_{i,2,h} \lesseqgtr \epsilon$. This translates to one shock having a larger impact than the other shock on the variable at that horizon h. The latter type of restriction adds another possibility to incorporate economic intuition into the identification.

In general, checking whether the sufficient for condition (a) is satisfied boils down to checking whether the function $IQ_{i,j,h}(q_{11})$ is monotonic in $q_{11} \in [0,1]$. The following two algorithms describe how to assess whether the sufficient condition (a) is satisfied for the two mentioned types of inequality constraints, namely $\eta_{i,1,h}(q_{11})$ or $D_{i,h}(q_{11})$. The algorithms merely state the necessary steps and further details are provided in Appendix 1.A.

Recall, that the structural impulse response of variable i to the first shock at horizon h is given by $\eta_{i,1,h} = e_i' C_h B_z q_1$, where e_i is the first column of the identity matrix I_2. Let

$$C_h B_z = \begin{pmatrix} \Theta_{1,1,h} & \Theta_{1,2,h} \\ \Theta_{2,1,h} & \Theta_{2,2,h} \\ \vdots & \vdots \\ \Theta_{k,1,h} & \Theta_{k,2,h} \end{pmatrix}.$$

Furthermore, let $\hat{B}_z^c$ and $\hat{C}_h$ be consistent estimators for the underlying quantities. Finally, let Q^{max} be the rotation matrix obtained by the baseline Max-Share framework without any inequality constraints and only under the condition that $q_{11}^{max} \geq 0$. Then $\Theta_{i,1,h}$ and $\Theta_{i,2,h}$ can be computed with the reduced form parameters estimates $\hat{C}_h$ and $\hat{B}_z = \hat{B}_z^c Q^{max}$.

Algorithm 1.3 assesses whether the inequality constraint on $\eta_{i,1,h}$ satisfies the sufficient condition (a).

Algorithm 1.3.

1. *Carry out the baseline Max-Share identification under the condition* $q_{11} \geq 0$ *and record the sign of* q_{21}^{max}.

2. *Compute* $\Theta_{i,1,h} q_{11}^{max}$ *and* $\Theta_{i,2,h} q_{21}^{max}$. *If the signs of* $\Theta_{i,1,h} q_{11}^{max}$ *and* $\Theta_{i,2,h} q_{21}^{max}$ *are the opposite,* $\eta_{i,1,h}$ *is strictly monotonic in* $q_{11} \in [0, 1]$.

Algorithm 1.4 describes how to assess whether $D_{i,h}(q_{11}) = \eta_{i,1,h} - \eta_{i,2,h}$ is monotonic in $q_{11} \in [0, 1]$. It contains more steps than Algorithm 1.3 as an additional normalization on the second column of Q^{max} is necessary. Further details can again be found in Appendix 1.A.

Algorithm 1.4.

1. *Carry out the baseline Max-Share identification under the condition* $q_{11} \geq 0$ *and record the sign of* q_{21}^{max}.

2. *Get* q_2 *by setting* $q_{22} = q_{11}^{max}$ *and* $q_{12} = -q_{21}^{max}$.

3. *Compute* $\eta_{i,1,h} = e_i' C_h B_z q_1$ *and denote its sign.*

4. *Compute* $\eta_{i,2,h} = e_i' C_h B_z q_2$. *If* $\eta_{i,2,h}$ *has the opposite sign of* $\eta_{i,1,h}$ *change the sign normalization of* q_2 *(multiply it with* -1*).*

5. *Compute* $\Theta_D = \Theta_{i,h,1} - \Theta_{i,h,2}$ *and* $\Theta_S = \Theta_{i,h,1} + \Theta_{i,h,2}$. *Record the signs of the two quantities.*

6.1 *If* q_{21}^{max} *is positive and given* q_2 *from Step* 4, $D_{i,h}(q_{11})$ *is monotonic if:*

   (i) the signs of Θ_D *and* Θ_S *are the same and* $q_{22} = -q_{11}^{max}$,

   (ii) or if the signs Θ_D *and* Θ_S *are the opposite and* $q_{22} = q_{11}^{max}$.

6.2 *If* q_{21}^{max} *is negative and given* q_2 *from Step* 4, $D_{i,h}(q_{11})$ *is monotonic if:*

   (i) the signs of Θ_D *and* Θ_S *are the same and* $q_{22} = q_{11}^{max}$,

   (ii) or if the signs Θ_D *and* Θ_S *are the opposite and* $q_{22} = -q_{11}^{max}$.

Note that monotonicity is just sufficient for condition (a) to hold, and that the circumstances under which condition (a) is satisfied without monotonicity of $IQ_{i,1,h}(q_{11})$ can also be found in Appendix 1.A.

Condition (b) of Proposition 1.3 requires the restriction to be set such that $q_{11} \lesseqgtr q_{11}^*$. For the simple sign restriction on the impulse responses, this implies that one response of the identified shock at one horizon has to be known. For the difference restriction, one needs to gauge the difference in the responses of the shocks correctly, i.e. the true ϵ^* has to be known. In practice such knowledge has to be drawn from economic theory, previous results or pure economic intuition. The simulation study presented in Appendix 1.B illustrates the correct choice of ϵ such that condition (b) is satisfied.

Yet, if the true ϵ^* is unknown an inequality restriction can nevertheless help to reduce the Max-Share bias. Let $\delta_{i,j,h} = \eta_{i,j,h}(q_{11}^*) - \eta_{i,j,h}(q_{11})$ be the bias of the structural impulse response of variable i at horizon h to shock j and let $\epsilon^{max} = IQ_{i,j,h}(q_{11}^{max})$. In the following, $\downarrow$ and $\uparrow$ denote the one-sided limits, e.g. $\epsilon \downarrow \epsilon^*$ means that ϵ approaches ϵ^* from above. The proof to Proposition 1.5 is again relegated to Appendix 1.A.

Proposition 1.5 (Reducing the Bias)**.** *Let $q_{11} \in [0,1]$. Let $IQ_{h,i,1}(q_{11})$ be a monotonic function in q_{11}. Suppose an inequality restriction is placed on $IQ_{h,i,1}(q_{11}) \gtreqless \epsilon$ in* (1.4.2)*. If $\eta_{i,j,h}(q_{11})$ is a monotonic function in q_{11}, then for $\epsilon \downarrow \epsilon^*$ the absolute bias $|\delta_{i,j,h}| \to 0$. Similarly, if $\epsilon \uparrow \epsilon^*$ then $|\delta_{i,j,h}| \to 0$.*

Proposition 1.5 states that if one approaches ϵ^*, i.e. the economic intuition is more precise, the Max-Share bias in the impulse responses decreases under the specified monotonicity conditions. Note that these are merely sufficient conditions and the bias also decreases if the underlying functions are monotonic only for the relevant, yet unknown, range of q_{11}. Proposition 1.5 also implies that once you overshoot with ϵ beyond ϵ^*, the bias increases again. The extent to which it increases is governed by the functional form of $IQ_{h,i,1}(q_{11})$ and $\eta_{i,j,h}(q_{11})$, and thus if one wants to guarantee a decrease in the bias, setting ϵ rather conservatively is advised.

In practice, one again has to assess whether $IQ_{h,i,1}(q_{11})$ and $\eta_{i,j,h}(q_{11})$ are monotonic in $q_{11} \in [0,1]$. For the two types of inequality constraints considered in the paper, Algorithm 1.3 is employed to assess $\eta_{i,j,h}(q_{11})$ and also if $IQ_{h,i,1}(q_{11}) = \eta_{i,j,h}(q_{11})$. Likewise Algorithm 1.4 is utilised if $IQ_{h,i,1}(q_{11}) = D_{i,h}(q_{11})$.

In practice, one can also report results for different values of ϵ over a range of values for which one is confident that ϵ^* is contained. However, this de-sharpens

the identification the larger the range of ϵ values. Thus, the augmented Max-Share approach is most useful when a rather strong intuition for the correct ϵ^* exists.

The theoretical results established in sections 1.4.1 and 1.4.2 are supported by a simulation study which is described in Appendix 1.B. Firstly, the result of Proposition 1.2 is highlighted and it is demonstrated that the basic Max-Share approach correctly identifies the two shocks that are related to the instruments if one of the shock contributes exclusively to the FEV of a specific variable. Yet, the simulations also highlight that the basic Max-Share approach fails in the absence of exclusive contribution. The simulation then proceed to show that an inequality constraint, like in (1.4.2), removes the bias under certain conditions. Different ϵ values are simulated to demonstrate that the bias decreases as you approach the true margin ϵ^*.

1.4.3. Estimation and Inference

In point-identification schemes frequentist bootstrap procedures are an often used tool for conducting inference. In this case, I recommend to rely on the bootstrap proposed by Jentsch and Lunsford (2022), which is based on the heteroskedasticity robust bootstrap proposed by Brüggemann et al. (2016). This approach relies on estimating Z and Σ to obtain an estimate for B_{11}^c (see e.g. Piffer and Podstawski, 2018).

Yet, as the set-identification resulting from the FEV bounds, described in section 1.3, is handled with Bayesian techniques, I use these Bayesian techniques also for the Max-Share framework. The combination of the proxy VAR with the Max-Share framework can also be handled with the Bayesian Algorithm 1.1 described in section 1.3.3. In combination with the Max-Share framework, steps 2 and 3 of Algorithm 1.1 are simply replaced with carrying out the Max-Share optimization. Note that this reduces the robust Bayesian approach to conventional Bayesian inference. Step 3 in Algorithm 1.1, which constitutes the core of the robust Bayesian inference, computes the boundaries of the identified set and in the point-identification case the identified set is a singleton. Hence, computing the bounds reduces to the computation of the point estimates.

1.5. Empirical Illustrations

This section illustrates the usefulness of the above described econometric frameworks. The objective is to identify a conventional and an unconventional monetary policy (MP) shock during the Volcker and Greenspan era prior to the zero-lower bound

(ZLB) period. From today's standpoint further investigating this time period is of potential interest, given that we are likely away from the ZLB for the coming years and are in need for policy tools which successfully yield the simultaneous stabilization of real aggregates and inflation under these conditions.

The time period between the 1970s and the millennium was characterised by changes in the objectives of monetary policy makers as well as changing strategies to achieve those objectives. During the Volcker era, the two main goals to stabilize output and unemployment were appended by the goal to decelerate and to stabilize inflation as well. In doing so, the federal reserve also undertook a shift in their strategic operations. As pointed out by Goodfriend and King (2005), one key emphasis of monetary policymakers was to ensure the credibility of their policy decisions. The credible commitment to disinflation should reduce the potential for real contractions due to tightening monetary policy. Such commitment policies are ideologically comparable to the forward guidance policy since the ZLB period. Yet, credibility in the late 70s is more broadly understood as sticking to a clear monetary policy objective, i.e. disinflation, rather than specifying a clear path for a policy instrument to achieve the objective.

As also laid out by e.g. Goodfriend and King (2005), the federal reserve did not focus on a single specific policy instrument during the Volcker and Greenspan era. Although the federal funds rate remained a crucial factor in monetary policy decisions, during the Volcker period the federal reserve also heavily targeted monetary aggregates instead. However, even before the Volcker era, monetary aggregates received emphasis in monetary policy decisions, and even after Volcker, the federal reserve did not explicitly refrain from money growth targeting. Yet, e.g. Thornton (2006) argues that the monetary policy decision nevertheless focused on the federal funds rate for major parts of the considered time period, especially since 1982. Hence, monetary policy throughout the last three decades of the 20th century was a mix of federal funds rate targeting, a focus on the growth of monetary aggregates together with the new commitment and credibility strategy.

The objective of this analysis is to differentiate between the effects of the federal funds rate targeting and the other unconventional policy measures that were implemented during this period. The unconventional measures - from today's perspective - are the reserve targeting and the commitment efforts, which the unconventional MP shock should reflect. This resembles recent approaches to dissect monetary policy shocks into conventional and unconventional policy shocks for the ZLB period. The VAR literature which aims to identify MP shocks in the 70s and 80s, primarily

assesses whether monetary policies' effectiveness changed during that time period, see e.g. Boivin and Giannoni (2006), Clarida et al. (2000), Sims and Zha (2006) or Barakchian and Crowe (2013), and not whether the effects differed for different types of MP shocks. Other studies that estimate the effects of general MP shocks for that periods are e.g. Leeper et al. (1996), Bernanke et al. (2005) and of course the well-known work by Christiano et al. (1999), Uhlig (2005) and Romer and Romer (2004).

In order to identify the two MP shocks, I resort to two previously used proxies in the literature. Following Stock and Watson (2012) I use the Romer and Romer (2004) monetary policy surprises and the monetary policy shock from Sims and Zha (2006). These two instruments are selected as they are available in monthly frequency from 1969:M1 to 1996:M12. This time period encompasses the interesting Volcker era, the majority of the Greenspan years as well as most of the build up of the great inflation. The vector y_t in the following applications is thus:

$$y_t = \begin{pmatrix} \text{FFR}_t \\ \Delta \log(\text{M2}_t) \\ \text{5YEAR}_t \\ \text{UER}_t \\ \Delta \log(\text{NBR}_t) \\ \Delta \log(\text{CPI}_t) \\ \Delta \log(\text{IP}_t) \end{pmatrix}. \tag{1.5.1}$$

Hence, for the specified time period, the federal funds rate (FFR), the 5-year treasury bill interest rates (5YEAR) and the unemployment rate (UER) enter the VAR model in levels. Furthermore, the VAR contains the growth rates of industrial production (IP), the consumer price level (CPI), the non-borrowed reserves (NBR) and the growth rate of the money stock measured by M2. A plot of the data which is used in the analysis can be found in Figure 1.C.11 of Appendix 1.C.

In terms of strength of the instruments, I conduct the weak instrument test described in Lunsford (2015). Following Stock et al. (2002) and Stock and Yogo (2005), he constructs critical values for a weak instrument test in the proxy VAR. The F-statistic for the Sims and Zha (2006) proxy is 34.94 and free from any weak instrument concerns. The Romer and Romer (2004) shock series has an F-statistic of 8.68. According to the critical values by Lunsford (2015), less than 10% of the weak proxy bias remains with a probability greater than 95%. Yet, this test is set up to assess

the predictive power of a single instrument and it is unclear if the drawn conclusions carry over to multiple instrument applications.

To collect further evidence that the instruments have jointly sufficient predictive power for the underlying shocks of interest I regress the residuals of the policy-relevant variables on both proxies. This strategy is, for example, also employed by Lakdawala (2019) in a two instruments setting. I check the predictive power of the instruments for the residuals of the federal funds rate, the 5-year treasury bill interest rate and the non-borrowed reserves. The latter was targeted especially under Volcker to control the monetary aggregates and the medium-term 5-year interest rate is especially sensitive to longer-term commitments of the federal reserve to the future path and objectives of their policy. The F-statistic for the federal funds rate residuals is 122.97, for the 5-year treasury bill yield it is 16.63 and for the non-borrowed reserves it is 9.25. The proxies have considerable predictive power for the residuals of these three parameters of monetary policy, yet the relation to the reserves might be a bit weaker than that observed for the interest rates, potentially due to the fact that explicit reserve targeting was not carried out over the entire time period considered here. In summary, I conclude that the proxies do not suffer from a weak instrument problem but seem to have a bit more predictive power for the federal funds rate targeting and the credibility effects of unconventional monetary policy than for the reserve targeting.

In the literature focusing on the ZLB period, unconventional monetary policy shocks are distinguished from conventional ones with different identification strategies. Baumeister and Benati (2013) and Kapetanios et al. (2012) employ sign restrictions with one key exclusion restriction at impact. They identify the unconventional shock as a pure spread shock, which only affects the spread between the long- and short-term government bond yield but not short-term interest rates, i.e. the federal funds rate at impact. Such a shock is likely to pick up the large scale assets purchase (LSAP) programs which affected expectations about the future state of the economy what in turn altered longer term interest rates. Yet, it also should reflect the central banks forward guidance (FG) communication for the same reasons. The exclusion restrictions builds on the fact that during the ZLB period the short-term interest rate was bounded to zero.

A similar zero restriction is imposed by Lakdawala (2019) in a proxy VAR approach, who identifies a conventional monetary policy shock and a FG shock assuming that the FG shock has no direct effect on the federal funds rate at impact. He bases the exclusion restriction on the definition of FG, which should only lay out the future

path of the policy instrument. Yet, given the particularities of the ZLB period, the shock might also reflect effects of other policy instruments that did not immediately alter the federal funds rate.

Imposing such an exclusion restriction as in Lakdawala (2019) or in the papers by Baumeister and Benati (2013) and Kapetanios et al. (2012) is not appropriate to identify the unconventional policy shock in the sample period considered in this analysis. Firstly, the federal funds rate fluctuated distinctly and was not near the ZLB. Furthermore, the unconventional policies differ from those in the ZLB period. On the one hand, the commitment efforts of the federal reserve are more holistic and aimed at inducing credibility for stabilizing policy now and in the future. Hence, they certainly convey information about the future stance of monetary policy but could also induce current actions of the federal reserve. On the other hand, targeting the reserves should also lead to fluctuations in the federal funds rate. A delay between setting reserves targets and changes in the federal funds rate is possible, yet not necessary.

Restricting the correlation structure between the instruments and the shocks like e.g. proposed by Piffer and Podstawski (2018) or Giacomini et al. (2022) is not possible either. A priori it is unclear which proxy contains more explanatory power for which of the two monetary policy shocks. Therefore, alternative identification strategies are required. The next two subsections illustrate how to tackle the identification problem with the augmented Max-Share approach and with a set of agnostic sign restrictions which are accompanied by bounds on the FEV decomposition.

1.5.1. Max-Share

The key identification assumption I make is that out of the two MP shocks the conventional federal funds rate shock contributes the most to the FEV of the federal funds rate in the short run. I maximize the contribution in the first year after the shocks' impact, hence up to $H = 12$. As shown by Figure 1.C.13 in Appendix 1.C, the results do not change when choosing a shorter maximization horizon, i.e. over the first four months with $H = 4$. Given the different types of monetary policy strategies reflected by these shocks I deem this assumption reasonable. Yet, at least in the medium run the federal funds rate is undoubtedly highly endogenous, as the policymakers react to the observed fluctuations in the economy. Therefore, this assumption would be less credible without ruling out the other confounding shocks via the proxy VAR.

As described by Proposition 1.2 the Max-Share framework will only correctly identify the underlying structural shocks if the conventional MP shock exclusively

contributes to the FEV of the federal funds rate. This is not a reasonable assumption from an economic perspective and the results of the baseline approach are likely to be biased. Hence, an additional inequality constraint is needed in order to remove the bias. I base the choice of the inequality constraint on the work of Swanson (2021), who estimates factors of monetary policy surprises for both the ZLB and the pre-ZLB period. During the ZLB he obtains three factors which he labels as a federal funds rate factor, a FG factor and a LSAP factor. For the pre-ZLB periods he only obtains two factors as no LSAPs were conducted during that time. He then estimates the immediate effects of these factors on different economic variables. I use the effects of the interest rate factor and the FG factor on treasury bills with different maturities in the pre-ZLB period as a guidance for the relative size of the impact effects of the two estimated MP shocks in this analysis. This assumes that the FG factor picks up the residual variation in the monetary policy surprises in the pre-ZLB period not explained by the federal funds rate factor, and thus reflecting the unconventional policy.

Table 4 and Panel (B) in the paper by Swanson (2021) shows that the federal funds rate factor and the FG factor affect the treasury yields differently depending on their term structure. The federal funds rate factor affects the short-term rate more pronounced compared to the FG factor. Yet, the longer-term yields are more strongly affected by the FG factor, a relationship that clarifies the longer the maturity horizon becomes. Economically this pattern is reasonable as commitment policies should significantly affect rates with longer horizons as they give more information about future actions of the policymakers. Furthermore, it is reasonable that policies, which do not invoke a strong and immediate shift in the federal funds rate, do not alter short-term interest rates as strongly because these short-term rates are closely tied to the federal funds rate. A similar pattern of how the term structure is affected by different monetary policy factors is also found by Gürkaynak et al. (2005). From the five yields considered by Swanson (2021), three are available for the time period considered in this analysis. The 6-month, 5-year and 10-year yields are candidates for constructing an inequality constraint. Hence, I impose the inequality constraint stipulating that the unconventional MP shock has a more pronounced impact on the longer term yields compared to the conventional MP shock. Conversely, for the 6-month yield the opposite is true, i.e. that the conventional MP shock has the more pronounced impact.

In order to assess which of these possible constraints is suitable for eliminating the Max-Share bias, I include these three yields into the VAR model separately. Meaning

that for the 10-year and 6-month yield I replace 5YEAR_t in (1.5.1). Then for each variable I resort to Algorithm 1.4 to assess whether the inequality constraint on them satisfies condition (a) of Proposition 1.2. Only the 5-year and 10-year yields satisfy this condition and I choose the 5-year rate as the baseline specification, yet Figure 1.C.12 in Appendix 1.C displays the estimates for the 10-year rate which do not differ substantially from the baseline results. The following list illustrates the outcomes of Algorithm 1.4 for the 5-year yield.

Employing Algorithm 1.4

1. The baseline Max-Share identification establishes that $q_{21}^{max} \leq 0$.
2. Normalize q_2 by setting $q_{22} = q_{11}^{max}$ and $q_{12} = -q_{21}^{max}$.
3. $\eta_{5YEAR,1,0} > 0$.
4. $\eta_{5YEAR,2,0} > 0$. The signs are the same, and thus stick with the normalization of q_2 according to step 2.
5. $\Theta_D = \Theta_{5YEAR,1,0} - \Theta_{5YEAR,2,0} > 0$ and $\Theta_S = \Theta_{5YEAR,1,0} + \Theta_{5YEAR,2,0} > 0$. Note that $\Theta_D > 0$ indicates that the imposed restriction is binding. $\Theta_S > 0$ is just a technical condition which indicates whether the underlying function that is constrained is monotonic or not.

6.2 $q_{21}^{max} < 0$:

(i) The signs Θ_D and Θ_S are the same and $q_{22} = q_{11}^{max}$,

where shock $j = 1$ depicts the conventional MP shock and $j = 2$ the unconventional MP shock.

In order to specify the margin ϵ for my application the results by Swanson (2021) need to be scaled to the current application. The unrestricted Max-Share approach gives a response of the 5-year rate to the unconventional MP shock of roughly 20 basis points (bp) at impact. I scale the estimates of Table 4 and Panel (B) of Swanson (2021) such that the FG factor estimate matches this 20 bp response. The resulting difference between Swanson's scaled FG factor and FFR factor estimate on the 5-year yield is conservatively 9bp. To account for the estimation uncertainty in the underlying estimates as well as the imprecision introduced through the scaling, I add and subtract 2bp and present the results for this range of ϵ values. The resulting

range for the margin ϵ is 7bp $\leq \epsilon \leq$ 11bp, and thus at impact the responses of the 5-year treasury bill rate to unconventional shock are constrained to exceed the response of the conventional shock by this margin.

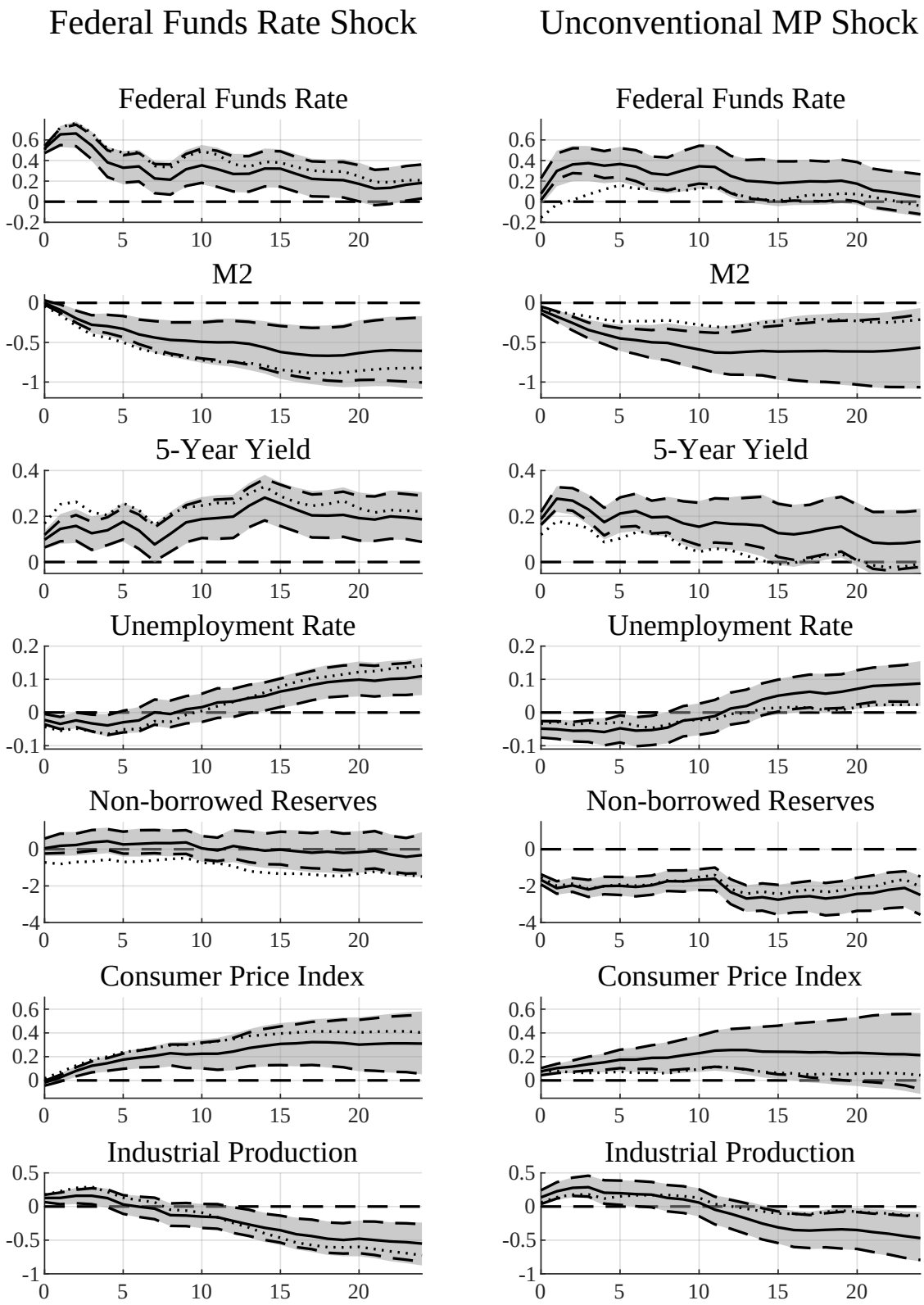
In the application the maximization problem defined by (1.4.2) is employed, maximizing the contribution of the conventional monetary policy shock to the FEV of the federal funds rate up to twelve months after impact and subject to the inequality constraint $IQ(q_{11}) = \eta_{5YEAR,2,0} - \eta_{5YEAR,1,0} > \epsilon$.

In Figure 1.1 the solid black line is the median target response for $\epsilon = 9$bp and the dashed black lines depict the respective 68% credible regions. The Bayesian quantities are derived with Algorithm 1.1, where step 2 and 3 are simply replaced by carrying out the augmented Max-Share identification. The dotted black line depicts the median responses of the baseline Max-Share framework without the inequality constraint based on ϵ. The shaded area depicts the range of the 68% credible regions for the different ϵ values within the considered range.

To discuss certain known features of the used proxies first, it should be noted that the responses of the price level are not in line with economic theory and show the well-known price puzzle. This is not unexpected as the Romer and Romer (2004) series is known to evoke this phenomenon as e.g. described by Ramey (2016). Another noteworthy feature of the Romer and Romer (2004) shock series is its expansionary effect on real activity in the first year after impact. However, my results only display initial expansionary effects which are limited to the first quarter after the shocks hit the economy. These unintuitive results remain when including additional variables like e.g. commodity prices. Furthermore, an inequality constraint on the CPI that enforces a negative response of the price level fails as the constrained Max-Share optimization problem does not have a solution. The price puzzle appears to be a feature of the underlying data and similarly to Lakdawala (2019), who is also confronted with the price puzzle in his analysis, I focus on the responses of the other real variables instead.

The remaining responses are consistent with economic intuition. The federal funds rate increases due to contractionary MP shocks and it does so more for the federal funds rate shock. For the unconventional MP shock the response of the federal funds rate is delayed, however, the initial response to a one standard deviation shock is nevertheless approximately between 0 and 0.2. As determined by the inequality constraint, the unconventional MP shock has a more pronounced effect on the 5-year treasury yield on impact, yet the effects seem to be more persistent for the conventional shock. Regarding the monetary aggregates, I observe a decline in the monetary base M2 for both MP shocks. Yet, a distinct drop in reserves is observed

Figure 1.1: Impulse Responses to the two MP Shocks - Max-Share



Note: The solid black line is the median target response for $\epsilon = 9$bp. The dashed black lines depict the respective 68% credible regions. The dotted lines depict the median responses of the baseline Max-Share approach without an additional inequality constraint. The grey shaded areas depict the range of the 68% credible region over a grid of ϵ values between 7bp and 11bp.

only for the unconventional shock. This is in line with the explicit targeting of the monetary aggregates which should be attributed to the unconventional MP shock. From a policy directive during the Volcker era one can infer that the goal of the federal reserve was to decelerate the growth of reserve aggregates in conjunction with the monetary base targets.

When examining the real variables, it can be seen that the contractionary MP shocks have adverse effects on unemployment rates which seem to be slightly more pronounced for the federal funds rate shock. After the initial expansionary tendencies, the impulse responses show that both MP shocks decrease real activity measured by industrial production. The literature suggests small to medium effects of MP shocks on real activity for this time period, see e.g. Ramey (2016) for an overview. Hence, it is reasonable that the effects in this analysis appear moderate, especially as the overall effects of MP shocks are apportioned to two different shocks.

Comparing the baseline Max-Share results (dotted lines) with the augmented approach, reveals the direction in which the bias is corrected. For a large portion of the impulse responses the bias is not substantial in a sense that the baseline median still lies inside the credible regions of the augmented approach. However, some responses are noteworthy. Firstly, the inclusion of the inequality constraint credits a delayed but more distinct response of the federal funds rate to the unconventional MP shock what is in line with the economic intuition. Secondly, it attributes the response of the non-borrowed reserves entirely to the unconventional MP policy shock which reflects the explicit reserve targeting of the unconventional MP shock. Finally, the inequality constraint gives greater importance to the unconventional MP shock in the explanation for the responses of the monetary base M2. Moreover, the bias correction reveals the effect of the unconventional MP shock on the unemployment rate and real activity. For the later horizons the median from the baseline results is closer to zero in both cases, near the border of the credible region of the augmented approach. All directions in which the bias is corrected are in line with the notion of the MP shocks considered in this analysis and the respective economic intuition.

1.5.2. Bounds on the FEV

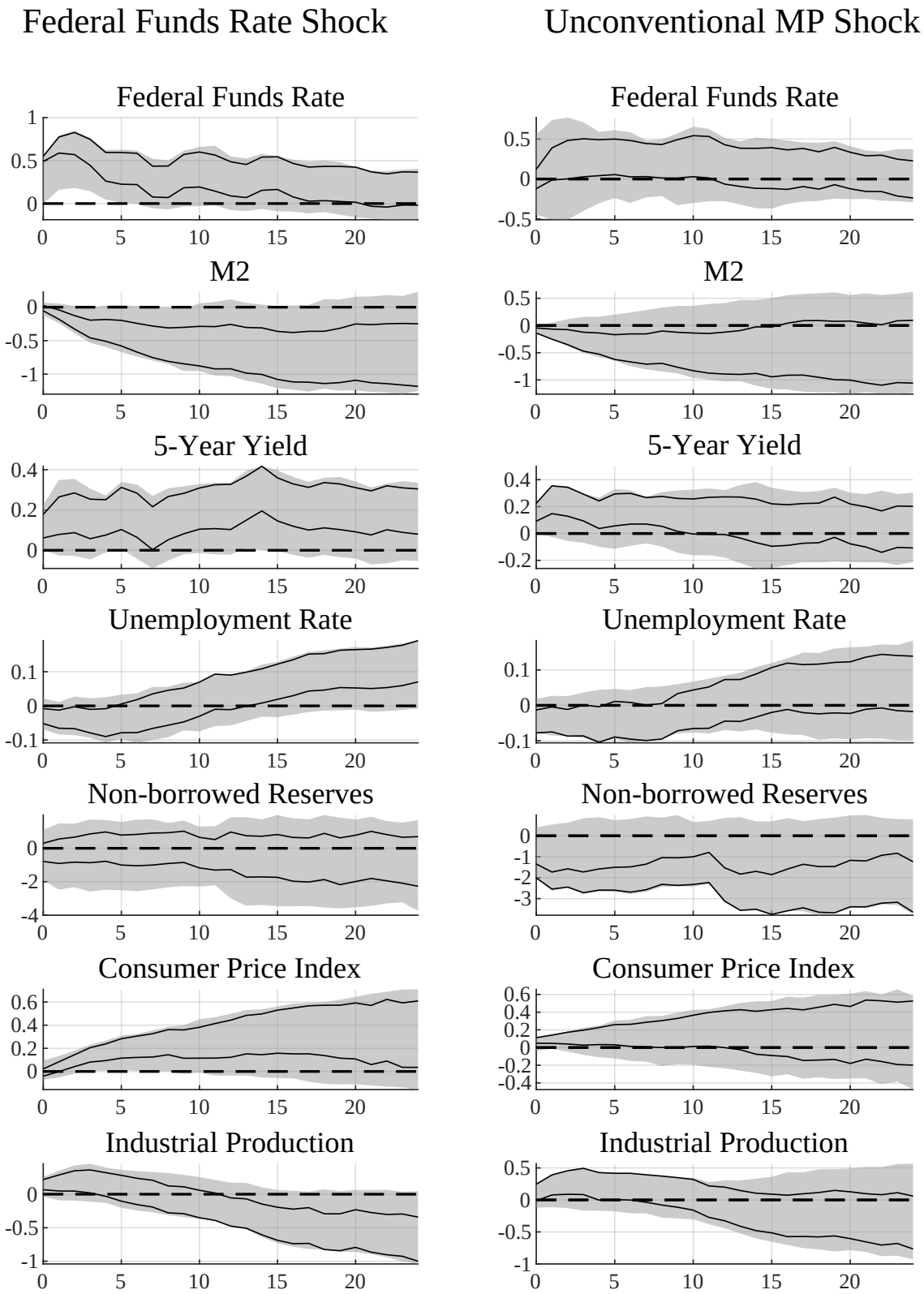
This subsection disentangles the conventional and unconventional MP shock with a small set of sign restrictions, paired with a targeted bound on the FEV. The agnostic sign restrictions induce that a contractionary conventional MP shock affects the response of the federal funds rate and the 5-year yield positively at impact. The contractionary unconventional MP shock decreases the initial response of M2 growth

and also increases the 5-year treasury yield at impact. This is a set of very agnostic sign restrictions which does not include any restriction on relative impulse responses, in contrast to the previous augmented Max-Share identification. Such restrictions are excluded on purpose in order to showcase that a targeted bound on the FEV helps to sharpen the identification even with such an agnostic set of baseline restrictions. Note that again the price puzzle is not corrected with sign restrictions on the price level. Given the data and the already instituted sign restrictions, no rotations exist which admit these negative effects of the price level to the two MP shocks.

The bound on the FEV is derived from the notion of the two different MP shocks in this analysis. The unconventional MP shock should reflect the monetary policy instruments other than the federal funds rate targeting. Previously mentioned studies identify unconventional MP shocks with the exclusion restriction that these shocks have no immediate impact on the federal funds rate. This is equivalent to imposing that at $h = 1$ the unconventional MP shock does not contribute to the FEV of the federal funds rate. As previously stated, this strict assumption is unlikely to hold true for the time period considered in this analysis as the interest rates were not at the ZLB. Yet, the commitment efforts and reserve targeting do not target the federal funds rate immediately, and thus only modest immediate effects on it are expected. Hence, the contribution of the unconventional MP shock to the FEV of the federal funds rate is likely to be small. I therefore impose bounds such that $\Omega^z_{FFR,2}(h = 1) \leq 0.05$, where shock $j = 2$ again depicts the unconventional MP shock. However, the choice of 5% maximum contribution is not based on any theoretical considerations, and thus Figure 1.C.14 in Appendix 1.C also provides estimates for $\Omega^z_{FFR,2}(h = 1) \leq 0.1$ in order to assess the robustness of the results.

Figure 1.2 depicts the 68% robust Bayesian credible regions for the identification with only sign restrictions and for the combination of sign restrictions and the bound on the FEV. The Bayesian quantities are derived by Algorithm 1.1 with the approximation explained in Algorithm 1.2. The dashed lines depict the credible regions based on the inclusion of the FEV bound while the grey shaded areas reflect the identification with only sign restrictions. Posterior means are omitted for clarity. The impulse responses appear to be largely similar to those identified with the augmented Max-Share approach. Hence, the two different strategies seem to identify the same shocks. Again, the prize puzzle and initial expansionary effects of the contractionary MP shocks are present. Yet, after a period of months, one observes the expected decrease in industrial production and increase in the unemployment rate.

Figure 1.2: Impulse Responses to the two MP Shocks - Bound=0.05



Note: The solid black line the 68% robust Bayesian credible regions for the identification including the bounds constraint. The grey shaded areas depict the 68% robust Bayesian credible regions for the identification based purely on sign restrictions. Posterior medians are omitted for clarity of the plot.

Comparing the grey shaded areas based on the very agnostic sign restrictions with the dashed lines, reveals that the single targeted bound on the FEV helps to carve out the two monetary policy shocks more precisely, particularly the impulse responses of the unconventional MP shock. Hence, as also shown by Volpicella (2022), the bound constraints can be a powerful tool to sharpen the identification in set-identification schemes.

1.6. Conclusion

Identifying restrictions on the FEV can be used in multiple ways to disentangle the shocks in the proxy VAR framework. Firstly, bounds on the FEV, as introduced by Volpicella (2022), replace or accompany other inequality restrictions to disentangle the shocks in the proxy VAR. This chapter provides the general framework of the bounds applied to the proxy VAR and outlines how to conduct robust Bayesian inference in this setting.

Secondly, the structural parameters are sharply point-identified in the proxy VAR with the help of the Max-Share framework. For this approach, the chapter only considers the case of two shocks being identified using two instruments, which is a highly relevant scenario in practice. I state under which conditions the Max-Share approach successfully identifies the structural shocks. If the baseline Max-Share approach suffers from bias due to confounding shocks an augmentation with an inequality constraint helps to remove or reduce the bias. I provide sufficient conditions under which the inequality constraint successfully removes or reduces the bias. This chapter further illustrates how to assess these conditions for two candidates of inequality constraints - for an inequality constraint on one impulse response at one horizon and for a constraint on the difference of impulse responses of two shocks on a specified variable. Furthermore, it outlines how to conduct Bayesian inference for the proxy VAR and Max-Share combination.

Finally, an empirical illustration identifies two types of monetary policy shocks in the pre-ZLB period. In detail, I identify a conventional federal funds rate shock and an unconventional shock, which reflects the effects of unconventional policy measures like reserve targeting as well as early commitment efforts of the federal reserve. For the identification, I employ the augmented Max-Share framework as well as an agnostic sign restriction approach in conjunction with a targeted bound restriction on the FEV. Both identification strategies reveal that both types of monetary policy shocks affect real variables, such as industrial production and unemployment rates. However, for

the federal funds rate shock, these effects are accompanied with a distinct increase in the federal funds rate while the unconventional MP shock exhibits a distinct decline in non-borrowed reserves.

References

An, S. and F. Schorfheide (2007). "Bayesian Analysis of DSGE Models". *Econometric Reviews* 26. (2-4), 113–172.

Barakchian, S. M. and C. Crowe (2013). "Monetary Policy Matters: Evidence from New Shocks Data". *Journal of Monetary Economics* 60. (8), 950–966.

Barsky, R. B. and E. R. Sims (2011). "News Shocks and Business Cycles". *Journal of Monetary Economics* 58. (3), 273–289.

Baumeister, C. and L. Benati (2013). "Unconventional Monetary Policy and the Great Recession: Estimating the Macroeconomic Effects of a Spread Compression at the Zero Lower Bound". *International Journal of Central Banking* 9, 165–212.

Baumeister, C. and J. D. Hamilton (2015). "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information". *Econometrica* 83. (5), 1963–1999.

Ben Zeev, N. and E. Pappa (2017). "Chronicle of a War Foretold: The Macroeconomic Effects of Anticipated Defence Spending Shocks". *The Economic Journal* 127. (603), 1568–1597.

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.

Boivin, J. and M. P. Giannoni (2006). "Has Monetary Policy Become More Effective?" *The Review of Economics and Statistics* 88. (3), 445–462.

Brüggemann, R., C. Jentsch, and C. Trenkler (2016). "Inference in VARs with Conditional Heteroskedasticity of Unknown Form". *Journal of Econometrics* 191. (1), 69–85.

Carriero, A. and A. Volpicella (2024). "Max Share Identification of Multiple Shocks: An Application to Uncertainty and Financial Conditions". *Journal of Business & Economic Statistics* 0, 1–13.

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*. Vol. 1. Elsevier, 65–148.

Clarida, R., J. Gali, and M. Gertler (2000). "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory". *The Quarterly Journal of Economics* 115. (1), 147–180.

Dieppe, A. M., F. Neville, and G. J. Kindberg Hanlon (2019). "New Approaches to the Identification of Low-Frequency Drivers : An Application to Technology Shocks". *The World Bank - Policy Research Working Paper 9047*.

Faust, J. (1998). "The Robustness of Identified VAR Conclusions about Money". *Carnegie-Rochester Conference Series on Public Policy* 49, 207–244.

Francis, N. and G. Kindberg-Hanlon (2022). "Signing Out Confounding Shocks in Variance-Maximizing Identification Methods". *AEA Papers and Proceedings* 112, 476–80.

Francis, N., M. T. Owyang, J. E. Roush, and R. DiCecio (2014). "A Flexible Finite-Horizon Alternative to Long-Run Restrictions with an Application to Technology Shocks". *The Review of Economics and Statistics* 96. (4), 638–647.

Gertler, M. and P. Karadi (2015). "Monetary Policy Surprises, Credit Costs, and Economic Activity". *American Economic Journal: Macroeconomics* 7. (1), 44–76.

Giacomini, R. (2013). "The Relationship between DSGE and VAR Models". In: *Advances in Econometrics*. Vol. 32, 1–25.

Giacomini, R., T. Kitagawa, and M. Read (2022). "Robust Bayesian Inference in Proxy SVARs". *Journal of Econometrics* 228. (1), 107–126.

Goodfriend, M. and R. G. King (2005). "The Incredible Volcker Disinflation". *Journal of Monetary Economics* 52. (5), 981–1015.

Gürkaynak, R. S., B. Sack, and E. Swanson (2005). "Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements". *International Journal of Central Banking* 1. (1), 55–93.

Jentsch, C. and K. G. Lunsford (2022). "Asymptotically Valid Bootstrap Inference for Proxy SVARs". *Journal of Business & Economic Statistics* 40. (4), 1876–1891.

Kapetanios, G., H. Mumtaz, I. Stevens, and K. Theodoridis (2012). "Assessing the Economy-wide Effects of Quantitative Easing". *The Economic Journal* 122. (564), 316–347.

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

Komunjer, I. and S. Ng (2011). "Dynamic Identification of Dynamic Stochastic General Equilibrium Models". *Econometrica* 79. (6), 1995–2032.

Lakdawala, A. (2019). "Decomposing the Effects of Monetary Policy Using an External Instruments SVAR". *Journal of Applied Econometrics* 34. (6), 934–950.

Leeper, E. M., C. A. Sims, T. Zha, R. E. Hall, and B. S. Bernanke (1996). "What Does Monetary Policy Do?" *Brookings Papers on Economic Activity* 1996. (2), 1–78.

Lunsford, K. G. (2015). "Identifying Structural VARs with a Proxy Variable and a Test for a Weak Proxy". *Federal Reserve Bank of Cleveland, Working Paper No. 15-28*.

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), 1212–47.

Montiel Olea, J. L. and J. Nesbit (2021). "(Machine) Learning Parameter Regions". *Journal of Econometrics* 222. (1), 716–744.

Piffer, M. and M. Podstawski (2018). "Identifying Uncertainty Shocks Using the Price of Gold". *The Economic Journal* 128. (616), 3266–3284.

Ramey, V. (2016). "Macroeconomic Shocks and Their Propagation". In: *Handbook of Macroeconomics*. Vol. 2. Elsevier, 71–162.

Romer, C. D. and D. H. Romer (2004). "A New Measure of Monetary Shocks: Derivation and Implications". *American Economic Review* 94. (4), 1055–1084.

Sims, C. A. and T. Zha (2006). "Were There Regime Switches in U.S. Monetary Policy?" *American Economic Review* 96. (1), 54–81.

Stock, J. H. and M. Watson (2012). "Disentangling the Channels of the 2007-2009 Recession". *Brookings Papers on Economic Activity* Spring 2012, 81–135.

Stock, J. H., J. H. Wright, and M. Yogo (2002). "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments". *Journal of Business & Economic Statistics* 20. (4), 518–529.

Stock, J. H. and M. Yogo (2005). "Testing for Weak Instruments in Linear IV Regression". In: *Identification and Inference for Econometric Models*. Ed. by Andrews, D. W. New York: Cambridge University Press, 80–108.

Swanson, E. T. (2021). "Measuring the Effects of Federal Reserve Forward Guidance and Asset Purchases on Financial Markets". *Journal of Monetary Economics* 118. (C), 32–53.

Thornton, D. L. (2006). "When Did the FOMC Begin Targeting the Federal Funds Rate? What the Verbatim Transcripts Tell Us". *Journal of Money, Credit and Banking* 38. (8), 2039–2071.

Uhlig, H. (2004a). "Do Technology Shocks Lead to a Fall in Total Hours Worked?" *Journal of the European Economic Association* 2. (2-3), 361–371.

— (2004b). "What Moves GNP?" *Econometric Society 2004 North American Winter Meetings* (No. 636).

— (2005). "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure". *Journal of Monetary Economics* 52. (2), 381–419.

Volpicella, A. (2022). "SVARs Identification Through Bounds on the Forecast Error Variance". *Journal of Business & Economic Statistics* 40. (3), 1291–1301.

Appendix 1.A Proofs and Remarks

Proof of Proposition 1.1

Proof. The proof of Proposition 1.1 closely follows the ideas of the proof of Proposition 3.1 in the article by Volpicella (2022). The contribution of shock j to the FEV of variable i at horizon h is given by:

$$\Omega^z_{i,j}(h) = q_j' R_{i,h} q_j,$$

where $R_{i,h}$ is a positive semidefinite symmetric $l \times l$ real matrix. As $R_{i,h}$ is symmetric it can be diagonalized such that:

$$P' R_{i,h} P = D,$$

where P is an orthogonal matrix and D a diagonal matrix with the real eigenvalues λ^{ih}_m of matrix $R_{i,h}$ as entries on the diagonal for $m = 1, ..., l$. The rest of the proof follows from the proof of Proposition 3.1 by Volpicella (2022). □

To aid with the proof of Proposition 1.2, I firstly define the conditions under which a shock does not contribute to the FEV of a specified variable.

Definition 1.1. *Let C_{ih} be the i-th row of the moving average coefficient matrix C_h and B^z_j the j-th column of the proxy impact matrix B_z. Shock j does not contribute to the FEV of variable i at horizon $\overline{h}$ if $C_{i,h} B^z_j = 0 \ \forall \ h \leq \overline{h}$.*

As C_0 is just the identity matrix I_k, $C_{i,0}$ is just a row of I_k. Hence, under Definition 1.1 the i-th element of B^z_j has to be zero.

For example, consider the case with two instruments in a three variable system. The second shock does not contribute to the FEV of the first variable. Definition 1.1 requires B_z to have the following structure:

$$B_z = \begin{pmatrix} * & 0 \\ * & * \\ * & * \end{pmatrix}.$$

Proof of Proposition 1.2

Proof. Suppose the two proxies z_1 and z_2 are valid and satisfy conditions (1.2.3) of the main text. Hence, the proxy VAR alone will identify the true proxy impact matrix B_z up to a rotation with an orthonormal matrix Q. For simplicity and without loss of generality assume that the initial rotation matrix Q is just the identity matrix such that B_{11}^c is the Cholesky decomposition of $B_{11}B_{11}' = \Sigma_{11} - B_{12}B_{12}'$. Furthermore, without loss of generality assume that the first two shocks of the system are related to the proxy variable and that the second shock does not contribute to the FEV of the first variable over horizon $\underline{h}$ up to $\overline{h}$. Definition 1.1 requires the structure of the proxy impact matrix to be:

$$B_z = \begin{pmatrix} * & 0 \\ * & * \\ \vdots & \vdots \\ * & * \end{pmatrix}.$$

As the first shock contributes exclusively to the FEV of the first variable, the Max-Share approach maximizes the contribution of the first shock to the FEV of the first variable.

Recall that the contribution of the first shock to the FEV of the first variable at horizon h can be written as:

$$\Omega_{1,1}^z(h) = q_1' R_{1,h} q_1.$$

Uhlig (2004b) shows that the sum over all $\Omega_{1,1}^z(h)$ from horizons $\underline{h}$ up to $\overline{h}$ can be expressed by:

$$\sum_{h=\underline{h}}^{\overline{h}} \Omega_{1,1}^z(h) = q_1' S q_1,$$

where

$$S = \sum_{h=0}^{\overline{h}} (\overline{h} + 1 - \max(\underline{h}, h))(e_1 C_h B_z)'(e_1 C_h B_z) = \sum_{h=0}^{\overline{h}} (\overline{h} + 1 - \max(\underline{h}, h)) S_h,$$

with e_1 being the first row of the identity matrix I_k. Hence, $e_1 C_h B_z$ is just the first row of $C_h B_z$.

For $h = 0$, C_0 is just the identity matrix and $e_1 C_0 B_z = (* \; 0)$, and thus

$$S_0 = \begin{pmatrix} * & 0 \\ 0 & 0 \end{pmatrix}.$$

If $h > 0$, due to Definition 1.1, $e_1 C_h B_z = (* \; 0)$ for $h < \overline{h}$, and thus

$$S_h = \begin{pmatrix} * & 0 \\ 0 & 0 \end{pmatrix} \quad \forall \; h < \overline{h}.$$

As S is just the weighted sum over all S_h, it follows that:

$$S = \begin{pmatrix} * & 0 \\ 0 & 0 \end{pmatrix}.$$

Uhlig (2004b) shows that finding the rotation vector q_1^* that maximizes the sum over the contributions

$$q_1^* = \text{argmax } q_1' S q_1 \quad \text{s.t.} \quad q_1' q_1 = 1$$

amounts to finding the eigenvector associated with the largest eigenvalue of the matrix S. Due to the particular structure of the matrix S, which induced by the exclusive contribution, it can easily be seen that the eigenvector associated with the non-zero eigenvalue is just the first column of the identity matrix I_2. Hence, the rotation matrix Q^* is just the identity matrix I_2 up to a sign normalization of the columns. Thus, the maximum of the contribution to the FEV is achieved at the true proxy impact matrix parameters up to a sign normalization of the columns. □

This paragraph aims to illustrate that Proposition 1.2 also holds without the assumption that B_{11}^c is the Cholesky decomposition of $B_{11}B_{11}' = \Sigma_{11} - B_{12}B_{12}'$. If another solution to $B_{11}B_{11}' = \Sigma_{11} - B_{12}B_{12}'$ is chosen, it can be represented as a rotation of B_z with an initial rotation matrix Q_{init}:

$$B_{init} = B_z Q_{init} = \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \\ \vdots & \vdots \\ b_{k1} & b_{k2} \end{pmatrix} \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix}.$$

Again suppose that the first two shocks are the target shocks, from which the first shocks exclusively contributes to the FEV of the first variable. For $\underline{h} = \overline{h} = 0$ it can be nicely seen that in the process of the Max-Share optimization the initial rotation simply is reverted.

The first row of the rotated proxy impact matrix is $(b_{11}q_{11}\ b_{11}q_{12})$, and thus for $\underline{h} = \overline{h} = 0$:

$$S = \begin{pmatrix} b_{11}^2 q_{11}^2 & b_{11}^2 q_{11} q_{12} \\ b_{11}^2 q_{11} q_{12} & b_{11}^2 q_{12}^2 \end{pmatrix}.$$

It can be shown that the eigenvalues of S are $[b_{11}^2, 0]$ and a corresponding eigenvector to the non-zero eigenvalue is

$$q_1^* = \begin{pmatrix} q_{11} \\ q_{12} \end{pmatrix}.$$

Hence, the optimal rotation matrix due to the Max-Share identification is equal to the transposed initial rotation Q'_{init}, up to a sign normalization of the columns. Thus, the initial rotation is reverted as

$$B_{init}Q'_{init} = B_z Q_{init} Q'_{init} = B_z$$

and the true impact matrix parameters are identified. Again, up to a sign normalization of the columns.

Proof of Proposition 1.3

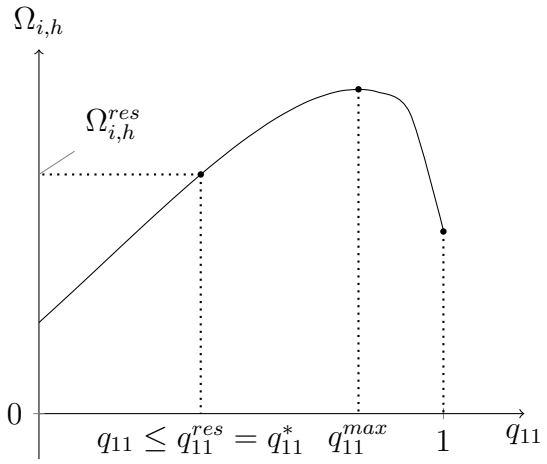
Proof. Let $q_{11} \in [0, 1]$. Without loss of generality assume that the first shock is the target shock in the maximization of the Max-Share approach. In general, the maximization problem of the Max-Share framework can be written as:

$$q_1^* = \text{argmax}\ q_1' S q_1 \quad \text{s.t.} \quad q_1' q_1 = 1, \tag{1.A.1}$$

where S is defined as above and is a symmetric and positive semidefinite matrix. Let:

$$S = \begin{pmatrix} s_{11} & s_{21} \\ s_{21} & s_{22} \end{pmatrix} \quad \text{and} \quad q_1 = \begin{pmatrix} q_{11} \\ q_{21} \end{pmatrix},$$

Figure 1.A.1: Constrained Maximization of $\Omega_{1,1}(h)$



where $q_{21} = \pm\sqrt{1 - q_{11}^2}$. Hence, the maximization problem in (1.A.1) can be written as:

$$q_1^* = \text{argmax } q_{11}^2 s_{11} + 2q_{11}q_{21}s_{21} + q_{21}^2 s_{22} \quad \text{s.t.} \quad q_1' q_1 = 1.$$

As S is positive semidefinite, $q_{11}^2 s_{11} \geq 0$ and $q_{21}^2 s_{22} \geq 0$. To achieve the maximum the middle term of the sum has to be positive. As $q_{11} \geq 0$, this term is positive if q_{21} and s_{21} have the same sign. Hence, $q_{21} = \pm\sqrt{1 - q_{11}^2}$ is implicitly normalized to be the same sign as s_{21} through to the maximization in the Max-Share framework.

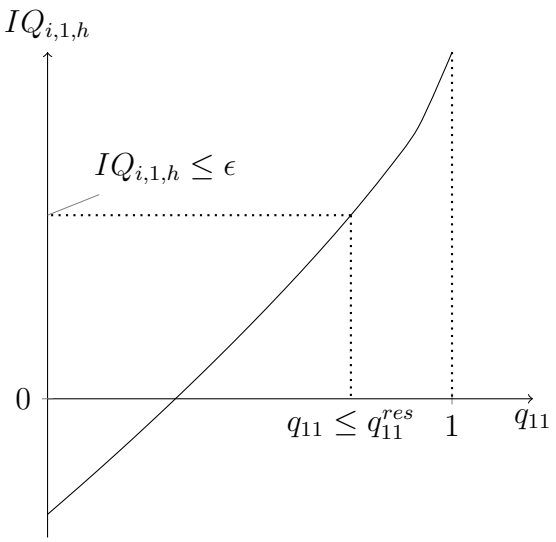
With the normalization of q_{21}, $\Omega^z_{1,1}(q_{11}) = q_1' S q_1$ has one extreme value at q_{11}^{max} and $\Omega^z_{1,1}(q_{11})$ increases monotonically at $q_{11} \in [0, q_{11}^{max}]$ and decreases monotonically at $q_{11} \in [q_{11}^{max}, 1]$.

Let the true rotation parameter, that yields the true structural parameters, be q_{11}^*. Suppose without loss of generality $q_{11}^* \leq q_{11}^{max}$. According to Condition (a) the inequality restriction $IQ_{i,1,h}(q_{11}) \lesseqgtr \epsilon$ implies a binding linear restriction on $q_{11} \leq q_{11}^{res}$, with $IQ_{i,1,h}(q_{11}^{res}) = \epsilon$. Conditions (b) requires that $q_{11}^{res} = q_{11}^*$, and thus $IQ_{i,1,h}(q_{11})$ implies the inequality constraint $q_{11} \leq q_{11}^*$. As $\Omega^z_{1,1}(q_{11})$ is a strictly increasing function at $q_{11} \in [0, q_{11}^{max}]$ the maximum of the maximization problem (1.A.1) augmented with $IQ_{i,1,h}(q_{11}) \lesseqgtr \epsilon$ (which implies $q_{11} \leq q_{11}^*$) is at q_{11}^*. Figure 1.A.1 illustrates the constrained maximization. The very same logic holds when $q_{11}^{max} \geq q_{11}^*$ and $IQ_{i,1,h}(q_{11}) \lesseqgtr \epsilon$ implies $q_{11} \geq q_{11}^{res}$. □

Proof of Proposition 1.4

Proof. Let $q_{11} \in [0,1]$ and $IQ_{i,1,h}(q_{11})$ be monotonic at $q_{11} \in [0,1]$. If $IQ_{i,1,h}(q_{11})$ is monotonically increasing it holds that $IQ_{i,1,h}(q_{11}^{low}) \leq IQ_{i,1,h}(q_{11}^{high})$ for all $q_{11}^{low} \leq q_{11}^{high}$.

Figure 1.A.2: Linear Constraint on q_{11}



Hence, a linear restriction $IQ_{i,1,h}(q_{11}) \leq \epsilon$ implies a linear restriction $q_{11} \leq q_{11}^{res}$ and $IQ_{i,1,h}(q_{11}) \geq \epsilon$ implies $q_{11} \geq q_{11}^{res}$. Figure 1.A.2 illustrates this relationship. The same logic holds true when $IQ_{i,1,h}(q_{11})$ is monotonically decreasing. □

Checking Condition (a) of Proposition 1.3

More intriguing than the proof of Proposition 1.4 is how to check it for different types of restrictions. The following describes how one can check whether an inequality constraint on $\eta_{i,1,h}$ or $D_{i,h}$ implies a linear constraint on q_{11} or not.

Firstly, one can check whether $\eta_{i,1,h}(q_{11})$ is monotonic. Let $q_{11} \in [0,1]$. Recall that the impulse response of variable i at horizon h to the first shock is $\eta_{i,1,h} = e_i' C_h B_z^c q_1$, with e_i being the first column of I_k and

$$C_h B_z^c = \begin{pmatrix} \Theta_{1,1,h} & \Theta_{1,2,h} \\ \Theta_{2,1,h} & \Theta_{2,2,h} \\ \vdots & \vdots \\ \Theta_{k,1,h} & \Theta_{k,2,h} \end{pmatrix}.$$

Hence, the impulse response function $\eta_{i,1,h}(q_{11})$ is given by:

$$\eta_{i,1,h} = \underbrace{\Theta_{i,1,h} q_{11}}_{\text{A}} + \underbrace{\Theta_{i,2,h} q_{21}}_{\text{B}}.$$

In practice, one can carry out the unrestricted Max-Share framework - with fixing q_{11} to be positive - and get q_{11}^{max} and q_{21}^{max} to calculate terms A and B. If the signs of A and B are the opposite $\eta_{i,1,h}(q_{11})$ is monotonic at $q_{11} \in [0,1]$, and thus a linear restriction on $\eta_{i,1,h}(q_{11})$ implies a linear restriction on q_{11}.

For the constraint on the difference in the impulse responses of variable i at horizon h to the two identified shocks $D_{i,h}(q_{11})$ one can also check whether the function $D_{i,h}(q_{11})$ is monotonic. The following describes the reasoning behind the steps of Algorithm 1.4 in the main text. The difference is:

$$D_{i,h} = \eta_{i,1,h} - \eta_{i,2,h} = e_i' C_h B_z^c q_1 - e_i' C_h B_z^c q_2.$$

This difference can also be written as:

$$\begin{aligned} D_{i,h} &= (\Theta_{i,1,h} q_{11} + \Theta_{i,2,h} q_{21}) - (\Theta_{i,1,h} q_{12} + \Theta_{i,2,h} q_{22}) \\ &= \Theta_{i,1,h}(q_{11} - q_{12}) + \Theta_{i,2,h}(q_{21} - q_{22}). \end{aligned}$$

Due to the properties of 2×2 orthonormal matrices $q_{22} = \pm q_{11}$ and $q_{12} = \pm q_{21}$. It must hold that if q_{12} and q_{22} have opposite signs, the sings of q_{11} and q_{21} are the same and vice versa. Hence, the relative sign of the entries of the second column q_2 of Q depend on the first column. Yet, whether q_{22} is negative or positive depends on the sign normalization of the second column. Dependent on the sign normalization of q_2 the difference can be written as a function of q_{11}:

$$D_{i,h}(q_{11}) = \underbrace{(\Theta_{i,1,h} - \Theta_{i,2,h})}_{\text{A}} q_{11} + \underbrace{(\Theta_{i,1,h} + \Theta_{i,2,h})}_{\text{B}} q_{21} \quad \text{if } q_{22} \geq 0 \tag{1.A.2}$$

$$D_{i,h}^{-}(q_{11}) = \underbrace{(\Theta_{i,1,h} + \Theta_{i,2,h})}_{\text{B}} q_{11} + \underbrace{(\Theta_{i,2,h} - \Theta_{i,1,h})}_{\text{-A}} q_{21} \quad \text{if } q_{22} \leq 0, \tag{1.A.3}$$

where again $q_{21} = \pm\sqrt{1 - q_{11}^2}$. The sign of q_{21} is determined by the Max-Share maximization, i.e. by q_{21}^{max}. The sign normalization, though, comes from the specification of the constraint. In order for $D_{i,h}$ to truly be a difference, $\eta_{i,1,h}$ and $\eta_{i,2,h}$ need to have the same sign. Hence, q_2 is normalized such that $\eta_{i,1,h} = C_h B_z q_1^{max}$ and $\eta_{i,2,h} = C_h B_z q_2$ have the same sign. From the normalization of q_2 one knows whether

the difference is written as (1.A.2) or (1.A.3). If one assesses (1.A.2) and q_{21}^{max} is positive, A and B need to have opposite sign in order for $D_{i,h}(q_{11})$ to be monotonic at $q_{11} \in [0,1]$. If q_{21}^{max} is negative the signs of A and B need to be the same. If one assesses (1.A.3) and q_{21}^{max} is positive, A and B need to have same sign in order for $D_{i,h}(q_{11})$ to be monotonic at $q_{11} \in [0,1]$. If q_{21}^{max} is negative the signs of A and B need to be the opposite.

If the sufficient condition of monotonicity does not hold for the chosen $IQ_{i,1,h}(q_{11})$ one can nevertheless assess whether condition (a) of Proposition 1.3 is satisfied.

Figure 1.A.3: Linear Constraint on q_{11} without Monotonicity

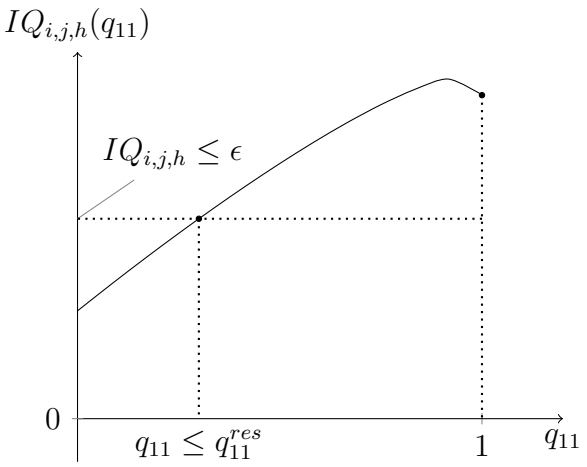


Figure 1.A.3 illustrates that $IQ_{i,j,h}(q_{11}) \leq \epsilon$ implies a linear restriction on q_{11} at $q_{11} \in [0,1]$ if the function value of $IQ_{i,j,h}(q_{11}) = \epsilon$ is unique, i.e. this equality is fulfilled for a single value q_{11}^{res} again at $q_{11} \in [0,1]$. Hence, if $IQ_{i,j,h}(q_{11}) \leq \epsilon$ does not constrain a monotonic function in q_{11}, checking whether the function value ϵ is unique at $q_{11} \in [0,1]$ allows to establish whether condition (a) of Proposition 1.3 is satisfied nevertheless.

Proof of Proposition 1.5

Proof. Let $q_{11} \in [0,1]$. Let placing a binding restriction on $IQ_{i,j,h}(q_{11}) \gtreqless \epsilon$ imply a linear restriction on $q_{11} \gtreqless q_{11}^{res}$, where q_{11}^{res} is again such that $IQ_{i,1,h}(q_{11}^{res}) = \epsilon$. $IQ_{i,j,h}(q_{11})$ is monotonic, and thus as $\epsilon \downarrow \epsilon^*$ one gets either $q_{11}^{res} \uparrow q_{11}^*$ or $q_{11}^{res} \downarrow q_{11}^*$, depending on whether $IQ_{i,j,h}(q_{11})$ is monotonically increasing or decreasing.

Solving the augmented maximization problem (1.4.2) from the main text yields q_{11}^{res} as the optimum. Hence, $|q_{11}^{res} - q_{11}^*| \to 0$ as $\epsilon \downarrow \epsilon^*$, meaning that the solution to (1.4.2) gets closer to the true value q_{11}^* as ϵ gets closer to its true value ϵ^*.

As $\eta_{i,j,h}(q_{11})$ is monotonic one gets either $\eta_{i,j,h}(q_{11}^{res}) \downarrow \eta_{i,j,h}(q_{11}^{*})$ or $\eta_{i,j,h}(q_{11}^{res}) \uparrow \eta_{i,j,h}(q_{11}^{*})$ as $|q_{11}^{res} - q_{11}^{*}| \to 0$, what is implied by $\epsilon \downarrow \epsilon^{*}$. Hence, the bias $\delta_{i,j,h} = |\eta_{i,j,h}(q_{11}^{res}) - \eta_{i,j,h}(q_{11}^{*})| \to 0$ as $\epsilon \downarrow \epsilon^{*}$. The same logic holds true for $\epsilon \uparrow \epsilon^{*}$. □

Appendix 1.B Simulation Results

The simulation study of this section serves the purpose of illustrating the Max-Share identification in the presence of two proxies. It illustrates the flaw in the baseline Max-Share approach and how the additional inequality constraint can be used as a remedy.

I simulate a trivariate system following Piffer and Podstawski (2018) who use the New Keynesian model by An and Schorfheide (2007) and Komunjer and Ng (2011). The model contains interest rates r_t, output x_t and inflation π_t. A TFP shock w_t^z, a government spending shock w_t^g and a monetary shock w_t^r are the structural shocks that hit the system. As pointed out by Giacomini (2013), calibrating the parameter gives the following DGP:

$$\begin{pmatrix} r_t \\ x_t \\ \pi_t \end{pmatrix} = \begin{pmatrix} 0.79 & 0 & 0.25 \\ 0.19 & 0.95 & -0.46 \\ 0.12 & 0 & 0.62 \end{pmatrix} \begin{pmatrix} r_{t-1} \\ x_{t-1} \\ \pi_{t-1} \end{pmatrix} + \begin{pmatrix} 0.61 & 0 & 0.69 \\ 1.49 & 1 & -1.16 \\ 1.49 & 0 & -0.75 \end{pmatrix} \begin{pmatrix} w_t^z \\ w_t^g \\ w_t^r \end{pmatrix}. \tag{1.B.1}$$

In contrast to Piffer and Podstawski (2018) I set the variance of the structural shocks to unity in order easily compute the actual contribution to the FEV decomposition. Hence, the structural shocks are drawn from a normal distribution with mean zero and unit variance and then used to simulate the data with equation (1.B.1). The instruments are constructed with:

$$\begin{aligned} m_{1t} &= \tau_1 w_t^z + (1-\tau_1) w_t^{(2)} + \tau_2 \nu_{1t} \\ m_{2t} &= (1-\tau_1) w_t^z + \tau_1 w_t^{(2)} + \tau_2 \nu_{2t}, \end{aligned}$$

where τ_1 governs the strength of the relation between the first to shocks and the instruments, while τ_2 governs the effect of the white noise disturbances ν_{1t} and ν_{2t}. I set $\tau_1 = 0.55$ and $\tau_2 = 0.01$ which leads to the proxies being sufficiently strong. The instruments are related to the respective structural shocks depending on the scenarios which are described below, i.e. $w_t^{(2)}$ will either be w_t^g or w_t^r.

The technology shock w_t^z is accountable for the major part of the FEV of the interest rate r_t. Up to horizon $H = 13$ the technology shock w_t^z contributes on average 82% to the FEV of the interest rate r_t. The government spending shock w_t^g does not contribute to the FEV of r_t at all and the monetary policy shock w_t^r contributes the remaining 18%. In this section, r_t is always the target variable in

Table 1.B.1: Two Shocks Simulated with Scenario A and B

DGP		Proxy + Max-Share		Proxy + Cholesky	
Scenario A					
0.61	0	0.609	0.004	0.609	0
		(0.007)	(0.004)	(0.007)	(0)
1.49	1	1.487	1.014	1.485	1.004
		(0.016)	(0.036	(0.012)	(0.03)
1.49	0	1.488	0.015	1.488	0.004
		(0.008)	(0.008)	(0.008)	(0.023)
Scenario B					
0.61	0.69	0.766	0.512	0.921	0
		(0.008)	(0.004)	(0.006)	(0)
1.49	-1.1	1.161	1.624	0.163	1.845
		(0.021)	(0.019	(0.022)	(0.016)
1.49	-0.75	1.25	1.578	0.426	1.6131
		(0.016)	(0.014)	(0.017)	(0.011)

Note: The table depicts the average of the estimated structural parameters over the 1,000 Monte-Carlo simulation iterations derived with the proxy VAR. The shocks are once disentangled with the baseline Max-Share approach and once with the Cholesky decomposition. The value in the parenthesis depicts the respective standard deviation.

the Max-Share framework. Hence, the underlying assumption is that the technology shock is the one that maximizes the contribution to the FEV of the interest rate. An interesting feature of this DGP is that the government spending shock does not contribute to the FEV of the interest rate and inflation at any horizon. This enables me to construct two scenarios:

> *Scenario A:* The proxies are constructed such that they are related to the technology shock w_t^z and the government spending shock w_t^g. Out of these two shocks, w_t^z contributes exclusively to the FEV of the interest rate r_t.
>
> *Scenario B:* The proxies are constructed such that they are related to the technology shock w_t^z and the monetary policy shock w_t^r. Both of these two shocks contribute to the FEV of the interest rate r_t.

Table 1.B.1 depicts the results for the two scenarios. The parameters are identified via maximization of the FEV after incorporating the information of the instruments as in (1.4.1) and without further inequality restrictions on relative magnitudes. As this

is more of a confirmation exercise in which cases the Max-Share approach succeeds and fails, I choose a large sample size of $T = 1,000$ with $M = 1,000$ Monte-Carlo iterations. The first two columns of the table show the true structural parameters of the DGP, the next two columns the combination of proxy VAR and Max-Share and the last two the identification via the Cholesky decomposition, i.e. $B_{11} = B_{11}^c$ is lower triangular. Looking at the results for Scenario A shows that after incorporating the proxies the shock that is not related to the proxies does pose a problem as a confounding shock. Only the two shocks that are related to the instruments are subject of the identification and as one of the shocks contributes exclusively to the FEV of the target variable the Max-Share approach is suitable to disentangle these two. Yet, this set-up with exclusive contribution implies a recursive structure for the contemporaneous impacts of the shock, and thus the Cholesky decomposition for the upper 2×2 block of the proxy impact matrix will also identify the true underlying structural parameters.

Hence, the potentially more interesting case is when both identified shocks contribute to the FEV of the target variable as in Scenario B. As seen in the panel for Scenario B, the Cholesky decomposition fails to identify the true impact matrix as there is no recursive structure between the two identified shocks. However, as expected also the combination of the proxy VAR with the Max-Share approach yields biased results because both shocks contribute to the FEV of the target variable. As pointed out by Dieppe et al. (2019), the Max-Share framework will be biased due to the confounding shock w_t^r.

To successfully disentangle the two shocks when both shocks contribute to the FEV, the Max-Share framework needs to be augmented. In this case I use an inequality constraint which restricts the difference in the contemporaneous responses of output $D_{2,0}$. In this DGP, a common, one standard deviation, expansionary technology shock has a more pronounced positive effect (1.49) than a common expansionary monetary policy shock (1.1). Augmenting the maximization problem such that this inequality constraint holds helps to disentangle the two shocks.

If the conditions of Proposition 1.3 hold, identification of the true underlying shocks is guaranteed. Algorithm 1.4 is used to assess condition (a):

Employing Algorithm 1.4

1. Carrying out the unrestricted Max-Share approach with $q_{11}^{max} \geq 0$ gives $q_{21}^{max} \geq 0$.

2. Normalize q_2 by setting $q_{22} = q_{11}^{max}$ and $q_{12} = -q_{21}^{max}$

3. $\eta_{2,1,0} = 1.161 \geq 0$.

4. $\eta_{2,2,0} = 1.612 \geq 0$. The signs are the same, and thus stick with the normalization of q_2 according to step 2.

5. $\Theta_D \leq 0$ and $\Theta_S \geq 0$.

6.1 $q_{21}^{max} > 0$:

(ii) The signs of Θ_D and Θ_S are the opposite and $q_{22} = q_{11}^{max}$.

Hence, $D_{2,0}(q_{11})$ is monotonic at $q_{11} \in [0, 1]$ and condition (a) of Proposition 1.3 is fulfilled.

Knowing the DGP makes it easy to also fulfil condition (b) of Proposition 1.3 by placing the constraint $D_{2,0} \geq \epsilon^* = 0.39$, which is a binding constraint. In order to impose such a constraint, one would need to argue that from prior knowledge or economic theory it is known that the common expansionary technology shock affects output more than the common expansionary monetary policy shock on impact. Without knowledge about the true DGP, the true margin ϵ^* needs to be gauged by the researcher. This simulation exercise reports results for different values of ϵ.

The full maximization problem for this particular simulation with the just mentioned restrictions is:

$$
\begin{aligned}
& q_1^* = \operatorname{argmax} \sum_{h=0}^{H} \Omega_{1,1}^{z}(h) \qquad (1.B.2) \\
& \text{s.t.} \\
& q_1' q_1 = 1, \\
& q_1' q_2 = 0, \\
& \eta_{2,1,0} - \eta_{2,2,0} \geq \epsilon, \\
& q_{11} \geq 0, \\
& q_{22} \geq 0.
\end{aligned}
$$

Table 1.B.2 depicts the results for this augmented Max-Share identification, denoted by Max-Share$^+$. The first two columns again depict the true DGP parameters, columns three and four give the results of the Max-Share$^+$ framework with $T = 1,000$ while the last two columns give the results for $T = 250$. Panel A shows the results for $\epsilon = 0.39$ which is the true margin by which $\eta_{2,1,0}$ exceeds $\eta_{2,2,0}$ in absolute terms. Panel B shows the results for $\epsilon = 0.2$ and Panel C the results for $\epsilon = 0$. The latter

Table 1.B.2: Max-Share with an Additional Inequality Constraint - Scenario B

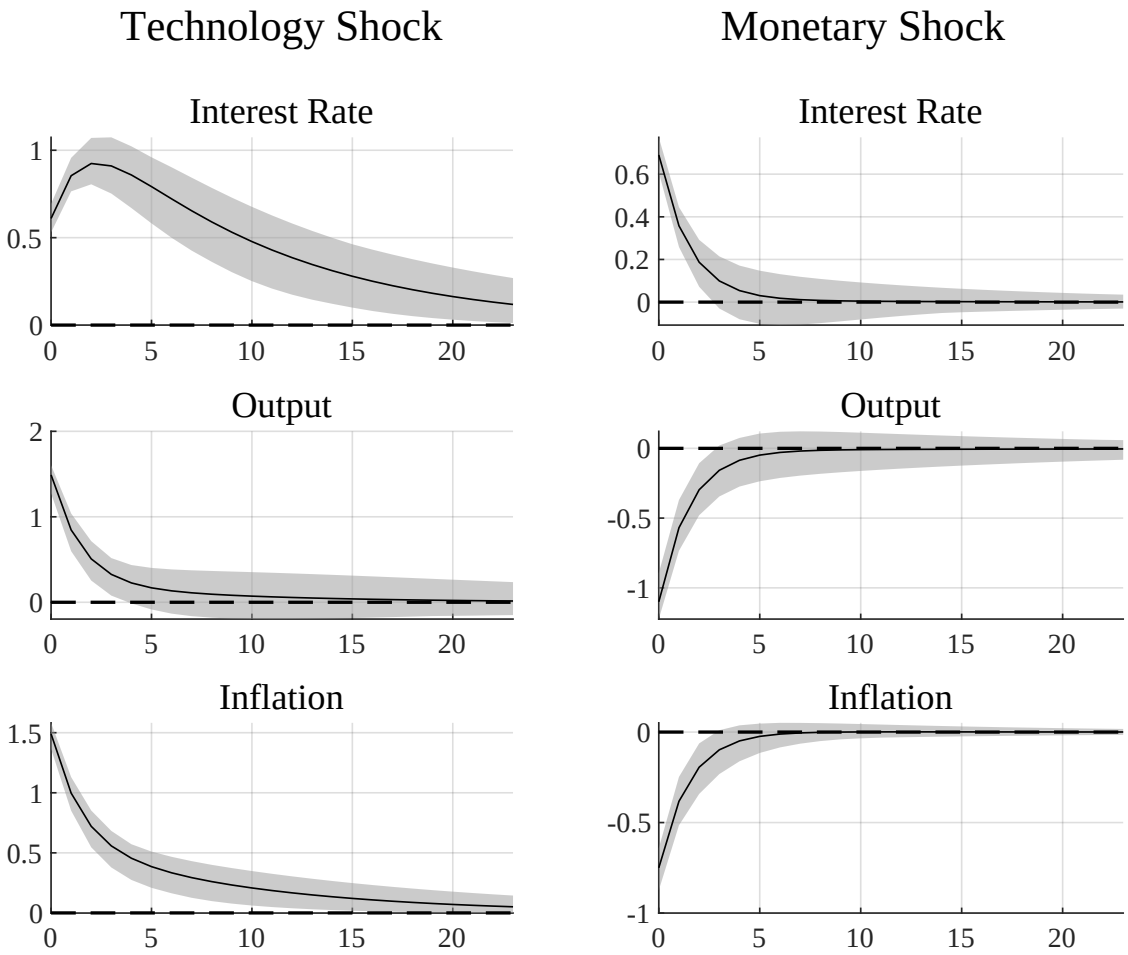
DGP		Max-Share$^+$, T=1000		Max-Share$^+$, T=250	
Panel A: $\epsilon = 0.39$					
0.61	0.69	0.613	0.688	0.612	0.688
		(0.029)	(0.026)	(0.057)	(0.052)
1.49	-1.1	1.486	-1.106	1.481	-1.101
		(0.044)	(0.044)	(0.087)	(0.087)
1.49	-0.75	1.487	-0.755	1.483	-0.755
		(0.031)	(0.038)	(0.062)	(0.075)
Panel B: $\epsilon = 0.2$					
0.61	0.69	0.659	0.644	0.658	0.643
		(0.027)	(0.027)	(0.055)	(0.054)
1.49	-1.1	1.406	-1.206	1.405	-1.205
		(0.044)	(0.045)	(0.075)	(0.074)
1.49	-0.75	1.432	-0.856	1.429	-0.858
		(0.032)	(0.037)	(0.062)	(0.072)
Panel C: $\epsilon = 0$					
0.61	0.69	0.706	0.592	0.706	0.591
		(0.026)	(0.027)	(0.053)	(0.055)
1.49	-1.1	1.311	-1.311	1.31	-1.309
		(0.037)	(0.037)	(0.075)	(0.074)
1.49	-0.75	1.362	-0.963	1.36	-0.965
		(0.032)	(0.036)	(0.063)	(0.07)

Note: The table depicts the average of the estimated structural parameters over the 1,000 Monte-Carlo simulation iterations. The estimates are derived with the Max-Share$^+$ framework in the proxy VAR with different values of ϵ. The value in the parenthesis depicts the respective standard deviation.

represents the case when restriction boils down to a ranking restriction on the initial impacts of the two shocks.

Comparing the estimates throughout the panels reveals that having the correct economic intuition with the inequality restriction ($\epsilon = 0.39$) removes all of the bias of the Max-Share approach. Yet, if the true margin is not met, part of the bias remains and it is larger the less close the true margin is met. This is true for both the technology as well as the monetary policy shock. Comparing the results in terms of the sample sizes shows that also for smaller sample sizes like $T = 250$ the correct

Figure 1.B.4: 95% Point Estimate Bands, $\epsilon = 0.39$ - Scenario B



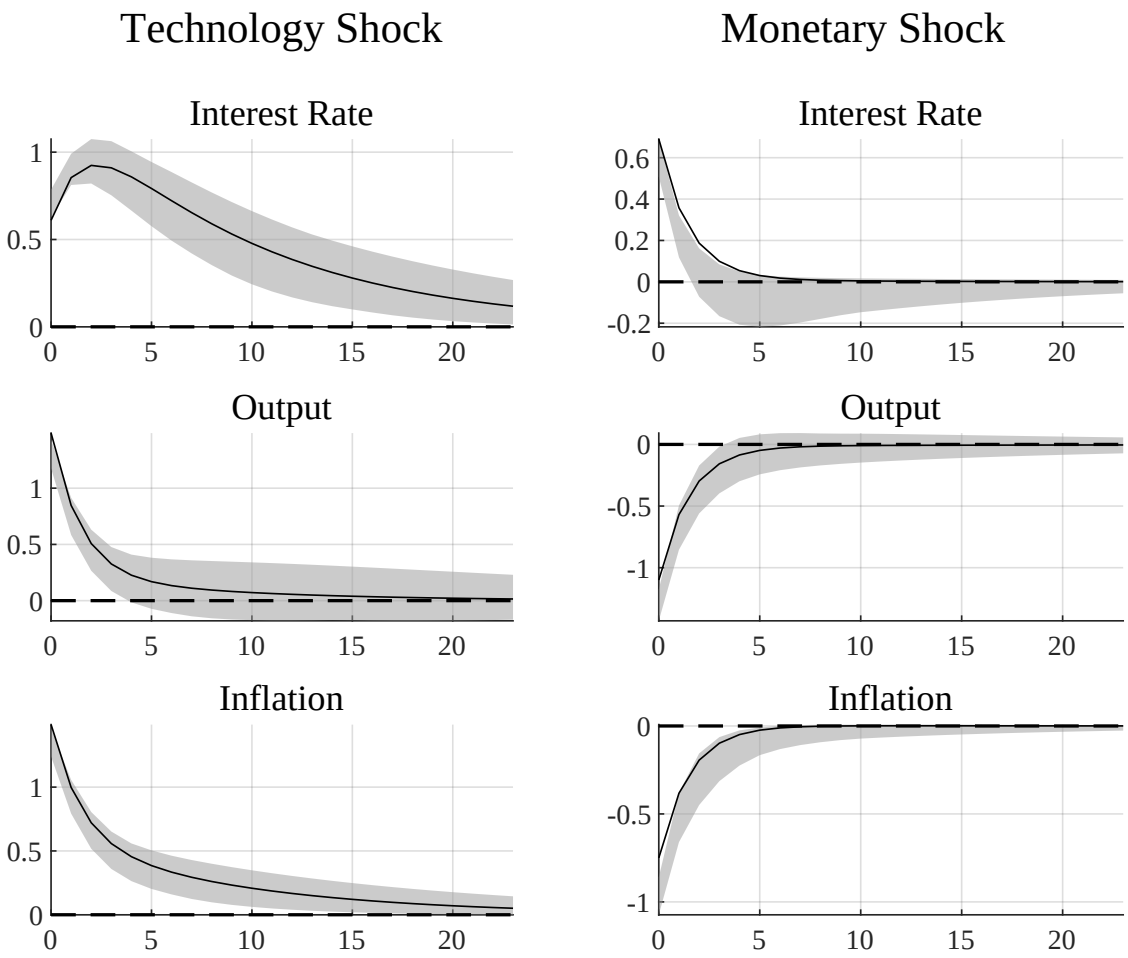
Note: The solid lines are the true impulse responses and the shaded areas span the 2.5% and 97.5% quantiles of the solutions found for the 1,000 simulation iterations and $T = 1,000$.

identification is achieved. Naturally, the bias also remains for smaller sample sizes when ϵ is not very close to the true margin.

Figure 1.B.4 to 1.B.6 show the 2.5% and 97.5% quantiles (dashed lines) of the estimated impulse responses identified by the augmented Max-Share throughout the 1,000 Monte-Carlo simulation rounds. The true impulse responses are depicted by the solid lines. Figure 1.B.4 presents the responses obtained with the $\epsilon = 0.39$, Figure 1.B.5 presents the results for $\epsilon = 0$ while Figure 1.B.6 depicts $\epsilon = 0.2$.

The figures are in line with the results of Table 1.B.2. The estimated responses in Figure 1.B.4 closely identify the true structural impulse response parameters while the responses in Figures 1.B.5 and 1.B.6 reflect the bias of the impact matrix parameters in the bias of the impulse responses especially at earlier horizons.

Figure 1.B.5: 95% Point Estimate Bands, $\epsilon = 0$ - Scenario B

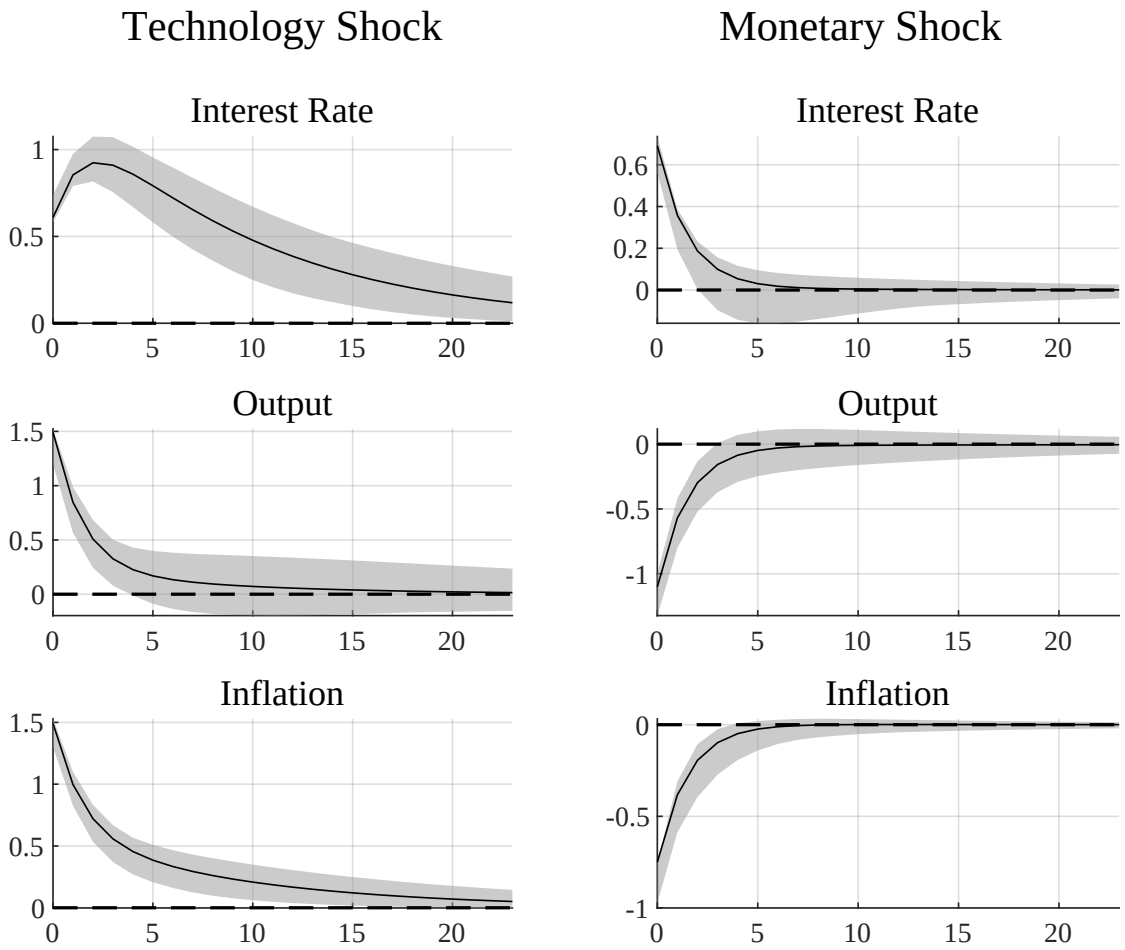


Note: The solid lines are the true impulse responses and the shaded areas span the 2.5% and 97.5% quantiles of the solutions found for the 1,000 simulation iterations and $T = 1,000$.

How the bias changes when getting closer to the true ϵ^* can be gauged by checking whether the sufficient conditions established in Proposition 1.5 are met. Table 1.B.3 depicts the outcome for the impulse responses of the different variable to the two shocks at different horizons. Based on this evaluation the bias is only guaranteed to decrease for the output responses to the technology shock at all considered horizon and for the IRFs of the interest rate to the monetary policy shock at almost all horizons. For the remaining IRFs the sufficient conditions are only satisfied for some horizons or not at all. Yet, these are only sufficient conditions and Tables 1.B.4 and 1.B.5, as well as Figures 1.B.7 to 1.B.10 show that the bias decreases when ϵ approaches ϵ^* anyway.

The bias of the structural IRFs at selected horizon and for different values of ϵ is depicted in Table 1.B.4 and 1.B.5 for $T = 1000$ and $T = 250$. Comparing the results

Figure 1.B.6: 95% Point Estimate Bands, $\epsilon = 0.2$ - Scenario B



Note: The solid lines are the true impulse responses and the shaded areas span the 2.5% and 97.5% quantiles of the solutions found for the 1,000 simulation iterations and $T = 1,000$.

along the columns shows that the bias is typically larger the further away ϵ is from ϵ^* and for earlier horizons of the responses. These patterns are consistent throughout both tables and both identified shocks.

The patterns found in Tables 1.B.4 and 1.B.5 are visualized in Figures 1.B.7 to 1.B.10 for all horizons. The different lines depict the bias of the structural impulse responses at different horizons and for different values of ϵ. Comparing the results throughout the panels shows that the bias is larger when ϵ is farther away from the true margin, despite the sufficient conditions established in Proposition 1.5 are not fulfilled. Only the responses of interest rates to the technology policy shock $\eta_{2,2,h}$ do not show the desired pattern for a few horizons. Thus, Proposition 1.5 unfortunately only yields limited information about the behaviour of the bias but can nevertheless serve as a rough guidance.

Table 1.B.3: Sufficient Condition for Bias Reduction - Scenario B

w_t^z	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
r_t																							○	○	○
x_t	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
π_t	○	○	○	○	○	○	○	○									○	○	○	○	○	○	○	○	○
w_t^r	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
r_t	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○			
x_t																									
π_t									○	○	○	○	○	○	○	○									

Note: The table displays whether the sufficient condition for the bias reduction established in Proposition 1.5 are met in Scenario B. A ○ means that the sufficient condition is met for the response of the according variable to the respective shock at the given horizon.

Table 1.B.4: Bias of the IRFs, $T = 1000$ - Scenario B

Variable	$h = 0$	$h = 6$	$h = 12$	$h = 18$
Technology Shock with $\epsilon = 0.39$				
r_t	0.0025	-0.0178	-0.0198	-0.0136
x_t	-0.0045	-0.0108	-0.0055	-0.0018
π_t	-0.0026	-0.0146	-0.0113	-0.0065
Technology Shock with $\epsilon = 0.2$				
r_t	0.0486	-0.0186	-0.0207	-0.014
x_t	-0.0844	-0.0131	-0.0061	-0.0022
π_t	-0.0584	-0.0163	-0.0117	-0.0067
Technology Shock with $\epsilon = 0$				
r_t	0.0958	-0.0233	-0.0235	-0.0154
x_t	-0.1795	-0.016	-0.0068	-0.0025
π_t	-0.1273	-0.0199	-0.0129	-0.0072
Monetary Shock with $\epsilon = 0.39$				
r_t	-0.0021	-0.0059	-0.0025	-0.0008
x_t	-0.0055	-0.0005	0.0009	0.001
π_t	-0.0051	-0.0029	-0.0007	-0.0001
Monetary Shock with $\epsilon = 0.2$				
r_t	-0.0462	-0.0546	-0.0278	-0.0138
x_t	-0.1056	-0.0088	-0.0025	-0.0007
π_t	-0.1062	-0.0249	-0.0115	-0.0057
Monetary Shock with $\epsilon = 0$				
r_t	-0.0982	-0.1084	-0.0557	-0.0282
x_t	-0.2105	-0.0184	-0.0067	-0.0029
π_t	-0.2135	-0.0491	-0.0234	-0.0119

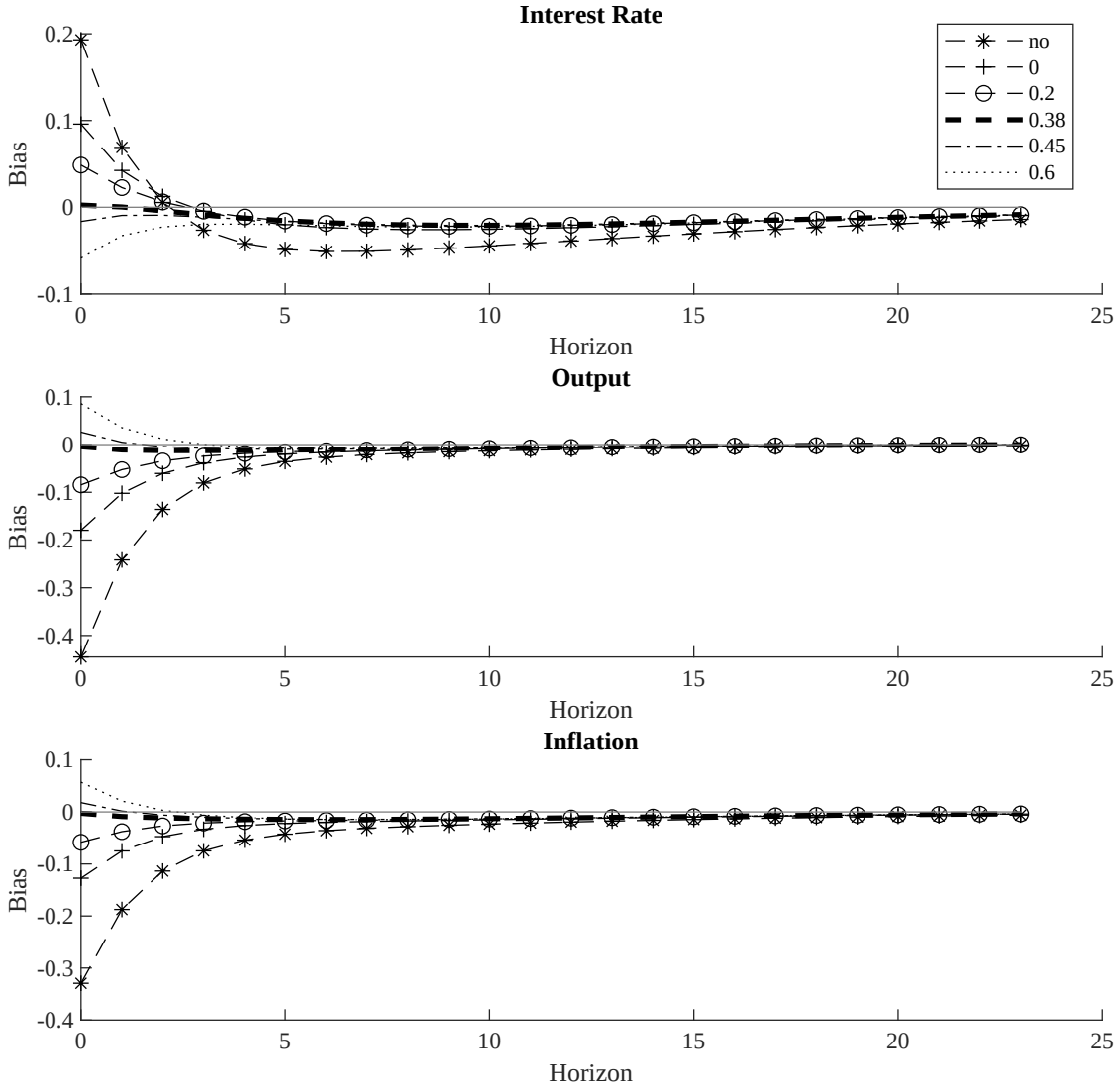
Note: The table depicts the average bias in the estimated structural parameters over the 1,000 Monte-Carlo simulation iterations. The estimates are derived with the Max-Share$^+$ framework in the proxy VAR with different values of ϵ.

Table 1.B.5: Bias of the IRFs to the Technology Shock, $T = 250$ - Scenario B

Variable	$h = 0$	$h = 6$	$h = 12$	$h = 18$
Technology Shock with $\epsilon = 0.39$				
r_t	0.0023	-0.0745	-0.0736	-0.0473
x_t	-0.0084	-0.0381	-0.0173	-0.0059
π_t	-0.0072	-0.0576	-0.0398	-0.0219
Technology Shock with $\epsilon = 0.2$				
r_t	0.0484	-0.0748	-0.0739	-0.0474
x_t	-0.0846	-0.0394	-0.0171	-0.0058
π_t	-0.0607	-0.059	-0.0399	-0.0219
Technology Shock with $\epsilon = 0$				
r_t	0.0956	-0.0795	-0.0764	-0.0485
x_t	-0.1803	-0.0421	-0.0176	-0.006
π_t	-0.1302	-0.0624	-0.0409	-0.0224
Monetary Shock with $\epsilon = 0.39$				
r_t	-0.0026	-0.0105	-0.003	-0.0001
x_t	-0.0016	0.0032	0.0049	0.0039
π_t	-0.005	-0.0046	-0.0003	0.0004
Monetary Shock with $\epsilon = 0.2$				
r_t	-0.0465	-0.0561	-0.0249	-0.011
x_t	-0.1053	-0.0041	0.0017	0.0021
π_t	-0.1082	-0.024	-0.0093	-0.0042
Monetary Shock with $\epsilon = 0$				
r_t	-0.0986	-0.1052	-0.0486	-0.0228
x_t	-0.2086	-0.011	-0.0012	0.0005
π_t	-0.2146	-0.0449	-0.0189	-0.0092

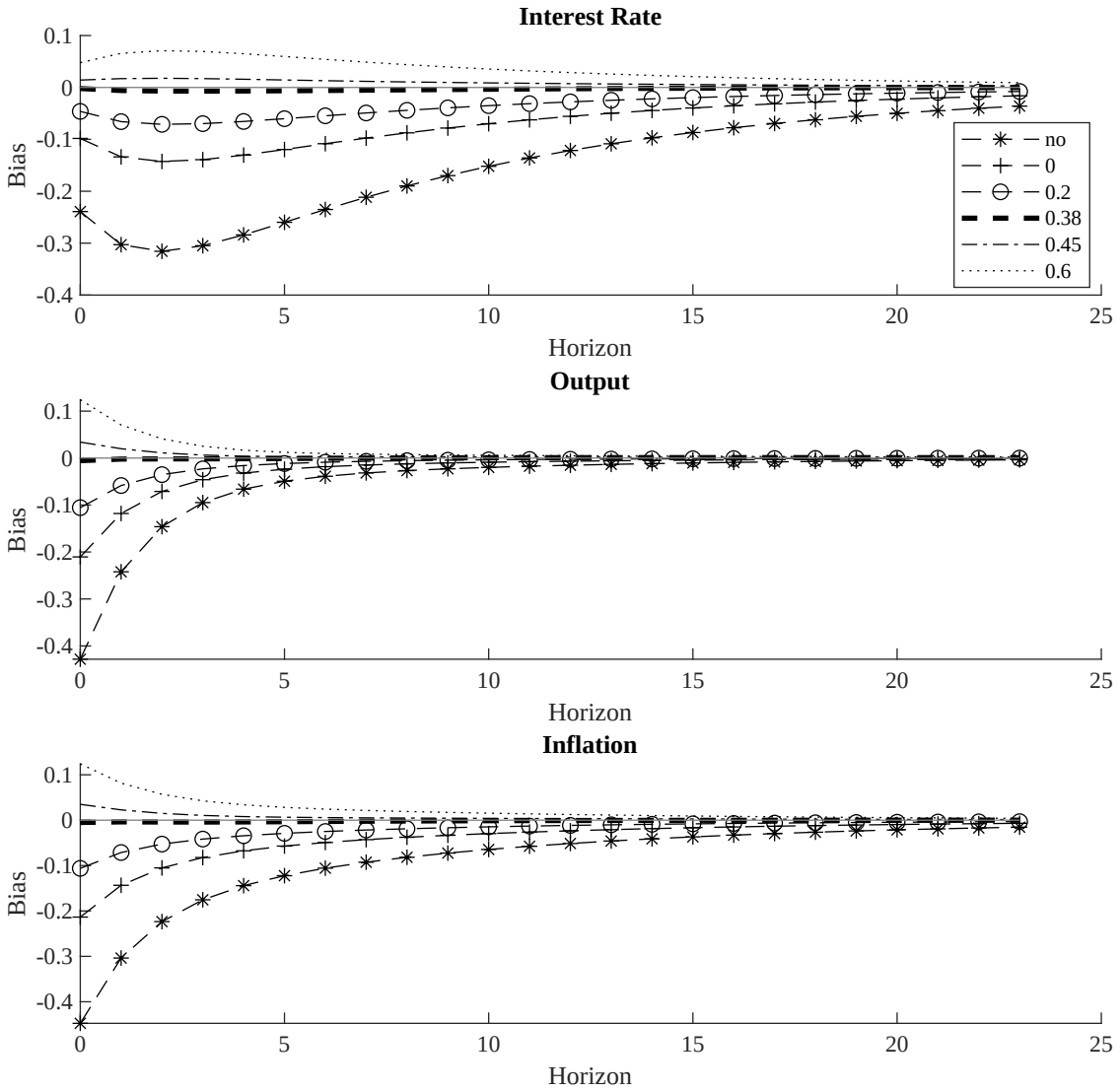
Note: The table depicts the average bias in the estimated structural parameters over the 1,000 Monte-Carlo simulation iterations. The estimates are derived with the Max-Share$^+$ framework in the proxy VAR with different values of ϵ.

Figure 1.B.7: Bias of the IRFs to the Technology Shock, $T = 1000$ - Scenario B



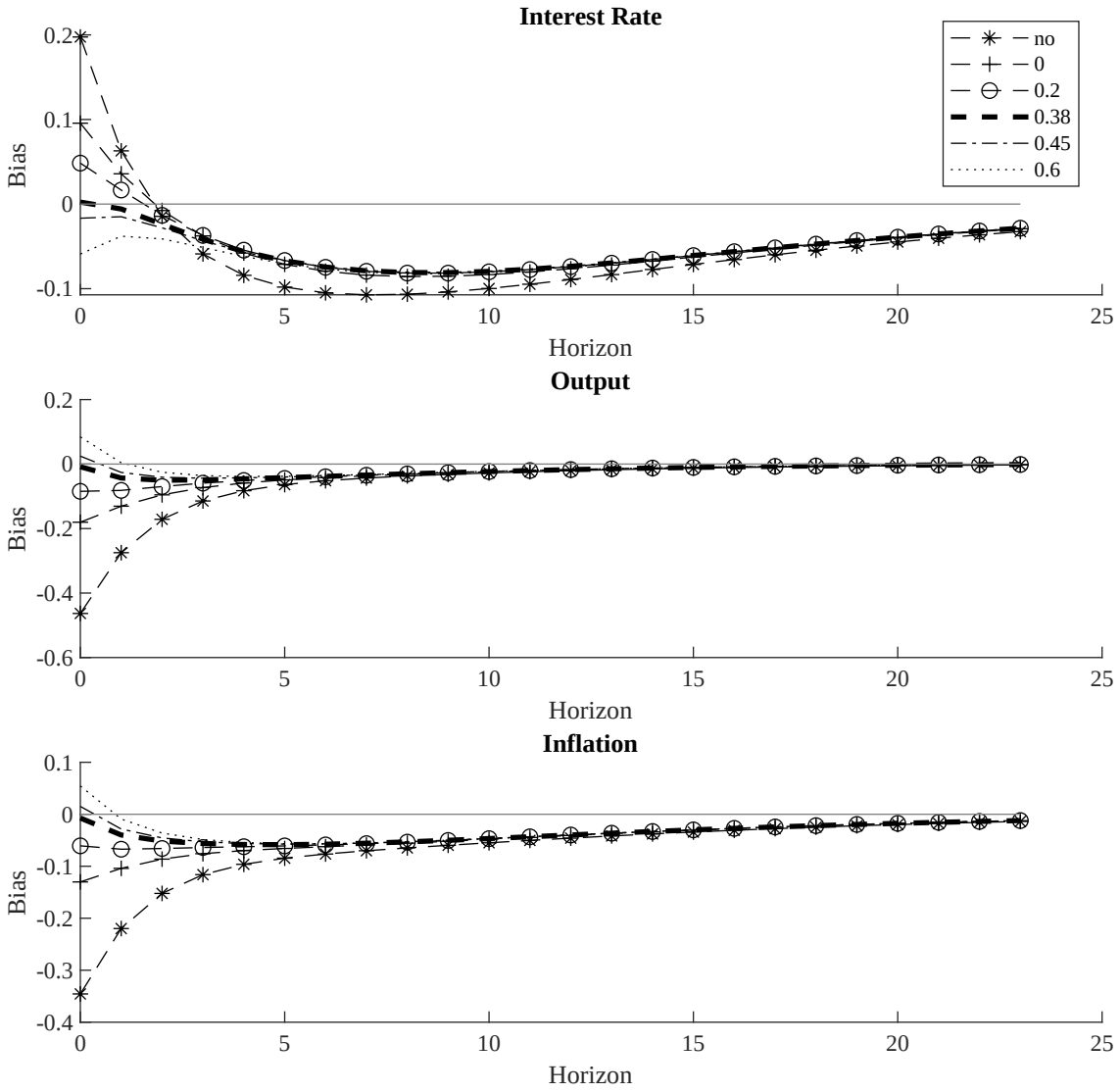
Note: The different lines depict the average bias of the impulse response functions over 1,000 Monte-Carlo simulations for different values of ϵ.

Figure 1.B.8: Bias of the IRFs to the Monetary Shock, $T = 1000$ - Scenario B



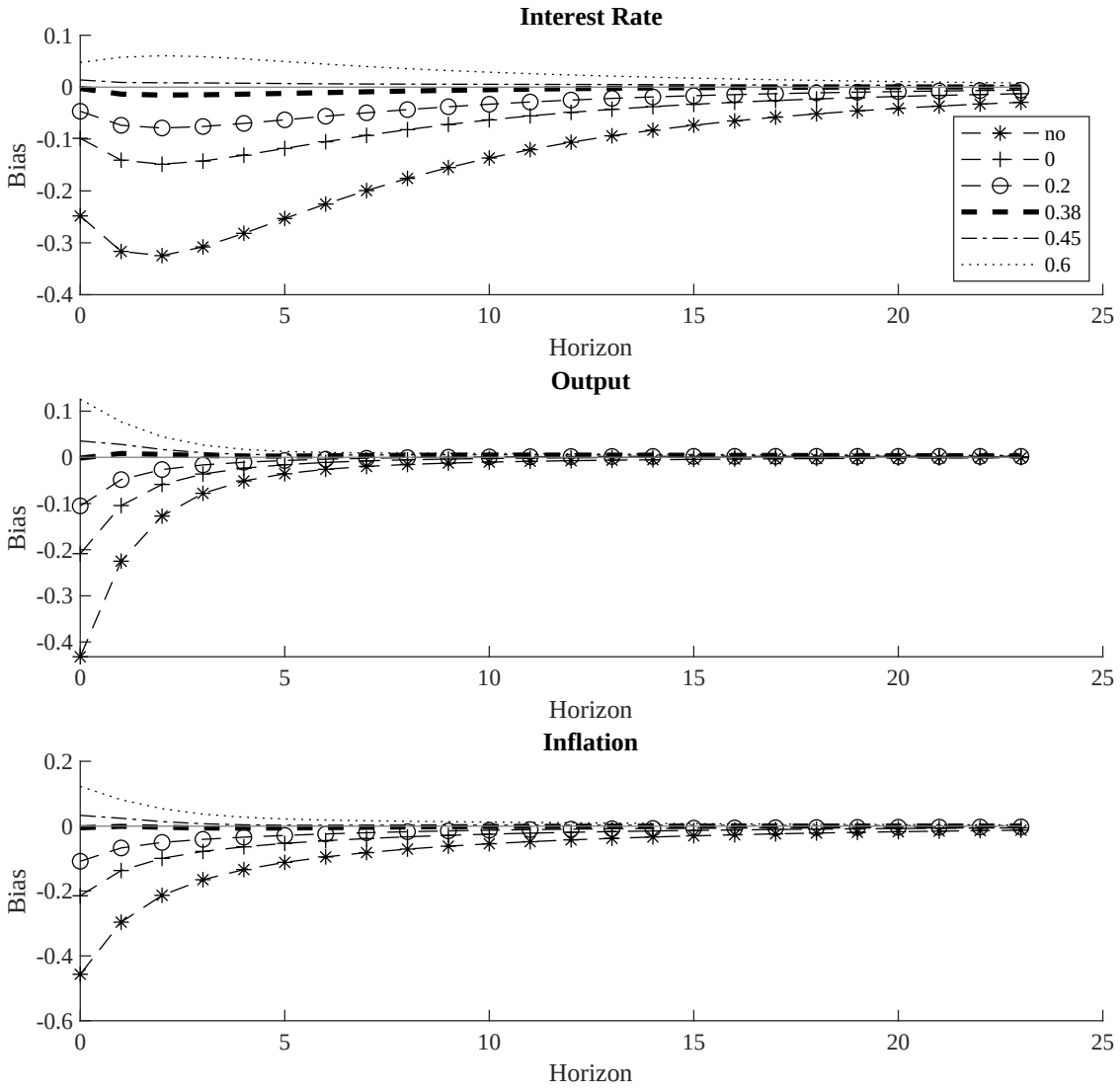
Note: The different lines depict the average bias of the impulse response functions over 1,000 Monte-Carlo simulations for different values of ϵ.

Figure 1.B.9: Bias of the IRFs to the Technology Shock, $T = 250$ - Scenario B



Note: The different lines depict the average bias of the impulse response functions over 1,000 Monte-Carlo simulations for different values of ϵ.

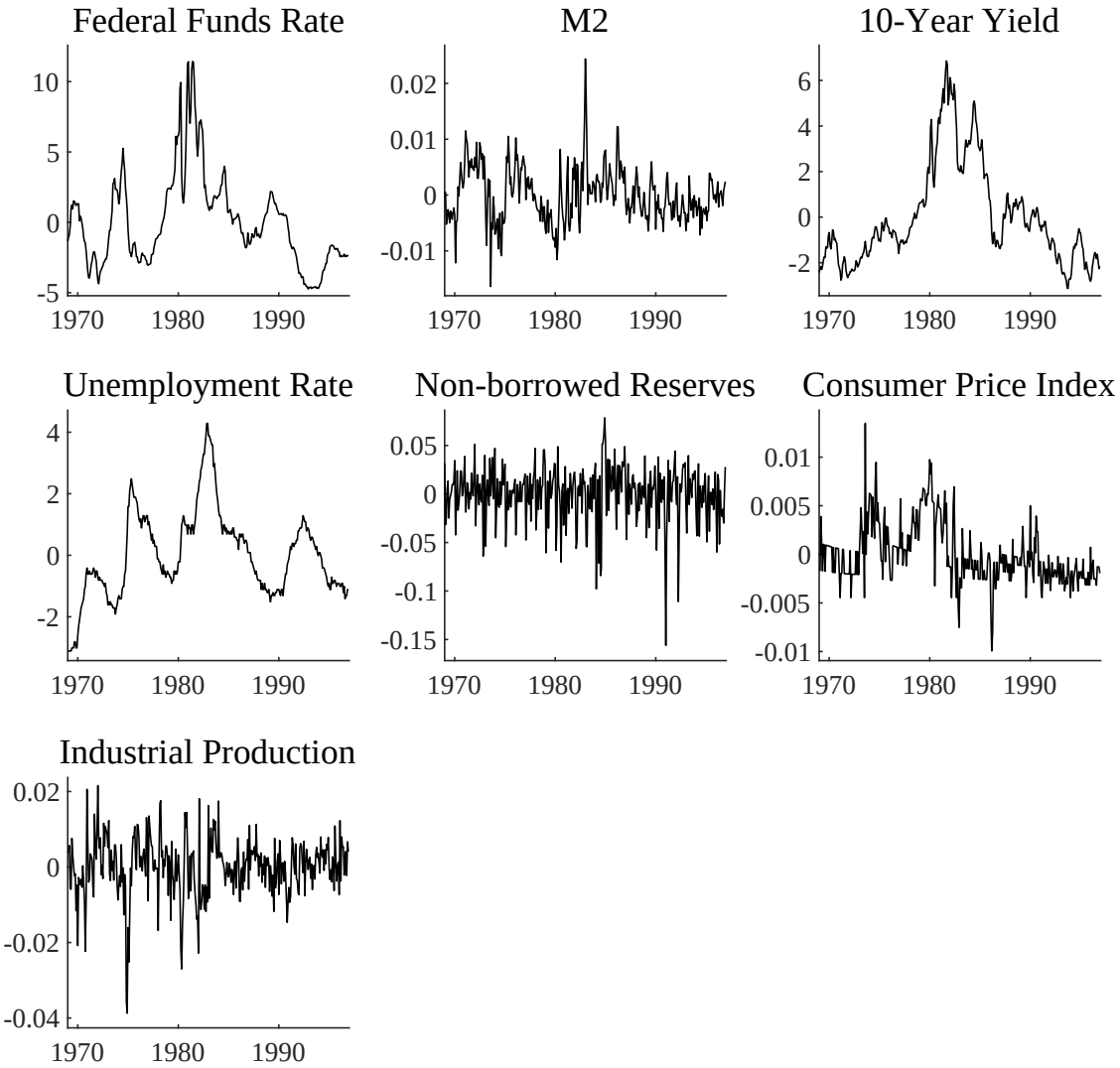
Figure 1.B.10: Bias of the IRFs to the Monetary Shock, $T = 250$ - Scenario B



Note: The different lines depict the average bias of the impulse response functions over 1,000 Monte-Carlo simulations for different values of ϵ.

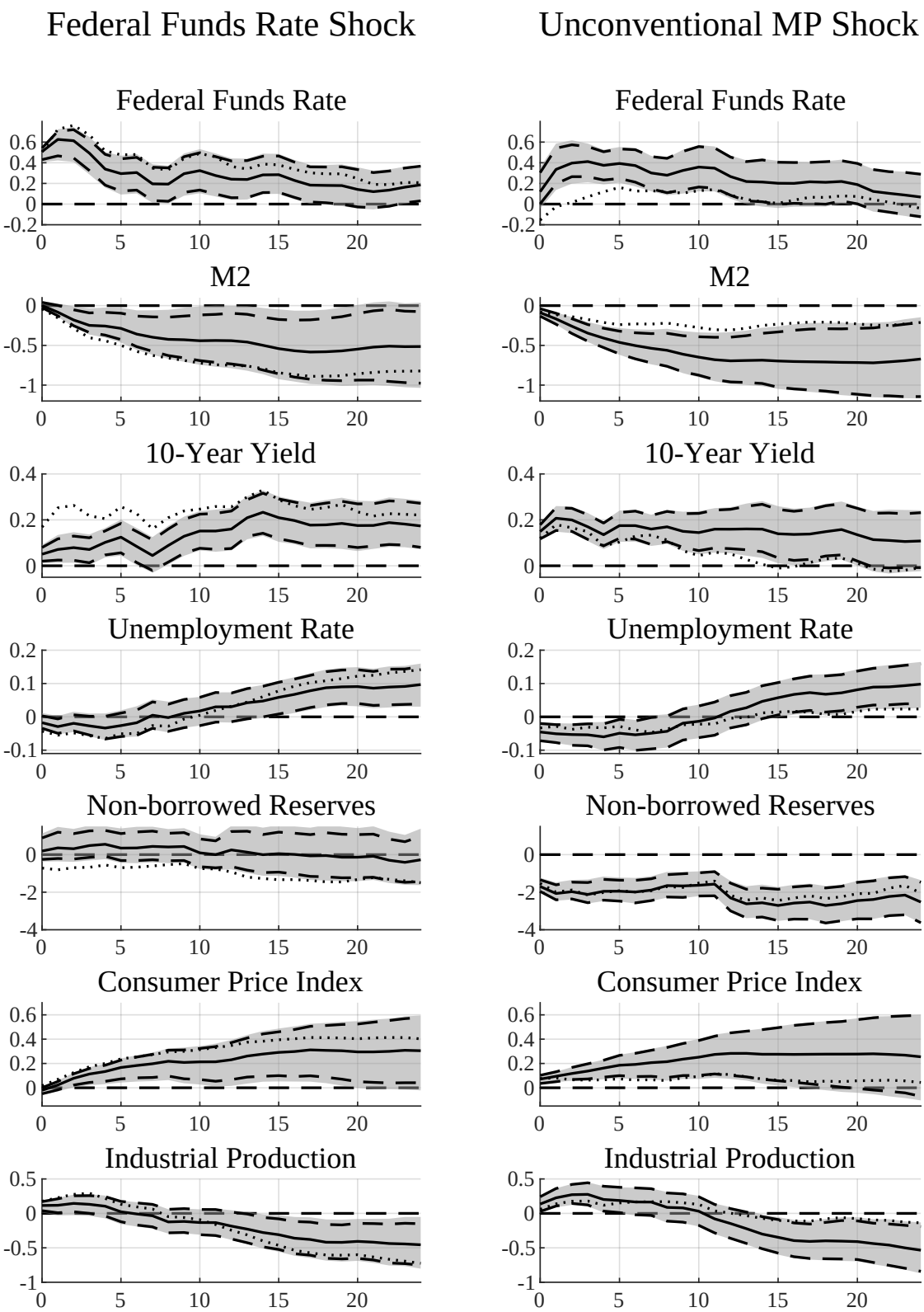
Appendix 1.C Additional Empirical Results

Figure 1.C.11: Data used in Section 1.5



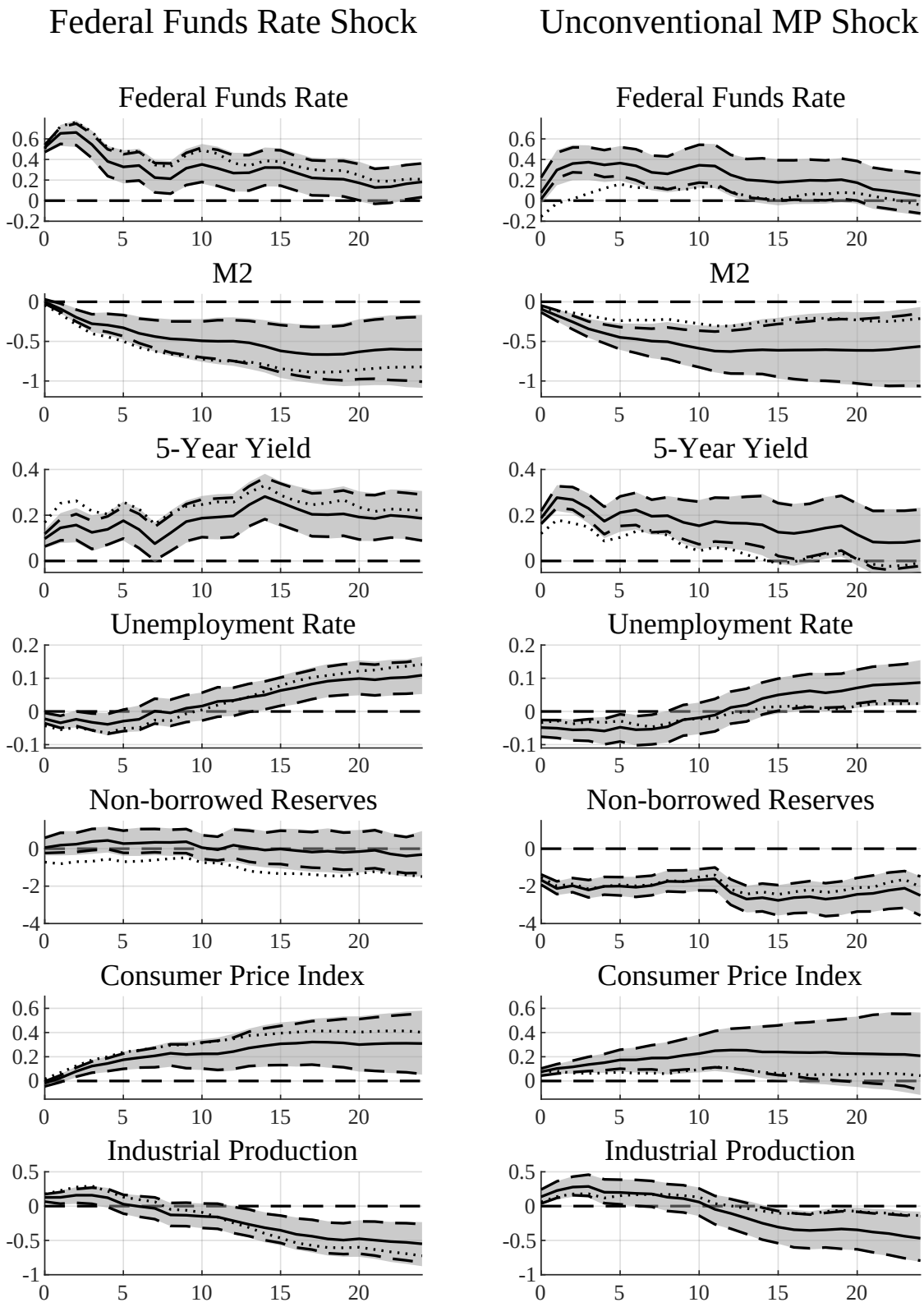
Note: The plots depict the data that is used in the empirical analysis.

Figure 1.C.12: Impulse Responses with the 10-year Treasury Yield



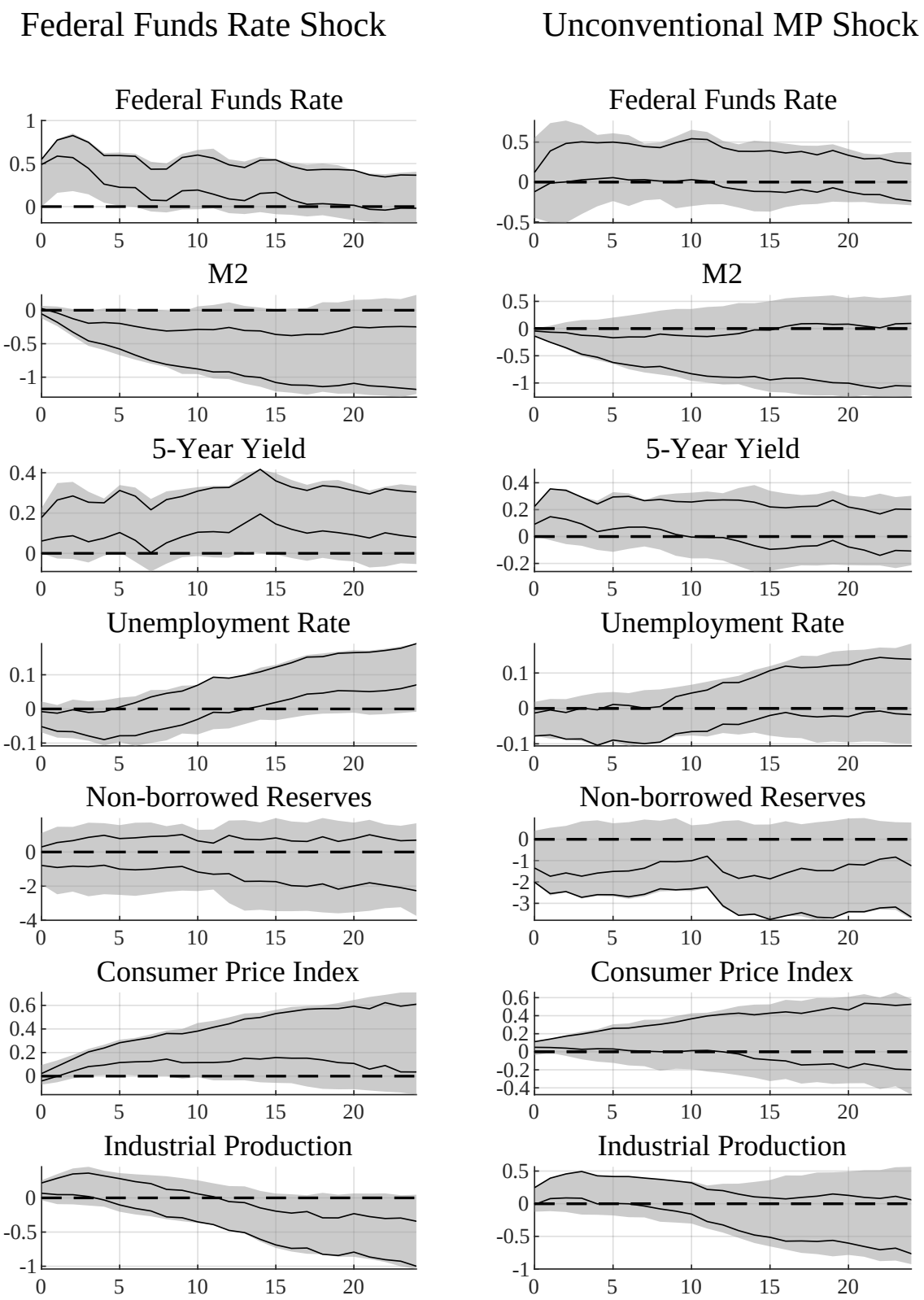
Note: The solid black line is the median target response for $\epsilon = 10$bp. The dashed black lines depict the respective 68% credible regions. The dotted lines depict the median responses of the baseline Max-Share approach without an additional inequality constraint. The grey shaded areas depict the range of the 68% credible regions over a grid of ϵ values. The estimates of Swanson (2021) are scaled to a 15bp response of the 10-year yield to the unconventional MP shock. The resulting range for ϵ is 8bp $\leq \epsilon \leq$ 12bp.

Figure 1.C.13: Impulse Responses with the 5-year Treasury Yield and $H = 4$



Note: The solid black line is the median target response for $\epsilon = 10$bp. The dashed black lines depict the respective 68% credible regions. The dotted lines depict the median responses of the baseline Max-Share approach without an additional inequality constraint. The grey shaded areas depict the range of the 68% credible regions over a grid of ϵ values between 7bp $\leq \epsilon \leq$ 11bp. The maximization horizon for the augmented share is $H = 4$, i.e. up to the first four months after the shock hits.

Figure 1.C.14: Impulse Responses to the two MP Shocks - Bound=0.1



Note: The solid black line the 68% robust Bayesian credible regions for the identification including the bounds constraint. The grey shaded areas depict the 68% robust Bayesian credible regions for the identification based purely on sign restrictions. Posterior medians are omitted for clarity of the plot. The bound on the FEV is $\Omega^z_{FFR,2}(h=1) \leq 0.1$.

Chapter 2

Financial and Real Uncertainty Shocks: Insights from Synthetic Proxy Variables

2.1. Introduction

Increasing fluctuations in uncertainty since the financial crisis raise the question of its role in business cycles. It is well-established that uncertainty is high during recessions (e.g. Bloom, 2014). However, it is not clear whether uncertainty is merely a consequence of economic fluctuations or a cause of economic downturns. The potential reverse causality between uncertainty and real activity makes the identification of exogenous uncertainty shocks difficult. Increasing uncertainty is likely to result in a decrease of economic activity but at the same time adverse developments in economic activity should also increase uncertainty. If one were to identify an uncertainty shock on the assumption that the shock increases uncertainty and decreases economic activity, the resulting shock would pick up both, the desired effect of uncertainty on real activity and the endogenous response of uncertainty to decreased economic activity. Hence, an identification strategy must handle this reverse causality in order to achieve the identification of uncertainty shocks.

In this study, we propose a novel combination of proxy variables and statistical identification via non-Gaussian shocks, which allows us to identify the effects of financial and real uncertainty shocks. On the one hand, the proxy VAR addresses the reverse causality issue because it relies on exogenous variation in uncertainty and estimates the effect of this exogenous variation on the economy. On the other hand, the statistical identification addresses another challenge in identifying uncertainty shocks. A priori, the effects of uncertainty shocks on the economy are not clear as theory describes different channels through which uncertainty affects the economy, some of which are contractionary, while others are expansionary, see e.g. Bloom (2014) for an overview. Hence, the typical restrictions on structural parameters are difficult to deduct from theory and already reflect a believe of the researcher which effects might prevail. The agnostic statistical identification via non-Gaussianity and independence criteria yields identification without the need to impose such restrictions.

In the context of the proxy VAR, we propose to identify multiple shocks with a single instrument by generating synthetic proxy variables which are non-linear transformations of the single available proxy. With this new methodology we identify two different uncertainty shocks with a single proxy variable. We show that financial uncertainty, measured by the stock market volatility (VXO), is largely driven by financial uncertainty shocks. However, financial uncertainty shocks have small and short-lived effects beyond financial markets. In contrast to that, real market uncertainty shocks have a more pronounced impact on real activity. At the same time, real

uncertainty, measured by the economic policy uncertainty (EPU) index proposed by Baker et al. (2016), is also an endogenous response to economic fluctuations as only a smaller part of the forecast error variance (FEV) of real uncertainty is explained by the uncertainty shocks.

We build on the work of Piffer and Podstawski (2018) to separate the uncertainty shocks from the remaining shocks. In detail, we use their identified uncertainty shock, rather than their proxy, as an uncertainty proxy in our application. We use the shock because their original proxy is also related to news shocks which would confound the identification of our uncertainty shocks. Hence, we use their estimated uncertainty shock series which is cleaned of the news shock as they disentangle the uncertainty shock from the news shock in their identification scheme. Furthermore, we employ the uncertainty shock identified by Piffer and Podstawski (2018) in particular as they are, to the best of our knowledge, the only ones to identify uncertainty shocks based on exogenous variation of uncertainty in a proxy VAR framework. In their application, they construct the proxy by examining changes in the gold price closely around unanticipated events and their uncertainty shock is estimated based on this exogenous variation. In order to preserve this feature in our own proxy VAR application, we use their estimated uncertainty shock.

We argue that the resulting uncertainty proxy is driven by both types of uncertainty shocks that we want to identify. For the identification of both uncertainty shocks with a single proxy z_t we propose to generate synthetic proxies which are the nonlinear transformation z_t^r. In our application we use $r = 2, 3$. Combining the baseline proxy with the synthetic proxy variables allows us to separate the financial and the real uncertainty shock from the remaining shocks of the system. For the synthetic proxies to be valid, we assume that the proxy z_t is mean-independent from the shocks not related to it. Hence, the proxy does not contain any information about the mean of the non-target shocks.

If the proxy VAR is used to identify multiple shocks, additional restrictions are necessary to disentangle the target shocks. We disentangle the two uncertainty shocks by assuming that the financial and real uncertainty shocks are independent and that at least one of them is non-Gaussian. The assumption that uncertainty shocks are non-Gaussian is reasonable and also supported by findings in the literature as e.g. Ludvigson et al. (2021) find that their identified uncertainty shocks exhibit skewness and excess kurtosis. Note that we do not need to impose that the uncertainty shocks are independent to the remaining structural shocks which are not identified. Not imposing independence between the uncertainty and the remaining shocks allows for the

possibility that uncertainty shocks affect higher moments (e.g. the variance) of other structural shocks.

The combination of proxy variables and identification via non-Gaussian shocks and independence suits the application of uncertainty shocks well. The proxy VAR deals with the reverse causality issue as the proxy is comprised of exogenous variation in uncertainty. Disentangling the shocks in the next step with a statistical approach has multiple advantages. Firstly, we do not need to impose restrictions on a priori unknown structural parameters. In particular, we do not impose restrictions on potentially noisy uncertainty measurements. Secondly, we explicitly acknowledge and use the non-Gaussian nature of uncertainty shocks. In combination with the proxy, we only rely on the independence of the two uncertainty shocks but allow for links to higher moments of the remaining structural shocks.

We assume that the uncertainty shocks are independent because of the type of events that evoke such uncertainty shocks. The analysis of Piffer and Podstawski (2018) and also the application of this chapter shows that uncertainty shocks occur in the wake of unexpected events like e.g. natural disasters, terrorist attacks, wars or phenomena like the Black Monday in 1987. We argue that the occurrence and timing of these exogenous and unexpected events are unrelated. Related consequences of these exogenous causes of uncertainty are an endogenous response to the independent exogenous events. Yet, these exogenous events might affect the variability in policymakers' decision-making, and thus might affect higher moments of policy related shocks like fiscal or monetary policy shocks.

As the shocks are disentangled statistically the labels have to be added ex post. For the mandatory labelling of the shocks, we compare the shock series themselves and identify certain uncertainty-relevant exogenous historical events. Depending on the types of events that occurred at the timing of the spikes in our two uncertainty shock series, we can attribute the label of a financial or real uncertainty shock. The empirical results show that the suspected events sort themselves into the two shocks based on whether they are related to financial markets or not. Hence, the resulting labelling is unambiguous. One advantage of the event-based labelling is that we again do not have to rely on looking at e.g. structural impulse response functions (IRFs) or FEVs which are not easily assignable to a specific uncertainty shock based on theoretical considerations. Hence, this strategy allows us to learn something about the key structural parameters of interest without imposing any restrictions on them.

To the best of our knowledge, we are the first ones to employ a proxy VAR to identify the effects of different types of uncertainty shocks. The paper by Ludvigson et al.

(2021), which sparked a lot of research in this field, accompanies sign restrictions with restrictions on the uncertainty shocks which have to exceed a certain threshold around selected historical events. We show that our labelling of the shocks is in line with their identification strategy. Angelini et al. (2019) identify the shocks under heteroskedasticity, namely different volatility regimes, paired with exclusion restrictions on the impact matrices in the respective regimes. Carriero et al. (2018) use a VAR with stochastic volatility and identify the two uncertainty shocks as the volatility components of the financial and macroeconomic variables. Shin and Zhong (2020) use a similar approach with stochastic volatility and additionally impose sign restrictions and lastly, Carriero and Volpicella (2024) identify the uncertainty shocks with restrictions on the contribution of the uncertainty shocks to FEV of uncertainty measurements.

The findings of the aforementioned papers do not allow for a definitive conclusion regarding the effects of different types of uncertainty shocks on the economy. While Ludvigson et al. (2021) and Shin and Zhong (2020) find that mainly financial uncertainty shocks lead to a decline in real activity, the results of Angelini et al. (2019) suggest the opposite, namely, that real uncertainty shocks have a more pronounced negative effect on real activity. Carriero et al. (2018) and Carriero and Volpicella (2024) find that both financial and real uncertainty shocks have a substantial negative effect on real activity. Our result are very much in line with Angelini et al. (2019). Real activity decreases in response to a real uncertainty shock, yet real activity does not decrease significantly in response to financial uncertainty shocks. Comparing the relative magnitudes of the responses to the two uncertainty shocks shows that the real uncertainty shock seems to have a stronger effect on labour market related variables, such as the average weekly working hours. Conversely, the financial uncertainty shock seems to increase credit spreads more than the real uncertainty shock. In line with the findings by Ludvigson et al. (2021), we find that real uncertainty is also an endogenous response to other business cycle shocks as a large part of the FEV in the real uncertainty measure is explained by non-uncertainty shocks. Yet, variation in financial uncertainty seems to be explained to a large extent by the uncertainty shocks, mainly by the financial one.

In the following, section 2.2 introduces the basic proxy VAR and the concept of synthetic proxies, while section 2.3 describes the identification strategy and estimation via non-Gaussian shocks and independence criteria. Section 2.4 presents a small simulation study to assess the finite sample properties of the synthetic proxy VAR

estimator. Then section 2.5 discusses the results of the empirical application to the identification of two types of uncertainty shocks.

2.2. The Proxy VAR

Equations (2.2.1) and (2.2.2) depict the k-dimensional structural VAR(p) model.

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

$$u_t = B_0 \varepsilon_t, \tag{2.2.2}$$

with parameter matrices $A_1, \ldots, A_p \in \mathbb{R}^{k \times k}$, an intercept term $\nu \in \mathbb{R}^k$, an invertible matrix $B_0 \in \mathbb{B} := \{B \in \mathbb{R}^{k \times k} | \det(B) \neq 0\}$, a k-dimensional vector of time series $y_t = [y_{1t}, \ldots, y_{kt}]'$ and a k-dimensional vector of serially uncorrelated structural shocks $\varepsilon_t = [\varepsilon_{1t}, \ldots, \varepsilon_{kt}]'$ with mean zero and unit variance. We assume that (2.2.1) is a stable process such that the parameter matrices $A_1, \ldots, A_p$ satisfy $\det(I - A_1 c - \cdots - A_p c^p) \neq 0$ for $|c| \leq 1$.

We partition ε_t into the shocks that are identified by the proxy variable $\varepsilon_t^{(1)}$, to which we refer as target shocks, and into non-target shocks $\varepsilon_t^{(2)}$. In the empirical application of this paper we identify two shocks, and thus the shocks are $\varepsilon_t = [\varepsilon_t^{(1)\prime}, \varepsilon_t^{(2)\prime}]'$ with $\varepsilon_t^{(1)} = [\varepsilon_{1t}, \varepsilon_{2t}]'$ and $\varepsilon_t^{(2)} = [\varepsilon_{3t}, \ldots, \varepsilon_{kt}]'$. In principle, this can be generalized to more than two shocks but to match the empirical application, the econometric framework is written down for two shocks in this section.

The two target shocks are related to the single proxy variable z_t. The standard proxy variable assumption implies that a proxy variable is valid if it is correlated with the target shocks and uncorrelated with all non-target shocks, see Definition 2.1.

Definition 2.1 (Valid proxy for $\varepsilon_t^{(1)}$)**.**
Proxy z_t is valid for the two target shocks $\varepsilon_t^{(1)}$ if z_t is relevant and exogenous

1. *Standard relevance:* $E\left[z_t \varepsilon_t^{(1)\prime}\right] = \underset{1\times 2}{\alpha} = [\alpha_1, \alpha_2] \neq \underset{1\times 2}{0}$
2. *Standard exogeneity:* $E\left[z_t \varepsilon_t^{(2)\prime}\right] = \underset{1\times(k-2)}{0}$

To identify the two shocks with a single proxy we propose to construct additional synthetic proxy variables z_t^r with $r = 2, \ldots, m$ from a given proxy variable z_t.

Analogous to a traditional proxy variable, a synthetic proxy is valid if it is correlated with the target shock and uncorrelated with all non-target shocks, see Definition 2.2.

Definition 2.2 (Valid synthetic proxy for $\varepsilon_t^{(1)}$)**.**
Given that z_t is relevant and exogenous, the artificial proxy z_t^r with $r = 2, \dots, m$ is a valid proxy if

1. *Synthetic relevance:* $E\left[z_t^r \varepsilon_t^{(1)\prime}\right] = \underset{1\times 2}{\alpha^{(r)}} = \left[\alpha_1^{(r)}, \alpha_2^{(r)}\right] \neq \underset{1\times 2}{0}$
2. *Synthetic exogeneity:* $E\left[z_t^r \varepsilon_t^{(2)\prime}\right] = \underset{1\times(k-2)}{0}$

The relevance of the synthetic proxy is related to the non-Gaussianity of the target shock. For example, suppose a linear relationship between the proxy variable z_t and target shock, $z_t = \psi\varepsilon_{1t} + \eta_t$, with η_t being mean zero and independent of ε_{1t} and $\psi \neq 0$. Such a linear relationship is assumed in most Bayesian proxy VARs, see e.g. Caldara and Herbst (2019), Angelini et al. (2023), Braun and Brüggemann (2022), or Giacomini et al. (2022). Further suppose $z_t^2 = \theta\varepsilon_{1t} + \xi_t$, again with ξ_t being mean zero and independent of ε_{1t}. One can show that $E(z_t^2\varepsilon_{1t}) = \theta E(\varepsilon_{1t}^2) + E(\varepsilon_{1t}\xi_t)$ and this implies that $\theta = E(z_t^2\varepsilon_{1t}) \neq 0$ if $E[\varepsilon_{1t}^3] \neq 0$. Similar arguments apply for $m \geq 3$. Hence, with a strong case for non-Gaussian shocks the synthetic relevance condition is a reasonable assumption.

Assuming that the standard and synthetic exogeneity conditions hold at the same time is technically a stronger assumption than imposing only the standard exogeneity. More general, by the definition of mean independence all exogeneity assumptions hold if the non-target shocks are mean independent w.r.t. the proxy. Hence, if $E[\varepsilon_t^{(2)}|z_t] = 0$, meaning the proxy contains no information on the expected value of the non-target shocks $\varepsilon_t^{(2)}$, also the synthetic exogeneity conditions hold.

We argue that this mean independence assumption is economically justified for most proxy variables. In particular, the economic reasoning underlying most proxy exogeneity assumptions is that a given proxy variable contains no information on other non-target shocks. In addition, the standard linear proxy variable approach assumes a data generating process in which $z_t = \psi\varepsilon_t^{(1)} + \eta_t$ with Gaussian noise η_t which is uncorrelated with the target shock. This directly implies that $E[\varepsilon_t^{(2)}|z_t] = 0$, and thus this implies exogenous synthetic proxies z_t^r for $r = 2, \dots, m$. Importantly, we do not assume that the proxy variable and the non-target shocks are independent. In particular, we allow that higher moments of the non-target shocks $\varepsilon_t^{(2)}$ contain information on the expected value of z_t. For example, the variance of the non-target shocks can contain information on the expected value of z_t. This is important for the application of this paper which uses an uncertainty proxy variable that might be related to second moments of $\varepsilon_t^{(2)}$.

In practice, the number of synthetic proxies that can be utilized depends, firstly, on how many structural shock one wants to identify. Identification is discussed in the next section which shows that in order to identify two structural shocks we need at least one additional synthetic proxy. Furthermore, the number and which proxies one can use depends on their strength. Like for the traditional proxy variables the synthetic proxies need to be strong enough to yield reliable results. Note that the strength of the synthetic proxies can be assessed as for the baseline proxy with the standard tools that are employed in the literature. In the empirical study presented in section 2.5, we rely, for example, on the weak proxy test described by Lunsford (2015). In anticipation of the empirical application, the two synthetic proxies z_t^2 and z_t^3 satisfy this condition.

2.3. The Synthetic Proxy VAR Estimator

This section derives the synthetic proxy VAR estimator which identifies and estimates two target shocks using a (synthetic) proxy variable and the assumption of independent and non-Gaussian target shocks.

Let $e_t(B)$ denote the structural innovations such that

$$e_t(B) := B^{-1}u_t,$$

for any invertible matrix B. For $B = B_0$, the innovations are equal to the structural shocks, i.e. $e_t(B_0) = B_0^{-1}u_t = \varepsilon_t$. To simplify the notation, the structural innovations which depend on B are just denoted by e_t. Without loss of generality, let $e_t^{(1)}$ denote the first two innovations, $e_t^{(1)} = [e_{1t}, e_{2t}]'$, and $e_t^{(2)}$ denote the last $k-2$ innovations $e_t^{(2)} = [e_{3t}, \dots, e_{kt}]'$. Identification of the structural VAR is given if one finds conditions to ensure that $B = B_0$, and thus $e_t = \varepsilon_t$.

We assume that z_t is a valid proxy for the first two structural shocks with valid synthetic proxies which contain information beyond the information in z_t.

Assumption 2.1. *The proxy z_t is a relevant and exogenous proxy for the first two shocks with relevant and exogenous synthetic proxies z_t^r with $r = 2, ..., m$. The synthetic proxy contains information beyond the information in z_t, such that*

$$\frac{\alpha_2^{(r)}}{\alpha_1^{(r)}} \neq \frac{\alpha_2}{\alpha_1}.$$

Assumption 2.1 ensures that the synthetic proxies are not redundant and offer new information for the identification of the structural shocks. In Proposition 2.1 we show that the ratios compared in Assumption 2.1 need to differ in order for the proxy VAR to separate the target and non-target shocks. This assumption rules out proxy variables which only contain two values, e.g. representing a positive or negative correlation of the proxy and the target shock for a given date. This excludes, for example, simple binary proxy variables as $z_t = z_t^r$. For other proxy variables, the informational content they express changes with the transformation. E.g. for the squared baseline proxy z_t^2 any possible sign information is lost but it now picks up the nonlinear relation between the proxy and the underlying structural shocks.

The following proposition shows that a valid proxy z_t for the target shocks can be used together with the synthetic proxies z_t^r to separate the target shocks from the non-target shocks. Yet, the target shocks are only identified up to a linear combination and further restrictions are necessary to disentangle them. Proposition 2.1 mimics the results of Mertens and Ravn (2013) in the GMM structural VAR framework. Mertens and Ravn (2013) show that the information from the proxies alone is not sufficient to disentangle the two target shocks related to the proxy.

Proposition 2.1. *If z_t and z_t^r are valid proxy variables and z_t^r satisfies Assumption 2.1, the moment conditions*

$$0 = E[e_t^{(1)} e_t^{(2)\prime}], \qquad 0 = E[z_t e_t^{(2)\prime}], \qquad 0 = E[z_t^r e_t^{(2)\prime}]$$

identify $\varepsilon_t^{(1)}$ up to a linear combination, meaning $e_t^{(1)} = V_{11}\varepsilon_t^{(1)}$ for some invertible matrix V_{11}.

Proof. See Appendix 2.A. □

As in the result by Mertens and Ravn (2013) the separation of target and non-target shocks is possible if one has as many proxies as target shocks. Hence, for two target shocks one needs the baseline proxy z_t and one additional synthetic proxy z_t^r. In the empirical application we use $r = 2, 3$. This allows us to estimate the results for each synthetic proxy individually but we can also estimate the results using both proxies at the same time. This means that, in the latter case, we have more proxies than target shocks what enables us to use overidentification tests based on Sargan (1958) and Hansen (1982) in order to assess the excess moment conditions.

Proposition 2.1 shows that the target shocks in $\varepsilon_t^{(1)}$ are only identified up to an unknown linear combination. To disentangle the target shocks we use the assumption

that the target shocks in $\varepsilon_t^{(1)}$ are mutually independent and at least one of them is non-Gaussian. Assumption 2.2 lists the underlying identification assumptions.

Assumption 2.2.

1. *The target shocks are mutually independent.*
2. *At most one target shock has zero skewness and zero excess kurtosis.*
3. *The target shocks have zero mean, unit variance, and finite third and fourth moments.*

Importantly, we only assume independence and non-Gaussianity for the target shocks and do not impose any restrictions on the higher-order (in)dependency and non-Gaussianity of the non-target shocks.

The following proposition derives identifying moment conditions for a valid proxy z_t and valid synthetic proxies z_t^r under Assumption 2.2.

Proposition 2.2. *If z_t satisfies Assumption 2.1 and the target shocks $\varepsilon_t^{(1)}$ satisfy Assumption 2.2, the moment conditions*

$$0 = E[e_t^{(1)} e_t^{(2)\prime}], \qquad 0 = E[z_t e_t^{(2)\prime}], \qquad 0 = E[z_t^r e_t^{(2)\prime}], \tag{2.3.1}$$

$$1 = E[e_{1t}^2], \qquad 1 = E[e_{2t}^2], \qquad 0 = E[e_{1t} e_{2t}], \tag{2.3.2}$$

$$0 = E[e_{1t}^2 e_{2t}], \qquad 0 = E[e_{1t} e_{2t}^2], \tag{2.3.3}$$

$$0 = E[e_{1t}^3 e_{2t}], \qquad 0 = E[e_{1t} e_{2t}^3], \qquad 1 = E[e_{1t}^2 e_{2t}^2] \tag{2.3.4}$$

identify $\varepsilon_t^{(1)}$ up to sign and permutation.

Proof. Follows from Proposition 2.1 and Keweloh (2021). □

The moment conditions in (2.3.1) represent the proxy VAR conditions, the moment conditions in (2.3.2) are the usual unit variance normalization and uncorrelatedness assumption while the conditions in (2.3.3) and (2.3.4) ensure that the target shocks are independent.

To implement the GMM estimator the reduced form model is estimated by least squares or maximum likelihood in the first step. Let $\hat{u}$ be the $(T-p) \times n$ matrix of estimated reduced form residuals. Then with the identification result from Proposi-

tion 2.2 we use the following GMM estimator using the synthetic proxy variables to identify the structural parameters:

$$\hat{B} := \underset{B \in \mathbb{B}}{\arg\min} \begin{bmatrix} g_z(B, \hat{u}) \\ g_{e^{(1)}}(B, \hat{u}) \end{bmatrix}' W \begin{bmatrix} g_z(B, \hat{u}) \\ g_{e^{(1)}}(B, \hat{u}) \end{bmatrix}$$
$$\text{s.t. } g_e(B, \hat{u}) = 0,$$

with

$$g_z(B, \hat{u}) = \left[\tfrac{1}{T} \textstyle\sum_{t=1}^{T} z_t e_t^{(2)\prime}, \tfrac{1}{T} \sum_{t=1}^{T} z_t^r e_t^{(2)\prime} \right]',$$

$$g_{e^{(1)}}(B, \hat{u}) = \begin{bmatrix} \frac{1}{T} \sum_{t=1}^{T} e_{1t}^2 e_{2t} \\ \frac{1}{T} \sum_{t=1}^{T} e_{1t} e_{2t}^2 \\ \frac{1}{T} \sum_{t=1}^{T} e_{1t}^3 e_{2t} \\ \frac{1}{T} \sum_{t=1}^{T} e_{1t}^2 e_{2t}^2 - 1 \\ \frac{1}{T} \sum_{t=1}^{T} e_{1t} e_{2t}^3 \end{bmatrix}, \quad g_e(B, \hat{u}) = \begin{bmatrix} \frac{1}{T} \sum_{t=1}^{T} e_{1t}^2 - 1 \\ \cdots \\ \frac{1}{T} \sum_{t=1}^{T} e_{kt}^2 - 1 \\ \frac{1}{T} \sum_{t=1}^{T} e_{1t} e_{2t} \\ \cdots \\ \frac{1}{T} \sum_{t=1}^{T} e_{(k-1)t} e_{kt} \end{bmatrix}.$$

Therefore, we minimize the proxy exogeneity conditions $g_z(B, \hat{u})$ and the independent target shocks condition $g_{\tilde{e}_1}(B, \hat{u})$ subject to the constrained $g_e(B, \hat{u}) = 0$ enforcing uncorrelated shocks with unit variance. Note that we use $r = 2, 3$ in the empirical application, and thus the proxy variable moment conditions $g_z(B, \hat{u})$ are:

$$g_z(B, \hat{u}) = \left[\tfrac{1}{T} \textstyle\sum_{t=1}^{T} z_t e_t^{(2)\prime}, \tfrac{1}{T} \sum_{t=1}^{T} z_t^2 e_t^{(2)\prime}, \tfrac{1}{T} \sum_{t=1}^{T} z_t^3 e_t^{(2)\prime} \right]'.$$

Consistency and asymptotic normality of the estimator follow from standard assumptions for GMM estimation, see Hall (2005), and require identification which is ensured by Proposition 2.2 if Assumptions 2.1 and 2.2 hold. As we are not interested in the parameters of $\hat{B}$ but in functions of them, we conduct inference with the residual-based moving block bootstrap for proxy VARs proposed by Jentsch and Lunsford (2022). They adapt the heteroskedasticity robust and asymptotically valid bootstrapping procedure proposed by Brüggemann et al. (2016) to the proxy VAR case. In the following empirical application, we weight the moment conditions with an identity matrix.

2.4. Finite Sample Performance

This section illustrates that the proposed synthetic proxy VAR estimator is capable of estimating the effects of two mutually independent target shocks using a single proxy

variable. In the simulation, the two target shocks represent uncertainty shocks that simultaneously affect the variance of all non-target shocks. Therefore, the simulated target and non-target shocks are not independent. However, by using information contained in the (synthetic) proxy variable, the proposed synthetic proxy VAR estimator is still able to consistently estimate the impact of the target shocks, even if the target and non-target shocks are not independent.

We simulate the structural VAR with two target shocks ε_{1t} and ε_{2t} representing uncertainty shocks and two non-target shocks ε_{3t} and ε_{4t}. In the simulated structural VAR, all shocks simultaneously affect all variables and we use

$$\begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ u_{4t} \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & \cos(-\theta) & -\sin(-\theta) \\ \sin(\theta) & \cos(\theta) & \sin(-\theta) & \cos(-\theta) \\ \cos(-\theta) & -\sin(-\theta) & \cos(\theta) & -\sin(\theta) \\ \sin(-\theta) & \cos(-\theta) & \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \end{bmatrix},$$

with $\theta = -\pi/5$ which extends the bivariate simulations in Gouriéroux et al. (2017), Keweloh (2021), and Lanne and Luoto (2021) to a four-dimensional structural VAR.

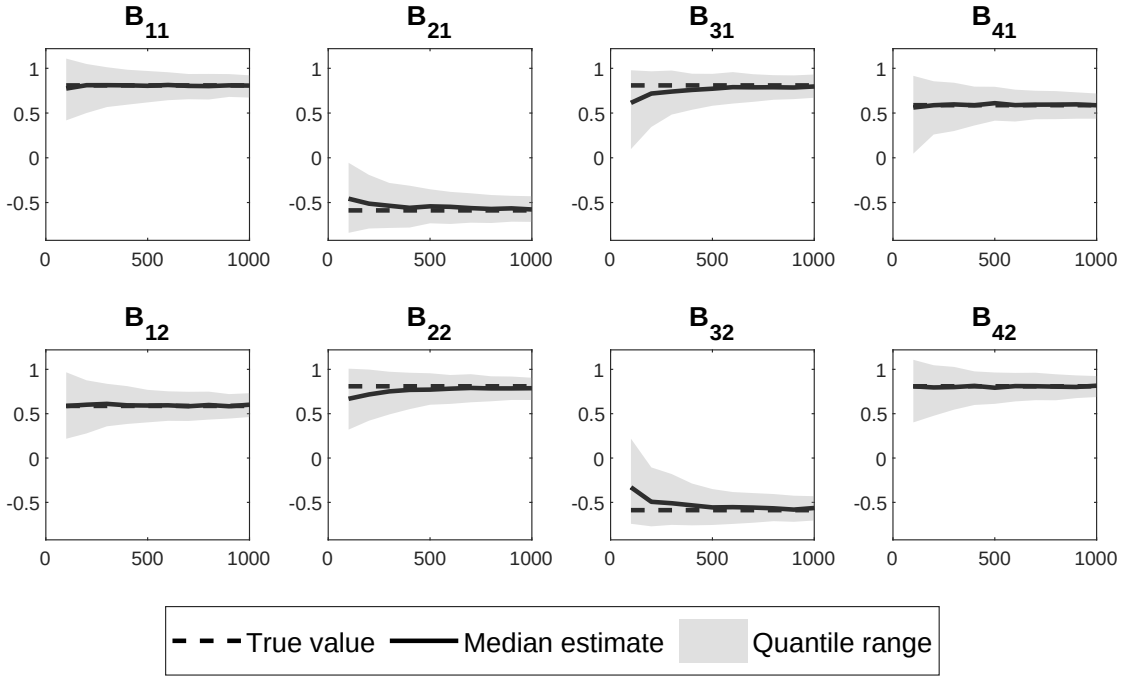
The target shocks ε_{1t} and ε_{2t} are drawn from a Pearson-type distribution with mean zero, unit variance, skewness 0.68, and kurtosis 5.33. The target shocks represent uncertainty shocks and simultaneously affect the variance of the two non-target shocks ε_{3t} and ε_{4t}. Specifically, in each period t the non-target shocks ε_{3t} and ε_{4t} are drawn from a normal distribution with mean zero and standard deviation $1+\varepsilon_{1t}^2+\varepsilon_{2t}^2$. Subsequently, we normalize all shocks to mean zero and unit variance for a given sample.

We generate a proxy variable using

$$z_t = \varepsilon_{1t} + \varepsilon_{2t} + \eta_t,$$

which is correlated with the two target shocks. The proxy noise term η_t is drawn from a normal distribution with mean zero and unit variance. By construction, the proxy variable is uncorrelated with the two non-target shocks. However, the uncertainty shocks ε_{1t} and ε_{2t} affect the proxy variable and the variance of the non-target shocks, and thus the proxy variable is not independent to the non-target shocks. The simulation set-up highlights that it is not necessary to impose independence among all structural shocks but only among the target shocks. Furthermore, it also highlights that we only need mean independence of the non-target shocks to the proxy variable

Figure 2.1: Finite Sample Performance using z_t^2 and z_t^3



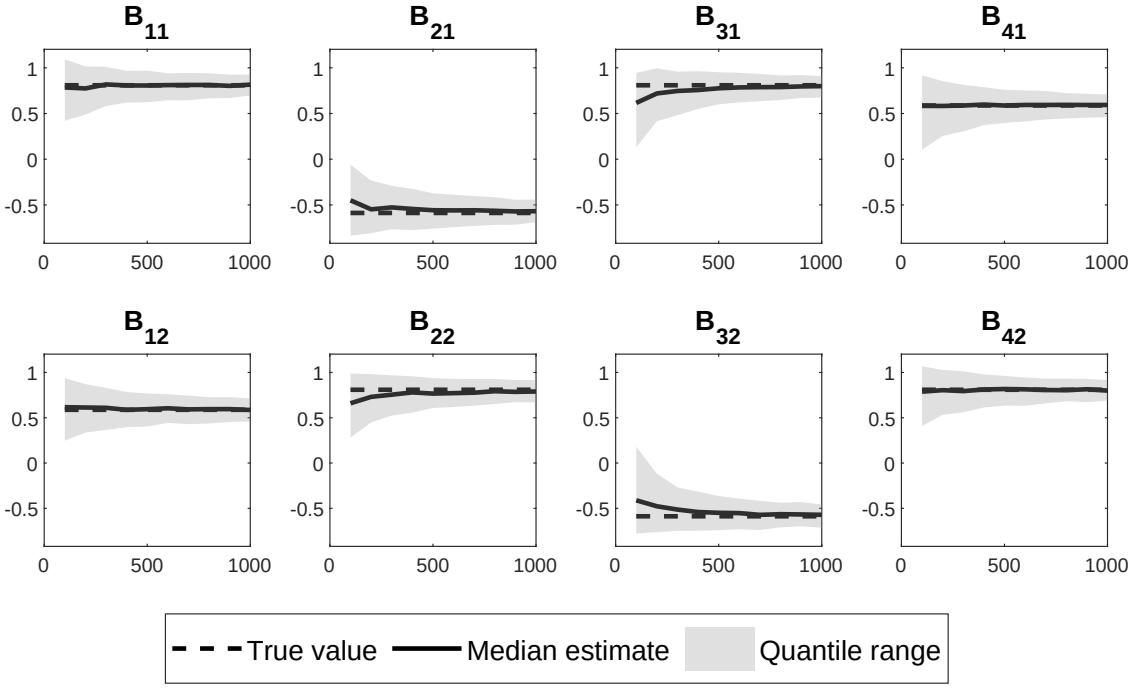
Note: The x-axis represents different sample sizes ranging from 100 to 1000 observations. For each sample size, we conduct 1000 simulations and calculate the median estimate as well as the area between the 16% and 84% quantiles of the estimated impact. The top row of the figure shows the results for the impact of the first target shock, while the bottom row shows the results for the second target shock. The true value, represented by the dashed horizontal lines, is compared to the median estimate and the uncertainty around it, represented by the shaded area.

z_t. This mean independence property yields exogenous synthetic proxy variables z_t^2 and z_t^3.

Note that this specification is neither identified using a traditional proxy VAR estimator since there is only a single proxy variable correlated with two structural shocks, nor is the system identified using a data-driven estimator based solely on independent and non-Gaussian shocks since the shocks are not independent. However, combining both approaches as proposed in the previous section allows for the identification and consistent estimation of the system.

Figure 2.1 shows the median and quantiles of the estimated impact of the two uncertainty shocks on all variables in $1,000$ simulation iterations and samples sizes $T = 100, \ldots, 1,000$ observations. The figure illustrates that the proposed estimator is able to estimate the impact of the two uncertainty shocks. For the estimation both synthetic proxies z_t^2 and z_t^3 are used. The finite sample bias and volatility of the estimator, measured by the quantile range, decrease with an increasing sample size.

Figure 2.2: Finite Sample Performance using only z_t^2



Note: The x-axis represents different sample sizes ranging from 100 to 1000 observations. For each sample size, we conduct 1000 simulations and calculate the median estimate as well as the area between the 16% and 84% quantiles of the estimated impact. The top row of the figure shows the results for the impact of the first target shock, while the bottom row shows the results for the second target shock. The true value, represented by the dashed horizontal lines, is compared to the median estimate and the uncertainty around it, represented by the shaded area.

For comparison Figure 2.2 shows the simulation results when only z_t^2 is used as a synthetic proxy. The results show that only using one synthetic proxy is sufficient to estimate the underlying structural parameters and that efficiency gains from including z_t^3 are marginal. Hence, if in practice only one synthetic proxy is available, e.g. due to instrument strength concerns, the synthetic proxy VAR estimator proposed in section 2.3 has good finite sample properties.

2.5. Identifying Uncertainty Shocks

We use the methodology developed in the previous sections to identify the effects of two different types of uncertainty shocks. One key challenge is that uncertainty is subject to the typical reverse causality issue. Uncertainty potentially affects real activity while it is also likely an endogenous response to economic fluctuations. It is a well-known fact that uncertainty is high during recessions (Bloom, 2014), yet, without

further investigation it is not clear whether uncertainty significantly contributes to the downturns or is merely a result of them.

In the empirical macroeconometric literature a prime candidate to deal with reverse causality are proxy variables. Ideally the proxy is constructed such that it only contains exogenous variation in uncertainty which can be utilized to estimate the causal effect on economic variables. To the best of our knowledge only Piffer and Podstawski (2018) study the effects of uncertainty in a proxy VAR framework. In our study we build on their results, i.e. their proxy and a more detailed description how the proxy is constructed follows in the next subsection.

Piffer and Podstawski (2018) identify general effects of uncertainty shocks. In contrast to that we distinguish between effects of financial market uncertainty and real market uncertainty shocks. We claim that the shock identified by Piffer and Podstawski (2018) contains information of both financial and real uncertainty and we use our methodology to break up the shock into these two independent sources of uncertainty. Hence, we identify multiple uncertainty shocks which so far required multiple instruments to do so. Our methodology allows us to use the single proxy variable we have at our disposal to identify these different types of uncertainty shocks in a proxy VAR framework. The main assumption is that the proxy is mean-independent to the other shocks in the system, i.e. the proxy does not contain any information about the mean of e.g. a monetary or fiscal policy shock. Importantly, we do not assume that the proxy and the non-uncertainty shocks are independent. In particular, we allow that higher moments of the non-uncertainty shocks contain information about the mean of the proxy variable, e.g. the variance of monetary or fiscal policy shocks may contain information about the mean of the uncertainty proxy. This flexibly allows for the interpretation that uncertainty shocks are linked to higher moments of other structural shocks, yet, without imposing this connection.

Compared to the other studies distinguishing between different types of uncertainty shocks, e.g. Ludvigson et al. (2021) or Angelini et al. (2019), we present the first analysis relying on instruments to deal with the reverse causality issue. In the proxy VAR framework with multiple instruments one needs additional restrictions to disentangle the shocks related to the instruments and our application is no exception. For these additional restrictions the researcher can choose from the plethora of different restrictions that are described in the literature. Mertens and Ravn (2013), for instance, rely on a recursive scheme, while Piffer and Podstawski (2018) impose that each shock is more strongly correlated to the instrument which primarily targets it. A recursive scheme is not applicable in our application as this would prevent

uncertainty from being both an exogenous cause and an endogenous response at the same time, rendering any claims to tackle reverse causality indefensible. The latter approach with correlation restrictions is not feasible either as the correlation structure of our synthetic proxies with the target shocks is not clear. Economic theory predicts expansionary as well as contractionary effects of uncertainty on the economy (see e.g. Bloom, 2014). Hence, standard restrictions, such as sign restrictions on impulse responses, are difficult to impose due to the ex ante unclear effect of uncertainty shocks on the economy.

An agnostic candidate to deal with this issue is statistical identification, which does not require to impose any structure on parameters of interest, yet, comes at the cost of needing to label the structural shocks ex post. Our statistical identification relies on assuming that the two uncertainty shocks are not only uncorrelated but also independent and sufficiently non-Gaussian. The assumption of non-Gaussian uncertainty shocks is intuitive. One observes many events that increase uncertainty - recent examples would be the Covid19 pandemic and the war in Ukraine - while there are likely fewer events that lead to a decrease in uncertainty. Hence, we would expect uncertainty shocks to exhibit some skewness and potentially fatter tails than a normal distribution. This non-Gaussian feature enables the statistical identification via independence which uses the information in the higher moments of the shocks. In the literature, Ludvigson et al. (2021) find that the uncertainty shocks they identify are non-Gaussian and we explicitly acknowledge this non-Gaussian feature of uncertainty shocks by accounting for it in our identification strategy.

In summary, we combine the economically intuitive information from the proxy with statistical identification. We separate the uncertainty shocks from the remaining structural shocks using the exogeneity of the proxies and decompose the uncertainty shock into two independent components. For our framework, with two types of uncertainty shocks, we need that at least one of the uncertainty shocks is non-Gaussian and in order to disentangle them we assume that the financial and real uncertainty shocks are independent.

2.5.1. Data and Model Specification

The analysis considers a VAR with 8 variables in monthly frequency. Firstly, stock market volatility (VXO) and the EPU index by Baker et al. (2016) act as different types of uncertainty measurements. Then we include the S&P500 (SP500), industrial production (IP), the effective federal funds rate (FFR), average weekly working hours in manufacturing (HOURS) as a labour market representative, the personal

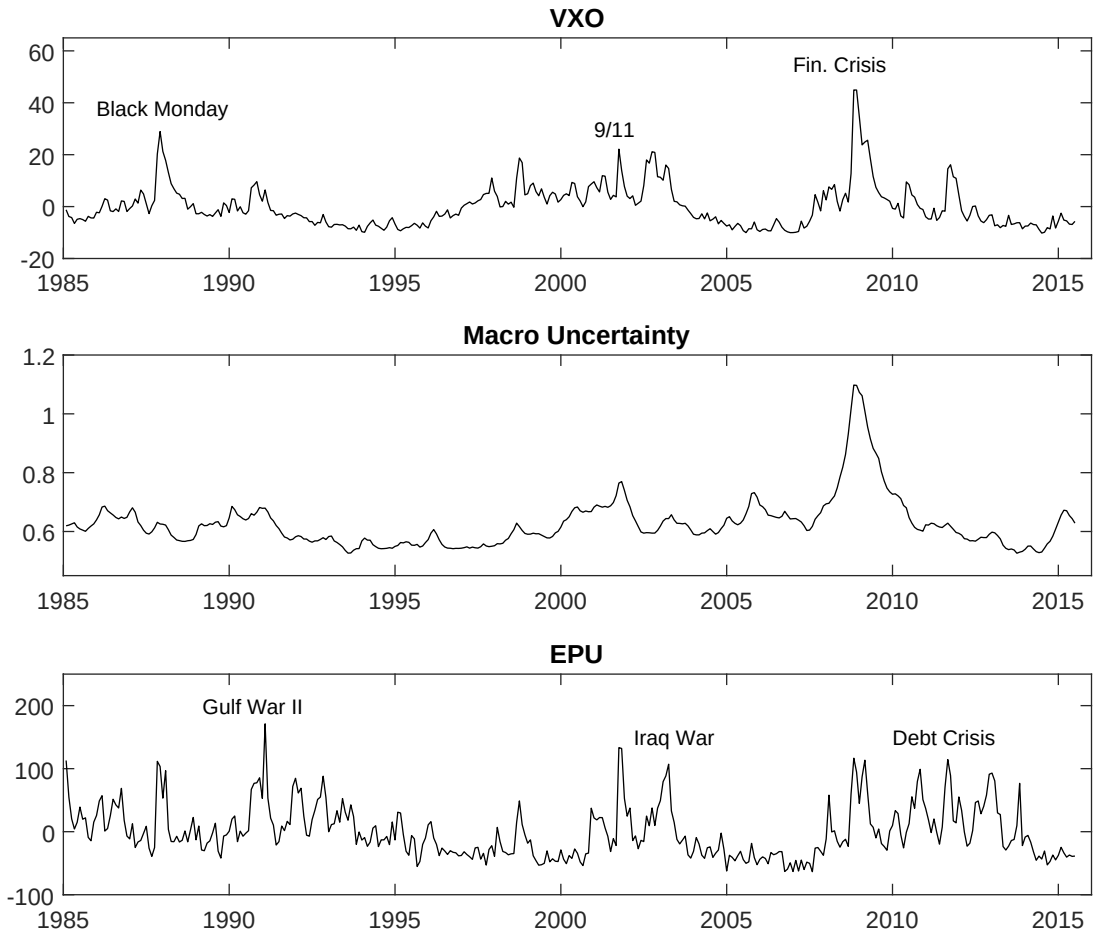
consumption price index (CPI) and Moody's Baa corporate bonds in relation to the 10-Year treasury rate as an indicator for credit spreads (CS). Apart from IP, CPI and SP500, which enter the model as log-differences, all other variables enter the model in levels. A plot of the data series can be found in Figure 2.B.1 of Appendix 2.B. We estimate our model including 12 lags and we use the sample period from January 1985 to June 2015 which corresponds to the availability of the proxy obtained by Piffer and Podstawski (2018). In principle, the synthetic proxy VAR estimator is also applicable if the proxy is only available for a subset of the entire time period of the data, yet we have chosen to match the time periods of the proxy and the data perfectly. This allows us to use the heteroskedasticity robust residual-based moving block bootstrap confidence intervals by Jentsch and Lunsford (2022) which are based on the bootstrap by Brüggemann et al. (2016). The robustness to heteroskedasticity is arguably important when dealing with uncertainty.

For the analysis we chose the EPU index over the Macroeconomic Uncertainty measurement by Jurado et al. (2015) as the EPU yields more distinct information. Figure 2.3 shows that the EPU index spikes during time periods where particular events took place, which can be linked to real uncertainty while the Jurado et al. (2015) index does not or only mildly react to those events. Hence, we believe that the EPU is a better indicator for the real uncertainty notion we have in mind.

2.5.2. The Proxy for Uncertainty Shocks

As previously stated, we build on the work of Piffer and Podstawski (2018) to obtain the proxy in our application. In order to identify their uncertainty shock, they construct a proxy based on changes in the gold price in a narrow window around unexpected historical events. Gold serves as a safe haven asset whose demand spikes during uncertain times. Hence, the relevance assumption is based on changes in the gold prices reflecting changes in uncertainty. They exclusively choose events that could not have been foreseen to ensure the exogeneity of their proxy. Furthermore, they only look at changes of the gold price in a short time interval around the events to rule out that the price changes might also reflect first repercussions of the economy to the particular event. For more details about the construction of the original proxy we refer the reader to the paper by Piffer and Podstawski (2018). They argue that in the chosen time window only the event itself could have evoked the change in the gold price. Yet, the price change could stem from an increase in uncertainty or from the news component of the event.

Figure 2.3: Different Measurements of Uncertainty



Note: The Macro Uncertainty index is based on Jurado et al. (2015) and is provided on the webpage of the authors. Analogously, the EPU index is based on Baker et al. (2016) and provided by the respective authors. The volatility index VXO is taken from the FRED database.

To avoid the effects of the news component, we do not use the original proxy in our application but the median target uncertainty shock that they estimate with their proxy. In their identification strategy they purge the news component from the uncertainty shock, which could be also be reflected in the original gold price proxy. Thus, they identify both an uncertainty shock and a news shock. Assuming that their identification strategy is valid, we use only the uncertainty shock as our proxy in order to avoid dealing with potential news shock contamination. Furthermore, the estimated shock potentially yields more information as it can also spike during events that were not considered in the construction of the proxy by Piffer and Podstawski (2018).

Table 2.1: Correlation between the Proxy and Uncertainty Measurements

Correlation	VXO	EPU	JLN
z_t	0.440	0.321	0.257
z_t^2	0.458	0.316	0.362
z_t^3	0.399	0.274	0.339

Note: VXO denotes stock market volatility, EPU the economic policy uncertainty index by Baker et al. (2016) and JLN the Macroeconomic uncertainty measurement by Jurado et al. (2015). z_t is the estimated median target uncertainty shock by Piffer and Podstawski (2018). The table depicts Pearson's correlation coefficients.

To assess the robustness of the choice of the median target shock as a proxy we also estimate the results with the upper and lower quartile uncertainty shock of Piffer and Podstawski (2018) as the baseline proxy z_t. The impulse responses estimated with the alternative proxies are unchanged compared to the results based on the median target shock. The corresponding results can be found in Appendix 2.B.

Since we want to disentangle financial and real uncertainty shocks, the general uncertainty shock we use as a proxy needs to have predictive power for both financial and real uncertainty. Underlying the shock that we are using as a proxy, are the changes in the gold price. The mechanism behind these changes is that people allocate assets away from e.g. risky and unpredictable stocks to the physically valued gold. This is a logical conclusion, given that financial market-related turmoil is a significant factor in this phenomenon. However, it is also applicable to periods where uncertainty is not evoked by financial market conditions. Historically, the gold price fluctuated during financial market downturns but it reaches its peak during times in which general uncertainty is high, such as during the recent pandemic or in the aftermath of the financial crisis starting in 2008.

In principle, the proxy should capture both variations in uncertainty. In practice, we support this claim by, firstly, calculating the correlation of the uncertainty shock proxy with different measurements of uncertainty to get a feeling whether the proxy is connected to measurements of both financial and real uncertainty. Table 2.1 depicts that measurements for the different types of uncertainty exhibit some notable positive correlation with the uncertainty shock proxy we are using. This same correlation is evident for the synthetic versions of the proxy. We conclude that the proxy contains information about the different types of uncertainty that are measured.

Secondly, we examine the relation of our proxy to the reduced form residuals in our VAR specification. Table 2.2 shows the F-statistics from regressing our proxy

on different reduced form residuals. The first 8 rows depict the F-statistics from bivariate regressions of z_t, z_t^2 or z_t^3 on the reduced form innovations u_{it} corresponding to a specific variable in our system. The last row, furthermore, depicts the F-statistic of the regression on all reduced form residuals. Based on the work of Stock et al. (2002) and Stock and Yogo (2005), Lunsford (2015) derives critical values corresponding to the latter F-statistics in order to test which strength is necessary to eliminate the asymptotic bias due to (weak) proxies. The F-statistics from the bivariate regressions serve as an informal way to assess the strength of the relation between the proxy and the innovations of specific variables. This informal assessment shows that the proxies help to explain variation for financial market related variables like the VXO, the S&P500 returns and credit spreads but they also explain variation in the EPU index. Yet, the proxies seem to have greater predictive power for the financial uncertainty shock, although one can also deduce predictive power for the real uncertainty shock. A formal assessment of the general strength of the proxies can be made by examining the F-statistics in the last row. The resulting critical values correspond to how likely it is that a certain portion of the bias remains. Based on these values, we conclude that all of our three proxies are deemed strong enough. Expressing it in terms of how Lunsford (2015) constructs the critical values for the test, we conclude that for a 1% significance level less than 5% of the asymptotic weak proxy VAR bias remains. Yet, the weak proxy test is designed for a single proxy variable and it is not clear whether the critical values exactly transfer to the case of multiple proxy variables. Nevertheless, based on the results of the individual proxies, together with the informal assessment described above, we consider all three proxies to have sufficient predictive power for our analysis.

Hence, both synthetic proxies with $r = 2, 3$ are suitable for use in the GMM estimator. The resulting overidentification enables us to conduct a J-test based on Hansen (1982) which comfortably does not reject the null hypothesis of valid moment conditions. Therefore, we use both synthetic proxies in the analysis, yet, estimates based on using only one synthetic proxy at a time are presented in Appendix 2.B.

2.5.3. Labelling the Shocks

The GMM estimator described in section 2.3 identifies the structural shocks statistically, meaning that no restrictions are imposed a priori which would allow to assign a certain label to the respective shocks. Hence, the labelling has to be done after estimating the model. In this section, we examine the estimated quantities coming from the synthetic proxy VAR estimator in order to sort out which shock corresponds

Table 2.2: Testing Relations of the Proxies to Specific Variables

F-Statistic	z_t	z_t^2	z_t^3
VXO	311.93	119.45	110.39
EPU	47.40	22.68	13.22
SP500	125.77	48.39	42.17
IP	19.16	0.04	0.11
FFR	2.51	4.46	3.56
HOURS	10.51	0.12	0.10
CPI	0.89	8.53	11.02
CS	70.55	51.40	52.16
Lunsford (2015)	61.56	16.47	15.56

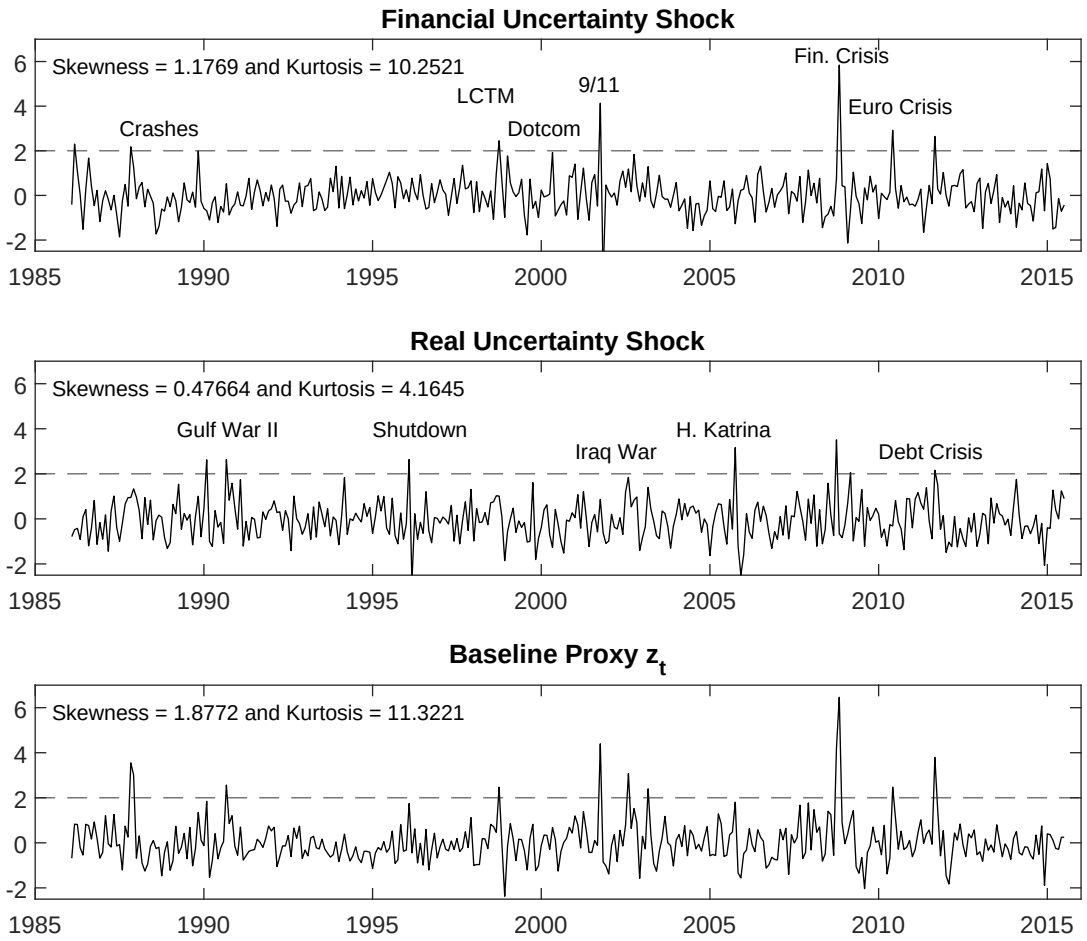
Note: The table depicts F-statistics of regressions of the proxies onto reduced form residuals of specific variables in our system u_{it}. The last columns depicts results for the weak proxy test of Lunsford (2015).

to the financial and which to the real uncertainty shock. This can, for example, be accomplished by examining the structural impulse response functions. However, it is not entirely clear how these differ for the two types of uncertainty shocks ex ante, and thus we rely on the estimated shock series themselves for the labelling. For uncertainty shocks this is particularly useful as most of the shocks' spikes should be linkable to specific historical events that occurred at that time. Ludvigson et al. (2021) actually build their identification out of event constraints which employ the timing of uncertainty shocks at specific events.

Figure 2.4 plots the two identified uncertainty shocks and the baseline proxy z_t. The first observation is that both shocks exhibit non-zero skewness and excess kurtosis. The respective magnitudes are given in Figure 2.4. Hence, as expected they show clear signs of non-Gaussianity. Moreover, the extreme values of the first shock, which we label as the financial market uncertainty shock, are more pronounced compared to the second, the real uncertainty shock. This higher maximum amplitude for a financial uncertainty shock is consistent with the more volatile nature of stock markets, on which extreme events evoke extreme reactions.

Yet, we base our labelling primarily on the timing of the spikes. For the financial uncertainty shocks we observe e.g. spikes in the late 80s. The spike at 1987:10 marks the Black Monday while the spike at 1989:10 is associated with the Black Fri-

Figure 2.4: Labelling of the Uncertainty Shocks



Note: The upper panel depicts the estimated first shock of system $\hat{\varepsilon}_{1t}$ related to the proxy variable, which we label as the financial uncertainty shock. The middle panel depicts the second estimated shock $\hat{\varepsilon}_{2t}$ related to the proxy variable, which is labelled as the real uncertainty shock. The bottom panel shows the baseline proxy z_t.

day mini-crash. The two days in October of the respective years mark severe stock market crashes and are clearly events that caused uncertainty on financial markets. Interestingly, the proxy z_t shows no spike during 1989:10 and only after breaking up the uncertainty shock this event becomes apparent. For the financial uncertainty shock we observe further spikes that (almost) exceed two standard deviations around the millennium. The date 1998:09 marks the month of the Long-Term Capital Management (LCTM) default, during which a large hedge fund received a substantial bailout from a group of banks. In 2000:04 the Dotcom bubble burst, followed by a significant downturn in the stock market and in 2001:09 the terrorist attacks on the World Trade Center had severe effects on the stock market. Even a part of its infras-

tructure was destroyed on that day. The largest shock can be observed in 2008:10 which is the month of the large stock market downturn caused by the financial crisis. Following the financial crisis the shocks in 2010:06 and 2011:08 exceeded two standard deviations. In both cases several events could be linked to these spikes. In June 2010, numerous important decisions were made by European institutions to tackle the European debt crisis with potential spillovers to investors on the U.S. stock markets. Furthermore, 2010:06 was the month of the first stock market flash crash. In 2011:08 the U.S. credit rating was downgraded, which was followed by a substantial stock market plummet. Yet, at the same time the U.S. debt ceiling crisis was unfolding, which likely contributed to the rating downgrade. Hence, it is unclear which of the two events is reflected in our shock series. As the shock from the upper panel already picks up other financial market related events we argue that the spike in 2011:08 can be attributed to the rating downgrade and the resulting stock market plummet. Conversely, the spike in the real uncertainty shock is attributed to the debt ceiling crisis. This would be supported by Ludvigson et al. (2021), who point out that the debt ceiling crisis of 2011 is a prime example of a macro or real uncertainty shock.

We observe further spikes in our real uncertainty shock series around the second Gulf war and the Iraq war. With regard to the second Gulf war we find a spike in 1990:08, the month of the Iraqi invasion of Kuwait. Additionally, we find a large shock in the build-up of the war, which we cannot link to a specific event. Interestingly, we observe a similar pattern for the Iraq war in 2003. We have a large shock in 2003:02 which is right before the start of the armed conflict, while we also observe a spike a couple of months prior. We interpret this as uncertainty in the build-up of the war, where there is uncertainty as to whether there will be a war or not. Attributing war-related events to the real uncertainty shock is again consistent with Ludvigson et al. (2021), who argue that wars are a contributor to real uncertainty. Apart from the wars, we have a substantial shock in 2005:09, which is the month in which Hurricane Katrina hit the south eastern United States. In addition to the destruction of a large part of the infrastructure in the affected states, the storm also caused substantial damage to oil production facilities in the Gulf of Mexico. Furthermore, we have a large spike in 1996:01 which marks the government shutdown during the Clinton era.

As for the financial uncertainty shock, the largest shock can, unsurprisingly, be observed during the financial crisis. However, the spike for the real uncertainty shock happens in 2008:09, the month of the Lehman collapse. Hence, we find the financial uncertainty shock in 2008:10 and the real uncertainty shock in 2008:09. In the study of Ludvigson et al. (2021) the timing of the two shocks is reversed. We argue that the

stock market plummet in 2008:10 aligns well with the notion of a financial uncertainty shock. The Lehman default and the policy decision to abandon the 'too big to fail' concept should have caused some real or policy uncertainty. Therefore, the timing of the shocks during the financial crisis fits the notion of the shocks we have in mind.

In conclusion, we find that the differently timed spikes and the corresponding events sort themselves well into two categories: those related to financial markets and those affecting the economy in general. Furthermore, we identify some new considerable spikes which were not originally reflected in the proxy but can be related to particular events and fit the labelling of the shocks.

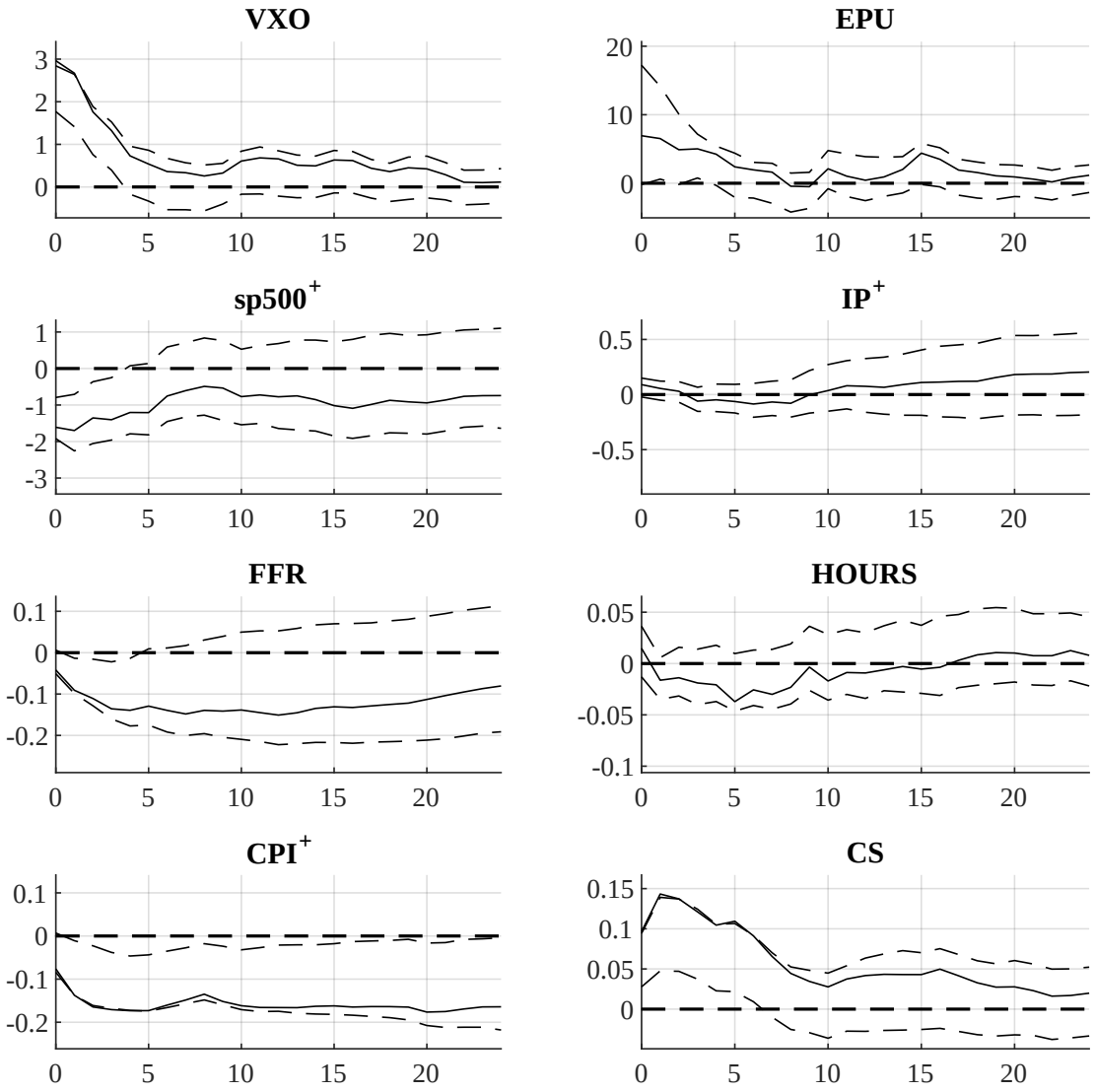
2.5.4. Impulse Responses

Figures 2.5 and 2.6 display the responses of the VAR variables to the estimated uncertainty shocks. Both figures depict the responses of a one standard deviation financial and real uncertainty shock, respectively. As expected, both uncertainty shocks increase uncertainty, measured by the VXO and the EPU. However, the response of the EPU to the real uncertainty shocks is not statistically significant.

An examination of the response of the financial variables reveals that they exhibit similar responses to both uncertainty shocks. In particular, we find that both shocks lead to an immediate sharp decline in the S&P500 returns. It is reasonable to conclude that both types of uncertainty shocks have negative effects on stock market returns as people, for instance, reallocate their portfolios away from stocks to safe haven assets. Moreover, we find that both shocks lead to higher credit spreads. Judging by the point estimates, the financial uncertainty shock has a more pronounced effect. This observation supports the labelling of the shocks. Increasing credit spreads are viewed as a key channel of uncertainty affecting the economy as they reflect increasing risk premia. This effect is predominantly attributed to financial uncertainty shocks and Shin and Zhong (2020) even use the ordering that financial uncertainty should affect credit spreads more than real uncertainty shocks as a ranking restriction in their identification strategy. Hence, our findings reinforce the hypothesis that widening credit spreads are caused by financial uncertainty shocks but, to a potentially lesser extent, also by real uncertainty shocks. Furthermore, our results indicate that both uncertainty shocks lead to a decline in interest rates and the confidence intervals suggest rather short-lived effects, again with potentially larger effects for the financial uncertainty shocks.

Comparing the responses for the set of 'real' market variables reveals larger disparities between the shocks. Firstly, we find a decrease in industrial production for

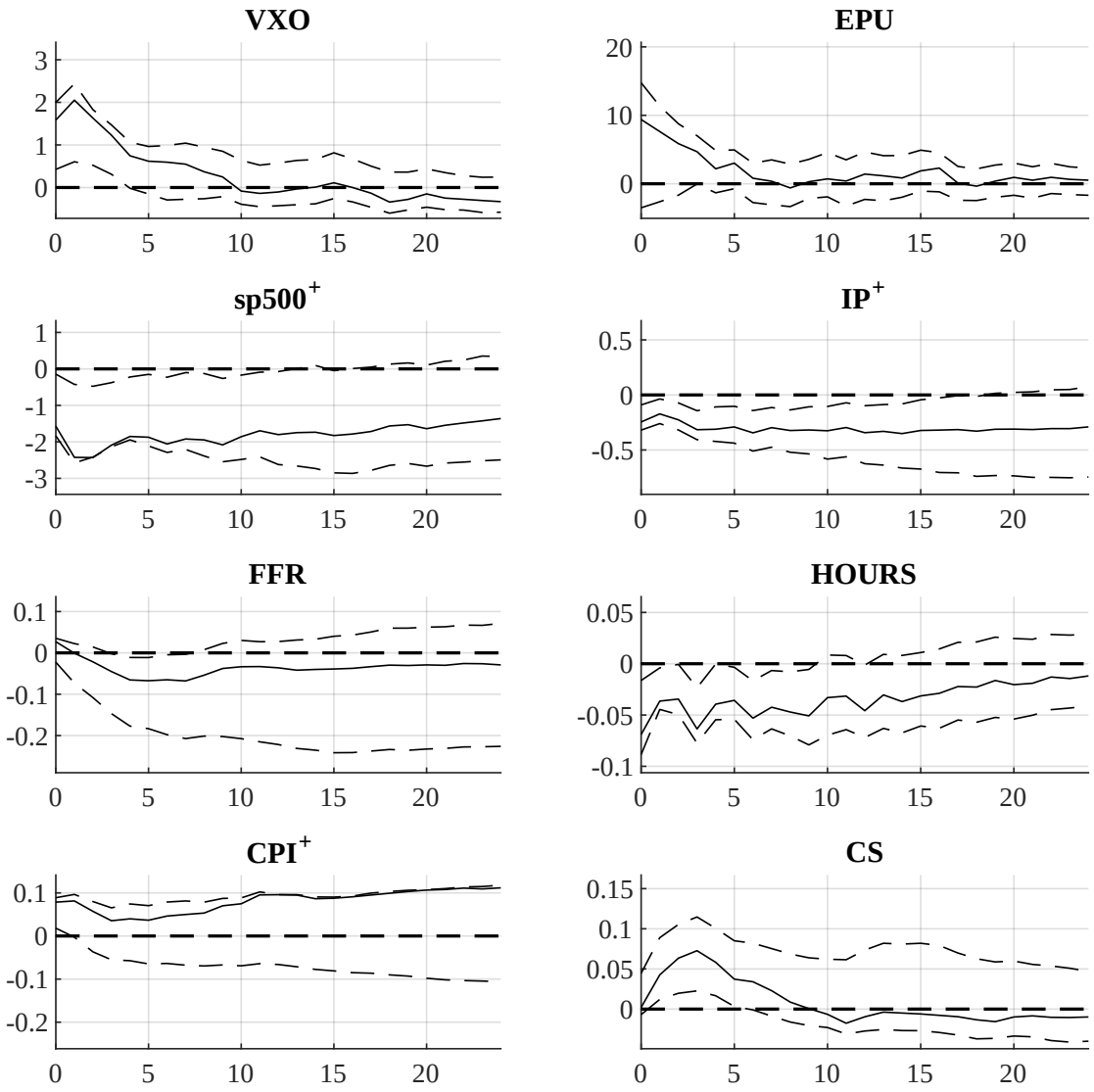
Figure 2.5: Impulse Responses to the Financial Uncertainty Shock



Note: The dashed lines are the 90% residual-based moving block bootstrap confidence intervals by Jentsch and Lunsford (2022) based on 1,000 bootstrap iterations. Responses marked with a + depict cumulative responses which are multiplied by 100.

the real uncertainty shock while the financial uncertainty shock does not affect real activity negatively. The reason for this can potentially be found in the effects of the uncertainty shocks on labour markets. We find a significant decrease in hours worked for the real uncertainty shock, both at impact and for almost a full year after the shock. In contrast, the financial uncertainty shock does not evoke any significant reaction of the labour market. This is also evident when the average weekly work-

Figure 2.6: Impulse Responses to the Real Uncertainty Shock



Note: The dashed lines are the 90% residual-based moving block bootstrap confidence intervals by Jentsch and Lunsford (2022) based on 1,000 bootstrap iterations. Responses marked with a + depict cumulative responses which are multiplied by 100.

ing hours are substituted with employment growth. Firms appear to immediately react with cuts in their labour usage to real uncertainty shocks which could explain the direct effect on industrial production. The respective figures can be found in Appendix 2.B. For financial uncertainty the credit spread channel seems to be more important, which consists of firms delaying investment due to increased costs and

risks. In general, the effects are rather short-lived which may be attributed to the urge of policymakers to resolve uncertain situations quickly and their efforts to do so.

Finally, our results suggest adverse effects on the price level. While the financial uncertainty shock results in a decrease in the price level, the real uncertainty shock increases prices at impact. As pointed out by Carriero and Volpicella (2024), the literature on uncertainty reports adverse effects on prices and disentangling uncertainty into different types could help to explain those findings.

In relation to the literature that disentangles uncertainty our results help to fill in some gaps. Angelini et al. (2019) estimate the effects for separate time periods with different volatility regimes. Their results for the time period which contains our sample period are very similar. Yet, we obtain the results for a larger scale model which helps to identify potential channels through which the types of uncertainty affect the economy. However, our results are at odds with those of Ludvigson et al. (2021) who find that real uncertainty does not affect real activity negatively. However, they estimate a small model with only three variables, a real uncertainty, a financial uncertainty and an output measurement. When we reduce our model to their variables and employ our proposed identification strategy our results look similar to theirs. Hence, there might be problems with informational deficiencies. This could explain why the labelling of our shocks is consistent with their identification strategy but yields different results. Carriero and Volpicella (2024) as well as Carriero et al. (2018) find significant and immediate effects on real activity to both types of uncertainty shocks. Yet, our proxy VAR approach potentially solves the reverse causality issue more successfully than their respective identification strategies. Remaining endogeneity could explain the more pronounced effects observed in their studies. Finally, our results are at odds with those of Shin and Zhong (2020) who obtain similar results to Ludvigson et al. (2021). As their model is similar to the one by Carriero et al. (2018) the same reasons might explain the differences to our results.

As previously mentioned, Appendix 2.B presents the results for different specifications. Firstly, it presents IRFs for different proxy variables, i.e. Figures 2.B.2 to 2.B.5 depict the results when only using z_t^2 or z_t^3 as a single synthetic proxy variable. Figures 2.B.6 to 2.B.9 show the results when using the upper and lower quartile shock of Piffer and Podstawski (2018) as the baseline proxy z_t. The results remain largely consistent with those presented in this subsection. Furthermore, Figures 2.B.10 and 2.B.11 present estimates when substituting hours worked with employment and, lastly, Figures 2.B.12 to 2.B.15 present results for lag orders $p = 6$ and $p = 18$ in the underlying VAR model. Once again, the results and the drawn conclusions are unchanged.

Table 2.3: FEV Explained by the Financial Uncertainty Shock

Variable	$h = 1$	$h = 6$	$h = 12$	$h = 24$
VXO	0.73 (0.90 / 0.42)	0.57 (0.73 / 0.26)	0.44 (0.56 / 0.20)	0.37 (0.46 / 0.15)
EPU	0.07 (0.61 / 0.00)	0.12 (0.57 / 0.02)	0.10 (0.45 / 0.03)	0.11 (0.39 / 0.04)
SP500	0.26 (0.52 / 0.10)	0.22 (0.45 / 0.09)	0.20 (0.37 / 0.09)	0.19 (0.35 / 0.09)
IP	0.03 (0.13 / 0.00)	0.06 (0.13 / 0.02)	0.07 (0.14 / 0.04)	0.06 (0.14 / 0.04)
FFR	0.08 (0.19 / 0.00)	0.16 (0.37 / 0.01)	0.12 (0.32 / 0.01)	0.08 (0.26 / 0.01)
HOURS	0.01 (0.07 / 0.00)	0.04 (0.11 / 0.01)	0.04 (0.12 / 0.01)	0.02 (0.16 / 0.01)
CPI	0.24 (0.41 / 0.01)	0.29 (0.37 / 0.05)	0.26 (0.32 / 0.06)	0.24 (0.28 / 0.07)
CS	0.51 (0.62 / 0.08)	0.47 (0.57 / 0.10)	0.30 (0.41 / 0.07)	0.18 (0.30 / 0.05)

Note: The table depicts the FEV decomposition of the respective variable at different horizons h. The values in parenthesis depict the upper and lower end of the 90% residual-based moving block bootstrap confidence intervals by Jentsch and Lunsford (2022) based on 1,000 iterations.

2.5.5. Forecast Error Variance Decompositions

The FEV decomposition results are largely in line with the findings from the results presented in section 2.5.4. Notably, they show that the real uncertainty shock does not explain much of the variation in credit spreads and interest rates. The financial uncertainty shock on the other hand does not explain a notable portion of the variation in real activity and the average weekly working hours. This supports the hypothesis that the labour market channel - uncertainty-induced layoffs or delayed hiring - plays a more important role in the propagation of uncertainty shocks to the real economy.

Upon examination of the FEV decomposition of the uncertainty measurements, one observes that the majority of the variation in the VXO is explained by the two uncertainty shocks, with the financial uncertainty shock representing the primary contributor. However, only a modest portion of the variation in the EPU is explained by the two shocks. We conclude that financial uncertainty, measured by the VXO, is largely driven by uncertainty shocks while real uncertainty, measured by the EPU, is to a large degree driven by the other business cycle shocks. This finding is in line with the literature (e.g. Ludvigson et al., 2021). Additionally, we find that real uncertainty

Table 2.4: FEV Explained by the Real Uncertainty Shock

Variable	$h = 1$	$h = 6$	$h = 12$	$h = 24$
VXO	0.21 (0.46 / 0.03)	0.31 (0.49 / 0.07)	0.24 (0.40 / 0.07)	0.19 (0.32 / 0.07)
EPU	0.14 (0.44 / 0.00)	0.16 (0.42 / 0.01)	0.13 (0.35 / 0.02)	0.11 (0.30 / 0.03)
SP500	0.25 (0.45 / 0.01)	0.27 (0.40 / 0.04)	0.22 (0.31 / 0.05)	0.22 (0.28 / 0.06)
IP	0.24 (0.56 / 0.06)	0.24 (0.45 / 0.10)	0.20 (0.38 / 0.10)	0.18 (0.33 / 0.10)
FFR	0.03 (0.11 / 0.00)	0.03 (0.30 / 0.01)	0.02 (0.29 / 0.01)	0.01 (0.28 / 0.01)
HOURS	0.16 (0.32 / 0.02)	0.18 (0.32 / 0.04)	0.19 (0.37 / 0.05)	0.14 (0.34 / 0.05)
CPI	0.25 (0.43 / 0.04)	0.20 (0.33 / 0.08)	0.19 (0.29 / 0.09)	0.17 (0.26 / 0.09)
CS	0.00 (0.18 / 0.00)	0.09 (0.36 / 0.03)	0.05 (0.29 / 0.03)	0.03 (0.23 / 0.02)

Note: The table depicts the FEV decompositions of the respective variable at different horizons h. The value in parenthesis depict the upper and lower end of the 90% residual-based moving block bootstrap confidence intervals by Jentsch and Lunsford (2022) based on 1,000 iterations.

shocks affect real activity. Financial uncertainty shocks affect the economy as well but mainly the financial sector which appears to transmit only modest effects to the rest of the economy. Furthermore, we find that at least for longer horizons variation in the uncertainty measures is only partly explained by the uncertainty shocks. Hence, also other economic fluctuations evoke uncertainty. This reflects the reverse causality relationship between uncertainty and other economic variables. Uncertainty shocks affect real activity, but economic fluctuations also evoke uncertainty.

2.6. Conclusion

In conclusion, we develop a novel identification strategy that allows for the identification of multiple shocks in a proxy VAR with only a single proxy variable. To do so, we construct synthetic proxy variables and show under which conditions they are valid proxies. The relevance assumption holds if the target shocks are non-Gaussian and exogeneity of the synthetic proxies holds given the assumption that the non-target shocks are mean-independent of the proxy variable z_t. The target shocks are

disentangled assuming that they are independent and at least one of them is non-Gaussian. Hence, we pair the identification via proxy variables with the agnostic statistical identification approach using independent non-Gaussian shocks.

We apply our methodology to identify and disentangle two uncertainty shocks, a financial and a real market uncertainty shock. Previous research has demonstrated that uncertainty shocks exhibit non-Gaussian characteristics and our identification approach exploits this feature. Apart from that, our identification strategy suits the empirical application well as we deal with potential reverse causality issues through the proxy VAR and avoid restrictions on a priori unclear effects of uncertainty through statistical identification.

The results suggest that real uncertainty shocks have a more pronounced adverse effect on real economic activity than financial uncertainty shocks. This can potentially be explained by the more pronounced effects on the labour market. Financial uncertainty affects stock returns and credit spreads, however, their impact on real economic activity is insignificant. Looking at uncertainty measurements, we find that real uncertainty is to a large degree driven by the endogenous response to business cycle or non-uncertainty shocks. In contrast, uncertainty originating on financial markets is primarily driven by the financial uncertainty shocks.

References

Angelini, G., E. Bacchiocchi, G. Caggiano, and L. Fanelli (2019). "Uncertainty Across Volatility Regimes". *Journal of Applied Econometrics* 34. (3), 437–455.

Angelini, G., G. Caggiano, E. Castelnuovo, and L. Fanelli (2023). "Are Fiscal Multipliers Estimated with Proxy-SVARs Robust?" *Oxford Bulletin of Economics and Statistics* 85. (1), 95–122.

Baker, S. R., N. Bloom, and S. J. Davis (2016). "Measuring Economic Policy Uncertainty". *The Quarterly Journal of Economics* 131. (4), 1593–1636.

Bloom, N. (2014). "Fluctuations in Uncertainty". *Journal of Economic Perspectives* 28. (2), 153–76.

Braun, R. and R. Brüggemann (2022). "Identication of SVAR Models by Combining Sign Restrictions with External Instruments". *Journal of Business & Economic Statistics* 41. (4), 1077–1089.

Brüggemann, R., C. Jentsch, and C. Trenkler (2016). "Inference in VARs with Conditional Heteroskedasticity of Unknown Form". *Journal of Econometrics* 191. (1), 69–85.

Caldara, D. and E. Herbst (2019). "Monetary Policy, Real Activity, and Credit Spreads: Evidence from Bayesian Proxy SVARs". *American Economic Journal: Macroeconomics* 11. (1), 157–192.

Carriero, A., T. E. Clark, and M. Marcellino (2018). "Measuring Uncertainty and Its Impact on the Economy". *Review of Economics and Statistics* 100. (5), 799–815.

Carriero, A. and A. Volpicella (2024). "Max Share Identification of Multiple Shocks: An Application to Uncertainty and Financial Conditions". *Journal of Business & Economic Statistics* 0, 1–13.

Giacomini, R., T. Kitagawa, and M. Read (2022). "Robust Bayesian Inference in Proxy SVARs". *Journal of Econometrics* 228. (1), 107–126.

Gouriéroux, 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.

Hall, A. R. (2005). *Generalized Method of Moments*. Oxford University Press.

Hansen, L. P. (1982). "Large Sample Properties of Generalized Method of Moments Estimators". *Econometrica* 50. (4), 1029–1054.

Jentsch, C. and K. G. Lunsford (2022). "Asymptotically Valid Bootstrap Inference for Proxy SVARs". *Journal of Business & Economic Statistics* 40. (4), 1876–1891.

Jurado, K., S. C. Ludvigson, and S. Ng (2015). "Measuring Uncertainty". *American Economic Review* 105. (3), 1177–1216.

Keweloh, S. A. (2021). "A Generalized Method of Moments Estimator for Structural Vector Autoregressions Based on Higher Moments". *Journal of Business & Economic Statistics* 39. (3), 772–782.

Lanne, M. and J. Luoto (2021). "GMM Estimation of Non-Gaussian Structural Vector Autoregression". *Journal of Business & Economic Statistics* 39. (1), 69–81.

Ludvigson, S. C., S. Ma, and S. Ng (2021). "Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response?" *American Economic Journal: Macroeconomics* 13. (4), 369–410.

Lunsford, K. G. (2015). "Identifying Structural VARs with a Proxy Variable and a Test for a Weak Proxy". *Federal Reserve Bank of Cleveland, Working Paper No. 15-28*.

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), 1212–47.

Piffer, M. and M. Podstawski (2018). "Identifying Uncertainty Shocks Using the Price of Gold". *The Economic Journal* 128. (616), 3266–3284.

Sargan, J. D. (1958). "The Estimation of Economic Relationships using Instrumental Variables". *Econometrica* 26. (3), 393–415.

Shin, M. and M. Zhong (2020). "A New Approach to Identifying the Real Effects of Uncertainty Shocks". *Journal of Business & Economic Statistics* 38. (2), 367–379.

Stock, J. H., J. H. Wright, and M. Yogo (2002). "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments". *Journal of Business & Economic Statistics* 20. (4), 518–529.

Stock, J. H. and M. Yogo (2005). "Testing for Weak Instruments in Linear IV Regression". In: *Identification and Inference for Econometric Models*. Ed. by Andrews, D. W. New York: Cambridge University Press, 80–108.

Appendix 2.A Proofs

Proof of Proposition 1

Proof. Step 1: Show that the proxy conditions imply that $e_t^{(2)}$ is a linear combination of the structural shocks $\varepsilon_t^{(2)}$.
With $e_t = B^{-1}u_t = B^{-1}B_0\varepsilon_t$ it follows that the innovations e_t are equal to a linear combination of the structural shocks ε_t. For any B let $e_t = V\varepsilon_t$ with

$$\begin{bmatrix} e_t^{(1)} \\ e_t^{(2)} \end{bmatrix} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_t^{(1)} \\ \varepsilon_t^{(2)} \end{bmatrix}.$$

If $V_{21} \neq 0$, the innovations $e_t^{(2)}$ are a linear combination of both the target shocks and the non-target shocks. Yet, the condition $0 = E[z_t e_t^{(2)\prime}]$ implies

$$\begin{aligned} 0 &= E[z_t e_t^{(2)\prime}] \\ 0 &= E[z_t \left(V_{21}\varepsilon_t^{(1)} + V_{22}\varepsilon_t^{(2)}\right)'] \\ 0 &= E[z_t \varepsilon_t^{(1)\prime} V_{21}'] \\ 0 &= E[z_t \varepsilon_t^{(1)\prime}] V_{21}' \\ 0 &= \alpha V_{21}'. \end{aligned} \tag{2.A.1}$$

As $\varepsilon_t^{(1)}$ is of dimension 2×1, equation (2.A.1) can be written as

$$0 = \alpha_1 V_{21,1}' + \alpha_2 V_{21,2}',$$

where $V_{21,i}$ denotes the i-th column of V_{21} with $i \in \{1, 2\}$. If

$$V_{21,1}' = -V_{21,2}' \frac{\alpha_2}{\alpha_1}, \tag{2.A.2}$$

$e_t^{(2)}$ is not only a linear combination of $\varepsilon_t^{(2)}$, but also of $\varepsilon_t^{(1)}$. However, for some r the additional synthetic proxy condition $0 = E[z_t^r e_t^{(2)\prime}]$ implies similarly to $0 = E[z_t e_t^{(2)\prime}]$:

$$0 = \alpha^{(r)} V_{21}',$$

and thus

$$\begin{aligned} 0 &= \alpha_1^{(r)} V_{21,1}' + \alpha_2^{(r)} V_{21,2}' \\ \implies V_{21,1}' &= -V_{21,2}' \frac{\alpha_2^{(r)}}{\alpha_1^{(r)}}. \end{aligned} \tag{2.A.3}$$

By Assumption 2.1 it holds that $\frac{\alpha_2^{(r)}}{\alpha_1^{(r)}} \neq \frac{\alpha_2}{\alpha_1}$, and thus equations (2.A.2) and (2.A.3) only hold if $V_{21}' = 0$, which implies that $e_t^{(2)}$ is a linear combination of $\varepsilon_t^{(2)}$ only and is not affected by $\varepsilon_t^{(1)}$. Hence, the baseline proxy and its synthetic proxy successfully separate the non-target shocks from the target shocks.

Step 2: Show that if $e_t^{(2)}$ is a linear combination of only the structural shocks $\varepsilon_t^{(2)}$ (meaning $V_{21} = 0$), the uncorrelatedness and unit variance assumption imply that $e_t^{(1)}$ is a linear combination of only the structural shocks $\varepsilon_t^{(1)}$ (meaning $V_{12} = 0$).
If V is invertible, $V_{21} = 0$ implies that V_{22} has full rank. If $V_{21} = 0$ and V_{22} has full rank, uncorrelatedness implies

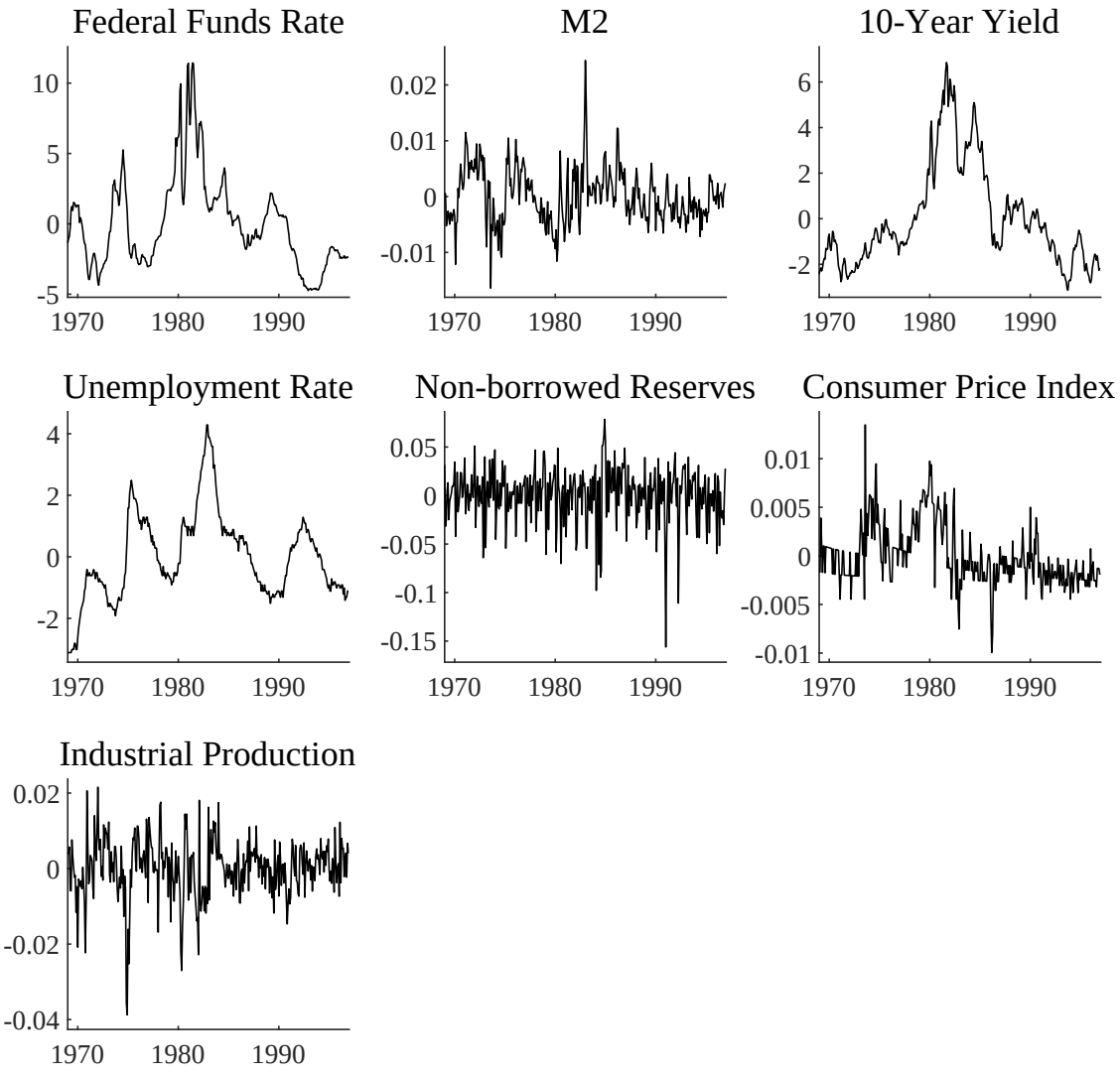
$$\begin{aligned} 0 &= E\left[e_t^{(1)} e_t^{(2)\prime}\right] \\ &= E\left[\left(V_{11}\varepsilon_t^{(1)} + V_{12}\varepsilon_t^{(2)}\right)\left(V_{21}\varepsilon_t^{(1)} + V_{22}\varepsilon_t^{(2)}\right)'\right] \\ &= E\left[\left(V_{11}\varepsilon_t^{(1)} + V_{12}\varepsilon_t^{(2)}\right)\left(V_{22}\varepsilon_t^{(2)}\right)'\right] \\ &= V_{11} \underbrace{E\left[\varepsilon_t^{(1)}\varepsilon_t^{(2)\prime}\right]}_{=0} V_{22}' + V_{12} \underbrace{E\left[\varepsilon_t^{(2)}\varepsilon_t^{(2)\prime}\right]}_{=I_{k-2}} V_{22}' \end{aligned} \tag{2.A.4}$$

Hence, (2.A.4) states that $0 = V_{12}V_{22}'$. As V_{22} is of full rank, and thus invertible, (2.A.4) only holds if $V_{12} = 0$. Therefore, $e_t^{(1)}$ is a linear combination of only the structural shocks $\varepsilon_t^{(1)}$. Thus, the proxy VAR successfully separates the target shocks from the non-target shocks but identifies only an unknown linear combination V_{11} of them.

□

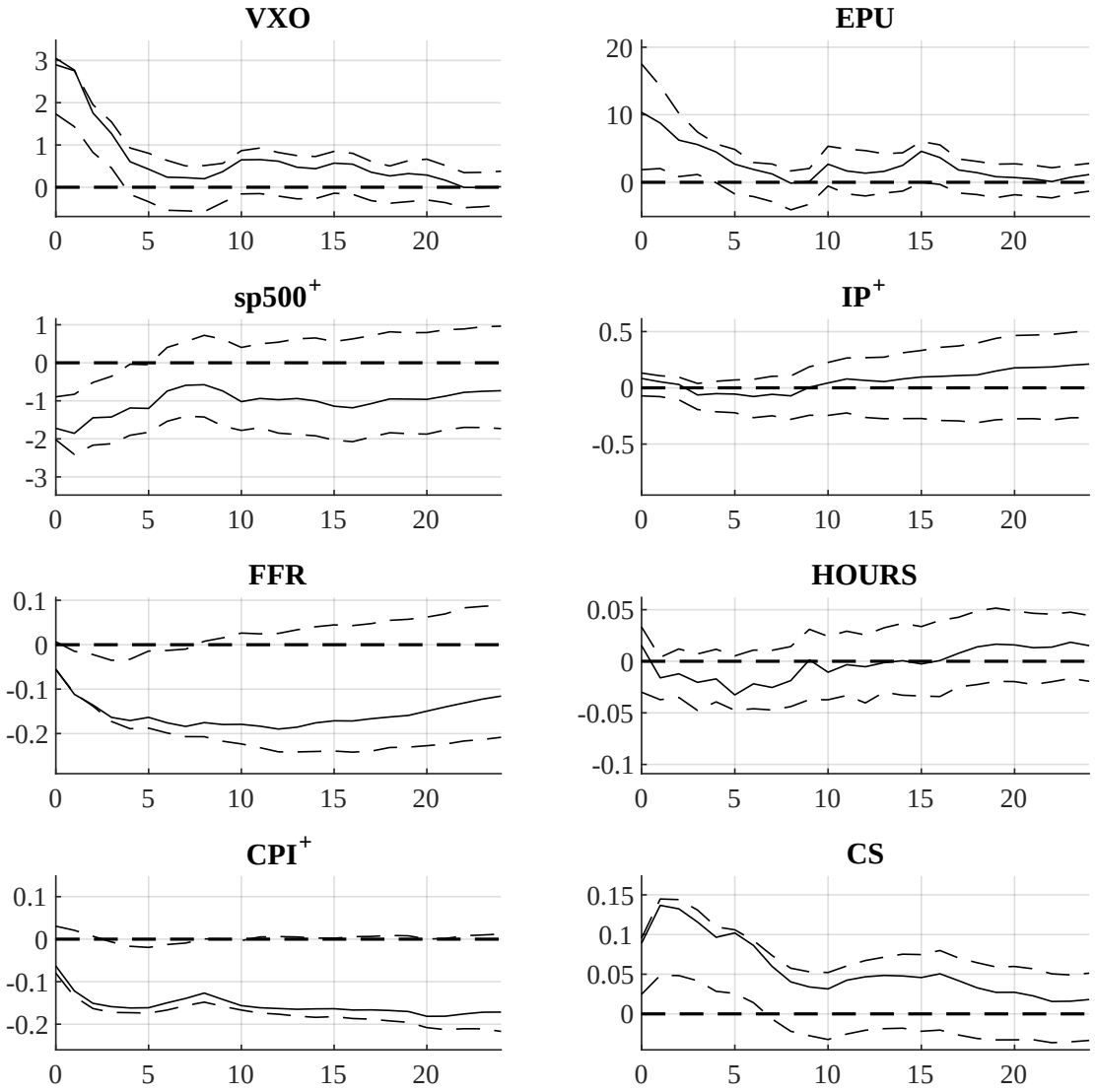
Appendix 2.B Additional Empirical Results

Figure 2.B.1: Data Used in the Baseline Analysis



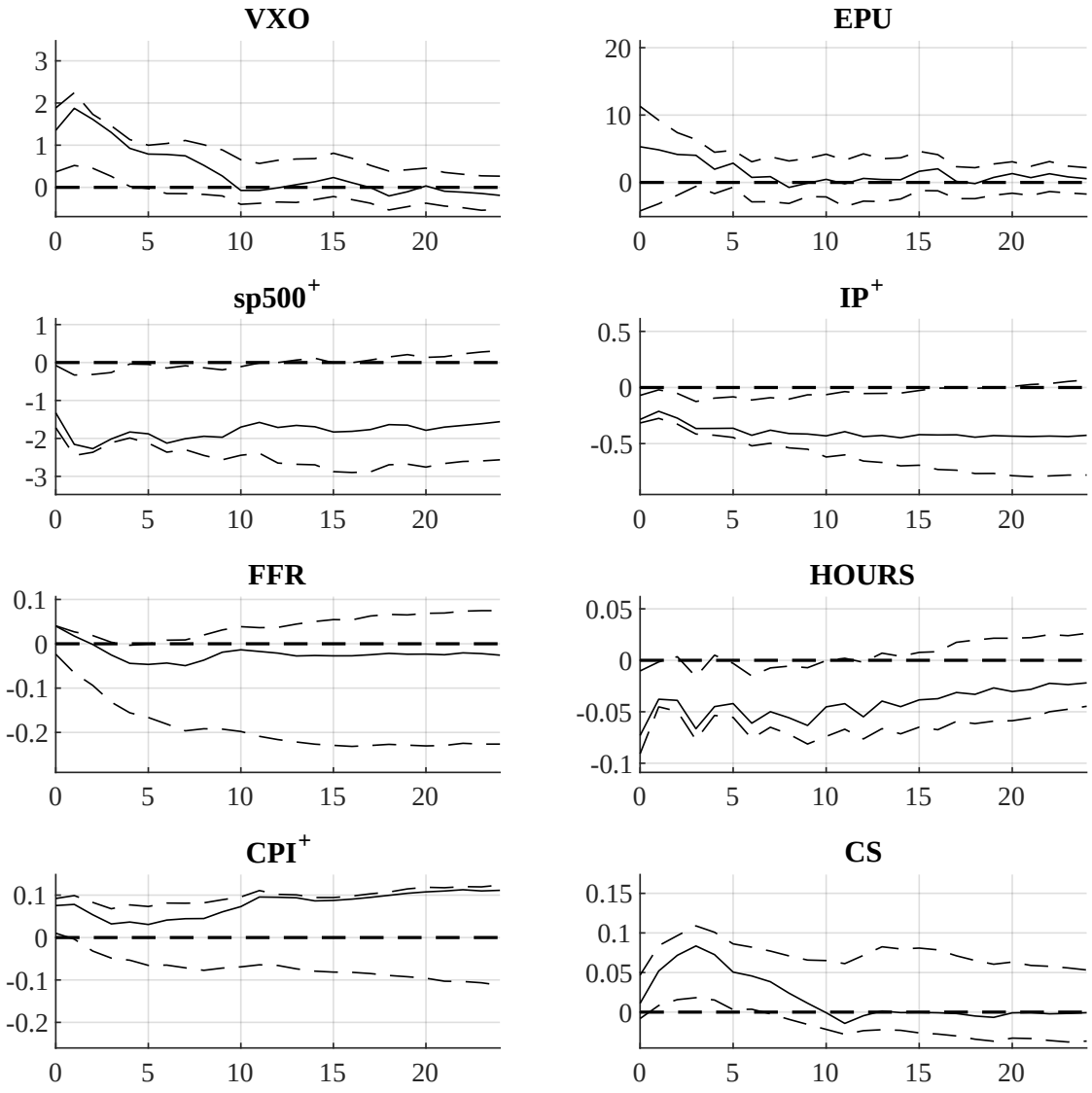
Note: The data for industrial production (IP), the consumer price index (CPI), the monetary base M2 (M2) and the S&P500 (SP500) depict log-differences, and thus monthly growth rates.

Figure 2.B.2: Impulse Responses to the Financial Uncertainty Shock using only z_t^2 as a Synthetic Proxy



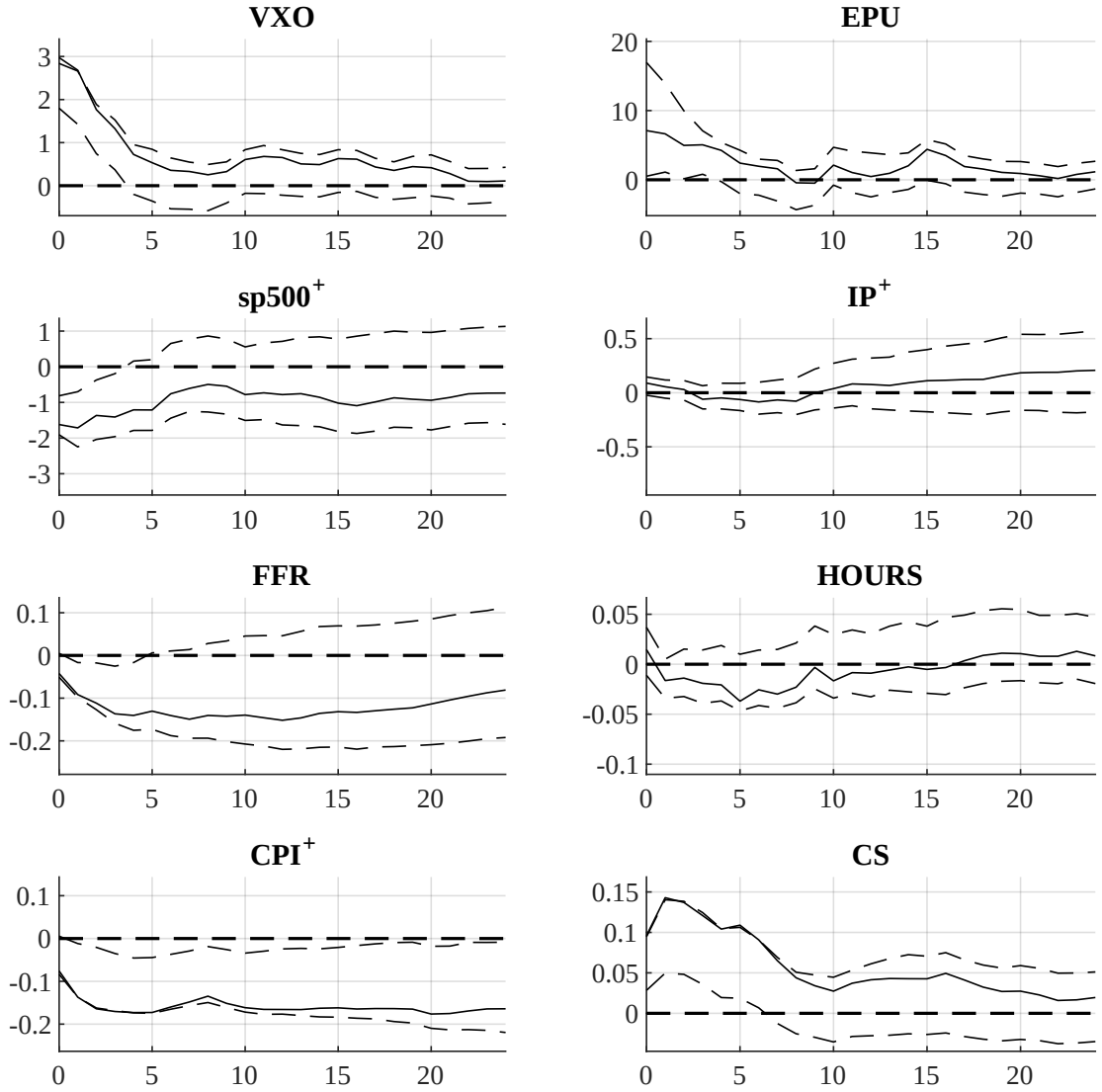
Note: The solid lines depict the point estimates of the structural impulse response functions using only z_t^2 as a synthetic proxy variable instead of z_t^2 and z_t^3. The dashed lines depict the 90% residual-based moving block bootstrap confidence intervals by Jentsch and Lunsford (2022) based on 1,000 iterations. The labelling of the shock is based on the same logic as described in section 2.5.3 of the main text.

Figure 2.B.3: Impulse Responses to the Real Uncertainty Shock using only z_t^2 as a Synthetic Proxy



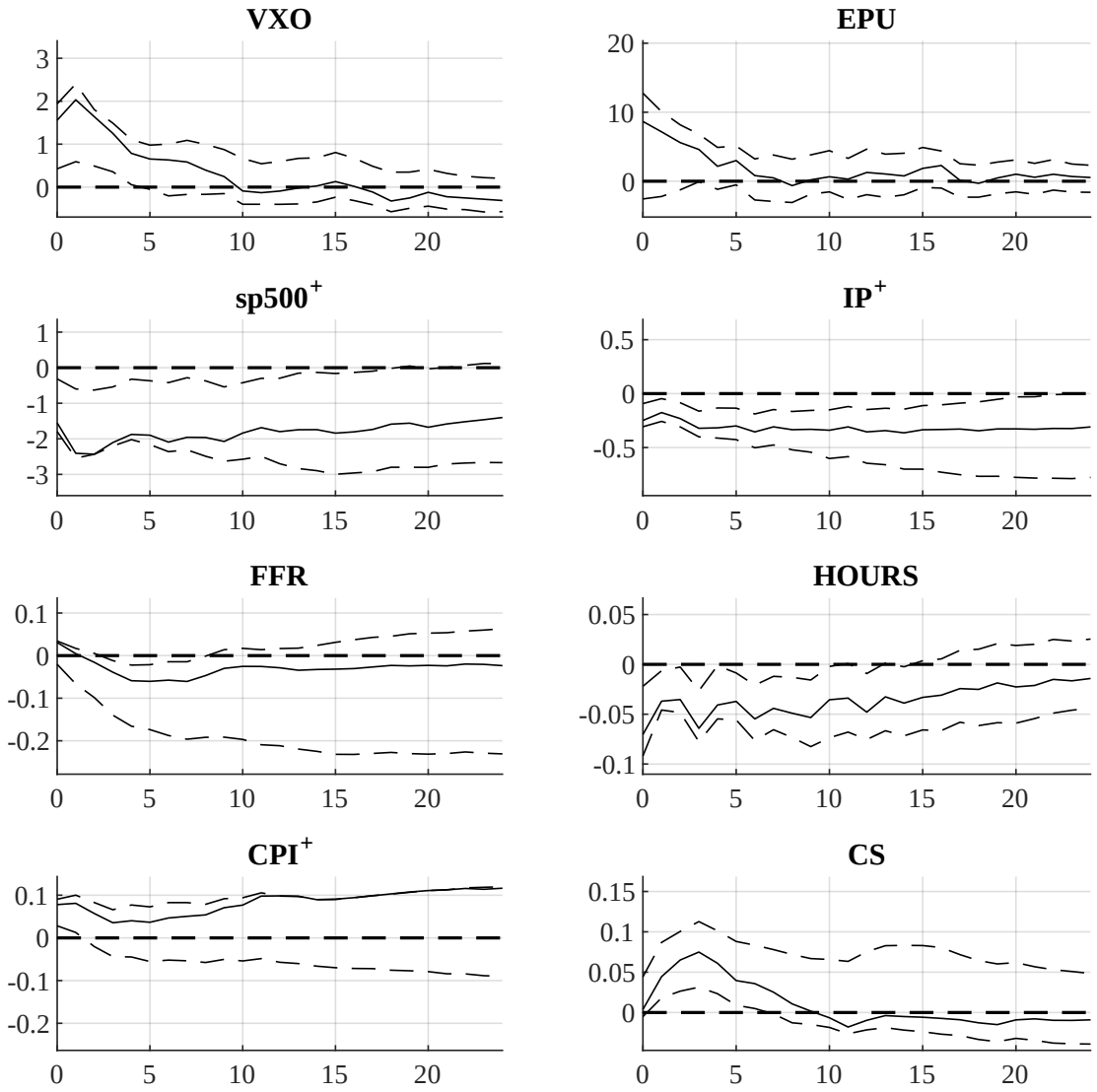
Note: The solid lines depict the point estimates of the structural impulse response functions using only z_t^2 as a synthetic proxy variable instead of z_t^2 and z_t^3. The dashed lines depict the 90% residual-based moving block bootstrap confidence intervals by Jentsch and Lunsford (2022) based on 1,000 iterations. The labelling of the shock is based on the same logic as described in section 2.5.3 of the main text.

Figure 2.B.4: Impulse Responses to the Financial Uncertainty Shock using only z_t^3 as a Synthetic Proxy



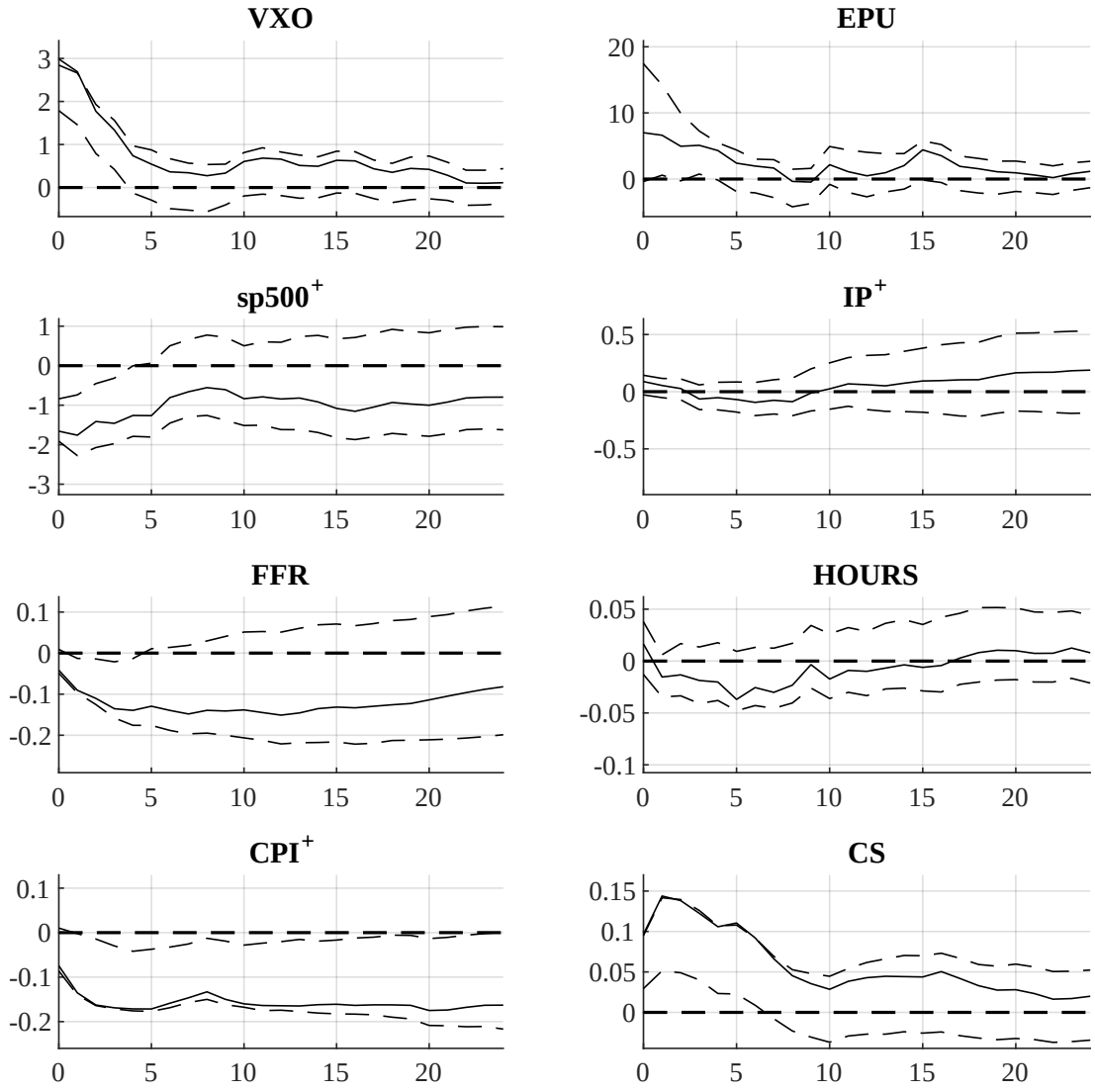
Note: The solid lines depict the point estimates of the structural impulse response functions using only z_t^3 as a synthetic proxy variable instead of z_t^2 and z_t^3. The dashed lines depict the 90% residual-based moving block bootstrap confidence intervals by Jentsch and Lunsford (2022) based on 1,000 iterations. The labelling of the shock is based on the same logic as described in section 2.5.3 of the main text.

Figure 2.B.5: Impulse Responses to the Real Uncertainty Shock using only z_t^3 as a Synthetic Proxy



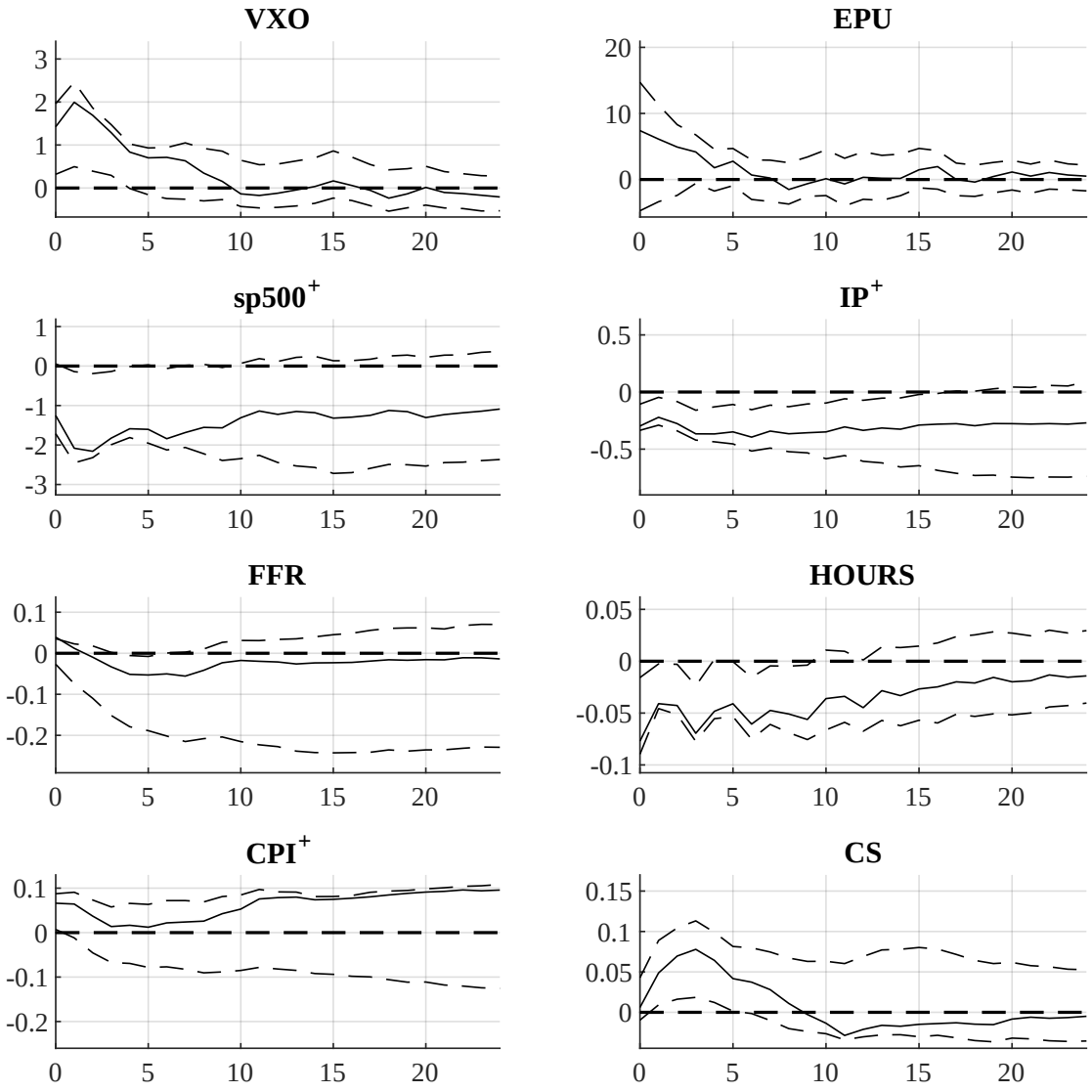
Note: The solid lines depict the point estimates of the structural impulse response functions using only z_t^3 as a synthetic proxy variable instead of z_t^2 and z_t^3. The dashed lines depict the 90% residual-based moving block bootstrap confidence intervals by Jentsch and Lunsford (2022) based on 1,000 iterations. The labelling of the shock is based on the same logic as described in section 2.5.3 of the main text.

Figure 2.B.6: Impulse Responses to the Financial Uncertainty Shock using the Lower Quartile Uncertainty Shock by Piffer and Podstawski (2018)



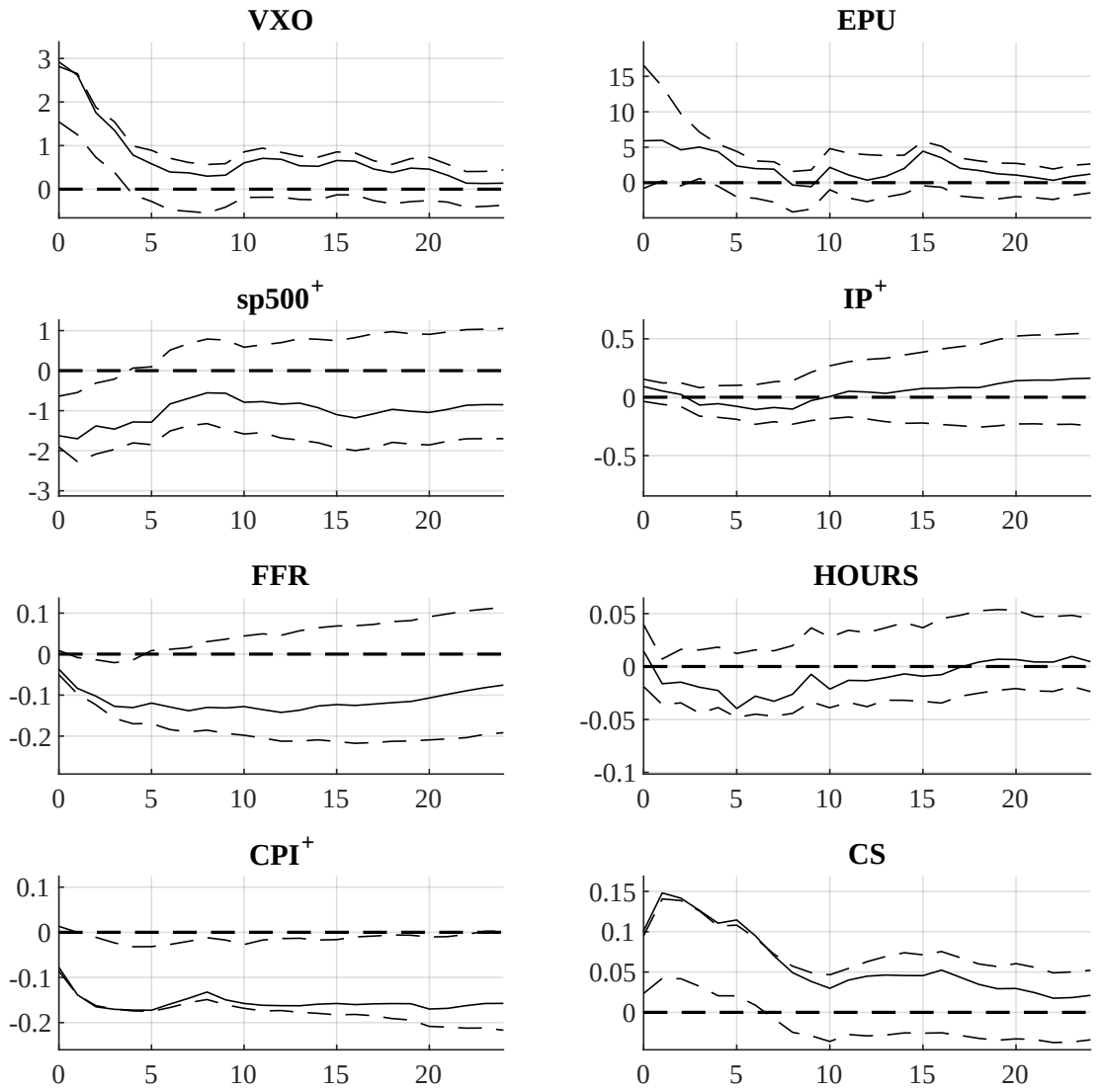
Note: The solid lines depict the point estimates of the structural impulse response functions using the 25% quantile of the uncertainty shock by Piffer and Podstawski (2018) instead of the median. The dashed lines depict the 90% residual-based moving block bootstrap confidence intervals by Jentsch and Lunsford (2022) based on 1,000 iterations. The labelling of the shock is based on the same logic as described in section 2.5.3 of the main text.

Figure 2.B.7: Impulse Responses to the Real Uncertainty Shock using the Lower Quartile Uncertainty Shock by Piffer and Podstawski (2018)



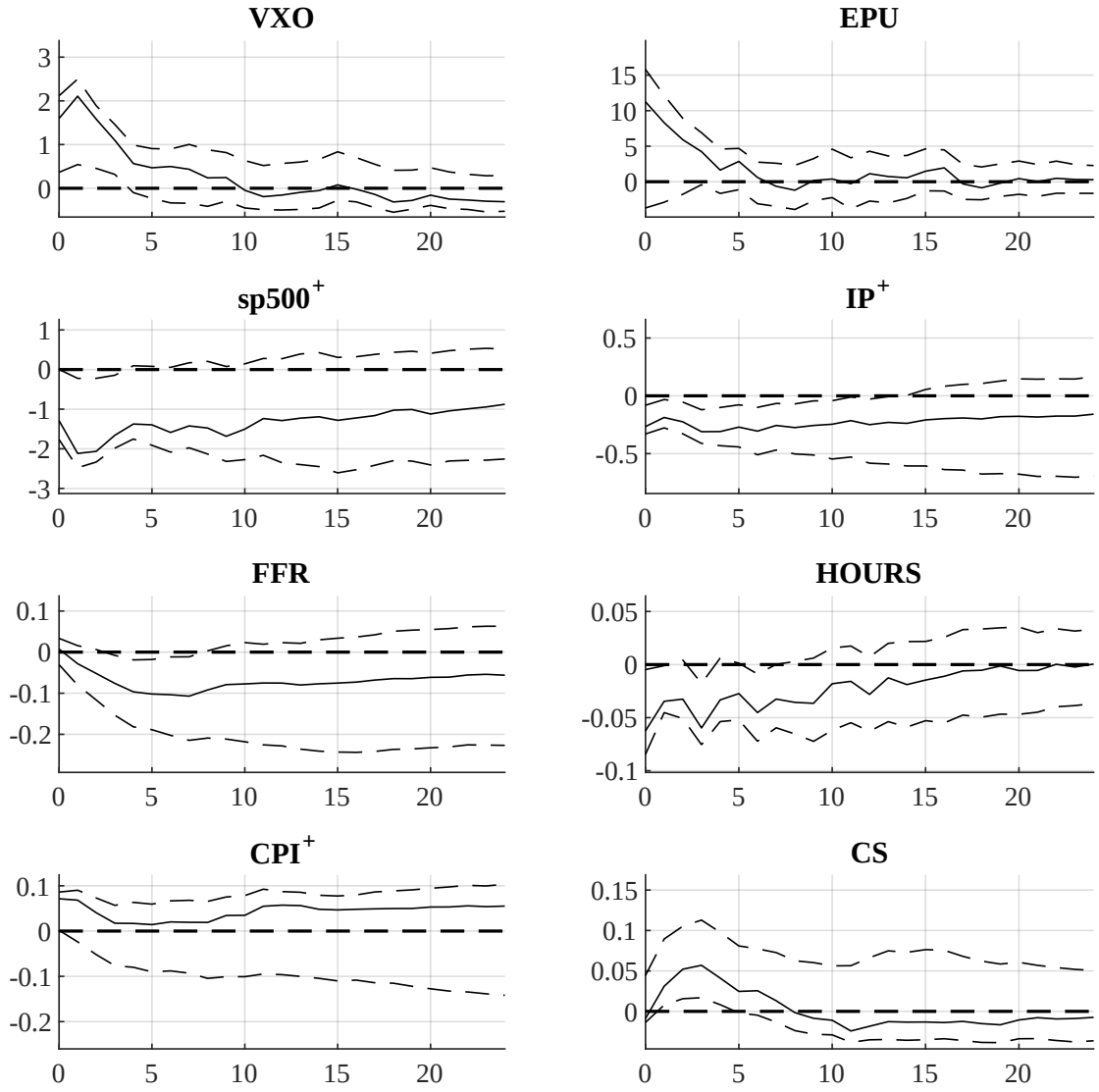
Note: The solid lines depict the point estimates of the structural impulse response functions using the 25% quantile of the uncertainty shock by Piffer and Podstawski (2018) instead of the median. The dashed lines depict the 90% residual-based moving block bootstrap confidence intervals by Jentsch and Lunsford (2022) based on 1,000 iterations. The labelling of the shock is based on the same logic as described in section 2.5.3 of the main text.

Figure 2.B.8: Impulse Responses to the Financial Uncertainty Shock using the Upper Quartile Uncertainty Shock by Piffer and Podstawski (2018)



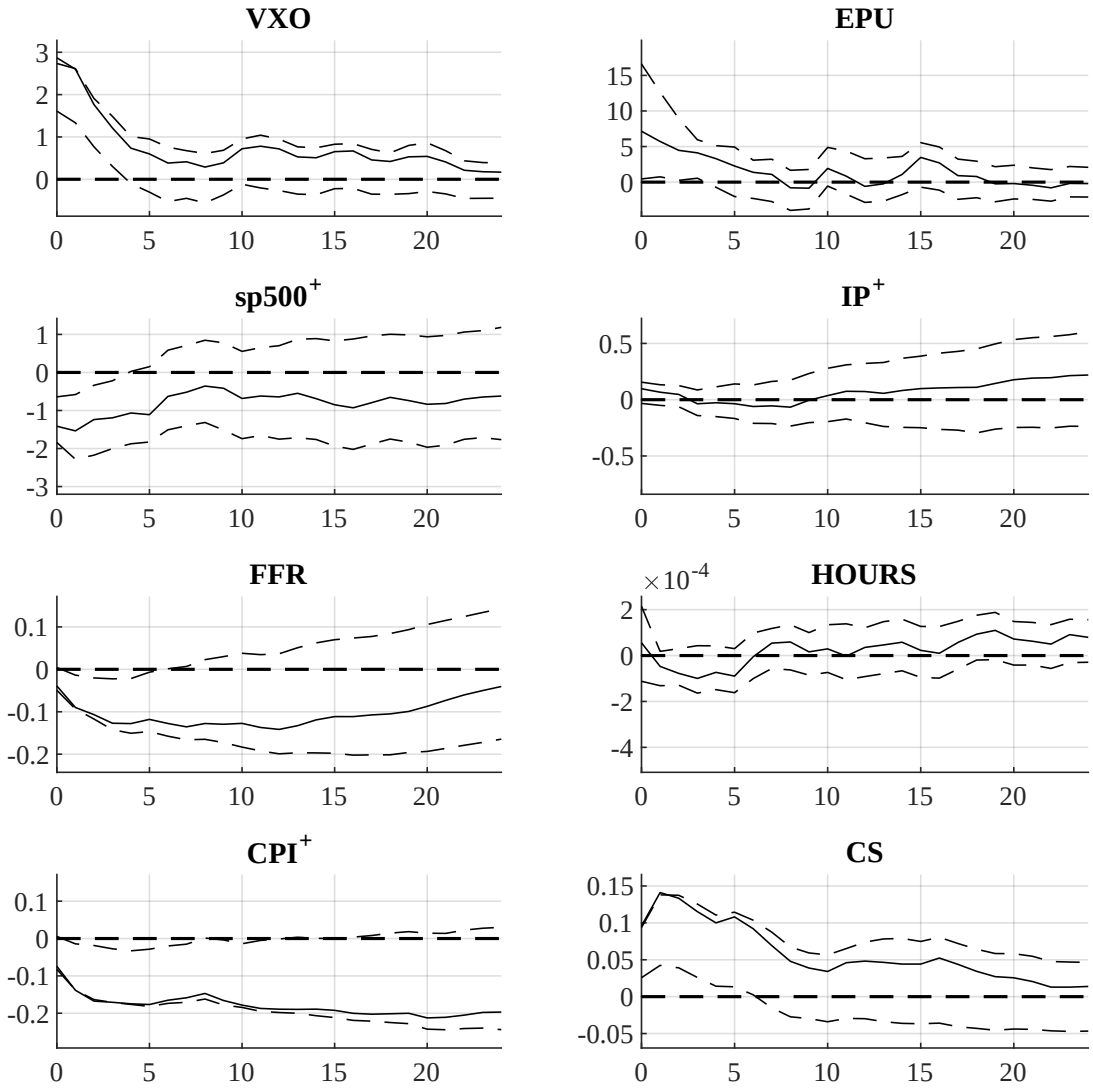
Note: The solid lines depict the point estimates of the structural impulse response functions using the 25% quantile of the uncertainty shock by Piffer and Podstawski (2018) instead of the median. The dashed lines depict the 90% residual-based moving block bootstrap confidence intervals by Jentsch and Lunsford (2022) based on 1,000 iterations. The labelling of the shock is based on the same logic as described in section 2.5.3 of the main text.

Figure 2.B.9: Impulse Responses to the Real Uncertainty Shock using the Upper Quartile Uncertainty Shock by Piffer and Podstawski (2018)



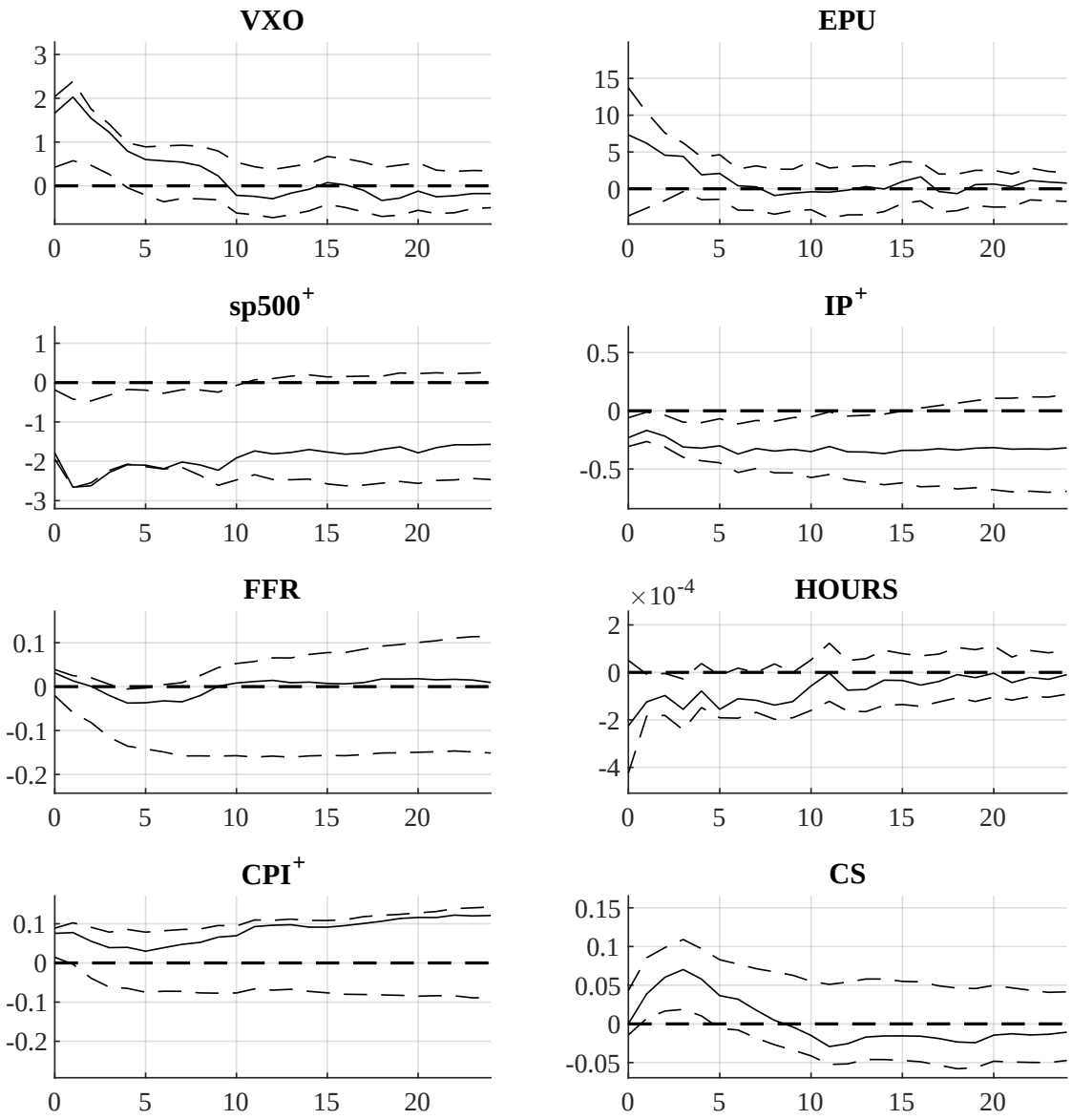
Note: The solid lines depict the point estimates of the structural impulse response functions using the 25% quantile of the uncertainty shock by Piffer and Podstawski (2018) instead of the median. The dashed lines depict the 90% residual-based moving block bootstrap confidence intervals by Jentsch and Lunsford (2022) based on 1,000 iterations. The labelling of the shock is based on the same logic as described in section 2.5.3 of the main text.

Figure 2.B.10: Impulse Responses to the Financial Uncertainty Shock using Employment instead of Hours Worked



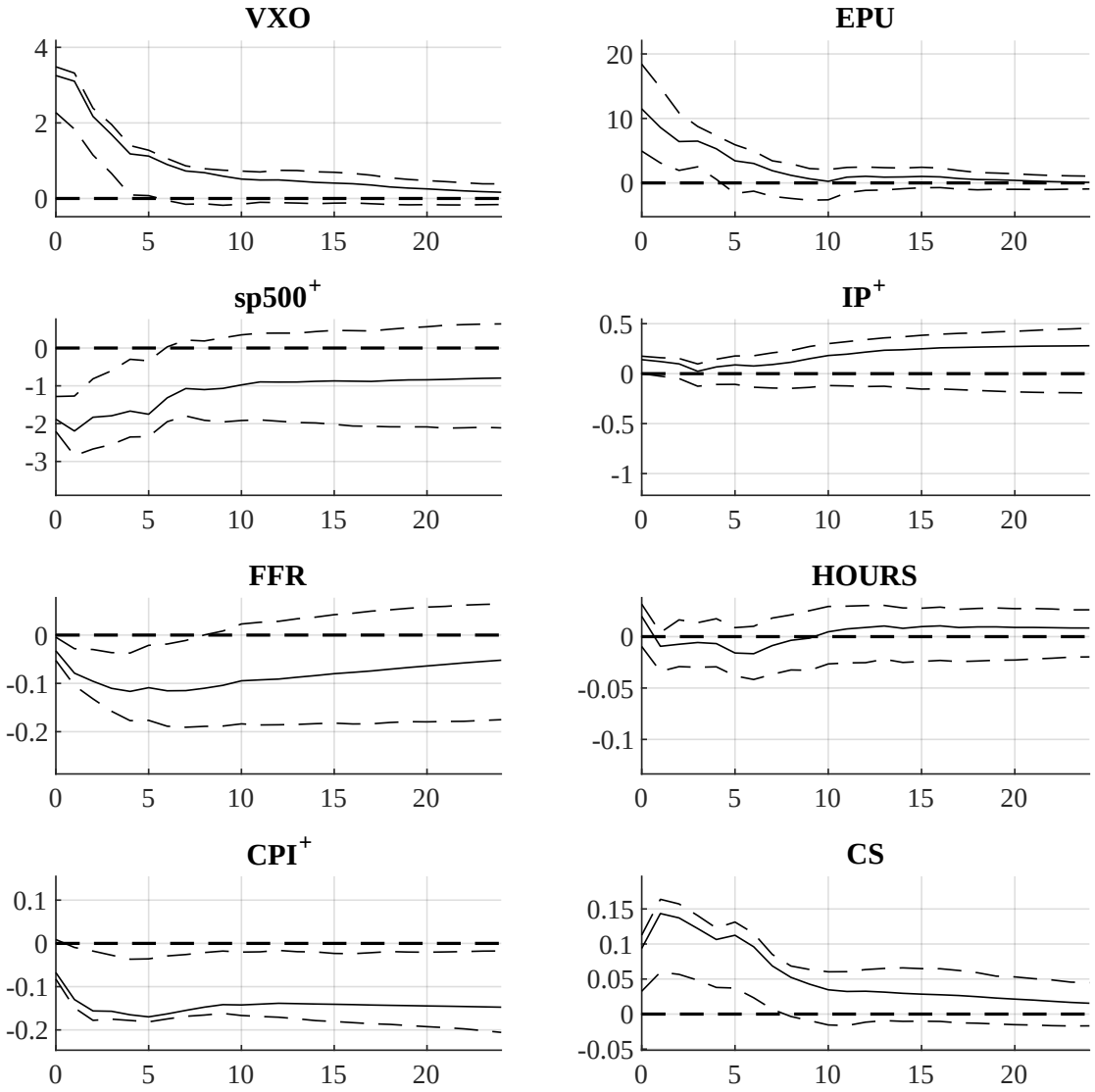
Note: The solid lines depict the point estimates of the structural impulse response functions when replacing HOURS with non-farm employment growth EMPL. The dashed lines depict the 90% residual-based moving block bootstrap confidence intervals by Jentsch and Lunsford (2022) based on 1,000 iterations. The labelling of the shock is based on the same logic as described in section 2.5.3 of the main text.

Figure 2.B.11: Impulse Responses to the Real Uncertainty Shock using Employment instead of Hours Worked



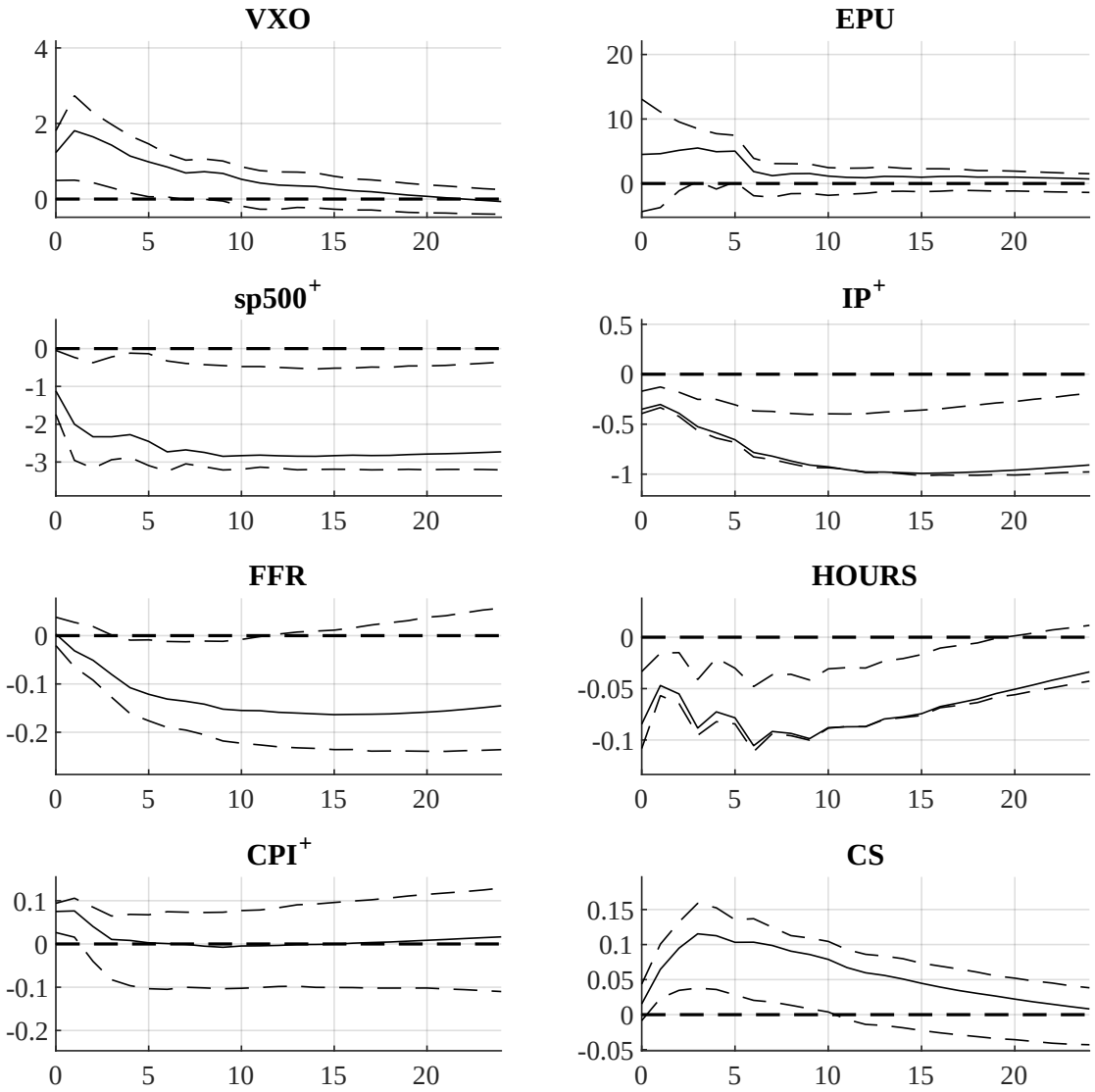
Note: The solid lines depict the point estimates of the structural impulse response functions when replacing HOURS with non-farm employment growth EMPL. The dashed lines depict the 90% residual-based moving block bootstrap confidence intervals by Jentsch and Lunsford (2022) based on 1,000 iterations. The labelling of the shock is based on the same logic as described in section 2.5.3 of the main text.

Figure 2.B.12: Impulse Responses to the Financial Uncertainty Shock with $p = 6$



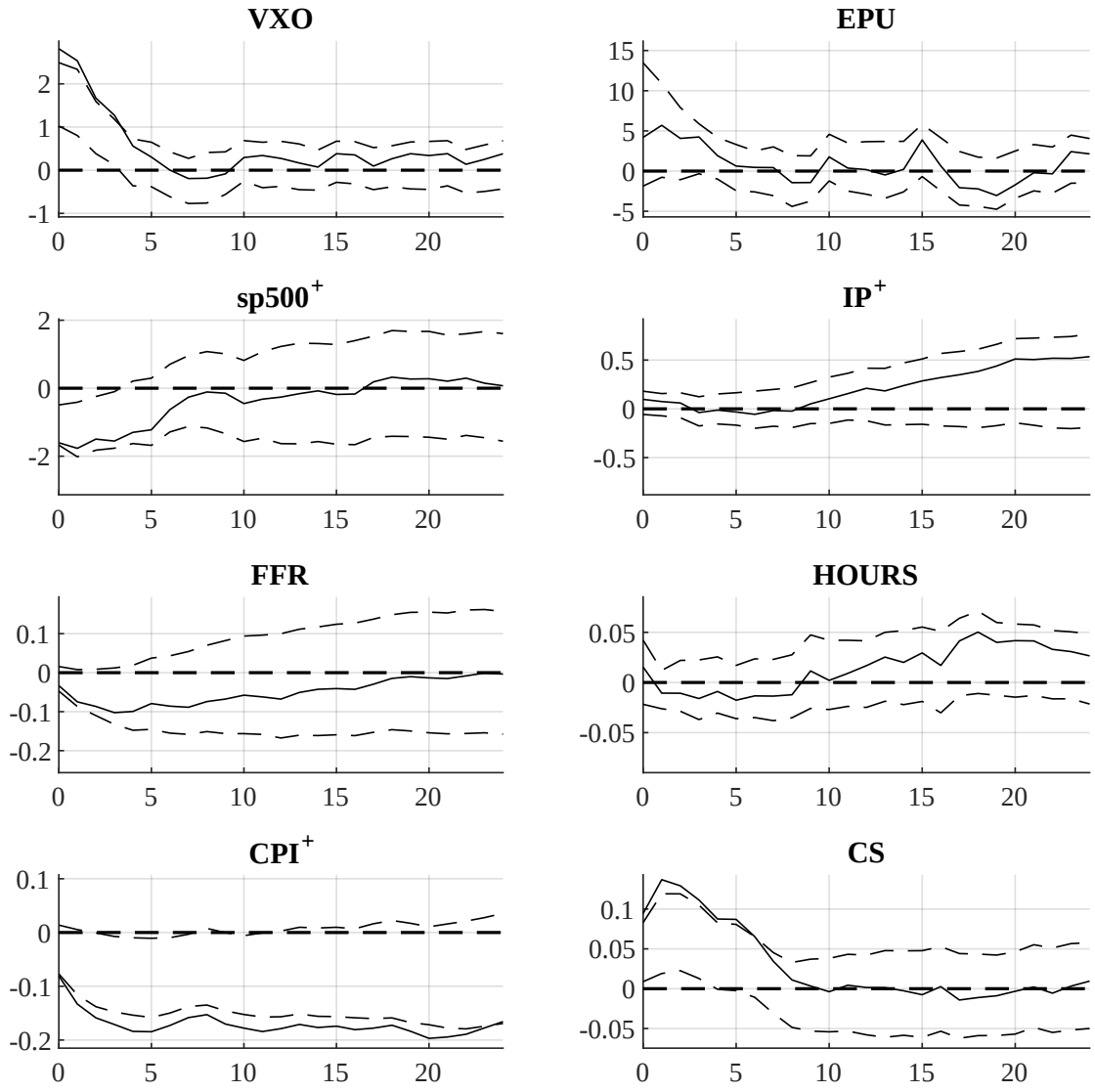
Note: The solid lines depict the point estimates of the structural impulse response functions for $p = 6$. The dashed lines depict the 90% residual-based moving block bootstrap confidence intervals by Jentsch and Lunsford (2022) based on 1,000 iterations. The labelling of the shock is based on the same logic as described in section 2.5.3 of the main text.

Figure 2.B.13: Impulse Responses to the Real Uncertainty Shock with $p = 6$



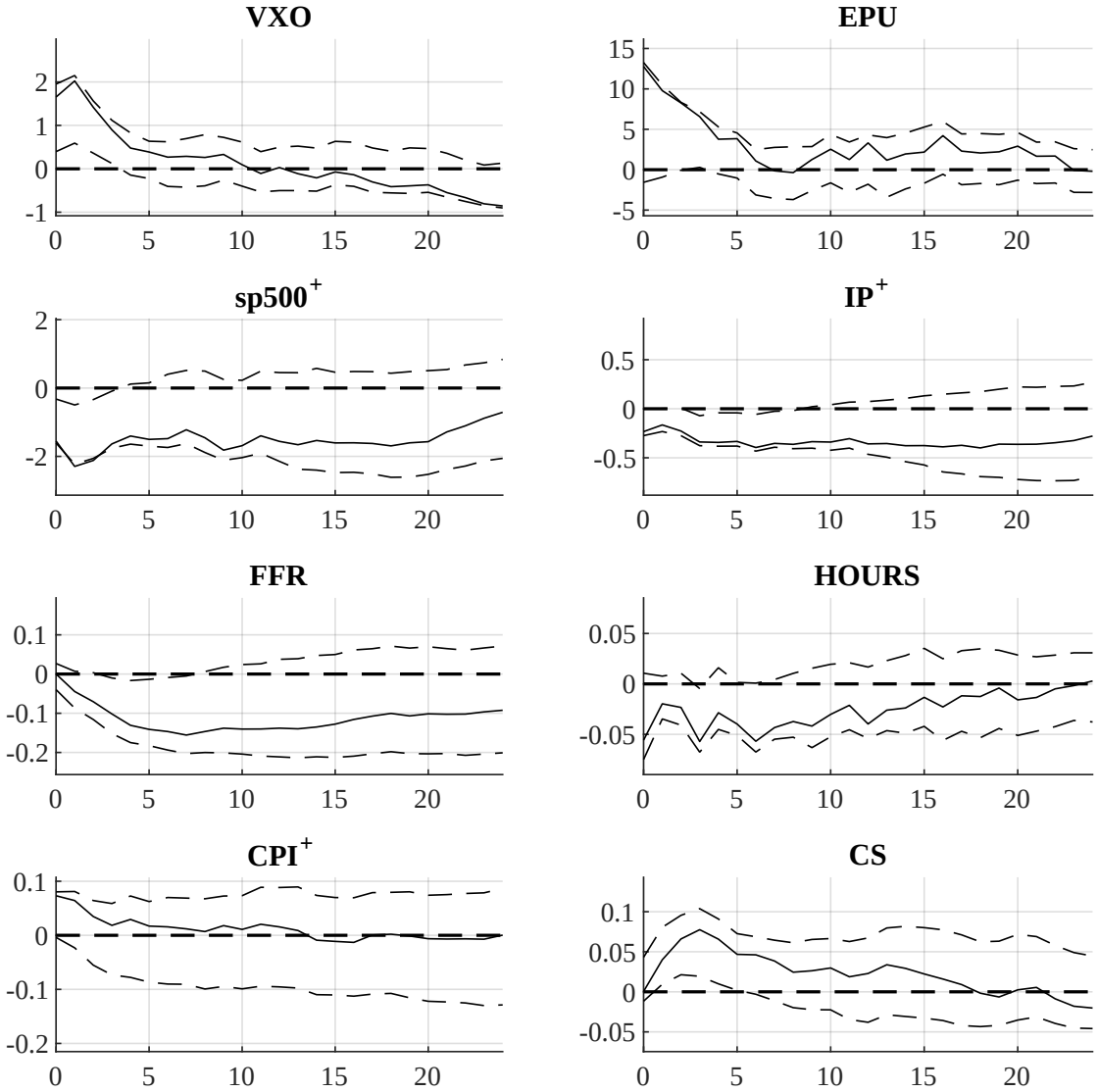
The solid lines depict the point estimates of the structural impulse response functions for $p = 6$. The dashed lines depict the 90% residual-based moving block bootstrap confidence intervals by Jentsch and Lunsford (2022) based on 1,000 iterations. The labelling of the shock is based on the same logic as described in section 2.5.3 of the main text.

Figure 2.B.14: Impulse Responses to the Financial Uncertainty Shock with $p = 18$



Note: The solid lines depict the point estimates of the structural impulse response functions for $p = 18$. The dashed lines depict the 90% residual-based moving block bootstrap confidence intervals by Jentsch and Lunsford (2022) based on 1,000 iterations. The labelling of the shock is based on the same logic as described in section 2.5.3 of the main text.

Figure 2.B.15: Impulse Responses to the Real Uncertainty Shock with $p = 18$



Note: The solid lines depict the point estimates of the structural impulse response functions for $p = 18$. The dashed lines depict the 90% residual-based moving block bootstrap confidence intervals by Jentsch and Lunsford (2022) based on 1,000 iterations. The labelling of the shock is based on the same logic as described in section 2.5.3 of the main text.

Chapter 3

Learning Structural Parameter Sets Accounting for Model and Estimation Uncertainty

3.1. Introduction

The defining feature of set-identified structural vector autoregressions (VARs) is that the solution is characterised, as the name suggests, by a set of structural parameters. Several structural models satisfy the imposed identification restrictions and this array of models is often called 'identified set'. The benefit of such set-identification schemes is that they allow for more agnostic identifying restrictions compared to point identification methods. However, the trade-off is between less strict assumptions and less sharp identification of the structural parameters.

Furthermore, set-identification introduces an additional source of uncertainty, namely the uncertainty about which structural models actually belong in the set. Hence, inference in set-identified structural VARs also needs to address this additional source of uncertainty, to which I will refer to as 'model uncertainty'. The other typical source of uncertainty arises from the estimation of the reduced form model parameters, which one also needs to account for in inference on the structural parameters. In the following this is referred to as 'estimation uncertainty'. As, for example, pointed out by Gafarov et al. (2018), inference is tricky from a frequentist perspective as there is no single point to conduct inference around. Instead, there are multiple admissible estimates within the identified set. Thus, typically Bayesian approaches are used in set-identified settings, see e.g. Rubio-Ramírez et al. (2010) and Arias et al. (2018) and Giacomini and Kitagawa (2021). However, these approaches can be computationally costly and do not yield correct frequentist coverage under general conditions.

This chapter presents a novel approach for full uncertainty quantification in set-identified structural VARs by learning the set of identified parameter vectors. In the structural VAR context, these parameter regions are e.g. the structural impulse response functions (IRFs) or the forecast error variance (FEV) decomposition. This learning approach is neither frequentist nor Bayesian; instead it builds on learning guarantees in order to make statements about the uncertainty inherent in the identified quantities. These learning guarantees inform the researcher how precise an object of interest is learned. In this chapter a parameter region is learned using a random sampling approximation. Montiel Olea and Nesbit (2021) rephrase this simple random sampling approximation as a supervised learning problem and derive the respective learning guarantees. The learning guarantees primarily depend on two hyperparameters, ϵ and δ, which are chosen by the researcher in order to induce a desired precision in the learning. Attached to those hyperparameters is the sample

complexity which is necessary to achieve the desired precision. In the case of the random sampling approximation, the sample complexity is the number of random sampling draws which is required to achieve the specified learning precision.

The precision is typically expressed by the probably approximately correct learning framework, abbreviated by PAC-learnability. To illustrate, let $[\lambda(S_U)]$ be the parameter region of interest and $[\hat{\lambda}_M]$ be the learned parameter region. Then the PAC-learnability statement is of the form $P(\mathcal{L}([\hat{\lambda}_M]; [\lambda(S_U)], W) < \epsilon) \geq 1 - \delta$, where $\mathcal{L}([\hat{\lambda}_M]; [\lambda(S_U)], W)$ is the misclassification probability and W is a probability distribution over the respective parameter space. A draw from the parameter region $[\lambda(S_U)]$ is misclassified if it is not in $[\hat{\lambda}_M]$ and the PAC-learnability statement controls this misclassification probability via ϵ. A probability with which this bound on the misclassification probability is true is specified with δ. Thus, depending on the choice of ϵ and δ the PAC-learnability framework states how precise the parameter region is learned. Montiel Olea and Nesbit (2021) provide the corresponding sample complexity to achieve this precision, i.e. the number of random sampling draws, and I apply their proof for the purpose of this chapter.

In the context of set-identified structural VARs, Montiel Olea and Nesbit (2021) learn the parameter region of interest for a fixed data realization, meaning that they neglect estimation uncertainty and only capture the model uncertainty with their baseline learning algorithm. In order to also account for estimation uncertainty I augment the learning procedure. In brief, the learning algorithm is preceded by a bootstrap for the reduced form parameters and, in the end, the union of the identified sets for each bootstrap replication is learned. I provide a clear asymptotic interpretation of the augmented learning approach based on the learning guarantees, which is again neither frequentist nor Bayesian, yet it relies on the frequentist school of thought. That is, there exists one true population parameter that one aims to identify. The asymptotic interpretation is based on the PAC-learnability framework, and thus reads similarly: *The probability that the structural parameters of interest are jointly inside the learned set is larger than* $1-\epsilon$*. This statement is true with probability greater than* $1 - \delta$. The statements derived from the PAC-learnability framework are of joint nature, meaning that the interpretation holds for the whole dimension of the structural parameter vector, e.g. for all horizons of a specific structural IRF. This leads to analogous statements about uncertainty compared to joint frequentist or joint Bayesian inference.

Furthermore, the augmented learning framework offers full uncertainty quantification while remaining computationally cheap. At the same time it is tractable as

it only requires the use of tools that a practitioner, working with structural VARs, should be familiar with. Thus, practitioners do not need to rely on a 'computational blackbox', but they can code the simple algorithm themselves and know exactly how the results are computed. In principle, the learning approach can be adapted to any available identification scheme or combinations of them, which makes the learning approach particularly flexible and once again adds to its simplicity.

In the literature, Bayesian approaches dominate the set-identified structural VAR inference. Initiated by Uhlig (2005), the 'single prior' Bayesian approaches gained popularity. Notable contributions in this strand of the literature are e.g. Rubio-Ramírez et al. (2010), Arias et al. (2018) and Inoue and Kilian (2022). The labelling as single prior approaches stems from the feature that only a single prior is specified over the space of models that potentially enter the identified set. Yet, Baumeister and Hamilton (2015) showed that this can result in the unintentional influence of an ex ante uninformative prior on the posterior distribution of the structural parameters, even asymptotically. In response, they propose to incorporate informative prior knowledge into the Bayesian framework in order to circumvent this problem. Yet, such knowledge might not be available in every empirical application, and thus general 'robust' Bayesian approaches which were developed by Giacomini and Kitagawa (2021), Giacomini et al. (2022) and Giacomini et al. (2021) are often employed to conduct inference in set-identified structural VARs.

Their main feature is that they do not specify a single prior over the space of potentially admissible models. Instead, they consider all possible priors. In doing so they rely heavily on numerical optimization of - often non-convex - problems in order to compute the boundaries of the identified sets. These optimization problems are subject to several constraints which depend on the imposed identifying restrictions and they have to be solved for every structural parameter individually. This implies that one has to solve an optimization problem for e.g. each specific IRF at each horizon. Furthermore, if one identifies multiple shocks, one has to conduct this procedure for each identified shock. Hence, these robust Bayesian approaches are computationally costly and prone to convergence issues in the numerical optimizations. However, the benefit is that one avoids the undesirable influence of uninformative priors on the posterior distribution and can construct robust credible regions. It should be noted, that these credible regions do not attain frequentist coverage under general assumptions, and thus excess coverage from a frequentist perspective is likely. Admittedly, the learning procedure introduced in this chapter also does not attain frequentist coverage, yet it does so without major computational costs.

It is possible to approximate the boundaries of the identified set in the robust Bayesian framework but the approximation is essentially the very same random sampling approximation which constitutes the supervised learning problem. Hence, to be precise one would need to factor in the statements from the learning guarantees into the interpretation of the robust Bayesian credible regions. I argue that it is simpler and more concise to then rely on the learning framework alone. An additional advantage of the learning approach is that it allows for joint uncertainty quantification across horizons, whereas inference deduced from the described robust Bayesian approaches is point-wise.

In terms of frequentist inference in set-identified structural VARs three approaches stand out. The projection based inference proposed by Gafarov et al. (2016) offers joint inference for structural parameters. Conceptually, the projection idea is comparable with the learning procedure. They construct a Wald ellipsoid for the reduced form parameters from which they derive a projection region for the structural parameters. In the augmented learning procedure, a set of bootstrapped reduced form parameters is used in order to learn the corresponding parameter region of interest. The primary challenge in projection based approaches is to correctly calibrate the projection region as it typically has excess coverage in a frequentist sense, yet, these calibrations are non-trivial. Based on the work of Kaido et al. (2019), the projection approach is calibrated such that it attains robust Bayesian credibility which does not yield frequentist coverage under general assumptions. Although the basic projection approach is quite general and computationally cheap, the calibration is tedious and, as Gafarov et al. (2016) state, comparable to the computational costs of the robust Bayesian approach. Consequently, the advantage of the learning procedure to the projection approach is similar to that of the robust Bayesian framework.

Other frequentist approaches include the delta-method inference proposed by Gafarov et al. (2018) which yields adequate point-wise frequentist coverage under certain regularity conditions. However, this approach is limited to the case of only a single shock being subject to inequality restrictions. In practice, it might be necessary to identify multiple shocks in order to distinguish similar shocks from each other which could pose as a difficulty with the less sharp identification from inequality restrictions. The minimum distance framework proposed by Granziera et al. (2018) also allows for the identification of multiple shocks, yet is again associated with computational costs. In return, one obtains joint confidence bands which are based on a Bonferroni correction. However, as is typical for Bonferroni frameworks the resulting confidence bands are conservative.

In contrast to existing methodologies for inference in set-identified VARs, the learning procedure proposed in this chapter offers joint uncertainty quantification without high computational costs. The learning procedure merely consists of a simple random sampling approximation combined with a bootstrap. Based on the learning guarantees, I provide an asymptotic interpretation for the learned bands which can be used to judge the uncertainty involved in the identification of the structural parameters. A simulation study sheds light on the finite sample properties of this augmented learning procedure. As the underlying interpretation is neither frequentist nor Bayesian, a specific choice of the hyperparameters ϵ and δ is not easy to link to typical nominal levels. Yet, the simulation results provide guidance on how to choose the hyperparameters in order to achieve a desired coverage in different settings.

Section 3.2 introduces how the learning approach is used to learn the union of parameter regions and adapts the key theoretical results presented by Montiel Olea and Nesbit (2021) to this case. Furthermore, section 3.2 shows how the described learning framework can be used for uncertainty quantification in set-identified VARs and establishes how to interpret the resulting learned bands for the structural parameters of interest.

Section 3.3 presents a simulation study which explores the finite sample properties of the augmented learning approach. In brief, the results suggest acceptable point-wise coverages rates which tend to show overcoverage in frequentist terms. This means that coverage rates are often above the nominal level of 95%, especially for larger sample sizes and tight tuning parameters. Yet, as a method to conduct joint uncertainty quantification, such point-wise overcoverages are likely to be observed. Upon examination of joint coverages, it becomes evident that with the correct setting of the hyperparameters of the learning approach one can achieve joint coverages that are close to the nominal coverage levels that are used in practice. In more detail, one can achieve roughly 90% for structural impulse responses (IRFs) and roughly 95% for forecast error variance (FEV) decompositions.

Section 3.4 presents an empirical comparison of the augmented learning procedure with existing methods of conducting inference in set-identified structural VARs. The empirical illustration is based on the same stylized and simple empirical illustration as in Montiel Olea and Nesbit (2021). They identify a monetary policy shock with a set of sign restrictions on structural impulse responses. The learned bands are wider than the frequentist delta-method intervals proposed by Gafarov et al. (2018) and the robust Bayesian credible regions proposed by Giacomini and Kitagawa (2021). Yet, these two methods yield point-wise inference, whereas the learned bands with the joint

interpretation are expected to be wider. In comparison with the joint Bonferroni confidence bands by Granziera et al. (2018) the learned bands are comparable in width. Finally, section 3.4 also replicates the empirical results presented by Piffer and Podstawski (2018) and quantifies the uncertainty in the estimation of the structural IRFs with the augmented learning procedure. The results reveal that the structural effects are still uncovered with the learned bands. Hence, the learned quantities are also informative in a proper empirical analysis.

3.2. Econometric Framework

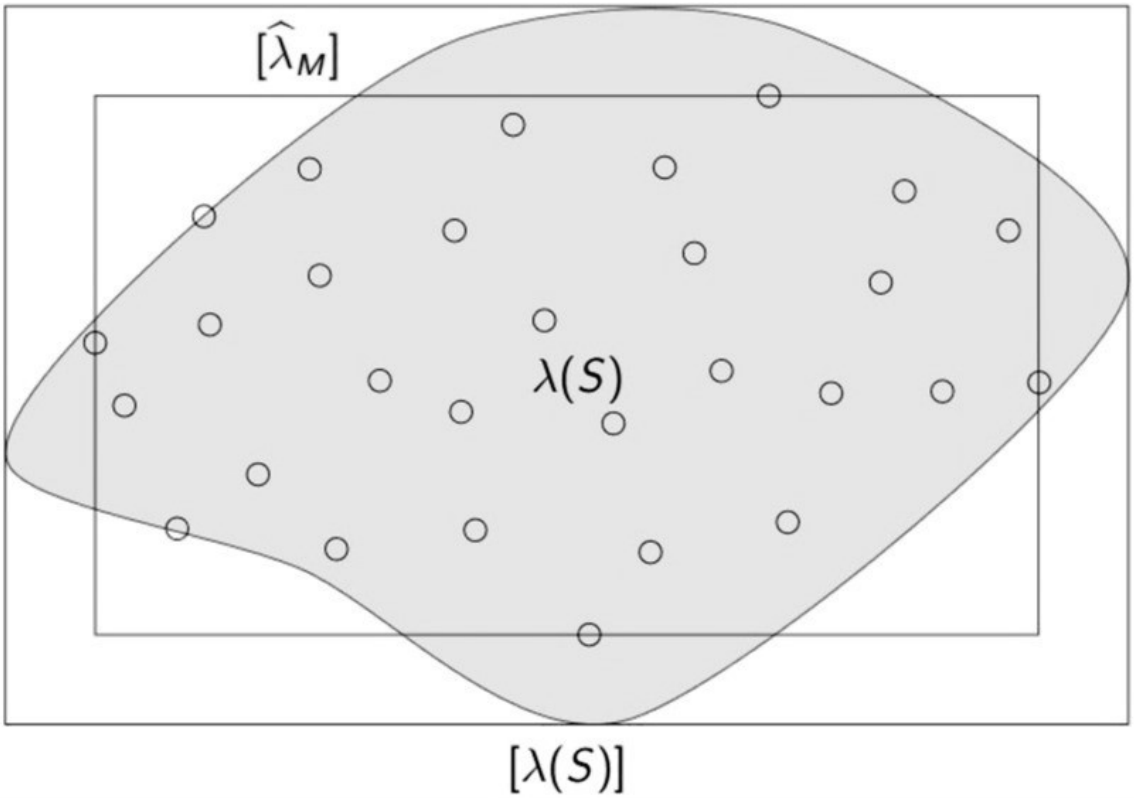
3.2.1. Learning the Union of Parameter Regions

This section introduces the learning framework which is based on the work of Montiel Olea and Nesbit (2021). They rephrase a random sampling approximation as a supervised learning problem and for comparability, I stick to the notation of their work. As described in the introduction, this learning framework relies on the standard PAC-learnability framework in order to provide the number of random sampling draws required to achieve a respective learning precision. These learning guarantees then enable theoretically founded statements about the precision with which the parameter region is learned.

Let θ denote the parameters of interest and let $\Theta \subseteq \mathbb{R}^b$ be the parameter space of $\theta \in \mathbb{R}^b$. $S \subseteq \mathbb{R}^b$ is the parameter region of interest one wants to learn. As this is a supervised learning algorithm it is known whether a draw of θ is in S, meaning that every draw θ has a label attached, which indicates whether $\theta \in S$. These labels are denoted by $l(\theta) = \mathbf{1}\{\theta \in S\}$. In practice, the quantities of interest are potentially not the parameters θ themselves but measurable functions of them, denoted by $\lambda(\theta)$ with $\lambda : \Theta \to \mathbb{R}^d$. To give an outlook, these $\lambda(\theta)$ will correspond to the structural IRFs or the FEV decompositions in the structural VARs. Hence, the parameter region that is ultimately learned is $\lambda(S)$ and the respective labels for the supervised learning algorithm are $l(\theta) = \mathbf{1}\{\lambda(\theta) \in \lambda(S)\}$.

As also described by Montiel Olea and Nesbit (2021), a key result from the statistical learning theory states that a set $\lambda(S)$ of a concept class Λ is learnable if it has finite Vapnik–Chervonenkis dimensions, d_{VC} (see Blumer et al., 1989). The d_{VC} are a measure of complexity of the underlying set and the result states that only finitely complex sets are learnable. Yet, the parameter region of interest does most likely not belong to such a simple concept class as even convex sets have infinite d_{VC}. Hence, Montiel Olea and Nesbit (2021) focus on learning the tightest band $[\lambda(S)]$

Figure 3.1: Learning the Tightest Band



Source: An earlier version of Montiel Olea and Nesbit (2021).

that contains $\lambda(S)$ (see Figure 3.1 for an illustration with $d = 2$). This tightest band has $d_{VC} = 2d$ and is therefore learnable. The tightest band is defined as:

$$[\lambda(S)] \equiv \bigtimes_{j=1}^{d} \left[\inf_{\theta \in S} \lambda_j(\theta), \sup_{\theta \in S} \lambda_j(\theta) \right]$$

where $\lambda_j(\theta)$ is the j-th element of $\lambda(\theta) \in \mathbb{R}^d$. Montiel Olea and Nesbit (2021) show that given a sample of draws $\boldsymbol{\theta}_M \equiv (\theta_1, ..., \theta_M)$, the algorithm denoted by $[\hat{\lambda}_M](\boldsymbol{\theta}_M)$ learns the parameter region of interest. The algorithm just picks the maximum and minimum parameter at each horizon over a specified number of samples from Θ:

$$[\hat{\lambda}_M](\boldsymbol{\theta}_M) := \bigtimes_{j=1}^{d} \left[\min_{m|l(\theta)=1} \lambda_j(\theta_m), \max_{m|l(\theta)=1} \lambda_j(\theta_m) \right],$$

where $\lambda_j(\theta_m)$ is the j-th element of $\lambda(\theta_m) \in \mathbb{R}^d$.

Yet, in order to achieve PAC-learnability the probability distribution over the parameter space of θ is crucial despite finite d_{VC} of the tightest band $[\lambda(S)]$. Consider,

for example, a probability distribution that puts arbitrarily high mass on the set $[\lambda(S)] \setminus \lambda(S)$, the white area between the set $\lambda(S)$ and its tightest band $[\lambda(S)]$ in Figure 3.1. In this case, the algorithm would output the empty set, although all draws are inside the tightest band that one wants to learn. To circumvent this, Montiel Olea and Nesbit (2021) show that all draws have to be from inside the parameter region $\lambda(S)$, to which they refer as 'learning from the inside'. Hence, the algorithm requires a specific number of draws from inside the parameter region. In Figure 3.1 these draws are depicted by the circles inside the shaded area.

Now suppose that not only a single set of interest S exists but N different sets S_n with $n = 1, ..., N$. This corresponds to N parameter regions $\lambda(S_n)$ when again considering a measurable function of the underlying parameters θ. In the structural VARs, these different parameter regions will represent the regions for N bootstrap iterations. The details follow in the next subsection. I denote the union of the individual sets $\lambda(S_n)$ with

$$\lambda(S_U) = \bigcup_{n=1}^{N} \lambda(S_n).$$

As one wants to learn this union, the respective draws need labels for the supervised learning algorithm. The labels for the union are defined by:

$$l_U(\theta) = \mathbf{1}\{\exists n \text{ such that } \lambda(\theta) \in \lambda(S_n) \text{ for } n = 1, ..., N\}.$$

According to the labels, the union is learned by the same learning algorithm, namely:

$$[\hat{\lambda}_M](\boldsymbol{\theta}_M) := \bigtimes_{j=1}^{d} \left[\min_{m|l_U(\theta)=1} \lambda_j(\theta_m), \max_{m|l_U(\theta)=1} \lambda_j(\theta_m) \right]. \tag{3.2.1}$$

Hence, learning the union is ultimately just a change in the definition of the parameter region that is subject of the learning procedure. Therefore, again only the tightest band containing the union $[\lambda(S_U)]$ can be learned from the inside.

In detail, the set $[\lambda(S_U)]$ is PAC-learnable. Let W be any probability distribution on Θ with $P(S_U) = 1$. To quantify the precision of the learning, the misclassification probability $\mathcal{L}([\hat{\lambda}_M]; [\lambda(S_U)], W)$ is evaluated, which represents the probability for a draw from inside of $[\lambda(S_U)]$ that is at the same time not in $[\hat{\lambda}_M]$. Such a draw would be misclassified based on the learned parameter region. Theorem 3.1 below is a loose replica of Theorem 3 in the paper by Montiel Olea and Nesbit (2021).

Theorem 3.1. *The algorithm in* (3.2.1) *learns* $[\lambda(S_U)]$ *irrespective of the shape of* $\lambda(S_U)$*. For any* $0 < \epsilon, \delta < 1$ *it holds that:*

$$P(\mathcal{L}([\hat{\lambda}_M]; [\lambda(S_U)], W) < \epsilon) \geq 1 - \delta, \tag{3.2.2}$$

for all probability distributions $W \in \mathcal{W}$ *with*

$$\mathcal{W}(S) \equiv \{W | W \text{ is a distribution on } \Theta \text{ and } P(S_U) = 1\}.$$

given that the number of draws $M \geq m^*(\epsilon, \delta)$ *admits*

$$m^*(\epsilon, \delta) \leq \min\{2d \ln(2d)/\delta, \exp(1)(2d + \ln(1/\delta))\}/\epsilon. \tag{3.2.3}$$

The function $m^*(\epsilon, \delta)$ represents the sample complexity of the learning algorithm, which is the minimum number of draws for which learning is achieved. This function is bounded from above by the two values depicted in (3.2.3), and thus if M exceeds this upper bound, learning is guaranteed with the precision governed by (3.2.2). The sample complexity depends, firstly, on the dimensionality d of the problem, with a higher dimensionality requiring more draws to achieve the learning guarantees. Secondly, the number of draws depends on ϵ and δ. Here ϵ controls the misclassification probability (the 'approximate' part of the PAC-learnability), while δ governs how likely it is that the bound on the misclassification probability holds (the 'probably' part of the PAC-learnability). Naturally, a tighter bound on the misclassification probability, a lower ϵ, increases the number of necessary draws. More confidence that the bound holds, i.e. a lower δ, does so as well. In practice, these parameters have to be set by the researcher, which is discussed in the next subsections based on the example of structural VARs.

Note that Montiel Olea and Nesbit (2021) also provide lower bounds for $m^*(\epsilon, \delta)$. However, these are omitted here for the sake of brevity, as the upper bound is always used as a guideline to guarantee learning, and the necessary conditions imposed by the lower bounds are automatically satisfied. The proof of the theorem by Montiel Olea and Nesbit (2021) is broad in the sense that the tightest band of any parameter region of interest is learnable, including $[\lambda(S_U)]$. Nevertheless, I provide a brief confirmation that the proof is adaptable to the learning of the union of identified sets in Appendix 3.A. In practice, the requirement that the probability distribution is such that $P(S_U) = 1$ is not a constraint. One simply only counts draws of θ with labels $l_U(\theta) = 1$. Finally, the application to full uncertainty quantification in struc-

tural VARs and the respective interpretation of the PAC-learnability statement are discussed in sections 3.2.2 and 3.2.3, respectively.

3.2.2. Set-identified Structural VARs

The learning framework is, for example, of use for uncertainty quantification in structural VARs. The baseline stationary structural VAR(p) model is depicted by equation (3.2.4), with y_t being a k-dimensional vector containing the variables of the system that are regressed on their lagged values up to order p,

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + B w_t. \tag{3.2.4}$$

The elements of the $k \times 1$ vector w_t denote the economically meaningful structural shocks with covariance matrix I_k and the elements of the $k \times 1$ white noise vector u_t are the reduced form innovations with $u_t = B w_t$. Hence, the impact matrix B maps between the reduced form innovations u_t and the structural shocks w_t.

The moving average representation of model (3.2.4) is denoted by:

$$y_t = \sum_{m=0}^{\infty} \Phi_m B w_{t-m},$$

where Φ_m are the moving average coefficients that give the responses of the system to the reduced form innovations m periods ago. The structural impulse response of variable i to shock j at horizon h is given by:

$$\eta_{ijh} = e_i' \Phi_h B e_j,$$

with e_i and e_j being the ith and jth column of the identity matrix I_k. Furthermore, the FEV decomposition yields structural parameters of interest as it tells the researcher how much a structural shock of interest contributes to the variation of a specific variable of the system. Let

$$y_{t+h|t} = \sum_{m=0}^{\infty} \Phi_{h+m} u_{t-m},$$

be the h-step-ahead forecast of y_t. Then the h-step ahead forecast error is denoted by:

$$y_{t+h} - y_{t+h|t} = \sum_{m=0}^{h-1} \Phi_m B w_{t+h-m},$$

while the respective error covariance matrix is given by:

$$\Omega(h) = \sum_{m=0}^{h-1} \Phi_m B B' \Phi_m' = \sum_{m=0}^{h-1} \Phi_m \Sigma \Phi_m'.$$

The contribution of shock j to the total forecast error variance of variable i at horizon h is then:

$$\Omega_{i,j}(h) = \frac{e_i'(\sum_{m=0}^{h-1} \Phi_m B e_j e_j' B' \Phi_m') e_i}{e_i'(\sum_{m=0}^{h-1} \Phi_m \Sigma \Phi_m') e_i}.$$

The key element of interest in the structural VAR is the impact matrix B which enables the computation of different structural parameters of interest, such as the IRFs or the FEV decomposition. Without loss of generality the normalization of $\mathbb{E}(w_t w_t') \equiv \Sigma_w = I_k$ yields $\mathbb{E}(u_t u_t') \equiv \Sigma = BB'$. Due to the symmetry of Σ this relationship represents a system of equations with more parameters than equations, and thus B is not identified with the simple unit variance normalization of the structural shocks. The system of equations $\Sigma = BB'$ can also be written as $\Sigma = BQQ'B'$ where Q is a $k \times k$ orthonormal matrix, also called rotation matrix, with the property $QQ' = I_k$.

To identify the impact matrix several forms of restrictions have been proposed in the literature, for example, restricting at least $k(k-1)/2$ elements of B to be equal to zero or other specific values. Such restrictions are called short-run or exclusion restrictions as they pin down the contemporaneous impulse responses. A straightforward solution for B is the Cholesky decomposition of Σ, which imposes a recursive structure on the relationship between the shocks and the variables of the system. A popular example for the use of such a recursive structure is the work by Christiano et al. (1999). However, such short-run restrictions impose strong assumptions on contemporaneous effects of the structural shocks which are difficult to defend in many applications, and thus further identification strategies have been proposed in the literature. In particular, set-identification schemes that impose less strict inequality restrictions on structural parameters are a popular approach.

In set-identification schemes, instead of a single solution for the impact matrix B, the solution is characterized by a set of impact matrices that satisfy certain conditions specified by the identification strategy. These conditions revolve around inequality restrictions on structural parameters in contrast to the aforementioned strict equality restrictions. A prominent example are sign restrictions on structural impulse responses as proposed by Uhlig (2005). A potentially non-exhaustive list of further set-identifying schemes inludes imposing inequality restrictions on the shock series themselves, as introduced by Ludvigson et al. (2021), or on the FEV decomposition, as proposed by Volpicella (2022). Amir-Ahmadi and Drautzburg (2021) describe ranking restrictions which impose that e.g. the impulse response of one variable has to exceed another. In the proxy VAR with multiple proxies, inequality or ranking restrictions can also be placed on the relationship between the proxies and the structural shocks as proposed by Piffer and Podstawski (2018). Furthermore, all the previously described set-identification strategies can also be used to disentangle the shocks of interest in the proxy VAR. Finally, one advantage of the different set-identification schemes is that they can easily be used in conjunction with each other, for example, it is straightforward to impose sign restrictions on impulse responses in conjunction with ranking restrictions.

As the set-identification schemes merely impose inequality restrictions several structural models, i.e. several impact matrices B, can satisfy these conditions. Typically, one possible solution to $\Sigma = BB'$ is computed and for convenience I choose the Cholesky decomposition of Σ which is denoted by Σ_c. This solution for the impact matrix is then rotated with orthonormal matrices to generate candidate solutions for the impact matrix $\Sigma_c Q$. These candidate solutions can then be assessed to determine whether the restrictions imposed by the identification strategy hold. Set-identification of structural VARs then amounts to finding the set of rotation matrices Q for which the specified restrictions are satisfied.

In the set-identified structural VARs, these rotation matrices are the ones that are randomly sampled in order to generate candidate solutions for the impact matrix. Hence, the Q matrices represent the $\theta \in \mathbb{R}^b$ parameters of the learning procedure with $b = k^2$. Ultimately, however, the structural parameters like IRFs or the FEV decomposition are of interest for the researcher, which are a function of the sampled parameters, and thus correspond to $\lambda(\theta) \in \mathbb{R}^d$. Suppose, one wants to learn the structural IRFs of one variable to a specific shock up to horizon H, then H is the dimensionality d of the learned parameter region. In this chapter the dimensionality d will always correspond to the horizon H up to which an IRF or the FEV decomposition

of a specific variable is learned. This allows for joint statements about the precision of the learning over all horizons up to H, which is how joint confidence bands are constructed most of the time. In principle, the learning approach can also be used to learn e.g. impulse responses of multiple variables jointly. This would simply increase the dimensionality of the problem, and hence the sample complexity.

As described in section 3.2.1 a key aspect of the supervised learning approach are the labels. In the structural VAR, the labels are determined by the identification strategy. A draw Q or θ is assigned the label $l(\theta) = 1$ if the imposed identifying restrictions are satisfied. Hence, in practice it is rather simple to determine the needed labels by assessing whether the restrictions are met. This is a common approach, for instance, in Bayesian algorithms. This means that the identified set, which characterizes the set of admissible Q matrices corresponds to S, the parameter regions of interest which are learned. The set of structural parameters of interest based on the set of Q matrices is then $\lambda(S)$. Learning this set allows for statements about the precision with which this set is learned via the PAC-learnability framework. Hence, one can gauge how precisely one learns the structural models that belong to $\lambda(S)$, which was previously referred to as model uncertainty.

Yet, the structural parameters of interest, i.e. the impulse responses and the FEV decomposition, are not functions solely of the rotation parameters θ, but also of the reduced form parameters of the underlying VAR model. Hence, the parameter region of interest is $\lambda(\theta, \Phi^{(H)}, \Sigma)$ where $\Phi^{(H)}$ collects the moving average coefficients over all horizons H. In the paper by Montiel Olea and Nesbit (2021), they learn the set $\lambda(S)$ while keeping the reduced form parameters fixed, e.g. at their maximum likelihood estimates. Thus, for their purpose they neglect the underlying estimation uncertainty inherent in these reduced form parameters but use the learning algorithm solely to make statements about model uncertainty. To account for estimation uncertainty, I augment the learning process with a preceding bootstrap for the reduced form for parameters.

Let $n = 1, ..., N$ bootstrap draws give N different pairs of reduced form coefficients $\Phi_n^{(H)}$ and Σ_n. For each pair of reduced form coefficients, a set of θ satisfies the imposed identification restrictions. These sets define, for a given bootstrap draw, the parameter regions of interest S_n and ultimately $\lambda(S_n, \Phi_n^{(H)}, \Sigma_n)$. The different parameter regions are likely not identical throughout the N bootstrap draws and the

goal is to learn the union $\lambda(S_U)$ of all the individual sets $\lambda(S_n, \Phi_n^{(H)}, \Sigma_n)$. This is achieved by obtaining the respective label for a draw of θ such that

$$l_U(\theta) = \mathbf{1}\{\exists(\Phi_n^{(H)}, \Sigma_n) \text{ such that } \theta \in S_n \text{ for } n = 1, ..., N\}.$$

Hence, if the imposed identification restrictions are satisfied for one of the bootstrapped reduced form parameters, the current draw is in the set $\lambda(S_U)$. Ultimately, the tightest band $[\lambda(S_U)]$ around the union is learned. From now on, $[\lambda(S_U)]$ denotes the union based on the preceding bootstrap. In the end, the learning works as in the baseline algorithm, namely by generating random draws, checking whether a draw is inside the union of the parameter regions and selecting the maximum and minimum values in each dimension throughout the admissible draws. Compactly, the learning algorithm is represented by:

$$[\hat{\lambda}_M](\boldsymbol{\theta}_M) := \bigtimes_{j=1}^{d} \left[\min_{m|l_U(\theta)=1} \lambda_j(\theta_m, \Phi_n^{(H)}, \Sigma_n), \max_{m|l_U(\theta)=1} \lambda_j(\theta_m, \Phi_n^{(H)}, \Sigma_n) \right]. \tag{3.2.5}$$

To employ the above described algorithm, a practitioner can stick to the following steps:

Algorithm 3.1.

1. *Generate N bootstrap samples and estimate the reduced form coefficients $\Phi_n^{(H)}$ and Σ_n. Store them.*
2. *Determine the emptiness of the set $\lambda(S_U)$.*
3. *Choose ϵ, δ and the dimensionality d and determine the number of draws M needed to guarantee PAC-learnability.*
4. *Draw rotation parameters θ until M draws are found for which $\theta \in S_n$ for at least one $n = 1, ..., N$.*
5. *For any valid combination of m and n compute the parameters of interest $\lambda(\theta_m, \Phi_n^{(H)}, \Sigma_n)$ and store them.*
6. *To describe the tightest band $[\hat{\lambda}_M]$ select $\min \lambda(\theta_m, \Phi_n^{(H)}, \Sigma_n)$ and $\max \lambda(\theta_m, \Phi_n^{(H)}, \Sigma_n)$ at each horizon h, as laid out in* (3.2.5).

Regarding the first point of the complete learning procedure, the choice of the bootstrap certainly affects the results. In principle, any appropriate time series bootstrap can be employed, yet, large and finite sample results may differ. I propose

to use the residual-based moving block bootstrap proposed by Brüggemann et al. (2016) as it is robust against heteroskedasticity of unknown form. As the convenient wild bootstrap is frequently used in practice, I also present simulation results for this bootstrap scheme, although, it does not replicate the distribution of Σ consistently, and thus overstates the uncertainty inherent in the estimation of the structural parameters. The implications of the choice of the bootstrap are further discussed in the next subsection. The simulation study shows the difference between the two just described bootstrap approaches in smaller samples. If the identification strategy involves proxy variables, for example, the adapted residual-based moving block bootstrap by Jentsch and Lunsford (2022) can be used to generate bootstrap estimates of the proxy VAR quantities which are used to compute the structural parameters of interest. For restrictions on structural shocks itself, such as e.g. in Ludvigson et al. (2021), yet another bootstrap approach is needed. With such restrictions one can follow the basic idea of Ludvigson et al. (2021), in which the dates for which the shocks are restricted are excluded from the bootstrap. Suppose at $t = \bar{t}$ the shock is restricted, then the residuals in the bootstrap are sampled such that the residual at time $\bar{t}$ remains in place. The subsequent extensive sampling of rotation matrices in order to construct quantiles of the resulting unknown distribution is replaced by the learning procedure and interpreted as such.

The second point of the algorithm checks whether there exist any draws of θ that fulfill the identification assumptions at all. In the literature, different procedures for assessing emptiness have been proposed in different contexts, see e.g. Amir-Ahmadi and Drautzburg (2021), Giacomini and Kitagawa (2021) or Volpicella (2022). Yet, one would need to assess emptiness for every bootstrap sample n separately, and thus I recommend to approximate emptiness in the usual way. Namely, by drawing a fixed number of rotation parameters θ and if none of them satisfy the identification restrictions for any bootstrap draw the set is deemed empty.

The choice of appropriate values for ϵ and δ is discussed in greater detail in the subsequent sections. With regard to the final three points of Algorithm 3.1, firstly, the rotation parameters θ are drawn in the usual way, e.g. employing the QR decomposition. Then, for each draw θ_m that satisfies the identifying restrictions, one has to compute and store the structural parameters of interest, yet, only for the bootstrap parameters $\Phi_n^{(H)}$ and Σ_n for which the restrictions are satisfied. In the last step, the maximum and minimum over all stored structural parameters in each dimension is selected. The dimensions correspond to the individual horizons h.

The resulting tightest band $[\hat{\lambda}_M]$ then allows for joint statements about the uncertainty left in the identification of the structural parameters, taking into account both model and estimation uncertainty. The interpretation of this tightest band is described in subsection 3.2.3.

3.2.3. Interpretation of the Learning Procedure

To draw conclusions about the remaining uncertainty in the learning procedure, I follow the frequentist school of thought. Suppose that given the data generating process (DGP), there are unique 'true' population parameters that need to be identified. This frequentist way of thinking allows, under certain assumptions, to make statements about how likely it is that the true population parameters are in the learned set, which allows the researcher to judge the empirical results more precisely. The true population parameters are denoted with an asterisk, i.e. the true impact matrix is $B^* = \Sigma_c Q^*$ with Q^* being the rotation matrix that yields B^*. Thus, in the learning notation, Q^* corresponds to θ^*. Ultimately, the objective is then to estimate the unique true IRFs or FEV decomposition. In this context, set-identification reflects the inability of the researcher to impose exact identifying restrictions. Hence, only a set that contains the true structural parameters is narrowed down.

To establish a clear interpretation of the learning guarantees, I impose following assumptions:

Assumption 3.1.

1. *The reduced form parameters of the DGP are estimated consistently.*
2. *The error covariance matrix Σ has full rank.*
3. *The bootstrap procedure for the reduced form parameters is consistent.*
4. *Given the true DGP, the rotation parameters θ^* yield the true structural population parameters.*
5. *The imposed identification assumptions are satisfied for θ^*.*

In this chapter, I stick to the basic structural VAR case, where the underlying DGP is modeled by a standard VAR(p) process. Yet, more flexible models, such as e.g. VARs with time-variation or other forms of nonlinearity are also covered. For the baseline structural VAR, Assumption 3.1.1 is easily satisfied since the parameters can be estimated consistently by either least squares or maximum likelihood estimation.

Assumption 3.1.2 ensures that Σ is positive definite, and in practice this means that no linear dependencies exist between the variables in the system. Assumption 3.1.2 also ensures that the structural parameters of interest, that are a function of Σ, are estimated consistently when Σ is. In settings with cointegration this assumption might be violated, but for this chapter I focus on the basic structural VAR framework without such cointegration relationships.

Assumption 3.1.3 requires that the bootstrap consistently replicates the distribution of the estimator for the reduced form parameters. This means that if the estimator consistently estimates the underlying population parameters, the bootstrap will do so as well. In the case of the structural VAR the residual-based moving block bootstrap proposed by Brüggemann et al. (2016) satisfies this assumption since it is consistent for both $\Phi^{(H)}$ and Σ. The simple wild bootstrap does not satisfy Assumption 3.1.3 and is only considered because it might be advantageous in smaller samples due to its overly conservative uncertainty quantification.

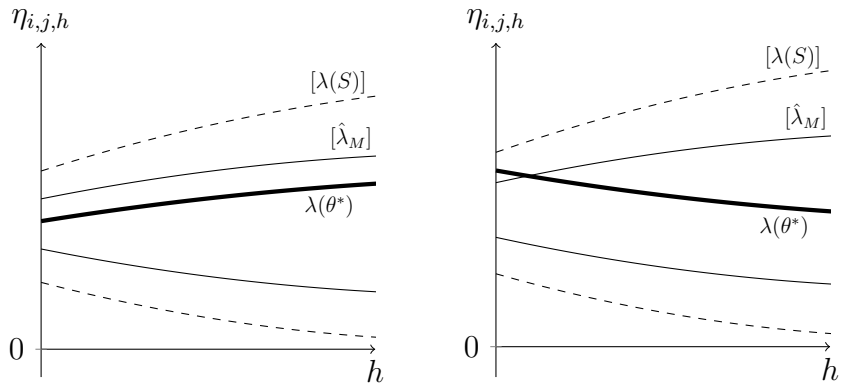
Assumption 3.1.4 again expresses the frequentist school of thought that one true population parameter is prevalent. Finally, assumption 3.1.5 translates to the economic intuition behind the identifying restrictions being correct in practice. Meaning that the imposed restrictions are satisfied by the true population parameters, e.g. that η^*_{ijh} is actually positive when a positive sign restriction is placed on it. In other words, this simply means that the identification strategy is valid, an assumption that is implicit in most structural VAR applications. Naturally, if the identification restrictions are wrong, one does not estimate the underlying parameters of interest in the first place.

Proposition 3.1. *Let Assumptions 3.1 hold and let $\lambda(\theta)$ represent either the structural impulse responses or the FEV decomposition. For $T \to \infty$, if the imposed identification restrictions are correct, meaning that θ^* satisfies them, then $P(\theta^* \in S_U) \to 1$, and thus $P(\lambda(\theta^*) \in \lambda(S_U)) \to 1$.*

Proposition 3.1 states that for $T \to \infty$ the true structural parameters, e.g. the true structural impulse response functions, are in the set which is learned by the augmented learning procedure. Thus, in this case the learning guarantees, which describe the precision with which the set is learned, allow for statements about how likely it is that the true structural parameters are learned. The proof of Proposition 3.1 is relegated to Appendix 3.A.

Corollary 3.1. *Following from Proposition 3.1, if the tightest band $[\lambda(S_U)]$ is learned as described in* (3.2.5) *and for $T \to \infty$, then $P(P(\lambda(\theta^*) \in [\hat{\lambda}_M]) \geq 1 - \epsilon) > 1 - \delta$.*

Figure 3.2: Learning Scenarios in the Limit



(a) Not misclassified (b) misclassified

Corollary 3.1 states that for $T \to \infty$ the true structural parameters lie within the learned set with a certain, but double-layered, probability. The formal reasoning behind Corollary 3.1 is given in Appendix 3.A. If Proposition 3.1 holds and the true parameters are within the set, two possible scenarios exist. Either θ^* is correctly classified as inside the set or it is misclassified. Figure 3.2 highlights these two possibilities for the example of the structural IRFs. As shown in panel (a) the impulse response is correctly classified if the response parameters are inside the learned band for each horizon $h = 0, ..., H$. Recall that the horizon H represents the dimensionality d in this application of the learning framework. Hence, the probability statement given by the PAC-learnability framework is of joint nature, allowing for statements about the responses up to horizon H simultaneously. For example, if the responses of a variable in the system to a structural shock of interest are learned up to horizon H, then the following asymptotic interpretation holds for all horizons simultaneously.

Asymptotic Interpretation: *The probability that the structural impulse responses of variable i are jointly inside the learned set is larger than $1 - \epsilon$. This statement is true with probability larger than $1 - \delta$.*

The interpretation for the FEV decomposition is analogous. Suppose one isolates the inner statement, the 'approximately' part, of the PAC-learnability framework $\mathcal{L}([\hat{\lambda}_M]; [\lambda(S_U)], W) < \epsilon$. Let e.g. $\epsilon = 0.05$, then if one were to repeat the analysis multiple times the true structural parameters would be misclassified in less than 5% of the resampling exercises. More than 95% of the time, the true structural

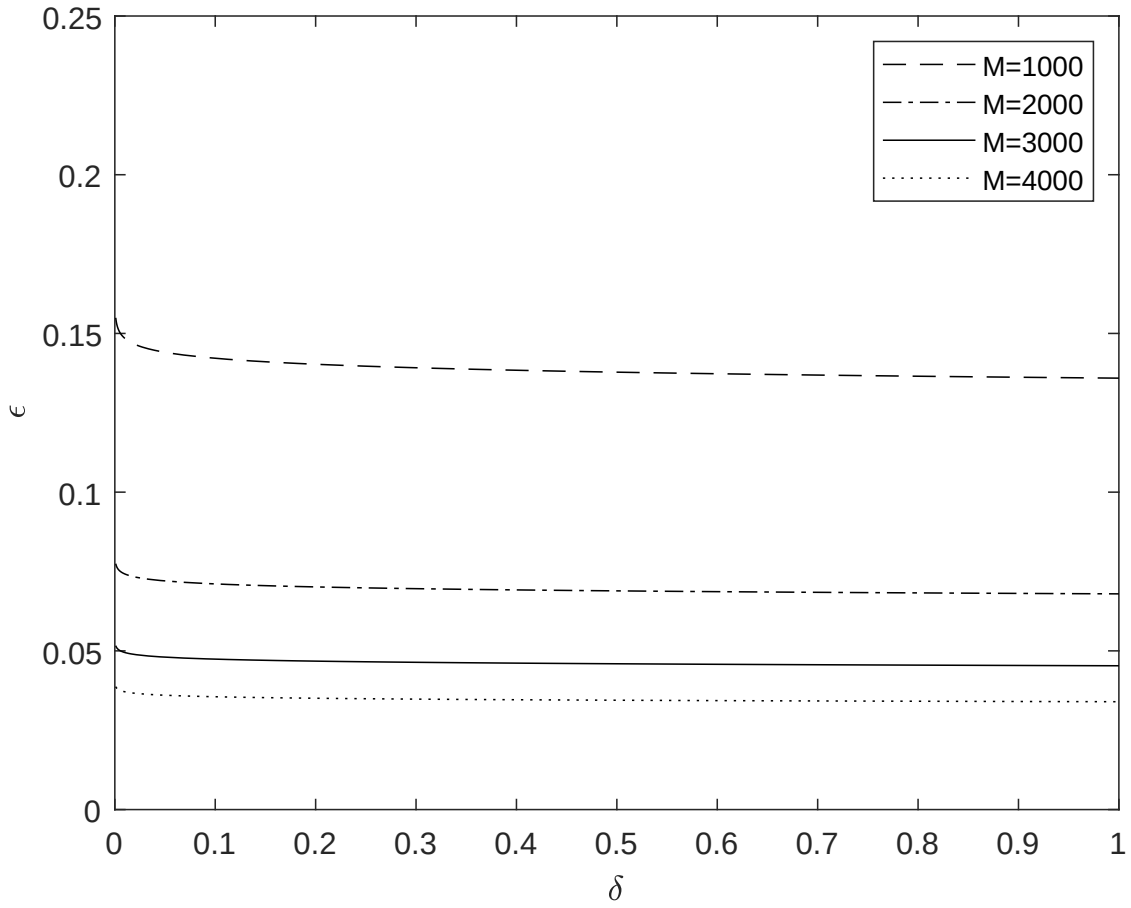
parameters are inside the learned band. This resembles the frequentist interpretation, and thus the statement from the partial PAC-learnability statement can be rephrased with a link to the classical frequentist interpretation. For $T \to \infty$, based on the 'approximately' part, one has a conservative frequentist band, however due to the 'probably' part of the statement one is not sure whether this frequentist interpretation holds.

Thus, this link to a frequentist interpretation is of limited use. First, due to the inequality $\mathcal{L}([\hat{\lambda}_M]; [\lambda(S)], W) < \epsilon$ one only gets the insight that one has a confidence band with a coverage of at least $1-\epsilon$. Second, even if ϵ can be compared to the classical level of significance α in isolation, this comparison breaks down in conjunction with δ. The additional uncertainty introduced through δ cannot easily be expressed in terms of α. Furthermore, a single number of draws M from inside the set admits different combinations of ϵ and δ, meaning that the interpretation is not unique. Thus, a purely frequentist interpretation and frequentist coverage is not attained by the learned bands. Yet, this is a feature that this chapter has in common with e.g. the general robust Bayesian inference of Giacomini and Kitagawa (2021), and thus also with the calibrated projection approach of Gafarov et al. (2016).

To employ this learning approach, several practical issues need to be addressed. Firstly, the finite sample properties of the augmented learning algorithm need to be assessed. Section 3.3 shows results of a Monte Carlo simulation study that computes coverage rates of the identified set for different typical sample sizes and different bootstrap procedures.

A second practical issue is the choice of ϵ and δ. As mentioned above these two hyperparameters cannot be readily related to a standard confidence level α or Bayesian quantiles. Yet, I suggest to set ϵ to typical significance levels, such as e.g. $\epsilon = 0.05$, since it can be interpreted in this fashion when isolated. However, one has to be aware that additional uncertainty is introduced through δ. The simulation study described in the next section also helps to set the two hyperparameters as it presents coverages for different values. Furthermore, the choice of particular values for ϵ and δ induces a specific number of draws M, but also other combinations of ϵ and δ can yield the same number of draws. Hence, as done in the paper by Montiel Olea and Nesbit (2021), it would be good practice to report the iso-draw curve which shows the different combinations of ϵ and δ for the given number of draws M, to present the reader the spectrum of admissible interpretations for a given number of draws M.

Figure 3.3: Iso-Draw Curves for Different Numbers of M



Notes: The lines depict the combinations of ϵ and δ that yield a fixed number of draws M.

Figure 3.3 gives the iso-draw curves for different values of M. The curves reveal that if you decrease ϵ by a small margin, for a given M, δ increases considerably, meaning that one is less certain that the bound on the misclassification probability holds. Hence, this relationship counteracts the conservativeness of the 'approximate' statement of the PAC-learnability framework. The misclassification probability is likely to be close to the chosen ϵ, since, for a given M, δ decays very quickly with a decrease in ϵ, making it unlikely that the misclassification probability is considerably smaller than the chosen ϵ value. This relationship is consistent throughout different numbers of draws as illustrated by the different iso-draw curves in Figure 3.3.

Another another practical consideration is that the learning procedure is a distribution free approach. In the end, it does not yield a point estimate or posterior median which is spanned by a band, but merely the maximum and minimum value of the learned set in each dimension. I propose to also report the learned set based only on the single reduced form parameter estimates. This is the set from the base-

line learning approach as featured in Montiel Olea and Nesbit (2021) which does not reflect estimation uncertainty. Hence, the learned bands from the augmented learning procedure proposed in this chapter span the baseline set and give an idea of the additional uncertainty that is introduced due to estimation uncertainty, similar to confidence bands around a point estimate in the frequentist world.

Finally, depending on the probability distribution W one has to be careful when interpreting the learned set. If the probability distribution W puts high mass on the center of the parameter region and no or only little mass near the extreme values in each dimension, the distance between the learned band $[\hat{\lambda}_M]$ and the actual tightest band $[\lambda(S_U)]$ might be large. Montiel Olea and Nesbit (2021) formalize this using the Hausdorff metric which measures the distance between two sets of a metric space. Looking at Figure 3.1 in section 3.2.1, the Hausdorff distance measures, roughly speaking, the largest distance between the outer and the inner rectangles. Hence, although the PAC-learnability statement holds and the resulting asymptotic interpretation is valid, the distance between the learned set and the true underlying set could be large. Montiel Olea and Nesbit (2021) show that for certain probability distributions W over the $\lambda(S_U)$, the worst-case relative Hausdorff difference, which normalizes the Hausdorff distance between 0 and 1, is controlled by the misclassification probability. The worst-case relative Hausdorff distance, as defined by Montiel Olea and Nesbit (2021), for the union of the parameter regions is:

$$\tilde{d}_H\left([\lambda(S_U)], [\hat{\lambda}_M])\right) \equiv \frac{d_H([\lambda(S_U)], [\hat{\lambda}_M])}{\sup\limits_{a \subseteq [\lambda(S_U)]} d_H([\lambda(S_U)], a)}, \tag{3.2.6}$$

with $d_H([\lambda(S_U)], [\hat{\lambda}_M])$ denoting the Hausdorff distance between sets $[\lambda(S_U)]$ and $[\hat{\lambda}_M]$, while a denotes any possible hyperrectangle contained in $[\lambda(S_U)]$. Following Montiel Olea and Nesbit (2021), suppose that the following assumptions regarding the probability distribution W hold.

Assumption 3.2.

1. *The projection of $\lambda(S_U)$ into its ith coordinate*

   $$p_i(\lambda(S_U)) \equiv \{x \in \mathbb{R} \mid x = \lambda_i,\ \lambda \in \lambda(S_U)\},$$

   is a bounded interval $[\underline{r}_i, \overline{r}_i]$.

2. *The distribution W on S_U is such that for any set $[a_i, b_i] \subseteq p_i(\lambda(S_U))$ and any coordinate i:*

$P(\lambda(S_U) \in [a_i, b_i]) = \frac{b_i - a_i}{\overline{r}_i - \underline{r}_i}$.

Assumptions 3.2.1 and 3.2.2 state that the parameter region is bounded and that the probability distribution is uniform over the projection in each coordinate. Proposition 3.2 applies the bound on the worst-case relative Hausdorff distance as established in Montiel Olea and Nesbit (2021) to the union of the parameter regions.

Proposition 3.2. *Let Assumption 3.2 hold. Then for any M it holds that*

$$\tilde{d}_H \left([\lambda(S_U)], [\hat{\lambda}_M]) \right) \leq \mathcal{L}([\hat{\lambda}_M]; [\lambda(S)], W),$$

and $\tilde{d}_H$ is defined in (3.2.6).

The proof follows from Proposition 1 in Montiel Olea and Nesbit (2021). Proposition 3.2 shows that if the underlying probability distribution is uniform in each dimension the Hausdorff difference is controlled by the misclassification probability. Together with Theorem 3.1 this means that if M exceeds the sample complexity induced by ϵ, δ and d, the worst-case relative Hausdorff difference is below ϵ with probability greater than $1 - \delta$.

In practice, however, it is neither easy to draw from a distribution that satisfies Assumption 3.2 nor is it guaranteed that such a distribution exists. I follow the recommendation of Montiel Olea and Nesbit (2021) and suggest to compute the marginal cumulative empirical distribution functions (ECDF) of the learned parameters in each dimension. Comparing the resulting ECDFs to uniform distributions helps to gauge whether a large Hausdorff distance is a potential concern. In section 3.4 I compute the ECDFs in an empirical example in order to illustrate how to assess the implications of Proposition 3.2.

The problem of potentially large Hausdorff distances is further investigated in the simulation study. If large Hausdorff distances were present, one would expect this to be reflected in the calculated coverages. However, since the coverages presented in section 3.3 are stable across horizons and close to typical nominal levels, a problem with large Hausdorff distances is no concern. And finally, one could also argue that if no draws are found in a particular region of the set one could conclude that the data do not support these values.

3.3. Simulation Study

This section presents a simulation study that investigates the finite sample properties of the augmented learning procedure by providing coverages for structural IRFs and

the FEV decomposition in a simulated structural VAR. In detail, I compute joint coverage rates, i.e. how often the learned band encompasses the true underlying set of structural parameters across all horizons jointly. Second, I compute point-wise coverage rates which count how often the set of structural parameters is contained in the learned band for each horizon individually, as e.g. also done by Granziera et al. (2018). Thus, for the point-wise coverages I pretend that the learned bands are used and interpreted like a frequentist or Bayesian point-wise interval. Based on the coverages computed below, one can draw conclusions on how to set the hyperparameters in the learning procedure by choosing which type and level of coverage is aimed at.

The data generating process (DGP) in the simulation resembles the empirical example in the baseline learning paper by Montiel Olea and Nesbit (2021). I take quarterly US data from 1960:Q1 to 2019:Q2 for the consumer price index (CPI), gross domestic product (GDP) and the federal funds rate (FFR), with CPI and GDP entering the VAR in log differences. The model is estimated with $p = 4$ and the resulting reduced form parameter estimates constitute the DGP for the simulation study. In detail, I draw residuals from a multivariate normal distribution with covariance matrix $\hat{\Sigma}$ and construct the simulated data with the estimated intercept $\hat{c}$ and coefficient matrices $\hat{A}_i$ where $i = 1, \ldots, 4$. The relevant matrices of the DGP are presented in detail in Appendix 3.B.

The coverage of the learned bands should vary across several dimensions. First, the sample size certainly influences the coverage of the augmented learning framework. To explore how the coverage varies for common sample sizes, I simulate $T = \{250, 500, 1000\}$, where $T = 250$ and $T = 500$ roughly represent typical sample sizes for quarterly and monthly data, respectively. $T = 1000$ is the upper end of sample sizes for applications with common macroeconomic data.

Second, the coverage varies with the tuning parameters that determine the number of draws M to achieve the learning guarantee. Here, ϵ and δ govern how the learned bands are interpreted and are thus the centerpiece of the augmented learning framework. As described in section 3.2.3, the number of draws M changes mainly with ϵ. Therefore, I propose to keep δ fixed at a small value, i.e. $\delta = 0.01$ in order to simplify the interpretation. Keeping δ fixed at a small value shifts more weight to the choice of ϵ and its interpretation. I simulate four different values for $\epsilon = \{0.01, 0.05, 0.1, 0.32\}$. Note that these do not correspond to a specific nominal coverage in conventional terms, but rather represent different gradations of how precisely the underlying set is learned.

Apart from ϵ and δ, the dimensionality of the learning procedure, i.e. the horizon up to which the structural parameters are learned, affects the number of draws M. Yet, if for a different number of draws, due to a change in the dimensionality d, ϵ and δ remain unchanged the same joint interpretation still holds. However, different numbers of draws M should influence the point-wise coverages since for each horizon a different number of draws is used to construct the interval. For example, as d decreases, the number of draws decreases and the learned band is likely to become narrower at each individual horizon, and thus point-wise coverages are likely to decrease. The baseline simulation chooses $d = H = 25$, meaning that the impulse responses are learned at impact and 24 periods after impact. For the FEV decomposition this means that the FEVs from horizon $h = 1$ up to $h = 25$ are learned. Increasing the dimensionality, increases the number of draws M accordingly. Hence, the baseline simulation results act as a lower boundary for the point-wise coverages which are obtained for higher dimensions. To assess the sensitivity of the point-wise coverage to lower dimensions, I also present results for $d = H = 13$.

Third, the coverage depends on the bootstrap that is incorporated into the learning framework. On the one hand, the type of bootstrap should influence the coverage rates and to assess this relationship I simulate once with the simple wild bootstrap and once with the residual-based moving block bootstrap proposed by Brüggemann et al. (2016). On the other hand, the number of bootstrap iterations should also have an effect on coverage rates. However, I propose simply keeping the number of bootstrap iterations fixed at a level that is typically deemed sufficient, i.e. the baseline simulations use 1,000 iterations, and tuning the coverage mainly with the ϵ parameter. This, in turn, helps the interpretation because a large number of bootstrap iterations maximizes the probability that the true structural parameters are in the learned set, what allows for the clear interpretations. Yet, in order to assess whether the augmented learning procedure is very sensitive along this dimension, I present simulation results for the residual-based moving block bootstrap with 500 bootstrap iterations in Appendix 3.B.

From the simulated DGP, a contractionary monetary policy shock is identified with a simple set of sign restrictions. As in Montiel Olea and Nesbit (2021), the initial response of CPI and GDP is restricted to be negative while the initial response of the FFR is positive. This pattern characterizes an identified set of structural models that satisfy these restrictions. This identified set has an upper and a lower boundary given the imposed restrictions, which in this simple example are easily computable for the structural IRFs with the mathematical program established by Gafarov et al. (2018).

This has the advantage that I can analytically compute the true boundaries of the identified set of the DGP and use these boundaries to evaluate coverages. However, for the FEV decomposition I have to rely on numerical techniques to compute the true boundaries of the identified set given the DGP parameters. Therefore, small deviations of the numerically computed parameters from the underlying true ones might influence the coverage results for the FEV parameters.

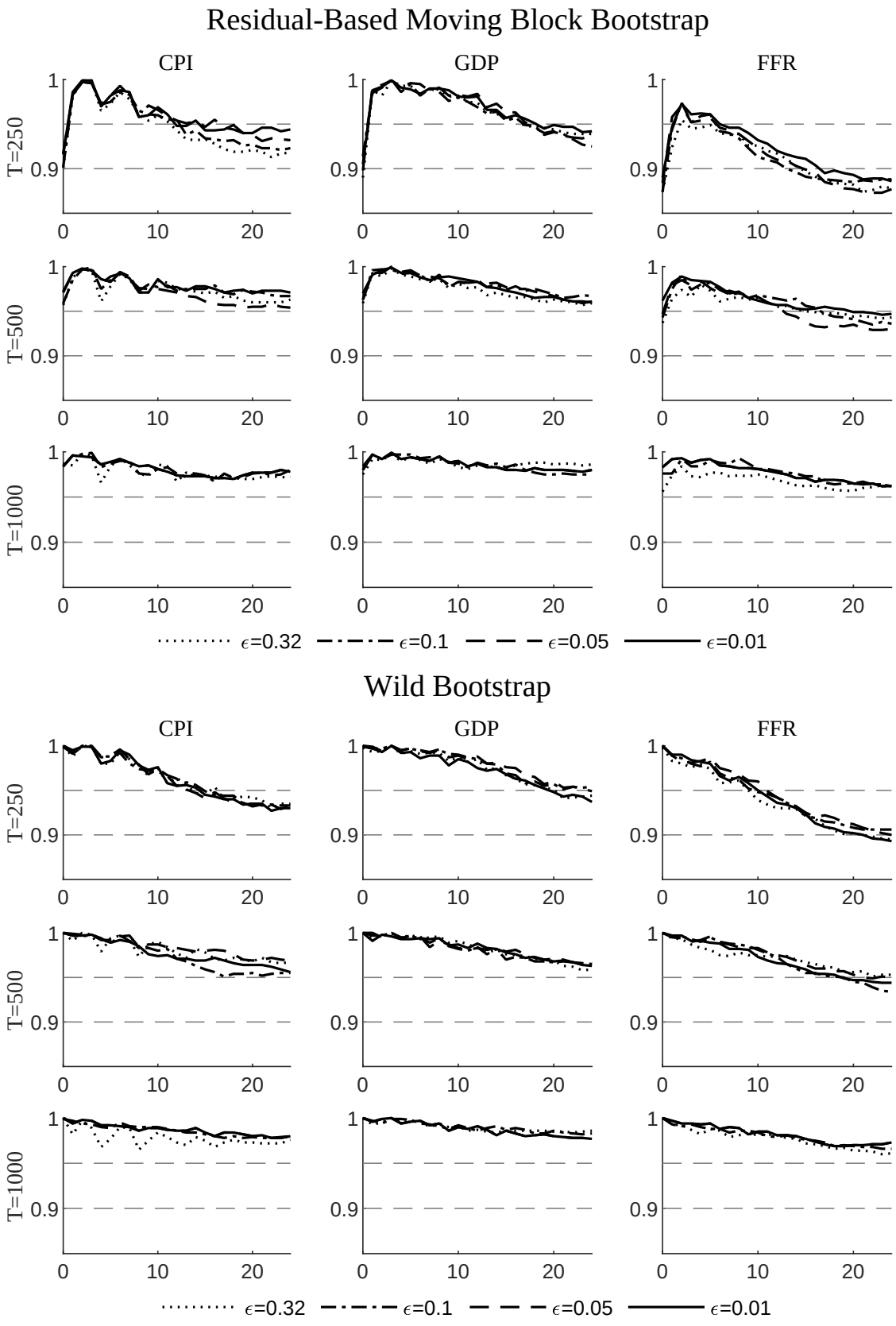
Furthermore, the coverages which are presented below always refer to the boundaries of the identified set, i.e. if the learned set encompasses the upper and lower boundary of the true identified set. Depending on the type of coverage that is computed, it does so either point-wise for each horizon individually or jointly over all horizons at the same time. As described in the previous section, the interpretation of the augmented learning procedure is tied to the frequentist way of thinking with a single true structural model. Yet, I do not fix such a true single structural model which lies within the identified set of the simulated DGP. In practice, this one true model and its position in the identified set is also unknown. If the underlying identification restrictions are correct, it could theoretically fall on the boundary or inside the set. Hence, coverages with respect to the true model could largely vary depending on the position of the true model in the set. Thus, the following coverages are computed with respect to the identified set and should be interpreted as such. In the end, they indicate how often the true identified set, which contains the true structural parameters, is covered. Hence, these coverages act as lower bounds for the coverage of any single structural parameter vector inside the set.

3.3.1. Impulse Response Functions

Figures 3.4 and 3.5 first show the point-wise coverages for the structural IRFs for $H = 25$ and $H = 13$, respectively. The upper panel shows the coverages for the residual-based moving block bootstrap while the lower panel depicts the results for the wild bootstrap. In the rows the results for different sample sizes are shown and the columns correspond to the three different variables of the system. For reference, the nominal coverage rates of 90% and 95% are sketched with the thin dashed lines. The different black lines give the point-wise coverages for the different ϵ values.

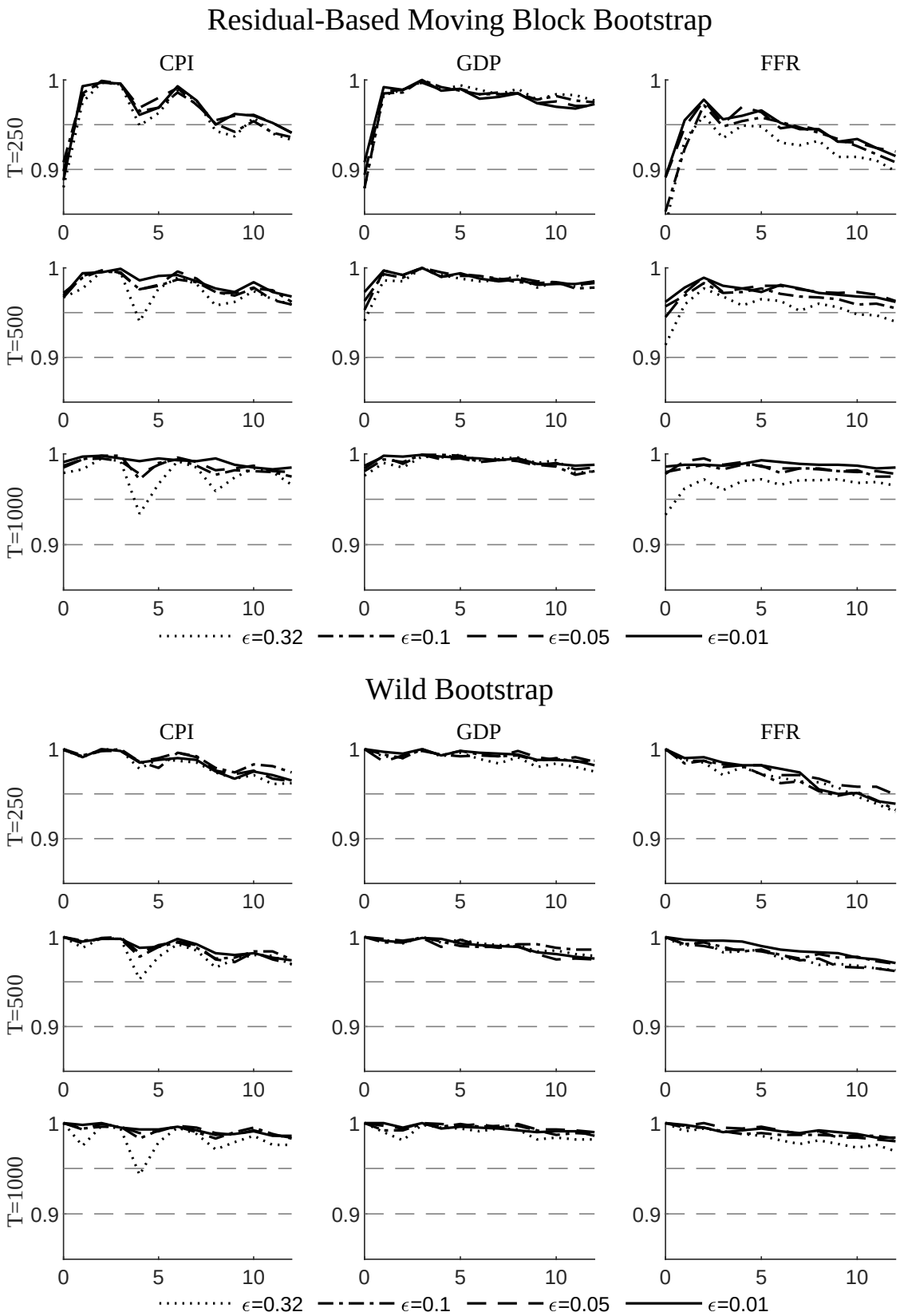
Note that the identified set is bounded from above at exactly zero at impact for the CPI and GDP because the impulse response is restricted to be negative. Similarly, the identified set for the FFR response at impact is bounded at zero from below as this response is restricted to be positive. The probability to get exactly zero for an impulse response based on a random draw of rotation matrices is zero, and thus

Figure 3.4: Coverages for the IRFs for H=25



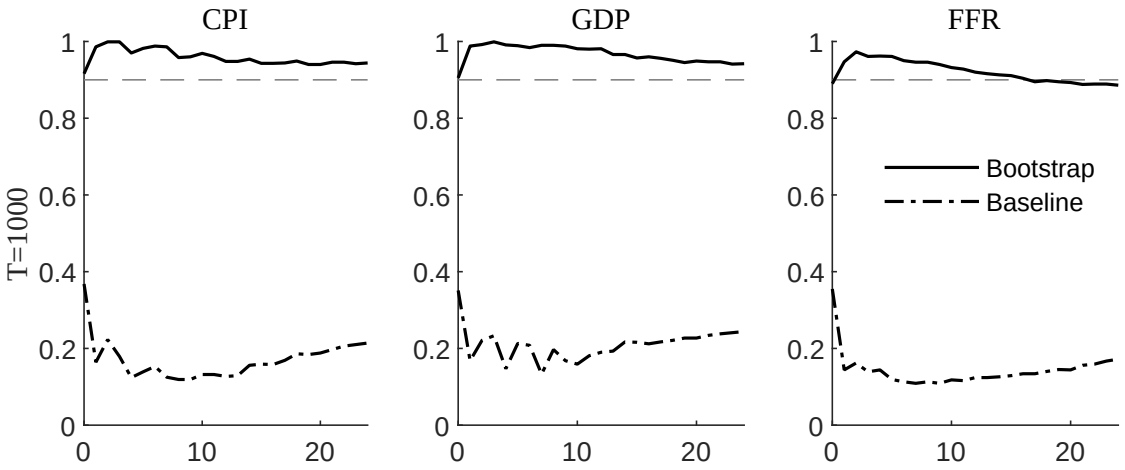
Notes: The upper panel depicts point-wise coverages for the augmented learning procedure based on 1,000 residual-based moving block bootstrap iterations while the lower panels are based on the wild bootstrap.

Figure 3.5: Coverages for the IRFs for H=13



Notes: The upper panel depicts point-wise coverages for the augmented learning procedure based on 1,000 residual-based moving block bootstrap iterations while the lower panels are based on the wild bootstrap.

Figure 3.6: Coverages for the Baseline Learning Algorithm



Notes: The solid black lines depict the point-wise coverages for of the augmented learning approach based on the residual-based moving block bootstrap. The dash-dotted lines depict the point-wise coverages for the baseline algorithm of Montiel Olea and Nesbit (2021) without a preceding bootstrap. Both coverages are computed for $T = 1000$, $H = 25$, $\delta = 0.01$ and $\epsilon = 0.01$.

the coverage of the impulse responses at impact is zero by construction. Therefore, the coverages for $h = 0$ that are depicted in Figures 3.4 and 3.5 represent one-sided coverages, meaning that for CPI and GDP only the lower boundary of the identified set is considered and accordingly the upper boundary for the FFR response. In the following, I describe the simulation results first and give a practical recommendation regarding the choice of the hyperparameters at the end of the subsection.

Figure 3.4 reveals that for the residual-based moving block bootstrap and $T = 1000$, coverage rates are consistently above a point-wise nominal coverage rate of 95%. This is true for all different ϵ values. Hence, the setting of the hyperparameters, and thus the number of draws M, seems to have only a modest influence on point-wise coverages. A similar picture emerges for $T = 500$, but with slightly lower coverage rates. Only for $T = 250$ coverage seems to vary more across ϵ values and variables with point-wise coverages dropping below 95% for the CPI and GDP at several horizons h or even below 90% for the FFR. The results for the wild bootstrap show the same pattern but as expected coverages are generally slightly higher. The one major difference is that for the wild bootstrap coverages at impact are close to 100% across all sample sizes, variables and hyperparameters. Overall, the relatively large point-wise coverage rates are not surprising as the underlying interpretation of the learning approach is of joint nature. Similar 'overcoverage' is observed by other methods to conduct joint inference, e.g. for the Bonferroni approach by Granziera et al. (2018). The upper panel of Figure 3.B.1 in Appendix 3.B shows that the results

are not overly sensitive to the number of bootstrap iterations. For 500 residual-based moving block bootstrap iterations, the coverages are only slightly lower and the same patterns emerge overall.

Figure 3.5 shows the results for $H = 13$. As described above the dimensionality of the parameter region could also affect point-wise coverages by altering the number of draws M which are used to learn the band in each dimension. Consistent with the findings in Figure 3.4 that the number of draws has only a small effect on the point-wise coverages, the results for $H = 13$ look similar to the results of $H = 25$. It does not matter whether the number of draws changes due to a change in the hyperparameters or due to a change in the dimensionality of the parameter region. If one compares the first horizons of Figure 3.4 with Figure 3.5, one observes similar coverages.

For comparison, Figure 3.6 shows the difference in the coverages between the augmented learning approach with the residual-based moving block bootstrap and the baseline learning algorithm of Montiel Olea and Nesbit (2021) which neglects the estimation uncertainty. The coverages are computed for $T = 1000$, $H = 25$, $\delta = 0.01$ and $\epsilon = 0.01$. The comparison highlights that in terms of coverage the preceding bootstrap is largely accountable for the coverages being close to the nominal levels used in practice.

Yet, since the bands from the augmented learning algorithm have a joint interpretation, the respective joint coverages are important. As the joint interpretation depends directly on the hyperparameters, joint coverages might also show more dependence regarding the choice of the hyperparameters. Table 3.1 depicts joint coverage rates for the two considered bootstrap schemes. Joint coverages are the proportion of Monte Carlo simulations for which the learned bands encompasses the identified set over all horizons jointly. Note that for $h = 0$ again only the lower or upper boundary is considered depending on the variable of the system.

Table 3.1 shows that joint coverages are closer to nominal coverage rates targeted in empirical applications. For the residual-based moving block bootstrap and for $T = 1000$ joint coverages are between 90% and 95% for the tight choice of $\epsilon = 0.01$ and decrease to roughly 90% for $\epsilon = 0.05$ and $\epsilon = 0.1$. Coverages go down to roughly 85% for $\epsilon = 0.32$. Hence, up to $\epsilon = 0.1$ joint coverages seem to be above or around the conventional nominal coverage of 90%. For smaller sample sizes the joint coverages are considerably lower, ranging from 80% to 90% for $T = 500$ and from 65% to 80% for $T = 250$. Joint coverages for the wild bootstrap are higher across all combinations of hyperparameters and sample sizes. Especially for the lower

Table 3.1: Joint Coverages of the Augmented Learning Framework - IRFs

CPI / GDP / FFR	$\epsilon = 0.01$	$\epsilon = 0.05$	$\epsilon = 0.1$	$\epsilon = 0.32$
	Residual-Based Moving Block Bootstrap			
$T = 250$	79.1/81.0/70.6	78.1/80.6/70.6	75.5/80.2/68.8	74.9/78.9/65.6
$T = 500$	87.3/89.2/85.8	86.8/88.6/81.5	86.6/89.1/81.5	82.0/86.7/80.0
$T = 1000$	93.2/94.0/92.3	90.1/91.9/90.5	89.2/91.9/88.2	85.8/91.5/85.2
	Wild Bootstrap			
$T = 250$	86.4/89.3/83.6	87.7/90.5/83.7	87.4/91.6/85.3	86.5/89.2/81.3
$T = 500$	90.1/92.8/91.4	92.6/93.3/91.0	90.8/92./90.1	89.5/93.3/91.5
$T = 1000$	93.7/94.9/94.6	93.6/94.9/94.4	92.8/95.5/93.5	87.7/94.2/90.3

Note: The table depicts the joint coverage rates of the learned bands out of 1,000 simulation rounds. The first number gives the coverage rate for the variable CPI and the second and third number the rates for the variables GDP and FFR, respectively.

sample sizes the joint coverages considerably exceed those based on the residual-based moving block bootstrap.

From the simulation results one can draw a practical recommendation regarding the setting of the hyperparameters ϵ and δ. Since the number of draws depends primarily on ϵ, I propose to fix δ at a low value to simplify the interpretation, e.g. $\delta = 0.01$. The choice of ϵ depends on the goal of the researcher. If certain point-wise coverages are aimed at, the choice of ϵ is important only for smaller sample sizes where a tighter value can help to increase coverage rates, but also for less tight ϵ values coverages are above or around standard nominal levels. Note that coverages are often above typical nominal levels because the learned bands admit a joint interpretation and are therefore conservative for point-wise evaluations.

Due to this point-wise overcoverage there is no need to rely on the even more conservative and asymptotically invalid wild bootstrap. The residual-based moving block bootstrap yields usable results also for smaller sample sizes. However, in situations where the latter bootstrap procedure is not applicable, one has to be aware of the additional overcoverage when one falls back to the wild bootstrap.

Though, if the target of the researcher are particular joint coverages, the recommendation for the choice of the hyperparameters changes. Even for larger sample sizes it could be beneficial to choose tight ϵ values to achieve coverages between 90% and 95%. For $\epsilon \geq 0.1$ joint coverages are likely to drop below 90%. For smaller sample

sizes one can also consider to use the more conservative wild bootstrap, which seems to yield coverages closer to the desired nominal level which are also stable across variables for $T = 250$.

Hence, as the point-wise coverages do not suffer from setting tight ϵ values but joint coverages profit, I suggest to set ϵ rather tightly to $\epsilon = 0.05$ or even $\epsilon = 0.01$. If only a limited number of observations are available for the analysis, it could be beneficial to also report learned bands based on the wild bootstrap to assess the robustness of the results.

3.3.2. Forecast Error Variance Decompositions

Similar to the previous subsection, Figures 3.7 and 3.8 depict the point-wise coverages but now for the FEV decomposition. The layout of the figures is identical to the previous coverage figures. The difference for the FEV parameters is that the true boundaries of the underlying identified sets are not computed analytically but with numerical techniques. Specifically, I solve the following optimization problem for each variable i and for each horizon h:

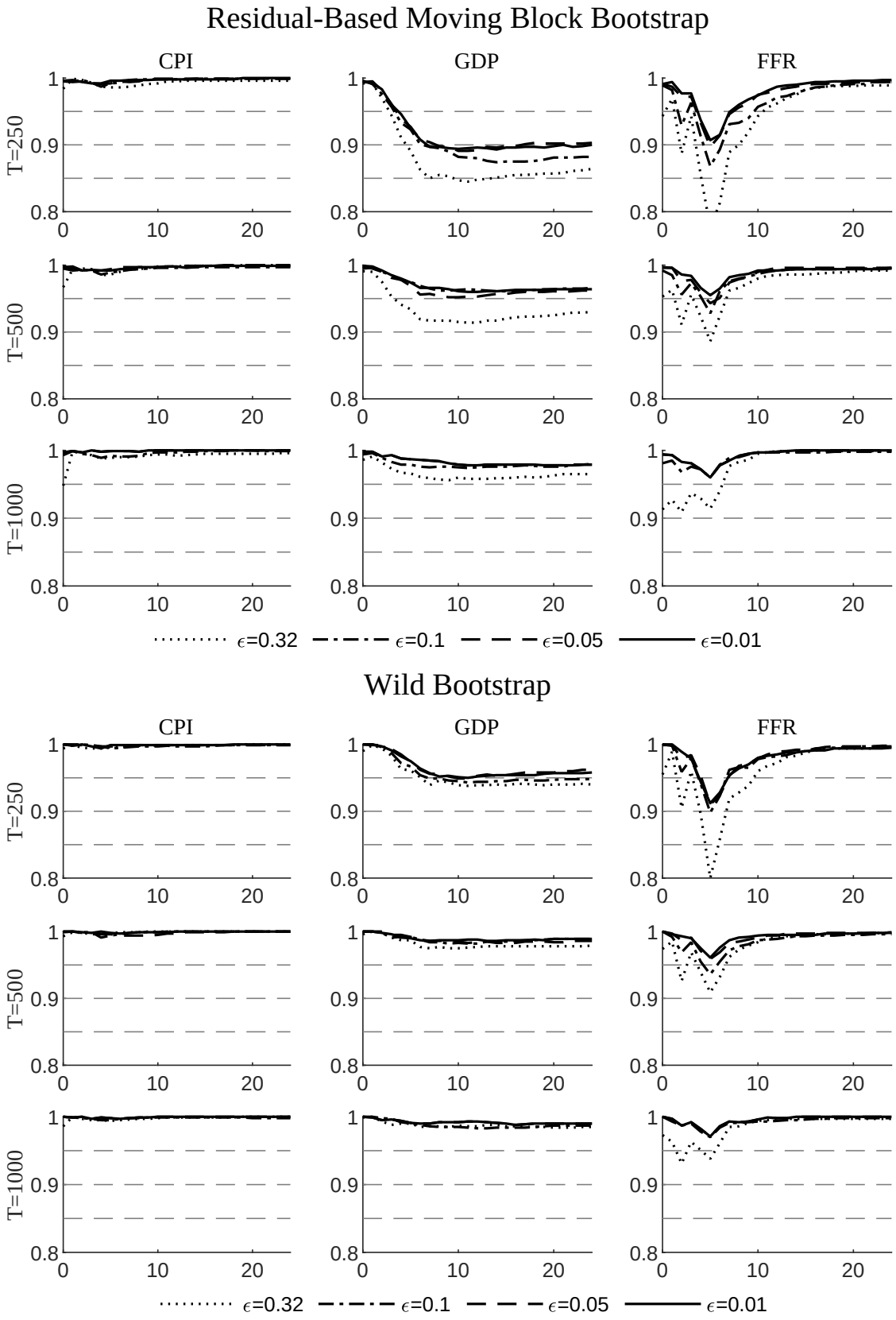
$$q^+ = \underset{q}{\operatorname{argmax}} \sum_{m=0}^{h} \Omega_i^+(h) \quad \text{s.t.} \quad q'q = 1 \text{ and the sign restrictions,}$$

with $\Omega_i^+(h)$ denoting the FEV corresponding to the DGP parameters $\Phi^{(H)+}$ and Σ^+. The lower bounds of the sets are computed analogously by minimizing $\Omega_i^+(h)$ subject to the same constraints. Plugging q^+ into $\Omega_i^+(h)$ yields the numerically computed boundaries of the true identified sets which are then used to evaluate the coverages.

Based on the numerical results, the restrictions imply boundaries for the true FEV sets that are numerically zero for different combinations of variables and horizons. As with the structural IRF sets it is very unlikely that a random draw of a rotation parameter yields a result so close to zero, and thus the coverage of the learned bands is again zero by construction. Hence, for these horizons, namely $h = 1$ and $h = 2$ for the CPI and $h = 1$ for GDP, the coverages are computed considering only the upper boundary of the FEV set. Again, the results are described first and the practical recommendation is given at the end of the subsection.

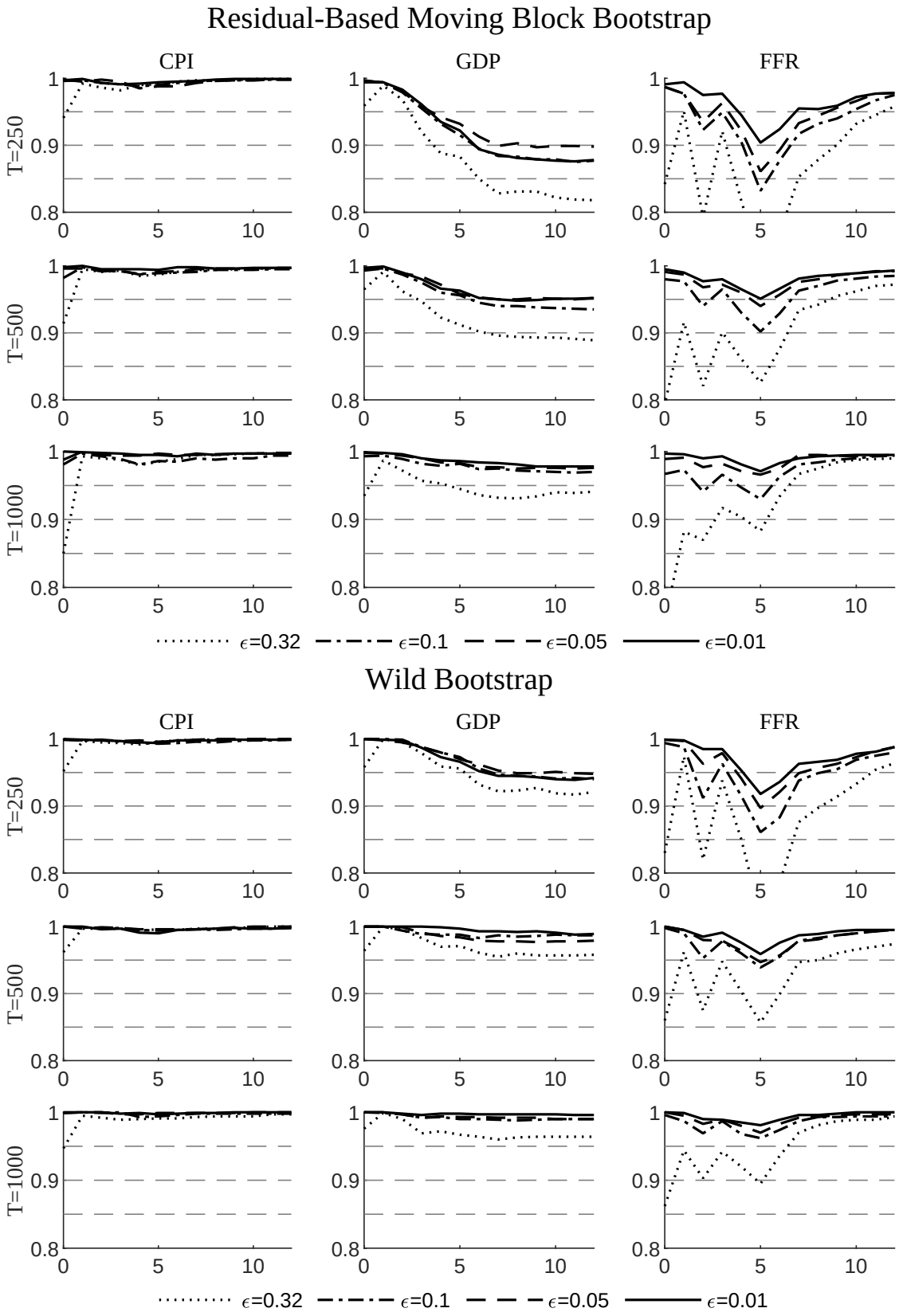
In contrast to the structural IRFs, one observes more differences across ϵ values and across variables. Looking at Figure 3.7 shows that point-wise coverages vary more. For tight ϵ values the coverage is close to 100% for the CPI but the pattern is different for the other two variables. For the GDP the coverage is close to 100% initially, but drops to roughly 90% for $T = 250$ and to roughly 95% for $T = 500$. One

Figure 3.7: Coverages for the FEV decomposition for H=25



Notes: The upper panel depicts point-wise coverages for the augmented learning procedure based on 1,000 residual-based moving block bootstrap iterations while the lower panels are based on the wild bootstrap.

Figure 3.8: Coverages for the FEV decomposition for H=13



Notes: The upper panel depicts point-wise coverages for the augmented learning procedure based on 1,000 residual-based moving block bootstrap iterations while the lower panels are based on the wild bootstrap.

also observes differences across hyperparameters and especially for $\epsilon = 0.32$ point-wise coverages are considerably lower. For the FFR the coverage dips in the first ten horizons and converges to 100% afterwards. Note that some coverages are so small that they are not shown in the figures for $T = 250$. For the visibility of the main results and the comparability with the other figures, I limit the axis of the respective panels to coverages above 80%. As for the GDP the dip is more pronounced for smaller sample sizes and less tight values for ϵ. For the wild bootstrap the very same pattern is observed, again with higher coverages across all variables and hyperparameters.

Compared to the structural IRFs, the coverage rates of the FEV decomposition are more sensitive to the number of bootstrap iterations. The lower panel of Figure 3.B.1 in Appendix 3.B shows the point-wise coverage rates for only 500 residual-based moving block bootstrap iterations. Although the patterns across the different variables remain intact, the coverages are sometimes considerably smaller for the GDP or the FFR, especially for smaller sample sizes and less strict hyperparameters. Hence, it is important to learn the parameters region based on at least 1,000 bootstrap iterations in order to ensure stable coverages for the FEV decomposition across the variables.

In contrast to the structural IRFs, point-wise coverages also differ more for $H = 13$. This is consistent with the observation that for the wild bootstrap coverages seem to depend more on the choice of hyperparameters, and thus on the number of draws M. If the dimensionality of the parameter region is decreased, i.e. the number of draws is reduced, point-wise coverages decrease accordingly. Yet, the same patterns across the different variables remain intact.

The patterns found for the point-wise coverages are also reflected in the joint coverages which are shown in Table 3.2. Joint coverages based on the residual-based moving block bootstrap are mostly above 95% for the CPI. For the GDP the joint coverage drops below 90% for $T = 250$ and joint coverages for the FFR are mostly below 90% for small sample sizes or $\epsilon \geq 0.1$. For $\epsilon = 0.32$ coverages are even below 80%. For $T = 500$ or $T = 1000$ and $\epsilon \leq 0.05$ the joint coverages are above 90% across all variables. As for the structural IRFs, the joint coverages based on the wild bootstrap are higher across the board and especially for smaller sample sizes. Even for $T = 250$ and $\epsilon \leq 0.05$ coverages are above or closely around 90% for all three variables.

The practical recommendation for the choice of ϵ for the FEV bands is analogous to the recommendation when IRFs are the parameters of interest, but the reasoning differs. The point-wise coverages actually differ considerably for different choices of

Table 3.2: Joint Coverages of the Augmented Learning Framework - FEVs

CPI / GDP / FFR	$\epsilon = 0.01$	$\epsilon = 0.05$	$\epsilon = 0.1$	$\epsilon = 0.32$
	Residual-Based Moving Block Bootstrap			
$T = 250$	98.0/87.1/87.4	97.9/88.1/85.7	97.4/86.0/80.8	96.1/82.1/66.1
$T = 500$	98.1/94.9/93.6	98.3/94.0/91.4	97.3/94.8/88.8	94.0/89.1/77.5
$T = 1000$	98.3/97.0/95.6	99.1/97.0/94.1	97.5/95.8/90.4	92.5/93.1/77.7
	Wild Bootstrap			
$T = 250$	99.2/94.6/90.0	99.0/94.2/89.0	99.2/93.3/85.4	98.1/90.7/70.7
$T = 500$	99.6/98.1/95.4	98.8/97.6/95.0	98.9/97.8/91.2	98.5/96.5/81.8
$T = 1000$	99.2/98.4/95.8	99.3/98.1/95.8	99.1/98.2/95.4	97.6/96.8/84.4

Note: The table depicts the joint coverage rates of the learned bands out of 1,000 simulation rounds. The first number gives the coverage rate for the variable CPI and the second and third number the rates for the variables GDP and FFR, respectively.

ϵ and I suggest setting $\epsilon = 0.01$ or $\epsilon = 0.05$ to ensure stable coverages across all variables. For smaller sample sizes $\epsilon = 0.01$ is preferable. The same reasoning applies to the joint coverages where $\epsilon = 0.01$ or $\epsilon = 0.05$ seems to give stable coverages above 90% across all variables. For smaller sample sizes $\epsilon = 0.01$ is again preferable, or one could also report results based on the wild bootstrap to assess the robustness of the results. Furthermore, note that the stability of the coverages across all variables induced by the tight ϵ values likely comes at the cost of excess coverage in terms of typical nominal levels for certain variables.

3.4. Empirical Illustrations

To illustrate the augmented learning framework, I present two different empirical illustrations. Firstly, I present the stylized empirical example on which also the simulation results in the previous section are based. This example is used to compare the learned bands with conventional frequentist and Bayesian inference. The second illustration replicates the empirical results of Piffer and Podstawski (2018) in order to demonstrate that the learned bands are informative in a realistic identification scenario.

3.4.1. Comparison with Conventional Inference

The first empirical illustration aims to compare the learned bands with existing methods of conducting inference. As mentioned above, for this purpose I revert back to the empirical example that was introduced in the simulation section. Again, the same FFR, CPI and GDP data are used to identify a monetary policy shock with the restrictions mentioned above. The learned bands are then compared to confidence intervals or credible regions which are based on existing inference procedures.

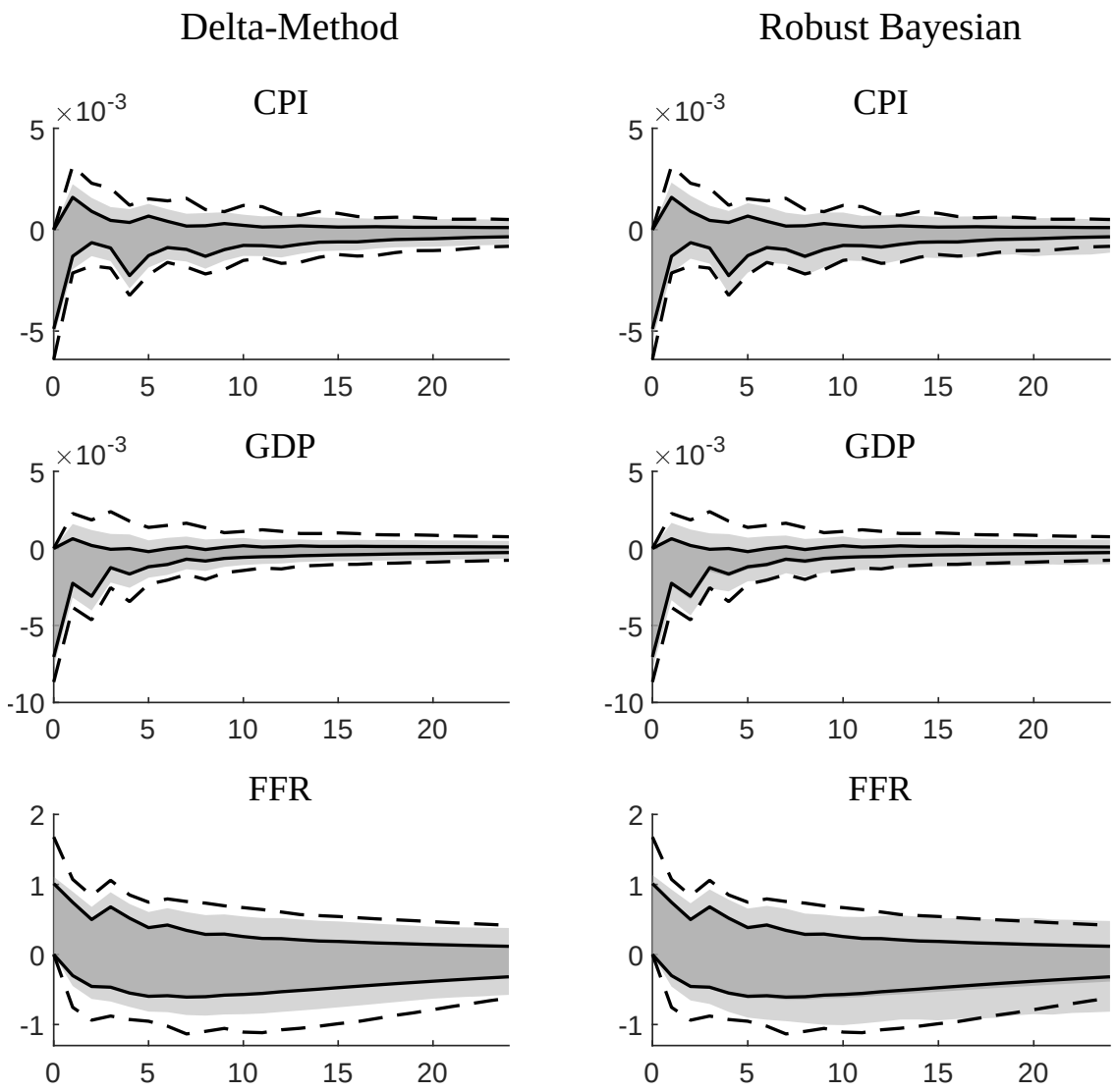
The simple example enables the comparison of the learned quantities with the delta method confidence intervals proposed by Gafarov et al. (2018), a frequentist approach with correct asymptotic coverage. These intervals are applicable when only a single shock is identified with restrictions on only that shock. Furthermore, I compute the robust Bayesian credible regions of Giacomini and Kitagawa (2021) as they are currently a popular way to conduct inference in set-identified structural VARs.

Comparing the learned bands with the existing methods to conduct inference allows to assess the informativeness of the learning procedure compared to these other approaches. The simulation results suggest high coverage rates from a point-wise perspective and if the learned quantities are too wide compared to the frequentist or Bayesian counterparts, the learning procedure might be of limited value to practitioners who want sharp results. Both the delta method and the robust Bayesian approaches are of point-wise nature, and thus the learned bands with the joint interpretation across all horizons are likely to be wider when each horizon is considered individually.

Figure 3.9 compares the learned bands from the augmented learning procedure with frequentist and Bayesian point-wise inference. The shaded areas depict the frequentist quantities on the left and the Bayesian quantities on the right while the lines depict the learned bands. The darker grey areas are the estimated identified set and the posterior mean bounds, respectively, and together with the lighter grey area they depict the confidence intervals or, in the Bayesian case, the credible regions. The solid lines depict the learned identified set based on the baseline algorithm of Montiel Olea and Nesbit (2021), while the dashed lines depict the bands from the augmented learning procedure. Both sets are learned with $\epsilon = 0.05$ and $\delta = 0.01$.

Comparing the solid line on the left with the darker grey shaded area reveals that the baseline algorithm without a preceding bootstrap learns the fixed data identified set precisely and problems with a potentially large Hausdorff distance are not present. Furthermore, comparing the 95% delta-method confidence intervals with the joint bands of the augmented learning algorithm shows that the learned bands

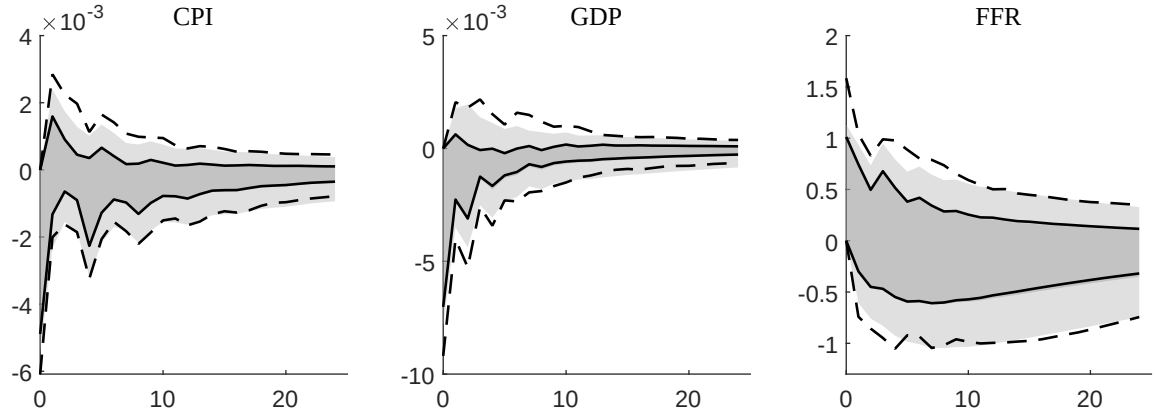
Figure 3.9: Comparing Learned to Frequentist and Bayesian Quantities



Notes: The lines represent the learned quantities. The solid lines depict the learned identified set based on the baseline algorithm by Montiel Olea and Nesbit (2021). The dashed lines depict the learned band based on the augmented algorithm with $\epsilon = 0.05$ and $\delta = 0.01$ and 1,000 residual-based moving block bootstrap iterations. On the left the darker grey area depicts the identified set and the lighter grey area is the 95% delta-method confidence intervals based on Gafarov et al. (2018). On the right the darker grey depicts the posterior mean bounds and light grey the according 95% robust Bayesian credible regions based on Giacomini and Kitagawa (2021).

exceed the point-wise frequentist intervals, but often not by a significant amount, especially for the later horizon. An analogous observation is made when comparing the learned and Bayesian quantities. The learned bands from the augmented learning procedure exceed the point-wise credible regions of the robust Bayesian approach, but they converge for later horizons. In terms of computational time, the robust

Figure 3.10: Comparing Learned Bands to Joint Inference



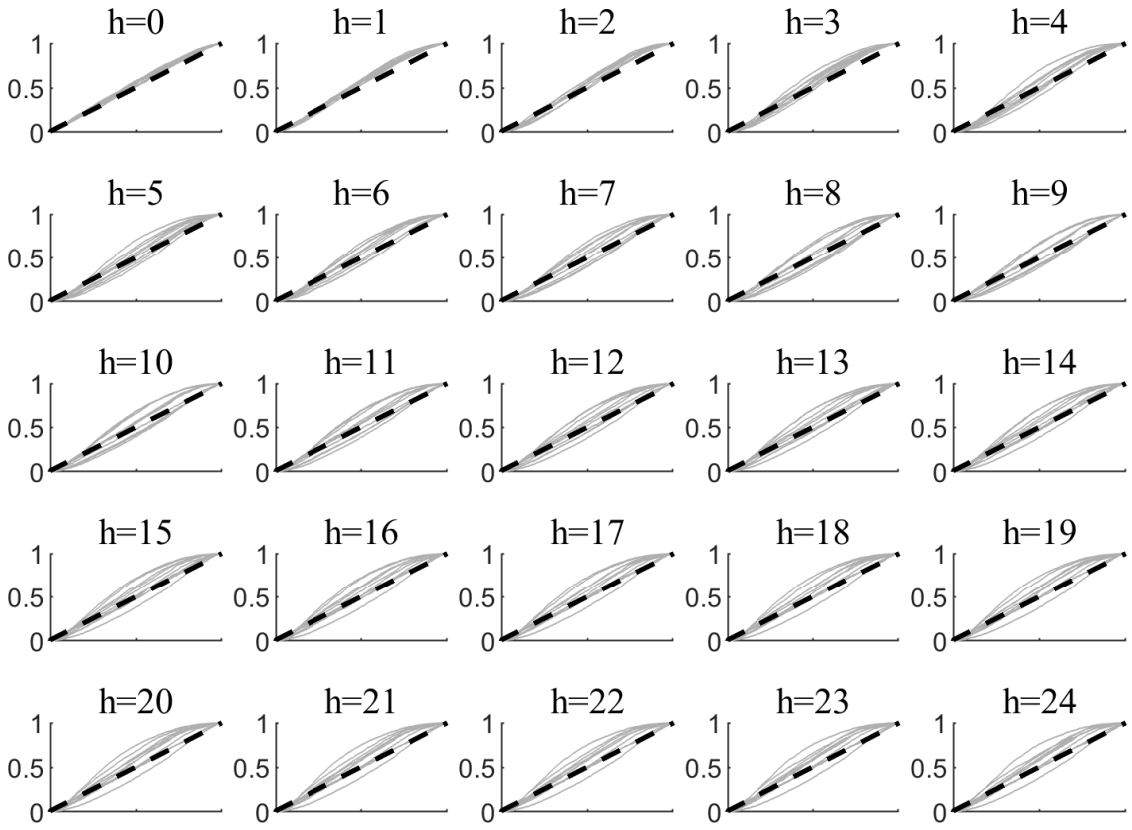
Notes: The lines represent the learned quantities. The solid lines depict the learned identified set based on the baseline algorithm by Montiel Olea and Nesbit (2021). The dashed lines depict the learned band based on the augmented algorithm with $\epsilon = 0.05$ and $\delta = 0.01$ and 1,000 residual-based moving block bootstrap iterations. The darker grey area depicts the estimated identified set and the light grey area the 95% joint confidence bands based on Granziera et al. (2018).

Bayesian credible regions for the three variables require roughly three hours when each boundary is computed in each horizon using numerical techniques. In contrast, the augmented learning approach requires about eight minutes for its calculations. Yet, the computational time of the robust Bayesian algorithm scales proportionally to the size of the underlying system while the computational time of the learning approach increases only marginally.

Figure 3.10 makes the same comparison with the joint frequentist bands based on Granziera et al. (2018), where the lines again represent the learned quantities. Both bands with joint interpretation are roughly comparable in their width. Looking at the simulations of Granziera et al. (2018) shows that their Bonferroni approach results in excess coverage when looking at the individual horizons in a point-wise manner. The same holds true for the simulation study in section 3.3, which shows that for $\epsilon = 0.05$ and $\delta = 0.01$ and for $T = 250$ observations some point-wise overcoverage is expected as well. Thus, in this particular stylized application the learning approach is roughly equivalent to the frequentist framework of joint inference. Yet, the learning approach has the edge over the frequentist approach in more general settings because it is easily adaptable to different identification schemes and because of its computational simplicity.

The concern about a large relative Hausdorff distance between the learned set and the one that should actually be learned can be addressed by looking at the marginal ECDFs of $\lambda(\theta_m)$ as described in section 3.2.3. A practical problem is that for any

Figure 3.11: ECDFs for the FFR Responses



Notes: The grey lines depict in each individual graph depict the marginal ECDF of the impulse response draws for ten randomly selected bootstrap parameters at the respective horizon. The dashed black line represents the CDF of the uniform distribution as a reference.

draw of θ_m one gets a cluster of different points $\lambda(\theta_m, \Phi_n^{(H)}, \Sigma_n)$ depending on for how many bootstrapped reduced form parameters the identification restrictions are satisfied. Hence, if one were to take all admissible values of $\lambda(\theta_m, \Phi_n^{(H)}, \Sigma_n)$ the resulting marginal ECDF would have a higher mass around points in the parameter region for which more individual parameter regions $\lambda(S_n)$ intersect. To circumvent this problem one can look at the ECDFs of the individual parameter regions $\lambda(S_n)$, namely the marginal ECDFs of all admissible $\lambda(\theta_m, \Phi_n^{(H)}, \Sigma_n)$ for a fixed n. Assumption 3.2 implies that the boundedness and uniform distribution of the projection in each dimension of the parameter region also holds for the subsets $\lambda(S_n) \subseteq \lambda(S_U)$. Thus, one can simply check whether one finds this uniform distribution empirically. If the underlying marginal ECDFs of the individual parameter regions are close to a uniform distribution one can then conclude that a large Hausdorff distance is unlikely to be a problem for the learned bands.

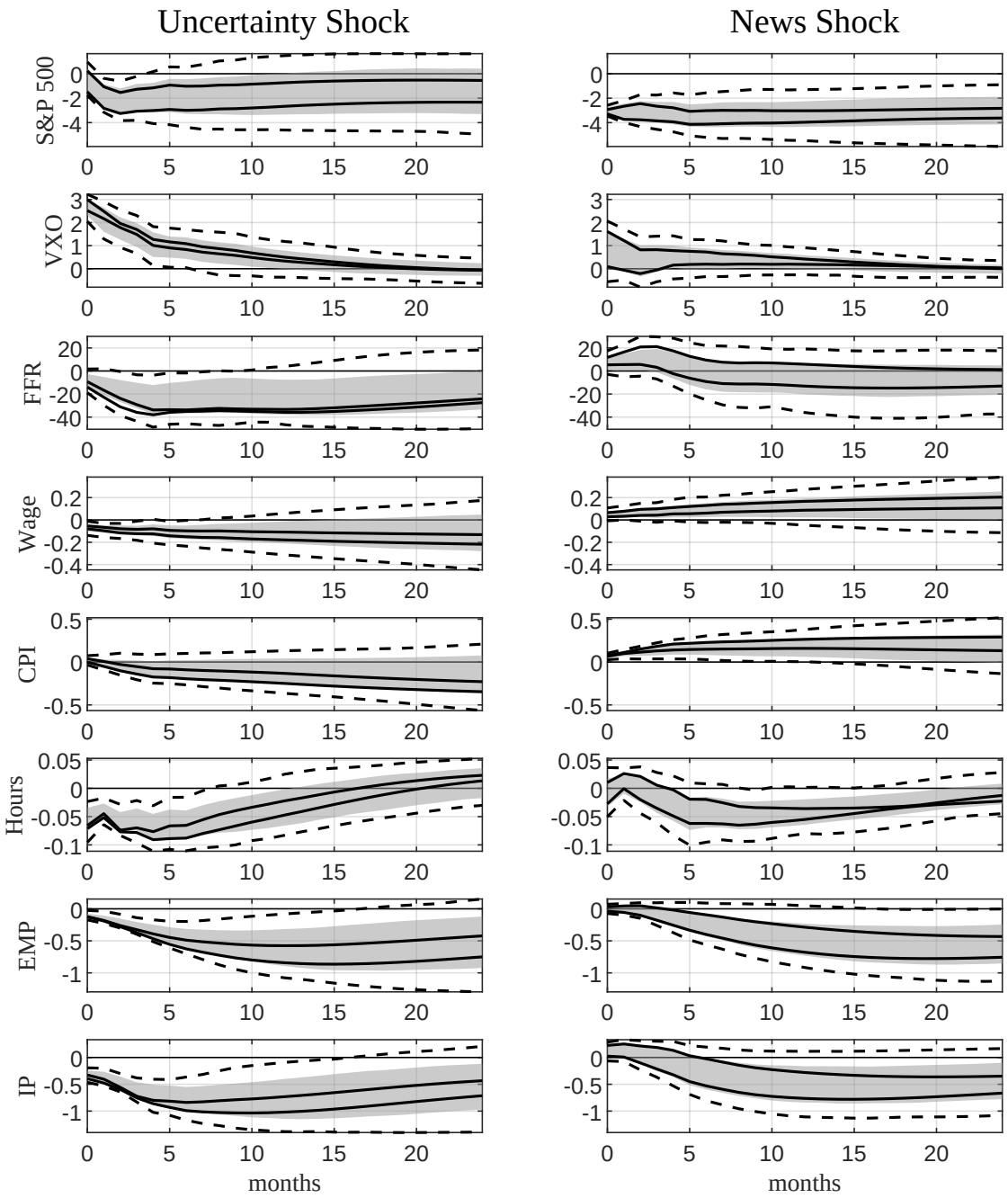
For illustrative purposes, Figure 3.11 shows only the marginal ECDFs of the FFR responses for ten randomly selected bootstrapped reduced form parameters. The ECDFs of the other two variables can be found in Figures 3.B.2 and 3.B.3 in Appendix 3.B. Looking at the grey lines and comparing them with the black dashed lines representing the uniform distribution, one can see that the marginal distribution of the impulse response draws are often rather close to a uniform distribution. Yet, for the later horizons the ECDFs diverge a bit more from the black dashed lines. The picture is similar for the other variables. Thus, a large relative Hausdorff distance is unlikely to be a problem, and if it is, then only for later horizons of the impulse responses. This is potentially also reflected in the decreasing coverage as h increases found in the simulations presented in section 3.3. However, the extent of the drop in the coverages is not large enough to conclude that their is a concerning problem with large Hausdorff differences.

3.4.2. Identifying an Uncertainty Shock

The second empirical illustration replicates the results of Piffer and Podstawski (2018). They identify an uncertainty and a news shock with a proxy VAR approach. They use two proxies that are not necessarily related only to their target shock, i.e. the uncertainty shock proxy might be also related to the news shock to some extent and vice versa. In order to disentangle the shocks in the proxy VAR they impose inequality restrictions on the correlation structure between the shocks and the proxies, leading to set-identification of the structural parameters. Specifically, they restrict the uncertainty shock proxy to have a higher correlation with the uncertainty shock than with the news shock and an analogous restriction is placed on the news shock proxy. The VAR system consists of S&P 500 returns, stock market volatility (VXO), the federal funds rate (FFR), wages, the consumer price index (CPI), average weekly working hours, employment (EMP) and industrial production (IP). The stock market returns, wages, the CPI, EMP and IP enter as log differences and subsequently these responses are cumulated and multiplied by a factor of 100.

The uncertainty shock proxy is based on changes in the gold price around exogenous events identified by the authors. The news shock proxy is constructed based on various news shocks from the literature. One difficulty is that while the uncertainty proxy is available at monthly frequency, the news shock proxy is only available at quarterly frequency. Piffer and Podstawski (2018) deal with this by aggregating the monthly reduced form residuals into quarterly frequency when computing the relevant proxy VAR quantities with the news shock proxy. In their ad hoc wild bootstrap

Figure 3.12: Impulse Responses to an Uncertainty and a News Shock



Notes: The solid black lines are the learned identified set and the dashed lines are the learned bands from the augmented learning procedure with $\delta = 0.01$ and $\epsilon = 0.05$. The grey areas are the bootstrap intervals from the original study by Piffer and Podstawski (2018).

scheme they also use these aggregated residuals for the respective parts of the proxy VAR. I follow their approach in this analysis and for more details I refer the reader to the paper by Piffer and Podstawski (2018).

The augmented learning framework is well equipped to handle this challenging identification setting. Yet, one has to fall back to the wild bootstrap. Due to the different frequencies of the proxies as well as the different availability of the sample periods of the base data and the instruments, the residual-based moving block bootstrap is not applicable. Thus, the wild bootstrap is used in the augmented learning procedure and based on the simulation results I choose $\epsilon = 0.05$ and $\delta = 0.01$ in order to obtain stable results. The probability that the true impulse responses of each variable are jointly within the learned (dashed) bands is at least 95% and this statement is true with a probability of at least 99%. The impulse responses are learned at impact and up to horizon $H = 24$.

Figure 3.12 depicts the learned bands (dashed lines) obtained from the augmented learning procedure, the learned identified set (solid lines) obtained with the baseline learning framework and the 95% solution bands (shaded areas) from the bootstrap approach by Piffer and Podstawski (2018). Note that here, unlike in the original paper by Piffer and Podstawski (2018), the responses corresponding to the transformed variables are reported as cumulative impulse responses. The results show that the learned bands with the joint interpretation still reveal the adverse effects of uncertainty shocks on the economy. One observes that industrial production decreases together with other real variables such as employment, the hours worked and wages in the short run. Furthermore, the results show a decline in stock market returns in conjunction with increased stock market volatility as measured by the VXO.

These results are revealed by the learned bands although they are wider than the intervals presented in the original study. Note that these shaded areas are not frequentist confidence intervals with a valid asymptotic interpretation but they are computed with a simplistic ad hoc bootstrap approach. This likely explains the large margin with which the learned bands exceed the shaded areas. Yet, this highlights that the original results of Piffer and Podstawski (2018) largely hold up to a more robust way of quantifying uncertainty in the structural parameters.

The responses show that uncertainty shocks have a negative impact on real activity, as indicated by the responses of IP and labor market related variables. Financial markets are also affected as uncertainty shocks lower stock market returns. The decline in the federal funds rate is likely to counteract a potential decline in investment. Yet, all of these effects appear to be rather short-lived. The results for the news shock

are less clear, however, they were also less sharp in the original study. Hence, joint bands from the augmented learning approach are expected to suggest no or minor effects in this case.

3.5. Conclusion

In conclusion, I provide a new method for uncertainty quantification in set-identified structural VARs that accounts for both model and estimation uncertainty. The augmented learning procedure serves as a general and computationally feasible alternative to existing frequentist or Bayesian approaches which are often limited in their applicability or are computationally costly. Furthermore, the learned bands allow for joint statements about the uncertainty involved in the identification of the structural parameters.

The framework augments and adapts the baseline learning algorithm by Montiel Olea and Nesbit (2021) which is used to learn parameter regions with a preceding bootstrap in order to account for estimation uncertainty. Statements about the uncertainty are derived from the learning guarantees which give the sample complexity required to learn the parameter region with a certain precision. I show that, under certain assumptions, the learning guarantees allow for a clear asymptotic interpretation of the learned bands.

The finite sample properties of the augmented learning framework are investigated in a Monte Carlo simulation study. The results show that from a point-wise perspective, coverages often exceed nominal levels typically chosen in empirical applications. Yet, since the bands quantify the uncertainty jointly, point-wise overcoverage is not surprising. In terms of joint coverages, the learning approach can achieve coverages close to typically used nominal levels, with the proper setting of the hyperparameters that govern the learning precision.

A stylized comparison with already existing methods to conduct inference in set-identified structural VARs shows that the learned bands are also informative. The bands are similar in width compared to the frequentist joint confidence bands of Granziera et al. (2018), but consistent with the finding of excess point-wise coverage in the simulations, they are wider compared to point-wise approaches. However, a real empirical application also supports the informativeness of the learned bands. In the replication of the empirical results of Piffer and Podstawski (2018) the bands still reveal the detrimental effects of uncertainty shocks on the economy.

References

Amir-Ahmadi, P. and T. Drautzburg (2021). "Identification and Inference with Ranking Restrictions". *Quantitative Economics* 12. (1), 1–39.

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.

Baumeister, C. and J. D. Hamilton (2015). "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information". *Econometrica* 83. (5), 1963–1999.

Blumer, A., A. Ehrenfeucht, D. Haussler, and M. K. Warmuth (1989). "Learnability and the Vapnik-Chervonenkis Dimension". *J. ACM* 36. (4), 929–965.

Brüggemann, R., C. Jentsch, and C. Trenkler (2016). "Inference in VARs with Conditional Heteroskedasticity of Unknown Form". *Journal of Econometrics* 191. (1), 69–85.

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*. Vol. 1. Elsevier, 65–148.

Gafarov, B., M. Meier, and J. L. Montiel Olea (2016). "Projection Inference for Set-Identified SVARs". *Manuscript, Columbia University*.

— (2018). "Delta-Method Inference for a Class of Set-Identified SVARs". *Journal of Econometrics* 203. (2), 316–327.

Giacomini, R. and T. Kitagawa (2021). "Robust Bayesian Inference for Set-Identified Models". *Econometrica* 89. (4), 1519–1556.

Giacomini, R., T. Kitagawa, and M. Read (2021). *Identification and Inference Under Narrative Restrictions*. Papers 2102.06456. arXiv.org.

— (2022). "Robust Bayesian Inference in Proxy SVARs". *Journal of Econometrics* 228. (1), 107–126.

Granziera, E., H. R. Moon, and F. Schorfheide (2018). "Inference for VARs Identified with Sign Restrictions". *Quantitative Economics* 9. (3), 1087–1121.

Inoue, A. and L. Kilian (2022). "Joint Bayesian Inference about Impulse Responses in VAR Models". *Journal of Econometrics* 231. (2). Special Issue: The Econometrics of Macroeconomic and Financial Data, 457–476.

Jentsch, C. and K. G. Lunsford (2022). "Asymptotically Valid Bootstrap Inference for Proxy SVARs". *Journal of Business & Economic Statistics* 40. (4), 1876–1891.

Kaido, H., F. Molinari, and J. Stoye (2019). "Confidence Intervals for Projections of Partially Identified Parameters". *Econometrica* 87. (4), 1397–1432.

Ludvigson, S. C., S. Ma, and S. Ng (2021). "Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response?" *American Economic Journal: Macroeconomics* 13. (4), 369–410.

Montiel Olea, J. L. and J. Nesbit (2021). "(Machine) Learning Parameter Regions". *Journal of Econometrics* 222. (1), 716–744.

Piffer, M. and M. Podstawski (2018). "Identifying Uncertainty Shocks Using the Price of Gold". *The Economic Journal* 128. (616), 3266–3284.

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.

Uhlig, H. (2005). "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure". *Journal of Monetary Economics* 52. (2), 381–419.

Volpicella, A. (2022). "SVARs Identification Through Bounds on the Forecast Error Variance". *Journal of Business & Economic Statistics* 40. (3), 1291–1301.

Appendix 3.A Proofs and Remarks

Remark on Theorem 3.1

The full proof for Theorem 3.1 can be found in the Appendix by Montiel Olea and Nesbit (2021) in which it is labelled Theorem 3. Here, I merely present how the learning of the union of identified sets fits in their proof and I refer the reader to their Appendix for the details.

The key difference in this chapter is in the definition of the set which is learned. They learn the set for a fixed data realization, meaning for a fixed pair of reduced form parameters. In the augmented learning process multiple pairs of reduced form parameters are present, denoted by $(\Phi_n^{(H)}, \Sigma_n)$ for $n = 1, ..., N$, and the union of the sets $[\lambda(S_U)]$ is learned. Let $[\hat{\lambda}_M]$ denote the learned set and to shorten the notation let $\lambda_n(\theta) = \lambda(\theta, \Phi_n^{(H)}, \Sigma_n)$. Hence, $\lambda_n(\theta)$ simply refers to the function of θ which corresponds to the n-th draw of bootstrapped reduced form parameters.

The quantity which is bounded by the PAC-learnability statement is the misclassification probability, i.e. the probability that a draw from $[\lambda(S_U)]$ is misclassified based on $[\hat{\lambda}_M]$. Naturally, the misclassification probability for learning the union of the sets is defined differently than the misclassification probability for a fixed data realization as in Montiel Olea and Nesbit (2021). If $[\lambda(S_U)]$ is learned the misclassification probability for a fixed θ is:

$$\begin{aligned}\mathcal{L}([\hat{\lambda}_M]; [\lambda(S_U)], P) &= P(\mathbf{1}\{\exists n \text{ s.t. } \lambda_n(\theta) \in [\hat{\lambda}_M]\} \neq \mathbf{1}\{\exists n \text{ s.t. } \lambda_n(\theta) \in [\lambda(S_U)]\}) \\ &= P(\exists n \text{ s.t. } \lambda_n(\theta) \in [\hat{\lambda}_M] \quad \text{and} \quad \forall n \ \lambda_n(\theta) \notin [\lambda(S_U)]) \quad (3.A.1) \\ &+ P(\forall n \ \lambda_n(\theta) \notin [\hat{\lambda}_M] \quad \text{and} \quad \exists n \text{ s.t. } \lambda_n(\theta) \in [\lambda(S_U)]). \quad (3.A.2)\end{aligned}$$

By definition of $[\hat{\lambda}_M]$, if $\exists n$ s.t. $\lambda_n(\theta) \in [\hat{\lambda}_M]$ then it follow that $\exists n$ s.t. $\lambda_n(\theta) \in [\lambda(S_U)]$) because $[\hat{\lambda}_M] =$. Hence, the term (3.A.1) is equal to 0 and:

$$\mathcal{L}([\hat{\lambda}_M]; [\lambda(S_U)], P) = P(\exists n \text{ s.t. } \lambda_n(\theta) \in [\lambda(S_U)] \setminus [\hat{\lambda}_M]). \quad (3.A.3)$$

An analogous definition of the misclassification probability can be found in the proof by Montiel Olea and Nesbit (2021) for the case of a single fixed data realization. With the definition in (3.A.3) the proof for both upper bounds proceeds as in the Appendix of Montiel Olea and Nesbit (2021), yet now the set $\lambda(S_U)$ is learned and the labels for the learning are obtained as described in the main text of this chapter. Hence, in the proof $\lambda(S)$ is replaced by $\lambda(S_U)$ and whether $\lambda(\theta)$ is an element of a

certain set comes down to assessing whether the conditions hold for at least one or for all $\lambda_n(\theta)$ depending on the logical operator.

Proof of Proposition 3.1

Proof. By assumption 3.1.1 the reduced form parameters of the underlying VAR(p) process are estimated consistently. Parameters marked with a '*' denote the true DGP parameters. Hence, for $T \to \infty$:

$$\hat{\Phi}^{(H)} \xrightarrow{p} \Phi^{(H)*}.$$

The same holds for $T \to \infty$ and the estimate of the error covariance matrix Σ:

$$\hat{\Sigma} \xrightarrow{p} \Sigma^*.$$

Assumption 3.1.2 requires $\text{rank}(\Sigma) = k$, meaning that Σ is positive definite. Therefore, its Cholesky decomposition Σ_C is a continuous mapping, and thus, for $T \to \infty$, by the continuous mapping theorem:

$$\hat{\Sigma}_C \xrightarrow{p} \Sigma_C^*.$$

By Assumption 3.1.3 the bootstrap replicates the distribution of the consistent estimator consistently. Hence, for $T \to \infty$:

$$\hat{\Phi}_n^{(H)} \xrightarrow{p} \Phi^{(H)*}, \text{ for } n = 1, ..., N.$$

Similarly, for $T \to \infty$:

$$\hat{\Sigma}_n \xrightarrow{p} \Sigma^*, \quad \text{for } n = 1, ..., N,$$

and thus,

$$\hat{\Sigma}_{C,n} \xrightarrow{p} \Sigma_C^*, \quad \text{for } n = 1, ..., N.$$

Let θ be fixed at θ^* and denote the parameters of Q and Q^*, respectively, where Q^* denotes the rotation matrix which yields the true impact matrix $B^* = \Sigma_C^* Q^*$. Then,

$$\hat{\eta}_{ijh,n} = e_i' \hat{\Phi}_{h,n} \hat{\Sigma}_{C,n} Q^* e_j,$$

is, by the continuous mapping theorem, estimated consistently as well, and thus:

$$\hat{\eta}_{ijh,n} \xrightarrow{p} \eta^*_{ijh}, \quad \text{for } n = 1, ..., N.$$

The same holds true for the FEV decomposition

$$\hat{\Omega}_{ij,n}(h) = \frac{e_i'(\sum_{m=0}^{h-1} \hat{\Phi}_{m,n}\hat{\Sigma}_{C,n}Q^* e_j e_j' Q^{*\prime}\hat{\Sigma}'_{C,n}\hat{\Phi}'_{m,n})e_i}{e_i'(\sum_{m=0}^{h-1} \hat{\Phi}_{m,n}\hat{\Sigma}_n\hat{\Phi}'_{m,n})e_i},$$

and hence

$$\hat{\Omega}_{ij,n}(h) \xrightarrow{p} \Omega^*_{ij}(h), \quad \text{for } n = 1, ..., N.$$

Thus, for $\lambda(\theta)$ being the structural impulse responses or the FEV decomposition:

$$\lambda(\theta^*, \hat{\Phi}_n^{(H)}, \hat{\Sigma}_n) \xrightarrow{p} \lambda(\theta^*, \Phi^{(H)*}, \Sigma^*),$$

where $\lambda(\theta^*, \Phi^{(H)*}, \Sigma^*)$ denotes the true structural impulse responses and the true FEV decomposition, respectively. By Assumption 3.1.5 the identification restrictions are satisfied for $\lambda(\theta^*, \Phi^{(H)*}, \Sigma^*)$, and thus the label is $l_U(\theta^*, \Phi^{(H)*}, \Sigma^*) = 1$, meaning that $\lambda(\theta^*, \Phi^{(H)*}, \Sigma^*) \in \lambda(S_U)$.

As a consequence, for $T \to \infty$ one gets $P(\lambda(\theta^*, \Phi^{(H)*}, \Sigma^*) \in \lambda(S_U)) \to 1$, and thus in the limit the true structural parameters are in the set that is learned. □

Proof of Corollary 3.1

Proof. By Proposition 3.1, for $T \to \infty$ the true structural parameters are in $\lambda(S_U)$. When learning $[\hat{\lambda}_M]$ according to (3.2.1), the PAC-learnability statement holds:

$$P(\mathcal{L}([\hat{\lambda}_M]; [\lambda(S_U)], W) < \epsilon) \geq 1 - \delta. \tag{3.A.4}$$

Suppose that after learning, an additional draw with probability distribution W from inside $\lambda(S_U)$ is made. Without loss of generality, let the draw be the true rotation parameters θ^*. According to (3.2.2) the probability to misclassify the draw is less than ϵ with probability greater or equal than $1 - \delta$. Hence, for $T \to \infty$, the probability that the true structural parameters is not in the learned set $[\lambda_M]$ is less than ϵ with probability greater or equal than $1 - \delta$. Hence,

$$P(P(\lambda(\theta^*, \Phi^{(H)*}, \Sigma^*) \in [\hat{\lambda}_M]) \geq 1 - \epsilon) \geq 1 - \delta.$$

□

Appendix 3.B Additional Simulation and Empirical Results

DGP for the Simulations

The DGP parameters for the simulation exercise come from the estimated three-variate VAR which includes the CPI, GDP and the federal funds rate. CPI and GDP enter the model as logged differences. The model is estimated with $p = 4$ lags, and thus the simulation DGP is:

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + A_3 y_{t-3} + A_4 y_{t-4} + u_t,$$

with the coefficient matrices being

$$A_1 = \begin{pmatrix} 0.2438 & 0.0502 & 0.0016 \\ 0.0989 & 0.2985 & 0.0006 \\ 2.2592 & 42.4630 & 0.7383 \end{pmatrix}, \; A_2 = \begin{pmatrix} -0.1049 & -0.0001 & -0.0007 \\ -0.1107 & 0.2876 & -0.0027 \\ -8.7941 & 21.1489 & -0.0815 \end{pmatrix},$$

$$A_3 = \begin{pmatrix} 0.1793 & -0.0459 & 0.0002 \\ 0.0323 & 0.0178 & 0.0013 \\ 16.0129 & -3.9416 & 0.4496 \end{pmatrix}, \; A_4 = \begin{pmatrix} 0.2872 & 0.0504 & -0.0008 \\ 0.0430 & 0.1303 & 0.0007 \\ -6.0571 & 1.4894 & -0.2156 \end{pmatrix},$$

the intercept is

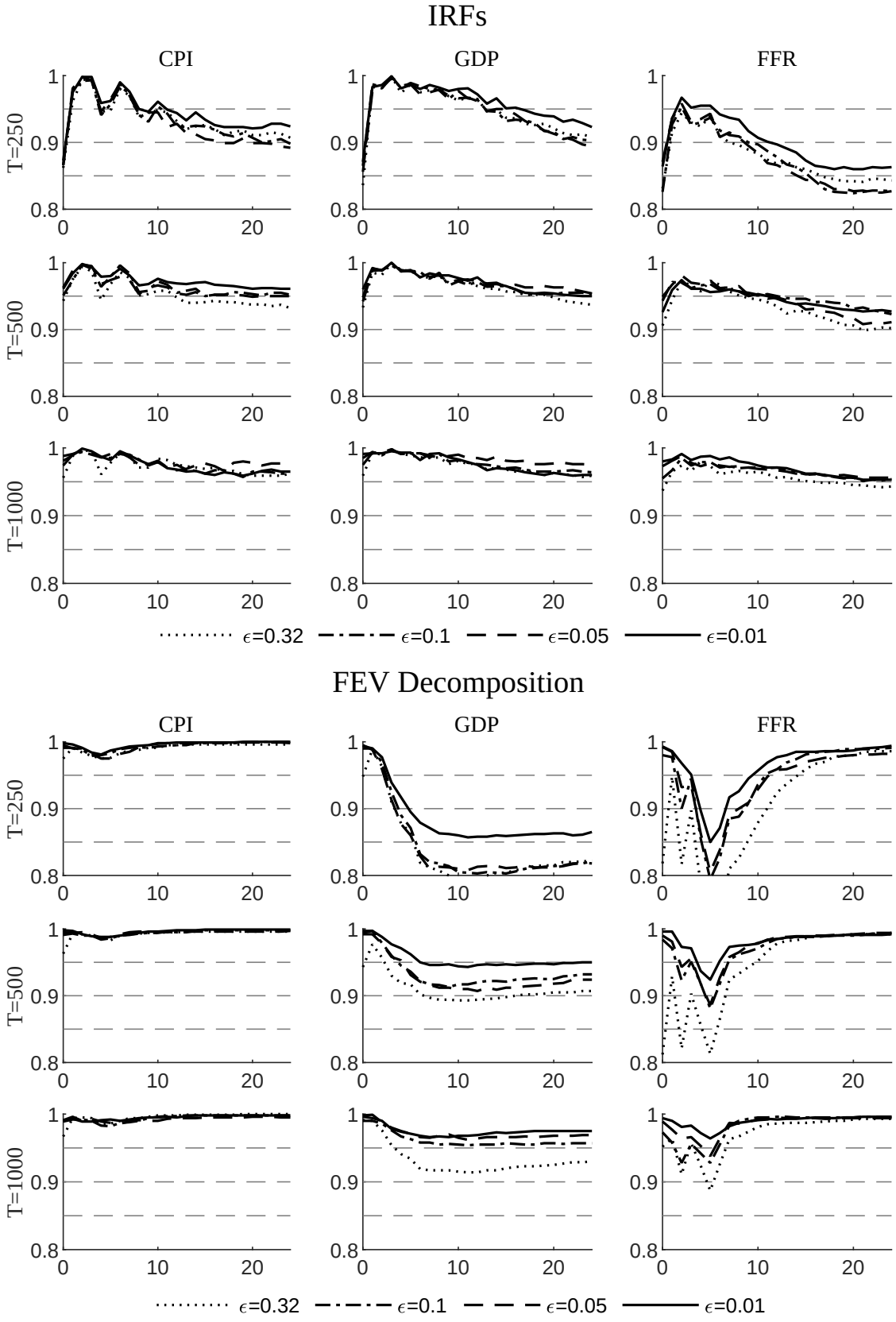
$$c = \begin{pmatrix} 0.0011 \\ 0.0037 \\ -0.4401 \end{pmatrix},$$

and the reduced form innovations are drawn from a multivariate normal distribution with the expected value being a vector of zeros and the covariance matrix

$$\Sigma = \begin{pmatrix} 0.0024 & 0.0009 & 0.0484 \\ 0.0009 & 0.0053 & 0.1838 \\ 0.0484 & 0.1838 & 109.6528 \end{pmatrix} \cdot 10^{-2}.$$

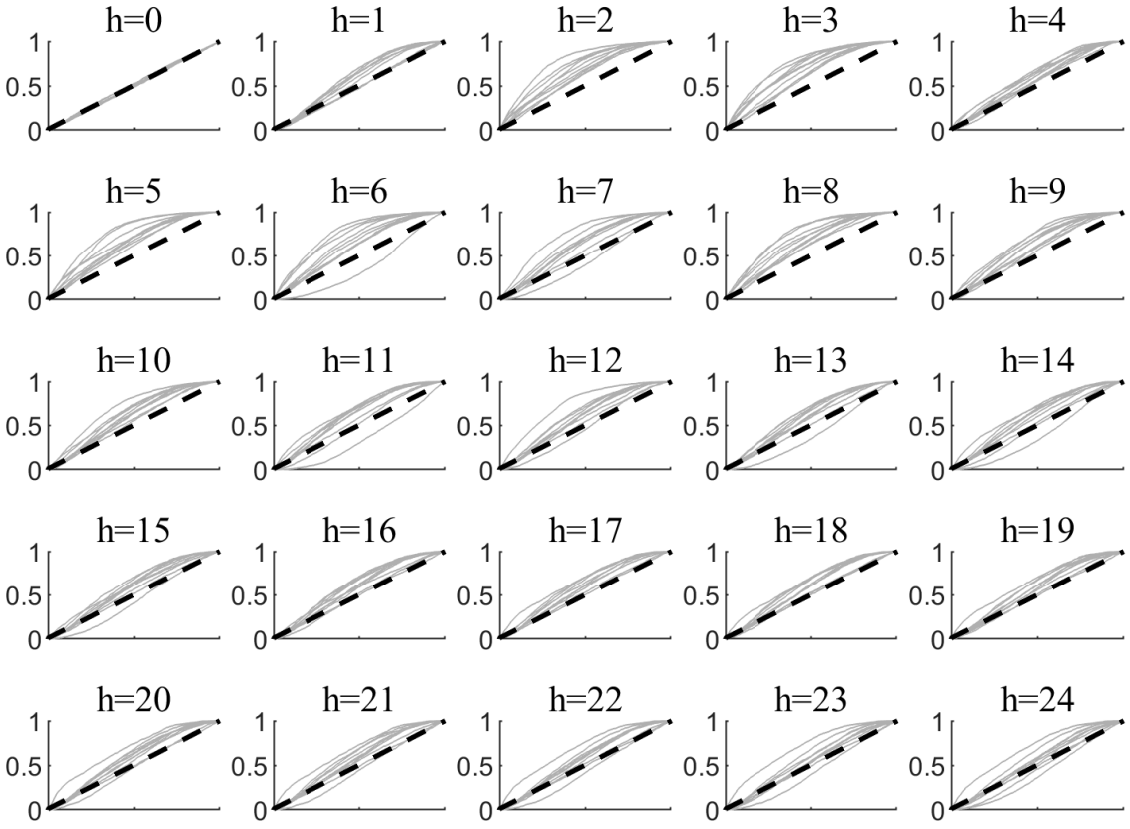
Note that the matrices here are rounded to the fourth digit while in the simulation study the unrounded estimates are used for the DGP.

Figure 3.B.1: Coverages with 500 Residual-Based Moving Block Bootstrap Iterations



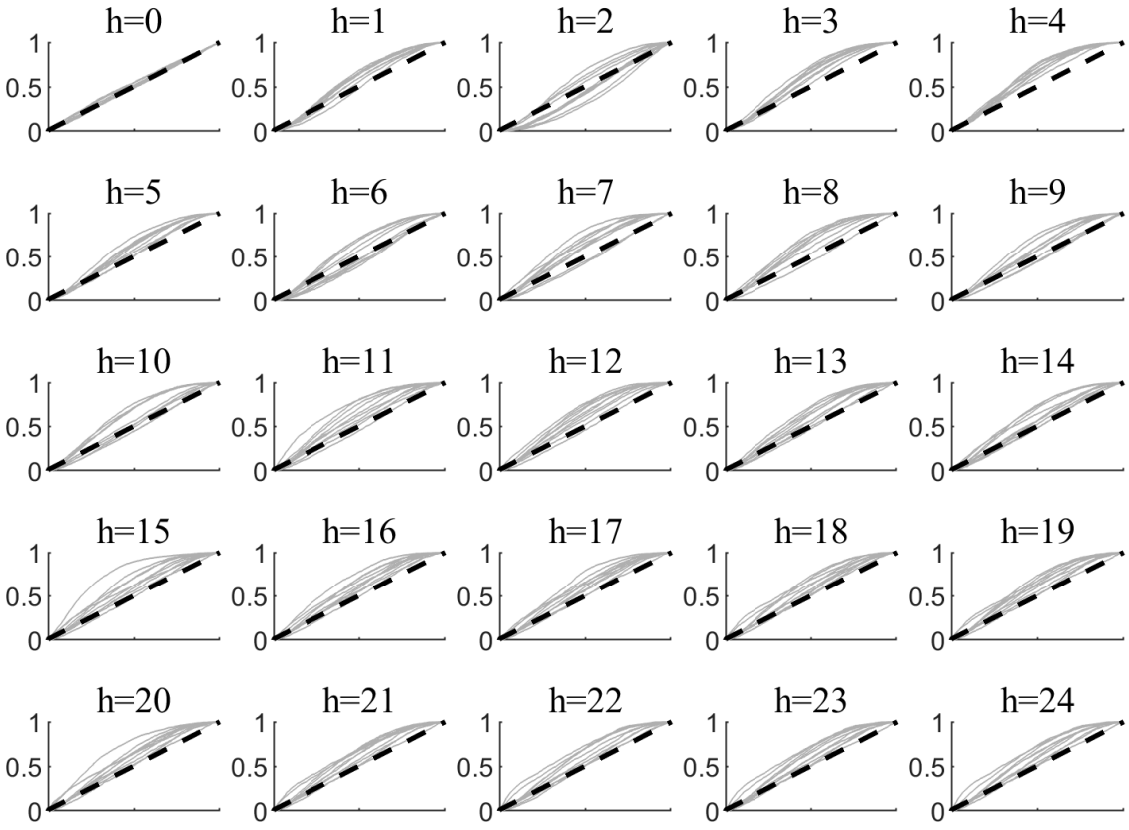
Notes: The upper panel depicts point-wise coverages for the IRFs and the lower panel for the FEV decomposition for the augmented learning procedure based on 500 residual-based moving block bootstrap iterations.

Figure 3.B.2: ECDFs for the GDP Responses



Notes: The grey lines depict in each individual graph depict the marginal ECDF of the impulse response draws for ten randomly selected bootstrap parameters at the respective horizon. The dashed black line represents the CDF of the uniform distribution as a reference.

Figure 3.B.3: ECDFs for the CPI Responses



Notes: The grey lines depict in each individual graph depict the marginal ECDF of the impulse response draws for ten randomly selected bootstrap parameters at the respective horizon. The dashed black line represents the CDF of the uniform distribution as a reference.

Complete References

Amir-Ahmadi, P. and T. Drautzburg (2021). "Identification and Inference with Ranking Restrictions". *Quantitative Economics* 12. (1), 1–39.

An, S. and F. Schorfheide (2007). "Bayesian Analysis of DSGE Models". *Econometric Reviews* 26. (2-4), 113–172.

Angelini, G., E. Bacchiocchi, G. Caggiano, and L. Fanelli (2019). "Uncertainty Across Volatility Regimes". *Journal of Applied Econometrics* 34. (3), 437–455.

Angelini, G., G. Caggiano, E. Castelnuovo, and L. Fanelli (2023). "Are Fiscal Multipliers Estimated with Proxy-SVARs Robust?" *Oxford Bulletin of Economics and Statistics* 85. (1), 95–122.

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.

Baker, S. R., N. Bloom, and S. J. Davis (2016). "Measuring Economic Policy Uncertainty". *The Quarterly Journal of Economics* 131. (4), 1593–1636.

Barakchian, S. M. and C. Crowe (2013). "Monetary Policy Matters: Evidence from New Shocks Data". *Journal of Monetary Economics* 60. (8), 950–966.

Barsky, R. B. and E. R. Sims (2011). "News Shocks and Business Cycles". *Journal of Monetary Economics* 58. (3), 273–289.

Baumeister, C. and L. Benati (2013). "Unconventional Monetary Policy and the Great Recession: Estimating the Macroeconomic Effects of a Spread Compression at the Zero Lower Bound". *International Journal of Central Banking* 9, 165–212.

Baumeister, C. and J. D. Hamilton (2015). "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information". *Econometrica* 83. (5), 1963–1999.

Ben Zeev, N. and E. Pappa (2017). "Chronicle of a War Foretold: The Macroeconomic Effects of Anticipated Defence Spending Shocks". *The Economic Journal* 127. (603), 1568–1597.

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.

Bloom, N. (2014). "Fluctuations in Uncertainty". *Journal of Economic Perspectives* 28. (2), 153–76.

Blumer, A., A. Ehrenfeucht, D. Haussler, and M. K. Warmuth (1989). "Learnability and the Vapnik-Chervonenkis Dimension". *J. ACM* 36. (4), 929–965.

Boivin, J. and M. P. Giannoni (2006). "Has Monetary Policy Become More Effective?" *The Review of Economics and Statistics* 88. (3), 445–462.

Braun, R. and R. Brüggemann (2022). "Identication of SVAR Models by Combining Sign Restrictions with External Instruments". *Journal of Business & Economic Statistics* 41. (4), 1077–1089.

Brüggemann, R., C. Jentsch, and C. Trenkler (2016). "Inference in VARs with Conditional Heteroskedasticity of Unknown Form". *Journal of Econometrics* 191. (1), 69–85.

Caldara, D. and E. Herbst (2019). "Monetary Policy, Real Activity, and Credit Spreads: Evidence from Bayesian Proxy SVARs". *American Economic Journal: Macroeconomics* 11. (1), 157–192.

Carriero, A., T. E. Clark, and M. Marcellino (2018). "Measuring Uncertainty and Its Impact on the Economy". *Review of Economics and Statistics* 100. (5), 799–815.

Carriero, A. and A. Volpicella (2024). "Max Share Identification of Multiple Shocks: An Application to Uncertainty and Financial Conditions". *Journal of Business & Economic Statistics* 0, 1–13.

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*. Vol. 1. Elsevier, 65–148.

Clarida, R., J. Gali, and M. Gertler (2000). "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory". *The Quarterly Journal of Economics* 115. (1), 147–180.

Dieppe, A. M., F. Neville, and G. J. Kindberg Hanlon (2019). "New Approaches to the Identification of Low-Frequency Drivers : An Application to Technology Shocks". *The World Bank - Policy Research Working Paper 9047*.

Faust, J. (1998). "The Robustness of Identified VAR Conclusions about Money". *Carnegie-Rochester Conference Series on Public Policy* 49, 207–244.

Francis, N. and G. Kindberg-Hanlon (2022). "Signing Out Confounding Shocks in Variance-Maximizing Identification Methods". *AEA Papers and Proceedings* 112, 476–80.

Francis, N., M. T. Owyang, J. E. Roush, and R. DiCecio (2014). "A Flexible Finite-Horizon Alternative to Long-Run Restrictions with an Application to Technology Shocks". *The Review of Economics and Statistics* 96. (4), 638–647.

Gafarov, B., M. Meier, and J. L. Montiel Olea (2016). "Projection Inference for Set-Identified SVARs". *Manuscript, Columbia University*.

Gafarov, B., M. Meier, and J. L. Montiel Olea (2018). "Delta-Method Inference for a Class of Set-Identified SVARs". *Journal of Econometrics* 203. (2), 316–327.

Gertler, M. and P. Karadi (2015). "Monetary Policy Surprises, Credit Costs, and Economic Activity". *American Economic Journal: Macroeconomics* 7. (1), 44–76.

Giacomini, R. (2013). "The Relationship between DSGE and VAR Models". In: *Advances in Econometrics*. Vol. 32, 1–25.

Giacomini, R. and T. Kitagawa (2021). "Robust Bayesian Inference for Set-Identified Models". *Econometrica* 89. (4), 1519–1556.

Giacomini, R., T. Kitagawa, and M. Read (2021). *Identification and Inference Under Narrative Restrictions*. Papers 2102.06456. arXiv.org.

— (2022). "Robust Bayesian Inference in Proxy SVARs". *Journal of Econometrics* 228. (1), 107–126.

Goodfriend, M. and R. G. King (2005). "The Incredible Volcker Disinflation". *Journal of Monetary Economics* 52. (5), 981–1015.

Gouriéroux, 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.

Granziera, E., H. R. Moon, and F. Schorfheide (2018). "Inference for VARs Identified with Sign Restrictions". *Quantitative Economics* 9. (3), 1087–1121.

Gürkaynak, R. S., B. Sack, and E. Swanson (2005). "Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements". *International Journal of Central Banking* 1. (1), 55–93.

Hall, A. R. (2005). *Generalized Method of Moments*. Oxford University Press.

Hansen, L. P. (1982). "Large Sample Properties of Generalized Method of Moments Estimators". *Econometrica* 50. (4), 1029–1054.

Inoue, A. and L. Kilian (2022). "Joint Bayesian Inference about Impulse Responses in VAR Models". *Journal of Econometrics* 231. (2). Special Issue: The Econometrics of Macroeconomic and Financial Data, 457–476.

Jentsch, C. and K. G. Lunsford (2022). "Asymptotically Valid Bootstrap Inference for Proxy SVARs". *Journal of Business & Economic Statistics* 40. (4), 1876–1891.

Jurado, K., S. C. Ludvigson, and S. Ng (2015). "Measuring Uncertainty". *American Economic Review* 105. (3), 1177–1216.

Kaido, H., F. Molinari, and J. Stoye (2019). "Confidence Intervals for Projections of Partially Identified Parameters". *Econometrica* 87. (4), 1397–1432.

Kapetanios, G., H. Mumtaz, I. Stevens, and K. Theodoridis (2012). "Assessing the Economy-wide Effects of Quantitative Easing". *The Economic Journal* 122. (564), 316–347.

Keweloh, S. A. (2021). "A Generalized Method of Moments Estimator for Structural Vector Autoregressions Based on Higher Moments". *Journal of Business & Economic Statistics* 39. (3), 772–782.

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

Komunjer, I. and S. Ng (2011). "Dynamic Identification of Dynamic Stochastic General Equilibrium Models". *Econometrica* 79. (6), 1995–2032.

Lakdawala, A. (2019). "Decomposing the Effects of Monetary Policy Using an External Instruments SVAR". *Journal of Applied Econometrics* 34. (6), 934–950.

Lanne, M. and J. Luoto (2021). "GMM Estimation of Non-Gaussian Structural Vector Autoregression". *Journal of Business & Economic Statistics* 39. (1), 69–81.

Leeper, E. M., C. A. Sims, T. Zha, R. E. Hall, and B. S. Bernanke (1996). "What Does Monetary Policy Do?" *Brookings Papers on Economic Activity* 1996. (2), 1–78.

Ludvigson, S. C., S. Ma, and S. Ng (2021). "Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response?" *American Economic Journal: Macroeconomics* 13. (4), 369–410.

Lunsford, K. G. (2015). "Identifying Structural VARs with a Proxy Variable and a Test for a Weak Proxy". *Federal Reserve Bank of Cleveland, Working Paper No. 15-28*.

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), 1212–47.

Montiel Olea, J. L. and J. Nesbit (2021). "(Machine) Learning Parameter Regions". *Journal of Econometrics* 222. (1), 716–744.

Piffer, M. and M. Podstawski (2018). "Identifying Uncertainty Shocks Using the Price of Gold". *The Economic Journal* 128. (616), 3266–3284.

Ramey, V. (2016). "Macroeconomic Shocks and Their Propagation". In: *Handbook of Macroeconomics*. Vol. 2. Elsevier, 71–162.

Romer, C. D. and D. H. Romer (2004). "A New Measure of Monetary Shocks: Derivation and Implications". *American Economic Review* 94. (4), 1055–1084.

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.

Sargan, J. D. (1958). "The Estimation of Economic Relationships using Instrumental Variables". *Econometrica* 26. (3), 393–415.

Shin, M. and M. Zhong (2020). "A New Approach to Identifying the Real Effects of Uncertainty Shocks". *Journal of Business & Economic Statistics* 38. (2), 367–379.

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

Sims, C. A. and T. Zha (2006). "Were There Regime Switches in U.S. Monetary Policy?" *American Economic Review* 96. (1), 54–81.

Stock, J. H. and M. Watson (2012). "Disentangling the Channels of the 2007-2009 Recession". *Brookings Papers on Economic Activity* Spring 2012, 81–135.

Stock, J. H., J. H. Wright, and M. Yogo (2002). "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments". *Journal of Business & Economic Statistics* 20. (4), 518–529.

Stock, J. H. and M. Yogo (2005). "Testing for Weak Instruments in Linear IV Regression". In: *Identification and Inference for Econometric Models*. Ed. by Andrews, D. W. New York: Cambridge University Press, 80–108.

Swanson, E. T. (2021). "Measuring the Effects of Federal Reserve Forward Guidance and Asset Purchases on Financial Markets". *Journal of Monetary Economics* 118. (C), 32–53.

Thornton, D. L. (2006). "When Did the FOMC Begin Targeting the Federal Funds Rate? What the Verbatim Transcripts Tell Us". *Journal of Money, Credit and Banking* 38. (8), 2039–2071.

Uhlig, H. (2004a). "Do Technology Shocks Lead to a Fall in Total Hours Worked?" *Journal of the European Economic Association* 2. (2-3), 361–371.

— (2004b). "What Moves GNP?" *Econometric Society 2004 North American Winter Meetings* (No. 636).

— (2005). "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure". *Journal of Monetary Economics* 52. (2), 381–419.

Volpicella, A. (2022). "SVARs Identification Through Bounds on the Forecast Error Variance". *Journal of Business & Economic Statistics* 40. (3), 1291–1301.

Eigenabgrenzung

Das erste Kapitel, *Identifying Proxy VARs with Restrictions on the Forecast Error Variance*, und das dritte Kapitel, *Learning Structural Parameter Sets Accounting for Model and Estimation Uncertainty*, habe ich ohne Hilfe Dritter und nur mit den angegebenen Hilfsmitteln erstellt.

Das zweite Kapitel, *Financial and Real Uncertainty Shocks: Insights from Synthetic Proxy Variables* ist in Zusammenarbeit mit Sascha Keweloh, Postdoktorand an der TU Dortmund, entstanden. Mein Anteil an der Erstellung des Kapitels beträgt 60%.