

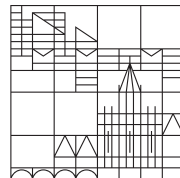
Three Essays on Identification in Structural Vector Autoregressive Models

Dissertation

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

vorgelegt von
Robin Braun

Universität
Konstanz



Sektion Politik – Recht – Wirtschaft
Fachbereich Wirtschaftswissenschaften
Konstanz, 2019

Tag der mündlichen Prüfung: 11. Juli 2019

1. Referent: Prof. Dr. Ralf Brüggemann
2. Referent: Prof. Dr. Winfried Pohlmeier
3. Referent: Jr.-Prof. Dr. Lyudmila Grigoryeva

Acknowledgments

This thesis would not have been possible without the support of many people, which I would like to acknowledge at this point.

First, I thank my supervisor Prof. Ralf Brüggemann for excellent academic guidance involving hours of discussions of my research, comprehensive proofreading and arranging financial support, to name just a few things.¹ I truly enjoyed the open and fun working atmosphere at the Chair of Statistics and Econometrics. I am also grateful to my second supervisor, Prof. Winfried Pohlmeier for his helpful feedback in seminars and for enlightening conversations. Ultimately, it was due to his excellent courses and seminars I got interested into econometrics. I also thank Prof. Lyudmila Grigoryeva for joining my supervision committee, Prof. Hedibert Lopes for hosting me at INSPER in São Paulo, and Verena Kretz for excellent administrative support.

At the University of Konstanz, I was privileged to work with a range of inspiring people, many of whom I consider friends today. A non-exhaustive list is given by Anna, Tamara, Moritz, Ruchira and Maurizio S. from the GSDS, Julie, Maurizio D. and Dominik from the Chair and Audronė, Ekaterina and Phillip from F3.

Of course, I am also deeply grateful to my families as well. To my father, Stefan Braun, my mother, Wiltrud Zeuner-Braun and to Doris Murillo for all their efforts in my education. To my sister Isabel and my brother Felix with Tanja and my nephew Phillip for wonderful moments in family. Finally, I attribute my last acknowledgment to Silvana for her love and support, and the joy she brought over the last years.

¹Financial support by the Graduate School of Decision Sciences, the German Academic Exchange Service and the German Science Foundation (grant number: BR 2941/3-1) is gratefully acknowledged.

Table of Contents

Summary	1
Zusammenfassung	4
1 Identification of SVAR Models by Combining Sign Restrictions with External Instruments	7
1.1 Introduction	8
1.2 Combining sign restrictions with external instruments	11
1.2.1 Proxy augmented SVAR model	12
1.2.2 Identifying multiple shocks with a combination of sign restrictions and instrumental variables	13
1.2.3 Identifying a single shock using information from sign restrictions and an external proxy	15
1.2.4 Bayesian estimation and inference	17
1.3 Empirical applications	21
1.3.1 The effects of oil price shocks	21
1.3.2 The effects of monetary policy shocks	25
1.4 Conclusion	30
References	32
Appendix 1.A Gibbs sampler for missing data	36
Appendix 1.B Data	37
1.B.1 Data for the oil price shock example	37
1.B.2 Data for the monetary policy shock example	37
Appendix 1.C Additional figures	39
2 Identification of Structural Vector Autoregressions by Stochastic Volatility	40
2.1 Introduction	41
2.2 Identification of SVARs by stochastic volatility	44
2.3 Maximum Likelihood estimation	48
2.3.1 Evaluation of the likelihood	48
2.3.2 EM Algorithm	51
2.3.3 Properties of the estimator	54

2.3.4	Testing for identification	56
2.4	Monte Carlo study	58
2.5	Interdependence between monetary policy and stock markets	61
2.5.1	Model and identifying constraints	62
2.5.2	Statistical analysis	64
2.6	Conclusion	69
	References	71
	Appendix 2.A Derivations and proofs	76
2.A.1	Autocovariance function of the second moments	76
2.A.2	Proof of proposition 1	77
2.A.3	Proof of corollary 1	78
2.A.4	Proof of corollary 2	78
	Appendix 2.B Estimation	79
2.B.1	Importance density	79
2.B.2	Newton Raphson method	80
2.B.3	EM algorithm	80
2.B.4	Derivatives complete data log-likelihood	86
2.B.5	Inference on structural impulse response functions	88
	Appendix 2.C Data and complementary results	89
2.C.1	Complementary results	91
3	The Importance of Supply and Demand for Oil Prices: Evidence from a SVAR Identified by non-Gaussianity	92
3.1	Introduction	93
3.2	Methodology	97
3.2.1	Constrained Dirichlet process mixture models	98
3.2.2	Bayesian inference	101
3.2.3	Assessing identification	103
3.3	Importance of supply and demand channels for oil price fluctuations	106
3.3.1	Model specification and identification analysis	107
3.3.2	Identifying supply and demand components in the non-Gaussian model	108
3.3.3	Incorporating prior knowledge about oil price elasticities	114
3.3.4	Structural analysis	116
3.4	Conclusion	125
	References	126
	Appendix 3.A Bayesian inference	130
3.A.1	Markov chain Monte Carlo algorithm	130
3.A.2	Dealing with label switching	134
	Appendix 3.B Time series plots	136
	Complete References	137
	Eigenabgrenzung	147

List of Tables

2.1	Cumulated MSEs at horizon $h = 5$	60
2.2	Model selection by information criteria	65
2.3	Tests of identification	67
2.4	Tests for overidentifying restrictions (EM-1)	68
2.5	Mixture approximation $\log \chi^2_{(1)}$ distribution (Kim et al.; 1998)	86
2.6	Tests on standardized structural shocks	91
2.7	Tests for overidentifying restrictions (EM-2)	91
3.1	Illustration of identification analysis based on higher moments	107
3.2	Identification analysis for the oil market SVAR based on higher moments of the structural shocks	108
3.3	Posterior probabilities to satisfy sign restrictions of Kilian and Murphy (2012)	110
3.4	Posterior probabilities to satisfy narrative restrictions of Antolín-Díaz and Rubio-Ramírez (2018)	113
3.5	Posterior probabilities to satisfy sign restrictions of Kilian and Murphy (2012), combined identification approach	117
3.6	Prior sensitivity analysis	124

List of Figures

1.1	Impulse responses in the oil market SVAR	24
1.2	Impulse responses in the monetary policy SVAR	28
1.3	Time series plots of oil market variables	37
1.4	Time series plots of monetary policy variables	38
1.5	Impulse responses in the monetary policy SVAR, alternative identifying restrictions	39
2.1	Standardized residuals	66
2.2	Impulse responses monetary policy shock	69
2.3	Time series plots	90
3.1	Identifying oil demand and supply curves by non-Gaussianity	94
3.2	Illustration of identification analysis based on predictive densities	106
3.3	Identification analysis for the oil market SVAR based on predictive densities of the structural shocks	109
3.4	Comparison of the first statistical shock with oil supply shocks from the literature	114
3.5	Oil supply elasticities: prior vs posterior distribution	115
3.6	Impulse responses	121
3.7	Historical decompositions	122
3.8	Forecast error variance decompositions	123
3.9	Time series plots	136

Summary

In order to make well-founded decisions, policy makers need precise point and interval forecasts of economic variables. The vector autoregressive (VAR) model is a popular tool for this purpose, predicting joint dynamics of multiple time series based on linear functions of past observations. However, VARs are reduced form models and therefore can only be used for certain descriptive purposes or for unconditional forecasting of economic variables. For predictions conditional on a policy intervention, a structural model is required which identifies similar policy events in the past and measures their effects on the economy. Furthermore, only through the lens of a structural model, decision makers can interpret current fluctuations observed in the data from an economic point of view. A sound understanding of current driving forces is a prerequisite to design effective policy responses. Pioneered by [Sims \(1980\)](#), non-sample information can be used to impose identifying restrictions in the VAR context which lead to what is known as the structural VAR (SVAR) model. Nowadays, it has become one of the most popular models in empirical macroeconomics for structural analysis.

This thesis is concerned with the development of novel econometric techniques to identify SVAR models. Technically, this requires decomposing VAR forecast errors into uncorrelated components which can be interpreted as primitive exogenous driving forces, also known as structural shocks. Prominent examples of such forces are fiscal policy shocks, monetary policy shocks, oil price shocks or uncertainty shocks. Given that there exist many decompositions which yield the same reduced form dynamics, identifying restrictions are required to get meaningful results. Among the most popular approaches are the use of short- or long run restrictions on the effects of structural shocks, imposed either as exclusion or sign restrictions. Furthermore, instrumental variables estimation or distributional assumptions might be exploited for identification of structural shocks in the model. The textbook of [Kilian and Lütkepohl \(2017\)](#) provides an excellent overview on the current state in this literature.

In Chapter 1, which is joint work with Ralf Brüggemann, we contribute to the literature proposing to identify SVAR models by combining sign restrictions with external instruments.

Relying on narrative evidence and extraneous information, “proxy” variables are frequently constructed by macroeconomists to function as instrumental variable for certain structural shocks, e.g. monetary policy shocks or fiscal policy shocks. In this chapter, we develop a modeling framework which allows to identify structural shocks in two different ways. The first possibility we consider is identifying multiple shocks using either sign restrictions or an external instrument approach, always ensuring that all shocks are orthogonal. In the second case, we show how a combination of both approaches can be exploited to identify a single shock. In this scenario, the exogeneity assumption on the proxies can be dropped. Instead, the information in these variables is used to reduce the set of models identified by sign restrictions, by discarding all those models that imply structural shocks with no close relation to the external proxies. For both cases, we provide algorithms to conduct full Bayesian inference reflecting model and estimation uncertainty. We illustrate the usefulness of our method within two empirical applications. First, we identify supply and demand shocks as drivers of the global crude oil market. In particular, a supply shock is identified by instrumental variables, while two different demand shocks are disentangled using sign restrictions. In the second application, a monetary policy shock is identified by a combination of sign restriction and information in a monetary policy proxy. Both empirical application suggest that combining sign restrictions with external instruments is a promising way to sharpen results from SVAR models.

In Chapter 2, written in collaboration with Dominik Bertsche, we develop a new econometric methodology to identify SVAR models by heteroskedasticity. This identification approach exploits time variation in the second moments of VAR residuals to identify a unique set of structural shocks via distributional assumptions. In our work, we propose to model the heteroskedasticity in SVARs by a stochastic volatility model (SV-SVAR). We discuss full and partial identification of the model and develop efficient expectation maximization algorithms for maximum likelihood inference. Simulation evidence suggests that the SV-SVAR works well in identifying structural parameters also under misspecification of the variance process, particularly in comparison to alternative heteroskedastic SVARs proposed in the literature. We apply the model to study the interdependence between monetary policy and stock markets in the United States. Since shocks identified by heteroskedasticity may not be economically meaningful, we exploit the framework to test conventional exclusion restrictions as well as instrumental variables restrictions which are overidentifying in the heteroskedastic model. The tests find statistical evidence against employing conventional short run restrictions to identify a monetary policy shock. However, there is no evidence against using a certain long run restriction and no evidence against identification by one of two external instruments included into the analysis.

In Chapter 3, I conduct an empirical analysis providing new evidence on the relative importance of supply and demand shocks for fluctuations in oil prices. To estimate their effects, a structural vector autoregressive (SVAR) model for the global oil market is identified by non-Gaussianity. Similar to identification by heteroskedasticity, non-Gaussianity exploits distributional assumptions about the structural shocks to identify a unique set of structural shocks. The statistical identification approach is further refined by ruling out economically unreasonable oil price elasticities a priori. To include this prior information in a coherent way, a new Bayesian SVAR is developed where the unknown distributions of the structural shocks are modeled nonparametrically using Dirichlet process mixture models. Given the nonparametric approach, the model requires no user input prior to estimation and identification is robust with respect to the form of non-Gaussianity underlying the data. Furthermore, the approach allows for a straightforward assessment of identification by simply comparing a Gaussian density with the predictive densities estimated for each shock. The empirical findings indicate that oil supply shocks have been minor drivers of oil prices post 1985. In terms of contributions to the long term forecast error variance of oil prices, the model arrives at median estimates between 1% and 13% depending on the exact prior specifications.

References

- Kilian, L. and Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*, Themes in Modern Econometrics, Cambridge University Press.
- Sims, C. A. (1980). Macroeconomics and reality, *Econometrica* **48**(1): 1–48.

Zusammenfassung

Um fundierte Maßnahmen ergreifen zu können, benötigen politische Entscheidungsträger genaue Punkt- und Intervallvorhersagen für wirtschaftliche Variablen. Das vektorautoregressive (VAR) Modell ist ein beliebtes Werkzeug für diesen Zweck, das multiple Zeitreihen basierend auf linearen Funktionen früherer Beobachtungen vorhersagt. VARs sind jedoch Modelle in reduzierter Form und können daher nur zu bestimmten beschreibenden Zwecken oder zur unbedingten Prognose wirtschaftlicher Variablen verwendet werden. Um den Effekt einer wirtschaftspolitischen Maßnahme vorherzusagen, ist ein strukturelles Modell erforderlich, das ähnliche Ereignisse in der Vergangenheit identifiziert und ihre Auswirkungen auf die Wirtschaft misst. Darüber hinaus können Entscheidungsträger nur mit Hilfe eines strukturellen Modells die in den Daten beobachteten Schwankungen interpretieren. Ein solides Verständnis der aktuellen wirtschaftlichen Antriebskräfte ist Voraussetzung für die Ergreifung wirksamer Gegenmaßnahmen. Seit der wegweisenden Arbeit von [Sims \(1980\)](#) werden Nicht-Stichprobeninformationen verwendet, um im VAR-Kontext identifizierende Restriktionen aufzuerlegen. Daraus resultiert das sogenannte strukturelle VAR (SVAR) Modell, das heutzutage eines der beliebtesten Modelle in der empirischen Makroökonomie ist.

Diese Arbeit befasst sich mit der Entwicklung neuartiger ökonometrischer Techniken zur Identifikation von SVAR-Modellen. Technisch erfordert dies eine Zerlegen von VAR-Prognosefehlern in unkorrelierte Komponenten, die als exogene Antriebskräfte interpretiert werden können. In der Makroökonomie werden diese als strukturelle Schocks bezeichnet. Prominente Beispiele dafür sind fiskalpolitische Schocks, geldpolitische Schocks, Ölpreisschocks oder Unsicherheitsschocks. Da es viele Zerlegungen gibt, die zur selben reduzierten Form des SVAR führen, sind Restriktionen zur Identifikation des Modells erforderlich, um aussagekräftige Ergebnisse zu erhalten. Zu den beliebtesten Ansätzen gehört die Verwendung von kurz- oder langfristigen Restriktionen auf die Auswirkungen der strukturellen Schocks, die entweder als Ausschluss- oder Vorzeichenrestriktionen dem Modell auferlegt werden. Darüber hinaus können Instrumentvariablenschätzung oder Verteilungsannahmen

zur Identifikation von strukturellen Schocks genutzt werden. Das Lehrbuch von [Kilian and Lütkepohl \(2017\)](#) bietet einen hervorragenden Überblick über den aktuellen Stand dieser Literatur.

In Kapitel 1, das gemeinsam mit Ralf Brüggemann erarbeitet wurde, schlagen wir vor, SVAR-Modelle durch eine Kombination von Vorzeichenrestriktionen und Instrumentenvariablen-schätzung zu identifizieren. Basierend auf narrativer Evidenz und modellexternen Informationen werden häufig so genannte “Proxyvariablen“ konstruiert, die als Instrumentenvariable für bestimmte strukturelle Schocks eingesetzt werden. In diesem Kapitel entwickeln wir einen Modellierungsrahmen, in dem strukturelle Schocks auf zwei verschiedene Arten identifiziert werden können. In unserem ersten Ansatz werden mehrere Schocks gleichzeitig identifiziert, entweder mit Vorzeichenrestriktionen oder mit einem externen Instrumentenansatz. Dabei stellen wir sicher, dass alle Schocks weiterhin unkorreliert sind. Im zweiten Fall zeigen wir, wie eine Kombination beider Ansätze genutzt werden kann, um einen einzelnen Schock zu identifizieren. Hierbei kann man auf die Exogenität der Proxyvariablen verzichten. Stattdessen werden sie dazu verwendet, die Anzahl der durch Vorzeichenrestriktionen ermittelten Modelle zu reduzieren. Dies geschieht, indem all jene Modelle verworfen werden, die strukturelle Schocks ohne engen Bezug zu den externen Proxyvariablen implizieren. Für beide Fälle entwickeln wir Algorithmen, um bayesianische Inferenz durchzuführen, die sowohl Modell- als auch Schätzungsunsicherheiten widerspiegelt. Wir illustrieren unsere Methode in zwei empirischen Anwendungen. Zunächst identifizieren wir Angebots- und Nachfrageschocks des globalen Rohölmarktes. Hierbei identifizieren wir einen Angebotsschock anhand von Instrumentvariablen-schätzung, während zwei weitere Nachfrageschocks anhand von Vorzeichenrestriktionen entflechtet werden. In der zweiten Anwendung wird ein geldpolitischer Schock durch eine Kombination aus Vorzeichenrestriktionen und Informationen in einer geldpolitischen Proxyvariable identifiziert. Beide empirischen Anwendungen legen nahe, dass die Kombination von Vorzeichenrestriktionen mit Informationen zu Proxyvariablen ein vielversprechender Weg ist, die Ergebnisse von SVAR Modellen zu schärfen.

In Kapitel 2, das in Zusammenarbeit mit Dominik Bertsche geschrieben wurde, entwickeln wir eine neue ökonometrische Methodik, um SVAR-Modelle durch Heteroskedastizität zu identifizieren. Dieser Identifikationsansatz nutzt die zeitliche Variation im zweiten Moment von VAR-Prognosefehlern, um anhand von Verteilungsannahmen strukturelle Schocks zu identifizieren. In unserer Arbeit schlagen wir vor, die Heteroskedastizität in SVARs mit einem stochastischen Volatilitätsmodell (SV-SVAR) zu modellieren. Wir besprechen formal die Bedingungen für vollständige und teilweise Identifikation des Modells und entwickeln effiziente Algorithmen zur Maximum-Likelihood-Schätzung. Monte-Carlo-Simulationen deuten darauf hin, dass das SV-SVAR Modell gut dazu geeignet ist, strukturelle Parameter

zu identifizieren, auch unter falscher Spezifikation des Varianzprozesses. Insbesondere im Vergleich zu alternativen heteroskedastischen SVARs, die in der Literatur vorgeschlagen werden, schneidet das SV-SVAR Modell gut in den Monte-Carlo-Simulationen ab. Wir wenden das Modell an, um die Wechselbeziehung zwischen Geldpolitik und Aktienmärkten in den USA zu untersuchen. Da jene Schocks, die durch Heteroskedastizität identifiziert werden, möglicherweise nicht ökonomisch interpretierbar sind, nutzen wir das Modell, um herkömmliche Restriktionen zur Identifikation von SVAR Modellen statistisch zu testen. Unsere Ergebnisse deuten darauf hin, dass herkömmliche Restriktionen auf den kurzfristigen Effekt von geldpolitischen Schocks nicht herangezogen werden können. Es gibt jedoch weder statistische Anhaltspunkte gegen eine bestimmte Restriktion auf den langfristigen Effekt des geldpolitischen Schocks, noch gegen die Identifikation durch eine in der Analyse herangezogene Instrumentenvariable.

In Kapitel 3 führe ich eine empirische Analyse durch, welche die relative Bedeutung von Angebots- und Nachfrageschocks für Schwankungen in Rohölpreisen untersucht. Um ihre Auswirkungen abzuschätzen, wird ein SVAR für den globalen Ölmarkt durch Nicht-Gaussianität identifiziert. Ähnlich der Identifikationsstrategie in Kapitel 2, nutzt auch dieser Ansatz Verteilungsannahmen über strukturelle Schocks, um das SVAR Modell zu identifizieren. Der Identifikationsansatz wird weiter verbessert, indem ökonomisch unplausible Ölpreiselastizitäten von vornherein ausgeschlossen werden. Um diese Information kohärent in das Modell zu integrieren, entwickle ich ein neues bayesianisches SVAR. In diesem Modell werden die unbekanntenen Verteilungen der strukturellen Schocks mit Hilfe von nichtparametrischen Methoden modelliert. Der Vorteil eines nichtparametrischen Ansatzes ist, dass keine Vorabinformationen über die Form der Nicht-Gaussianität benötigt werden, und die Identifikation robust ist. Darüber hinaus ermöglicht der Ansatz eine einfache Beurteilung ob ein SVAR Modell durch Nicht-Gaussianität identifiziert ist. Hierzu reicht es aus, die geschätzten Dichten der strukturellen Schocks mit Werten einer Standardnormalverteilung zu vergleichen. Die empirischen Ergebnisse deuten darauf hin, dass Angebotschocks seit 1985 nur eine geringfügige Rolle für Ölpreise gespielt haben. Im Hinblick auf deren Beitrag zur langfristigen Prognosefehlerabweichung des Ölpreises, liefert das Modell geschätzte Werte zwischen 1% und 13%.

References

- Kilian, L. and Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*, Themes in Modern Econometrics, Cambridge University Press.
- Sims, C. A. (1980). Macroeconomics and reality, *Econometrica* **48**(1): 1–48.

Chapter **1**

Identification of SVAR Models by
Combining Sign Restrictions with External
Instruments

1.1 Introduction

Ever since [Sims \(1980\)](#), structural vector autoregressive (SVAR) models have become a popular tool in applied macroeconomics.¹ Starting from a reduced form vector autoregressive (VAR) model, which summarizes the joint dynamics of a vector of time series variables, applied researchers impose various restrictions to identify structural shocks. Conditional on a particular identification scheme, the effects and importance of different structural shocks are summarized by impulse responses, forecast error variance decompositions or historical decompositions. Applications of this method include e.g. the analysis of monetary policy shocks, demand and supply shocks, fiscal shocks, oil price shocks, and news shocks.

In this paper, we suggest to identify structural shocks in SVAR models by combining both, sign restrictions and the information in narrative series constructed to instrument a certain shock in the model (proxy variables). As we argue below, combining both approaches is useful because it mitigates some drawbacks occurring when using either sign restrictions or external instruments only.

To achieve identification different types of restrictions have been suggested in the literature. Popular traditional restrictions include short- and long-run restrictions on the effects of structural shocks, which have the disadvantage that they are often difficult to justify and not testable if they are just-identifying. Therefore, alternative methods for identification have been developed including sign restrictions and the use of external instruments.²

Sign restrictions have been introduced into the literature by [Faust \(1998\)](#), [Canova and De Nicoló \(2002\)](#) and [Uhlig \(2005\)](#) as an alternative to existing methods involving short and long-run restrictions. The obvious advantage is that the researcher may directly restrict the signs of responses to the structural shock of interest. Sign restrictions may be imposed on the contemporaneous response or on the responses at later response horizons. In the context of monetary policy shocks, for instance, sign restrictions have been used to avoid the so-called ‘price-puzzle’ by restricting the response of the price level to be non-positive for a certain period after a contractionary monetary policy shock (see e.g. [Uhlig \(2005\)](#) and [Arias et al. \(2019\)](#)) but leaving the response of interest (e.g. the response of output) unrestricted. Employing sign restrictions does not yield a unique model but leads to set identification only. As pointed out by [Fry and Pagan \(2011\)](#), special care has to be taken when summarizing results across multiple models. Another practical problem of sign restrictions is that they are often rather weak, resulting in a wide range of admissible models with impulse responses

¹For a general overview of different structural VAR models see e.g. [Kilian and Lütkepohl \(2017\)](#).

²Another strand of the literature uses statistical identification and identifies structural shocks by exploiting changes in the error term variance (see e.g. [Lütkepohl and Netšunajev \(2017\)](#) and [Bertsche and Braun \(2018\)](#) for recent contributions).

that are not very informative. Consequently, additional restrictions are often needed to narrow down the set of models (see e.g. [Kilian and Murphy \(2012\)](#)).

Another method for identifying structural shocks without imposing short- or long-run restrictions is based on external instrument variables. While the underlying economic shock of interest is unobservable to the researcher, there may be related time series available that act as instrumental variables (IV) for the unobserved structural shock. To achieve identification the researcher needs to find an instrument variable that is highly correlated with the structural shock of interest and uncorrelated with all other structural shocks in the system. Studies that have used external instruments include e.g. [Hamilton \(2003\)](#), [Romer and Romer \(1989, 2004\)](#), [Kilian \(2008b, 2008a\)](#), [Mertens and Ravn \(2012, 2013, 2014\)](#), [Stock and Watson \(2012\)](#), and [Gertler and Karadi \(2015\)](#). While conceptually appealing, the external IV approach has also potential drawbacks. First, the exogeneity of instruments is questionable in some applications (see e.g. the discussion in [Ramey \(2016\)](#) on the narrative measures of monetary policy shocks). Furthermore, the external IV approach hinges critically on the validity of instruments and is problematic when the instruments are weak. In this case, inference is non-standard, complicated and still under development (see [Montiel-Olea et al. \(2015, 2016\)](#)). Finally, if the IV approach is used to identify different economic shocks at a time as e.g. in [Stock and Watson \(2012\)](#), the different structural shocks are not necessarily orthogonal.

Consequently, using either sign restrictions or external instruments alone may not be optimal in practical work. In this paper, we therefore suggest ways to combine these two identification strategies in order to circumvent some of the mentioned problems. Our paper makes the following contributions. First, we suggest an econometric framework to combine identification from sign restrictions with identification from using external proxy variables. This is achieved by augmenting a standard SVAR system with equations for the proxy variables to relate them with the structural shocks. Second, we discuss estimation and inference in a full Bayesian setup, which accounts for both model and estimation uncertainty.

Our framework allows to combine identifying information from sign restrictions and external proxy series in two different ways. First, it can be used to identify some shocks by sign restrictions, while other shocks are identified via instrumental variables. One set of shocks is therefore set-identified and the others are point-identified. Our modeling framework nests the pure sign restriction and the pure IV approach as special cases. Compared to the existing methodology, our framework analyzes shocks identified from sign restrictions and external instruments jointly in a system approach and leads to structural shocks that are orthogonal by construction. In addition, our Bayesian framework can handle situations in which the external proxy variables are only weak instruments. As long as one uses a proper prior, the inference requires no adjustment to be valid. Second, we discuss how our

framework can be exploited to identify a single shock by simultaneously using information from sign restrictions *and* external proxy series, however, without assuming exogeneity of the latter. This case is highly relevant in empirical work where applied researchers are often in the situation that pure sign restrictions are not enough to achieve informative results and at the same time, no credibly exogenous instruments are available. For this scenario, we suggest using the external variables to further tighten the set of models obtained by sign restrictions. This combination is simple and intuitive. Essentially, we narrow down the set of admissible models by only keeping those models that imply structural shocks that show a close relation to the external proxy variable. As explained in the paper, this relation is measured by either correlations or variance contributions. The resulting identification scheme is set-identifying, and thereby could be also thought of adding additional sign restrictions based on the information contained in the external proxies. As we will show in an empirical application, these additional restrictions can help to avoid the wide and uninformative confidence intervals around impulse responses, a typical problem observed in applied sign restriction studies.

Methodologically, our first way to combine sign restrictions with external proxy variables is related to recent formulation of Bayesian Proxy SVARs ([Drautzburg; 2016](#); [Caldara and Herbst; forthcoming](#); [Miranda-Agrippino and Rey; 2018](#)). Among those, we are closest to [Drautzburg \(2016\)](#) in that we also use a system approach to conduct inference on the reduced form parameters before proceeding with the identification of the model. In comparison to these papers, however, our methodology is more general in the sense that we allow for additional shocks to be identified by pure sign restrictions. Our second way of combining sign restrictions with external proxy variables is related to studies that exploit non-model information for identification. For example, [Kilian and Murphy \(2012\)](#) use microeconomic evidence to construct elasticity bounds with the goal to further tighten the credible set obtained by sign restrictions. Furthermore, [Antolín-Díaz and Rubio-Ramírez \(2018\)](#) and [Ben Zeev \(2018\)](#) suggest to narrow down the set of models by discarding all models that lead to structural shocks with signs and historical decompositions that are at odds with narrative evidence. Our approach is also close to [Ludvigson et al. \(2017\)](#) who, similar to [Antolín-Díaz and Rubio-Ramírez \(2018\)](#) and [Ben Zeev \(2018\)](#), suggest to include information on unusual historical events to restrict the sign and size of structural shocks. In addition, they discard models whose shocks are not strongly correlated with the ‘synthetic proxy’ variables and therefore, is similar in spirit to what we suggest.³ In comparison, our paper is more general with respect to important modeling aspects: Our setup does not require to choose a threshold for the correlation with the proxy and can also be used to identify some shocks by instrumental variables only. Moreover, we properly take into account all sources of model and estimation

³See also [Uhrin and Herwartz \(2016\)](#) for a similar idea to narrow down set identified SVAR models.

uncertainty within our full Bayesian framework. Therefore, inference is straightforward and error bands for impulse responses are readily available, which makes our setup particularly useful for applied work. Finally, since we use information of external variables without assuming strict exogeneity, our paper is also related to a branch of the microeconomic literature that discusses (set) identification using weakly endogenous instruments (Nevo and Rosen; 2012; Conley et al.; 2012).

We illustrate the usefulness of our method in two empirical applications. We identify structural shocks on the oil market and monetary policy shocks. In both applications, we find that combining both, sign restrictions and external proxy series, leads to more informative and economically more sensible impulse response patterns.

The remainder of the paper is structured as follows. Section 1.2 introduces the econometric modeling framework and discusses identification and inference. Section 1.3 illustrates the suggested methods in applications to oil market shocks and to US monetary policy shocks. Section 1.4 summarizes and concludes.

1.2 Combining sign restrictions with external instruments

In the following, we develop a unified SVAR framework which enables us to combine information from sign restrictions and proxy variables to identify the effect of structural shocks. In Section 1.2.1 we describe a general econometric framework for this purpose, which is based on a SVAR model augmented by equations for proxy variables. Section 1.2.2 discusses identifying restrictions for the case where multiple shocks are identified by a combination of sign restrictions and instrumental variables, a scenario which nests both approaches. Section 1.2.3 discusses how information on sign restrictions and that of proxy variables can be exploited jointly to identify a single shock, making weaker assumptions about the relation between the SVAR shocks and the proxy variables. This idea is similar in spirit to Ludvigson et al. (2017) and Antolín-Díaz and Rubio-Ramírez (2018). Finally, we discuss how to coherently conduct inference in SVAR models subject to these restrictions in Section 1.2.4. Note that, whenever we use the term ‘proxy variable’ we mean any external series designed to be similar to a structural shock, while we use the term ‘instrumental variable’ only if the proxy is used for identification in the sense of classical IV moment conditions.

1.2.1 Proxy augmented SVAR model

Let $y_t = (y_{1t}, \dots, y_{nt})'$ be a $n \times 1$ vector of endogenous time series generated by the SVAR model:

$$y_t = c + \sum_{i=1}^p A_i y_{t-i} + u_t, \quad u_t \sim (0, \Sigma_u), \quad (1.2.1)$$

$$u_t = B\varepsilon_t, \quad \varepsilon_t \sim (0, I_n), \quad (1.2.2)$$

where (1.2.1) corresponds to the reduced form VAR(p) model with c being an $n \times 1$ vector of intercepts and the $n \times n$ matrices A_i for $i = 1, \dots, p$ capturing the impact of lagged vectors of time series up to horizon p . Equation (1.2.2) contains the structural relations of the model, linking the reduced form errors u_t to structural shocks ε_t linearly by the $n \times n$ structural impact matrix B , which implies a reduced form error covariance matrix $\Sigma_u = BB'$. It is well known that the structural model (1.2.2) is not identified from the data alone.⁴ Therefore, a set of restrictions must be imposed on B in order to uniquely pin down the coefficients of the structural model. We propose a simple method to combine identification from sign restrictions and proxy variables. For this purpose, let $m_t = (m_{1t}, \dots, m_{kt})'$ be a $k \times 1$ vector of proxy series designed to provide identifying information about a subset of $k < n$ structural shocks of the SVAR model. Our approach is based on augmenting equation (1.2.2) by equations for the proxies m_t :

$$\underbrace{\begin{pmatrix} u_t \\ m_t \end{pmatrix}}_{\tilde{u}_t} = \underbrace{\begin{pmatrix} B & 0_{n \times k} \\ \Phi & \Sigma_\eta^{1/2} \end{pmatrix}}_{\tilde{B}} \underbrace{\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix}}_{\tilde{\varepsilon}_t}, \quad \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim (0, I_{n+k}). \quad (1.2.3)$$

The augmented model has a measurement error interpretation similar to [Mertens and Ravn \(2013\)](#). The k proxy variables m_t are modeled as a linear function of the structural errors ε_t with $k \times n$ regression coefficients Φ , plus a zero mean measurement error η_t , which is assumed to be orthogonal to the structural shocks ε_t , i.e. $\eta_t \perp \varepsilon_t$. A $n \times k$ block of zeros ensures that the measurement error η_t is also orthogonal to the reduced form errors u_t and

⁴This can be seen that by simply multiplying matrix B by any $n \times n$ orthogonal matrix Q with property $QQ' = I_n$. We then end up with the same reduced form covariance matrix $\Sigma_u = (BQ)(BQ)' = BQQ'B' = BB'$.

avoids any impact on the dynamics of y_t . The augmented system has the reduced form covariance matrix

$$\Sigma = \text{Cov} \begin{pmatrix} u_t \\ m_t \end{pmatrix} = \begin{pmatrix} \Sigma_u & \Sigma_{mu} \\ \Sigma_{mu'} & \Sigma_m \end{pmatrix},$$

and through restrictions on Φ , identifying information can be imposed to pin down values of B in the spirit of an IV regression. In the following we discuss possible restrictions in detail and provide Bayesian algorithms for inference.

1.2.2 Identifying multiple shocks with a combination of sign restrictions and instrumental variables

We first describe a scenario where either instrumental variables or sign restrictions are used to identify distinct structural shocks. Without loss of generality, assume that out of n structural shocks, the researcher wants to identify the first $k \leq n$ via an IV and the last $q \leq (n - k)$ shocks via sign restrictions. Given the following partition of the structural shocks $\varepsilon_t = [\varepsilon'_{1t} : \varepsilon'_{2t} : \varepsilon'_{3t}]'$ this corresponds to identifying ε_{1t} and ε_{3t} via the IV and sign restrictions, respectively.

To state the sign restrictions explicitly, we follow the notation of [Arias et al. \(2018\)](#). Assume that the researcher has prior information on the contemporaneous impact of the i th structural shock on the endogenous variables y_t and therefore on elements of the i th column of the impact matrix B , denoted as $B_{\bullet i}$ in the following.⁵ Let S_j be a $s_j \times n$ selection matrix with rank $\text{rk}(S_j) = s_j$ and let e_j be the j th column of the identity matrix I_n . Then, we gather all sign restrictions in the augmented SVAR system (1.2.3) as

$$S_j B e_j > 0, \quad j = n - q + 1, \dots, n.$$

With respect to the first k shocks (ε_{1t}), which are to be identified via instrumental variables, assume that a set of k instrumental variables are available for this purpose, implying the following restrictions:

$$E(m_{it}, \varepsilon_{i,t}) \neq 0, \quad i = 1, \dots, k, \tag{1.2.4}$$

$$E(m_{it}, \varepsilon_{j,t}) = 0, \quad i \neq j. \tag{1.2.5}$$

⁵More generally, prior information can be available for any function $F(B, A_i)$ of the structural coefficients B and lag matrices $A_i, i = 1, \dots, p$, including for example restrictions on higher horizons of the impulse responses.

Equations (1.2.4) and (1.2.5) are known as relevance and exogeneity conditions, respectively. For local identification of the structural shock $\varepsilon_{i,t}$, it is thus required that the corresponding proxy variable m_{it} is correlated with this shock and uncorrelated with all other shocks of the SVAR system.

As discussed by [Mertens and Ravn \(2013\)](#) and [Stock and Watson \(2012\)](#), these conditions imply linear restrictions that identify the respective column of the structural impact matrix $B_{\bullet i}$ up to sign and scale. To be more specific about the exact form of the restriction, we follow [Mertens and Ravn \(2013\)](#) by partitioning $B = [\beta_1 : \beta_2 : \beta_3]$ as well as $\Phi = [\phi_1 : \phi_2 : \phi_3]$. Equations (1.2.4) and (1.2.5) imply that

$$\Sigma_{mu'} = E(m_t u_t') = E(m_t \varepsilon_t' B') = \phi_1 \beta_1'.$$

Further partitioning of $\Sigma_{mu'}$ and β_1 yields

$$\begin{bmatrix} \Sigma_{mu'_1} & \Sigma_{mu'_2} \end{bmatrix} = \phi_1 \begin{bmatrix} \beta'_{11} & \beta'_{12} \end{bmatrix}.$$

This, in turn, translates into the following linear restrictions for the matrix β_1 :

$$\beta_{11} = (\Sigma_{mu'_2}^{-1} \Sigma_{mu'_1})' \beta_{12}. \quad (1.2.6)$$

For $k = 1$, equation (1.2.6) identifies β_1 up to sign and scale, while the additional restriction $\beta_1' \Sigma_u^{-1} \beta_1 = 1$ normalizes the shock to one standard deviation. For $k > 1$, additional restrictions must be specified to achieve identification, see e.g. [Mertens and Ravn \(2013\)](#) and [Angelini and Fanelli \(2018\)](#).

Note that an alternative representation of these restriction can be obtained by simply setting zero restrictions in the augmented impact matrix \tilde{B} as stated in equation (1.2.3). In particular, equations (1.2.4) and (1.2.5) imply that in addition to the zero block already mentioned in \tilde{B} , also $\phi_2 = 0$ and $\phi_3 = 0$. See also [Angelini and Fanelli \(2018\)](#) for a more extensive discussion. The described combination of sign restrictions with external instruments can therefore be also thought of combining zero and sign restrictions.

Example 1. Consider a simple three variable macro model as in [Fry and Pagan \(2011\)](#), involving the output gap z_t , prices π_t and an interest rate i_t . Assume that the system is driven by a monetary policy shock ε_t^{mp} identified via an external instrument m_t , as well as cost push (ε_t^c) and demand shocks (ε_t^d) identified with standard sign restrictions. Specifically, the cost push shock is assumed to decrease output and increase prices and interests on impact, while the demand shock is assumed to increase all variables. These identifying assumptions can be stated as follows:

$$\begin{pmatrix} u_t^z \\ u_t^\pi \\ u_t^i \end{pmatrix} = \begin{pmatrix} b_{11} & - & + \\ b_{21} & + & + \\ b_{31} & + & + \end{pmatrix} \begin{pmatrix} \varepsilon_t^{mp} \\ \varepsilon_t^c \\ \varepsilon_t^d \end{pmatrix}, \quad \begin{pmatrix} \varepsilon_t^{mp} \\ \varepsilon_t^c \\ \varepsilon_t^d \end{pmatrix} \sim (0, I_3),$$

$$b_{11} = \frac{\text{Cov}(u_{1t}, m_t)}{\text{Cov}(u_{3t}, m_t)} b_{31}, \quad b_{21} = \frac{\text{Cov}(u_{2t}, m_t)}{\text{Cov}(u_{3t}, m_t)} b_{31}, \quad 1 = \beta_1 \Sigma_u^{-1} \beta_1'.$$

1.2.3 Identifying a single shock using information from sign restrictions and an external proxy

In our second scenario, we describe how sign restrictions and identifying information from external proxy variables are combined to trace down the effects of a single shock. This idea is also proposed by [Ludvigson et al. \(2017\)](#) who start their SVAR analysis with a set identified model based on sign restrictions. To further narrow down the set of admissible models and thereby reduce the model uncertainty, they discard all those models where the corresponding structural shock exhibits a correlation with the external proxy less than a certain threshold value \bar{c} . Only relevance of the external proxy (equation (1.2.4)) is assumed, but not its exogeneity (equation (1.2.5)). The assumptions needed are therefore weaker than in a pure instrumental variables approach.

Without loss of generality, assume that the goal is to identify the first structural shock ε_{1t} by a combination of the two identifying sources, that is sign restrictions and information from a proxy variable. Let us first gather the inequality restrictions and assume that some information on the sign of elements in $B_{\bullet 1}$ is available. As in Section 1.2.2, we can make the restriction explicit with the help of the selection matrix S_1 defined previously:

$$S_1 B e_1 > 0. \tag{1.2.7}$$

Besides the sign restrictions of equation (1.2.7), assume that a set of k external proxy variables m_t is available containing information on the structural shock, that is:

$$E(m_{it} \varepsilon_{1t}) \neq 0, \quad i = 1, \dots, k.$$

We propose the following restrictions that exploit different degrees of identifying information from the external proxy variable without imposing its exogeneity:

1. The correlation between the i th proxy and the structural shock is positive:

$$\text{Corr}(m_{it}, \varepsilon_{1t}) = \frac{E(m_{it} \varepsilon_{1t})}{\sqrt{\text{Var}(m_{it})}} > 0.$$

From an economic point of view, this means that we are confident that the proxy variable is at least positively correlated with the structural shock it has been designed for. Note that this restriction does little harm if the proxy is only loosely associated with the structural shock and might be interesting for variables, which are assumed to be weak instruments.

2. The correlation between the i th proxy and the structural shock exceeds \bar{c}_i :

$$\text{Corr}(m_{it}, \varepsilon_{1t}) = \frac{E(m_{it}\varepsilon_{1t})}{\sqrt{\text{Var}(m_{it})}} > \bar{c}_i.$$

This restriction has been applied by [Ludvigson et al. \(2017\)](#). Of course, it is more restrictive than only a sign restriction on the correlation in that it rules out more models from the set of admissible SVARs. However, choosing \bar{c}_i is difficult and hard to justify in practice.

To circumvent the problem of choosing a threshold \bar{c}_i , we discuss two other possibilities:

3. First, among all structural shocks of the SVAR model ($\varepsilon_{jt}, j = 1, \dots, n$), the shock to be identified ε_{1t} has the largest correlation with the i th proxy variable.

$$\text{Corr}(m_{it}, \varepsilon_{1t}) > \text{Corr}(m_{it}, \varepsilon_{jt}), \quad j = 2, \dots, n. \quad (1.2.8)$$

This restriction imposes that among all shocks in the SVAR, the shock to be identified shows the highest correlation with the external variables.

4. Second, among all structural shocks of the SVAR model ($\varepsilon_{jt}, j = 1, \dots, n$), the shock to be identified ε_{1t} explains most of the variation of the i th proxy variable m_{it} . To trace down the restriction, recall the regression equation for m_t :

$$m_t = \Phi\varepsilon_t + \Sigma_{\eta}^{\frac{1}{2}}\eta_t, \quad \eta_t \sim (0, I_k).$$

Since the regressors ε_t are orthogonal by assumption, the contribution of the j th structural shock to the variance of the i th proxy m_{it} is $\psi_{ij} = \phi_{ij}^2 / \text{Var}(m_{it})$. Therefore, the restriction is given as:

$$\psi_{i1} > \psi_{ij} \quad j = 2, \dots, n. \quad (1.2.9)$$

In words, imposing this restrictions rules out all SVAR models where other shocks explain more variation of the i th proxy m_{it} than the one to be identified (ε_{1t}).

Note that in order to implement the last two restriction, the other shocks $\varepsilon_{jt}, j = 2, \dots, n$ have to be identified. If they are not identified by economic reasoning, it is still possible to use a statistical identification, e.g. by orthogonal rotations of arbitrarily identified shocks. In that case, the prior implicitly used to statistically identify these shocks is of particular relevance must be chosen carefully. We further discuss this possibility in the empirical application.

Note that similar to the observation made in Section 1.2.2, the just described restrictions relating the external proxy and the structural model can be reinterpreted as imposing additional sign restrictions on the augmented SVAR model given in equation (1.2.3).

Example 2. *Reconsider Example 1 with the three variable macro model involving the output gap z_t , prices π_t and an interest rate i_t . Assume that we are particularly interested in identifying the monetary policy shock ε_t^{mp} but that no credible instrument is available in terms of exogeneity. However, information about the sign of the impact of the shock is readily available, as well as an imperfect external variable m_t designed to proxy the policy shock. To state specific restrictions, assume that a monetary policy shock has an immediate positive effect on the interest rate i_t and decreases prices π_t and output z_t on impact. Moreover, among all three structural shocks, ε_t^{mp} explains most of the variation of the proxy variable m_t . This can be stated as follows in the SVAR framework:*

$$\begin{pmatrix} u_t^z \\ u_t^\pi \\ u_t^i \end{pmatrix} = \begin{pmatrix} - & b_{11} & b_{12} \\ - & b_{12} & b_{22} \\ + & b_{13} & b_{23} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t}^{mp} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}, \quad \begin{pmatrix} \varepsilon_{1t}^{mp} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix} \sim (0, I_3),$$

$$\frac{\phi_1^2}{\text{Var}(m_{it})} > \frac{\phi_j^2}{\text{Var}(m_{it})}, \quad j = 2, 3.$$

1.2.4 Bayesian estimation and inference

In the following we outline how to conduct Bayesian inference in our framework to coherently summarize both, modeling and sampling uncertainty. As in Drautzburg (2016), we specify priors on the reduced form proxy augmented VAR model which allows usage of an efficient textbook style Gibbs sampler on these parameters. In a second step, we identify the model by generating orthogonal matrices that relate the reduced form parameters to the structural representation of the model and satisfy the parameter restrictions discussed previously.

Given our Bayesian setup, we assume a standard Gaussian likelihood function for the proxy augmented SVAR:

$$y_t = Ax_t + u_t, \quad \begin{pmatrix} u_t \\ m_t \end{pmatrix} \sim \mathcal{N} \left(0, \Sigma = \begin{pmatrix} \Sigma_u & \Sigma_{mu} \\ \Sigma_{mu'} & \Sigma_m \end{pmatrix} \right),$$

where $A = [c, A_1, \dots, A_p]$ and $x_t = [1, y'_{t-1}, \dots, y'_{t-p}]'$. We link the reduced form to the structural form through the contemporaneous impact matrix:

$$B = PQ, \quad (1.2.10)$$

where P is the Cholesky decomposition of the VAR block reduced form error variance such that $PP' = \Sigma_u$. The matrix Q is orthogonal, that is $Q'Q = I_n$, and links the reduced form model to the structural representation as in equation (1.2.10). We assume conditionally conjugate prior distributions for the reduced form coefficients Σ^{-1} and A , which take the form of Normal and Wishart distributions respectively:

$$\Sigma^{-1} \sim \mathcal{W}(d, \Psi), \quad (1.2.11)$$

$$\text{vec}(A) \sim \mathcal{N}(\alpha_0, V_0). \quad (1.2.12)$$

Before discussing the prior distribution for the orthogonal matrix Q , we first characterize the posterior of the reduced form model. Unfortunately, this density is of no known form. However, random draws of the posterior can be efficiently generated by a standard Gibbs sampler iterating between the conditional posteriors $p(\Sigma|Y, M, A)$ and $p(A|Y, M, \Sigma)$. To find these distributions, it is instructive to rewrite the model in a Seemingly Unrelated Regression (SUR) representation (Drautzburg; 2016). Conditional on presample values, the model can be written as:

$$\tilde{y} = Z\alpha + \tilde{u}, \quad \tilde{u} \sim \mathcal{N}(0, \Sigma \otimes I_T),$$

where $\alpha = \text{vec}(A)$, $X = [x'_1, \dots, x'_T]'$, $Z = [(I_n \otimes X)' : \mathbf{0}_{n(np+1) \times Tk}]'$, $Y = [y'_1, \dots, y'_T]'$, $M = [m'_1, \dots, m'_T]'$, $\tilde{y} = [\text{vec}(Y)', \text{vec}(M)']'$, $U = [u'_1, \dots, u'_T]'$, $\tilde{U} = [U', M']'$ and $\tilde{u} = \text{vec}(\tilde{U})$. The density function of the observables \tilde{y} is then given as:

$$p(\tilde{y}|\alpha, \Sigma) \propto |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \text{tr}(S\Sigma^{-1})\right),$$

where $S = \tilde{U}'\tilde{U}$. Based on Geweke (2005, Chapter 5), the conditional posterior distributions of SUR models with priors (1.2.11) and (1.2.12) are of the following convenient form:

$$\alpha|\tilde{y}, \Sigma \sim \mathcal{N}(\bar{\alpha}, \bar{V}), \quad (1.2.13)$$

$$\Sigma^{-1}|\tilde{y}, \alpha \sim \mathcal{W}((\Psi + S)^{-1}, d + T), \quad (1.2.14)$$

where $\bar{V} = (V_0^{-1} + Z'(\Sigma^{-1} \otimes I_T)Z)^{-1}$ and $\bar{\alpha} = \bar{V} \left(V_0^{-1} \alpha_0 + Z'(\Sigma^{-1} \otimes I_T) \bar{y} \right)$. Based on these conditionals, it is straightforward to implement a Gibbs sampler to numerically characterize the posterior distribution of the reduced form parameters.

In the remainder of the section, we discuss the prior and posterior of the orthogonal matrix Q , which according to equation (1.2.10), will provide a mapping to the structural parameters of the model. Furthermore, we will discuss algorithms to generate draws from its posterior distribution. The details will depend on whether we are identifying the model as in Section 1.2.2 or 1.2.3. We start with the case described in Section 1.2.2 where a subset of shocks ($q \leq (n - k)$) are set identified via sign restrictions and that, for simplicity, just one additional shock ($k = 1$) is point identified via instrumental variables. Given a (prior or posterior) draw of the reduced form covariance matrix Σ , the linear restrictions from equation (1.2.6) along with the normalization $\beta_1' \Sigma_u^{-1} \beta_1 = 1$ enables us to point identify β_1 , the first column of the structural impact matrix B .⁶ Given this observation, it is convenient to decouple the orthogonal matrix Q into a fixed (\bar{Q}) and a stochastic (orthogonal) component (\tilde{Q}):

$$Q = \tilde{Q} \bar{Q}.$$

Given a value of the augmented reduced form matrix Σ , \tilde{Q} will provide the rotation of P such that the first columns of B are β_1 . To get an explicit expression, let $\tilde{q}_k = P^{-1} \beta_1$. Then, we define $\tilde{Q} = [N_\perp, \tilde{q}_k]$ where N_\perp is any orthonormal basis for the null space of \tilde{q}_k' such that $N_\perp' \tilde{q}_k = 0$ and $N_\perp' N_\perp = I$. With respect to the stochastic matrix \bar{Q} , we define it as follows:

$$\bar{Q} = \begin{pmatrix} I_k & 0 \\ 0 & Q_{n-k}^* \end{pmatrix}.$$

where Q_{n-k}^* is random and given the structure of \bar{Q} , will not affect β_1 but only relate the reduced form parameters to those structural parameters that are set-identified. We specify a uniform prior distribution (Haar prior) $\pi(Q_{n-k}^* | \Sigma, A)$ over the space of $n - k$ dimensional orthogonal matrices subject to the sign restrictions.⁷ This enables us to easily draw from the (conditional) posterior distribution of Q as follows (Algorithm 1).

1. Given a draw of the reduced form parameters Σ , solve for β_1 using equation (1.2.6) along with the normalization $\beta_1' \Sigma_u^{-1} \beta_1 = I_k$ described in Section 1.2.2. Furthermore, compute P such that $PP' = \Sigma_u$.

⁶In case of $k > 1$, recall that additional restrictions will be necessary to identify β_1 (Mertens and Ravn; 2013).

⁷A candidate draw from this prior can be easily obtained, e.g. via a Householder based algorithm (Rubio-Ramírez et al.; 2010) or a sequence of random Givens matrices (Uhlig; 2005; Fry and Pagan; 2011).

2. Compute $\tilde{q}_k = P^{-1}\beta_1$ and define the $n \times n$ matrix $\tilde{Q} = [N_\perp, \tilde{q}_k]$ where N_\perp is an orthonormal basis for the null space of \tilde{q}'_k such that $N'_\perp \tilde{q}_k = 0$ and $N'_\perp N_\perp = I$.
3. Draw Q_{n-k}^* from the uniform distribution over the space of all $(n-k) \times (n-k)$ dimensional orthogonal matrices with an algorithm of your choice and define $\bar{Q} = \begin{pmatrix} I_k & 0 \\ 0 & Q_{n-k}^* \end{pmatrix}$.
4. Compute $Q = \tilde{Q}\bar{Q}$ and its associated candidate structural impact matrix $B = PQ$. Accept the draw of Q if all sign restrictions are satisfied.

If the task is to identify a single shock combining sign restrictions with the identifying information of a set of k non-exogenous proxies as in Section 1.2.3, all elements in B are set-identified. In this case, we can directly specify the uniform prior distribution $\pi(Q|\Sigma, A)$ on the n dimensional orthogonal matrix Q as usually done in the literature. Then, given a posterior draw of the reduced form parameters, a draw of the (conditional) posterior of Q can be obtained in a standard fashion (Algorithm 2):

1. Given a draw of the reduced form parameters Σ , draw a candidate matrix Q from the uniform distribution over the space of $n \times n$ dimensional orthogonal matrices with an algorithm of your choice. Furthermore, compute P such that $PP' = \Sigma_u$.
2. Compute the following quantities: $E(m_t \varepsilon'_t) = Q'P^{-1}\Sigma_{mu'}$ as well as
 - $\text{Corr}(m_{it}, \varepsilon_{1t}) = E(m_{it}, \varepsilon_{1t})/\sigma_i$ for $i = 1, \dots, k$ where $\sigma_i^2 = \Sigma_{m,ii}$.
 - $\phi_{ij} = E(m_{it}, \varepsilon_{1t})$ for $i = 1, \dots, k$ as well as the associated contributions to the variance of m_{it} , $\psi_{ij} = \phi_{ij}^2/\sigma_i^2$.
3. Compute the candidate structural impact matrix $B = PQ$ and accept the draw if all sign and proxy induced moment restrictions (computed in step 2) are satisfied.

To sum up, inference can be decoupled into drawing reduced form parameters and generating draws of rotation matrix that satisfy the proposed restrictions and therefore, we take proper account of both sampling and model uncertainty. On a high level, the Gibbs sampler algorithm can be summarized as follows:

1. (Sampling uncertainty) Iteratively draw a set of reduced form parameters from their conditional posterior distribution as in equations (1.2.13) and (1.2.14).
2. (Model uncertainty) Draw Q subject to the sign restrictions and the proxy induced moment restrictions by either Algorithm 1 or 2.⁸

⁸Note that the conditional posteriors of the reduced form parameters do not depend on Q . Therefore, if desired, step one can be also run in isolation before draws of Q are generated in step two.

Note that both approaches can also be combined to identify a subset of shocks with IV and sign restrictions, and further shocks with a combination of both approaches.

Given the recent critique by [Baumeister and Hamilton \(2015\)](#) on the implications of using the Haar prior in SVAR models, we highlight that one might also use different priors for $\pi(Q_{n-k}^*|\Sigma, A)$ and $\pi(Q|\Sigma, A)$. However, this complicates efficient drawing of Q as discussed in [Arias et al. \(2018\)](#). Alternatively, one might also apply the Robust Bayes approach by [Giacomini and Kitagawa \(2015\)](#) to conduct inference on the identified set that is robust to the choice of prior for Q_{n-k}^* and Q .

Another highly relevant point for the practitioner is that many instruments are available only for a rather short sample period. For example, the shock series used in [Gertler and Karadi \(2015\)](#) are only available starting in the late 80s and early 90s. If the sample of all macroeconomic data in the VAR is not adjusted to the same length, a missing data problem arises. More generally, it might occur that the external variable is not observed in every period of time. Missing data can be handled in a straightforward way within a Bayesian setting. It involves a simple imputation of the missing observations through an additional step in the Gibbs sampler. We refer to Appendix 1.A for the details.

1.3 Empirical applications

We demonstrate the usefulness of our methodological framework in two empirical applications. In Section 1.3.1, we use instrumental variables and sign restrictions separately to identify multiple shocks in an SVAR model for the international oil market. This corresponds to the type of restrictions described in Section 1.2.2. In our second application in Section 1.3.2, we identify a single shock by combining information from sign restrictions and an external proxy variable. More specifically, we use the restrictions introduced in Section 1.2.3 to identify a monetary policy shock and investigate its effects on key macroeconomic variables, in particular on the US credit market. Note that for both applications, we specify an uninformative prior with little influence on the posterior of the reduced form parameters.⁹

1.3.1 The effects of oil price shocks

In this section, we illustrate how the external instrument approach and sign restrictions can be combined to simultaneously identify a set of different oil market specific shocks. In particular, we illustrate how an oil supply shock may be identified from using a proxy variable related to exogenous disruptive oil supply shortages, while at the same time the demand shocks may be identified from a set of sign restrictions.

⁹We use $\alpha_0 = 0$, $V_0 = 10 \cdot I$, $\Psi = I$ and $d = n + k + 1$.

In recent years, structural VAR models have been popular tools in investigating and understanding the dynamics of oil price shocks and their effect on macroeconomic variables. Different identifying assumptions have been used to disentangle oil supply demand shocks. While some studies rely only on short-run exclusion restrictions (e.g. Kilian (2009), Stock and Watson (2016)), there are also a number of studies using sign restrictions on either contemporaneous impacts (see Kilian and Murphy (2012) and Baumeister and Hamilton (2015)) or on multiple horizons (see e.g. Peersman and Van Robays (2012)). As pointed out by Kilian and Murphy (2012), the use of sign restrictions alone may not be enough to properly identify the oil market shocks.

An alternative way to identify an oil supply shock has been suggested by Hamilton (2003) and Kilian (2008b). Both papers construct an ‘exogenous’ time series based on disruptive events, which are typically related to wars in oil producing countries. Kilian (2008b), for instance, computes a measure of an oil supply shock based on a counter-factual evolution of oil production. If this measure is truly exogenous, the series from Kilian (2008b) may also serve as an external instrument for an oil supply shock. The external instrument approach in the context of oil prices shocks approach has been used in e.g. Stock and Watson (2012) using quarterly US data. Their results suggest that both measures, Hamilton’s and Kilian’s series, may only be weak instruments for the underlying structural oil price shock and consequently using standard IV inference is potentially problematic.

Instead of relying on either sign restrictions or external instruments alone, we combine both approaches within our modeling framework. We identify an oil supply shock using a Kilian (2008b) type proxy variable as an instrument. At the same time, we identify the aggregate demand and the oil-specific demand shock using the sign restrictions suggested by Kilian and Murphy (2012). Handling all three shocks simultaneously has the advantage that all shocks are mutually orthogonal by construction. Moreover, the information in the proxy variable allows us to be agnostic on the signs of the responses to an oil supply shock and may also help to narrow down the set of admissible impulse responses.

We use the 3-variable oil-market VAR of Kilian and Murphy (2012) and start with a reduced form VAR for $y_t = (\Delta\text{prod}_t, \text{rae}_t, \text{rpo}_t)'$, where Δprod_t is the change in the log of oil production), rae_t is a measure of (world) real economic activity, and rpo_t is the real price of oil. The data for these three series have been taken from Kilian and Murphy (2012). We fit a reduced form VAR with $p = 12$ lags over the sample period 1973M02 to 2008M09 as in Kilian and Murphy (2012).¹⁰

We identify three structural shocks: An adverse oil supply shock ε_t^s , an aggregate demand shock ε_t^d and an oil-specific demand shock ε_t^{od} . The oil supply shock ε_t^s will be identified by an external instrument m_t . As an instrument we use a monthly series of exogenous oil price

¹⁰We have also tried a lag length of $p = 24$ as in Kilian and Murphy (2012) and $p = 3$ as suggested by the Akaike information criterion. The results are very similar to those reported for $p = 12$.

shocks constructed as in Kilian (2008b). Since the monthly series from Kilian (2008b) was not available to us, we have constructed the proxy series for the oil supply shock using our own calculations based on the US Energy Information Administration (EIA) production data and following the description in Kilian (2008b). We show the corresponding time series plot of the instrument in the lower right hand panel of Figure 1.3.¹¹ Following e.g. Stock and Watson (2012), we measure the strength of this instrument by regressing the instrument m_t on the reduced form VAR(12) residuals \hat{u}_t and by computing the corresponding F -statistic. Using the monthly series, we find an F -statistic of 8.13. This value is much larger than what Stock and Watson (2012) have found for the corresponding quarterly series indicating that the monthly instrument series may be not a weak instrument.¹² Note that we only use observations on the proxy for a sample between 1973M02 to 2004M09, which corresponds to the sample used in Kilian (2008b). Consequently, the instrument and the VAR time series are observed over different sample periods. As explained in Section 1.2.4, our method can easily handle ‘missing data’ on the proxy variable.

The aggregate demand ε_t^d and oil-specific demand shock ε_t^{od} are identified by a set of sign restrictions on their contemporaneous impact. Here we simply use the sign restrictions suggested in Table 1 of Kilian and Murphy (2012). Consequently, in the 3-variable oil market system our identification scheme can be represented by

$$\begin{pmatrix} u_t^{\Delta\text{prod}} \\ u_t^{\text{rea}} \\ u_t^{\text{rpo}} \end{pmatrix} = B \begin{pmatrix} \varepsilon_t^s \\ \varepsilon_t^d \\ \varepsilon_t^{od} \end{pmatrix},$$

where the impact matrix is given as

$$B = \begin{pmatrix} \text{IV} & + & + \\ \text{IV} & + & - \\ \text{IV} & + & + \end{pmatrix}. \quad (1.3.1)$$

The ‘IV’ in the first column of (1.3.1) indicates that the supply shock is identified from using m_t as an instrumental variable only, while the signs in column 2 and 3 show the impact sign restrictions for identifying the two demand shocks. No further restrictions have been imposed on the model.

¹¹Visual inspection of the series shows a close resemblance to Figure 7 of Kilian (2008b).

¹²We have rerun this regression using quarterly oil market data and quarterly versions of our exogenous shock series as well as Kilian’s quarterly series. The corresponding F -statistics are 1.1 and 0.05. Thus it seems that the weak instrument problem is a consequence of losing information by converting the external instrument series to quarterly frequency.

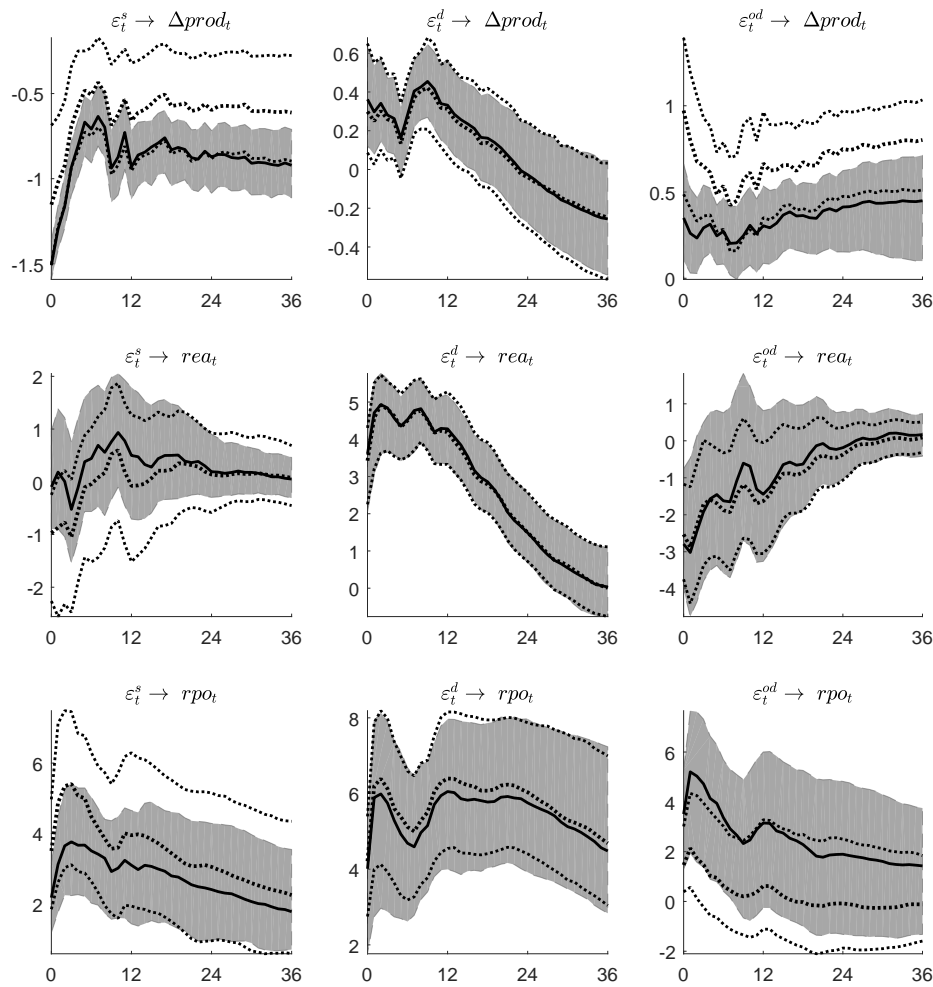


Figure 1.1: Impulse responses in the oil market SVAR. Dotted lines: Identification with sign restrictions only. Solid lines and shaded areas: Identification by IV for oil supply shock ε_t^s and sign restrictions for shocks to aggregate demand ε_t^d and oil specific demand ε_t^{od} . Sample period: 1973M02–2008M09.

Figure 1.1 compares the impulse responses up to horizon $h = 36$ obtained from combining IV and sign restriction identification to results obtained from using sign restrictions alone.¹³ The solid line is the posterior median and the shaded area correspond to 68% posterior credibility sets of impulse responses obtained from the combined IV and sign restriction identification. The dotted lines are obtained from using sign identification only and correspond to the model of Kilian and Murphy (2012).¹⁴

¹³The panels in the first row show accumulated responses on $\Delta prod_t$, and consequently show the effects on the log-level of production.

¹⁴In the sign restriction only identification, we follow Kilian and Murphy (2012) and use $(-, -, +)'$ for the first column of B in equation (1.3.1). Also note that we only use sign restrictions, i.e. no additional restrictions in the form elasticity bounds are imposed. The results shown in Figure 1.1 are based on a VAR with 12 lags.

A number of interesting results emerge. First, we observe that identifying the oil supply shock by the instrumental variable approach changes the magnitude of oil price responses to oil supply and oil-specific demand shocks. Compared to the sign restriction only model, we find a smaller increase of the oil price after an adverse oil supply shock ε_t^s and a more pronounced increase following an oil-specific demand shock ε_t^{od} . This change of impulse response pattern is reasonable according to [Kilian and Murphy \(2012\)](#). In fact, they find similar changes in responses once they impose their elasticity bounds. Second, it is interesting to see that identifying the oil supply shock by the proxy variable has also substantial impact on the oil-specific demand shock. In particular, compared to the sign restriction only model, the demand shock ε_t^{od} also leads to a much less pronounced increase in oil production. Taken together with the larger price response, this implies a much smaller impact elasticity of oil supply with respect to the real price of oil (b_{13}/b_{33}).¹⁵ Given the consensus of a small short-run oil supply elasticity, we find that the IV identification for the oil supply shock shifts the SVAR results to a much more reasonable region even without imposing elasticity bounds.

We also find that replacing the sign restrictions for the supply shock with IV identification leads to a response in economic activity which is no longer significantly different from zero. In other words, using the IV method we do not find support for the sign restriction on real economic activity imposed by [Kilian and Murphy \(2012\)](#). Compared to the sign restriction only model, the adverse supply shock identified by the IV approach also leads to a somewhat sharper drop in oil production. Together with the less pronounced increase in the real price of oil, this implies a somewhat larger demand elasticity (b_{11}/b_{31}).¹⁶ We point out, however, that the implied values are still around the lower bound of -0.8 discussed for quarterly data in [Baumeister and Peersman \(2013\)](#). Finally, we note that the aggregate demand shock ε_t^d is virtually unaffected by using an IV approach for the supply shock.

The oil market example illustrates the usefulness of combining the external instrumental approach for identification with sign restrictions in SVAR analysis. Exploiting the information from an instrument allows to relax some of the sign restrictions while at the same time leading to response patterns that are more in line with evidence on oil price elasticities.

1.3.2 The effects of monetary policy shocks

Since the seminal paper of [Sims \(1980\)](#), the effects of monetary policy shocks on economic activity have been extensively studied using SVAR models (see [Ramey \(2016\)](#) for a recent review of the literature). In the early literature, surprises to monetary policy have been

¹⁵As summary statistics of the posterior distribution of this elasticity we report the 16th and 84th quantile. For the sign only model we find [0.22; 3.14], while for the IV/sign restriction model we find [0.028; 0.42].

¹⁶The 16th and 84th quantile of the posterior distribution of this elasticity is -1.34 and -0.41 in our IV/sign restriction model.

identified by using a Cholesky decomposition of the reduced form VAR covariance matrix, with the policy instrument ordered below the real variables, see e.g. [Christiano et al. \(1999\)](#) or [Bernanke et al. \(2005\)](#). This identifying assumption implies that the central bank can respond instantaneously to movements in the real sector of the economy, while the real variables may only respond with one lag to the policy shock. Such an identification is in line with macroeconomic models subject to nominal rigidities ([Christiano et al.; 2005](#)). However, it is not yet clear how the effects of monetary policy shocks can be identified under the presence of fast moving financial variables such as credit costs or equity prices. A simple recursiveness assumption is unrealistic no matter of the ordering, since it can be assumed that both, monetary policy and financial markets respond immediately to any innovation in the system. Therefore, alternative identification schemes have emerged in recent years that avoid the recursiveness assumption. One strand of the literature draws on sign restrictions on the impulse responses with respect to the policy shock ([Uhlig; 2005](#); [Faust; 1998](#)). These restrictions are derived from conventional wisdom, such that a monetary policy tightening should be associated with an increase in the interest rates but not in consumer prices nor liquidity. Unfortunately, because of the implied set identification, this identification procedure often leads to wide confidence intervals around impulse responses such that results are often not informative enough to allow for policy conclusions. An alternative branch of the literature uses narrative measures of monetary policy shocks for identification. Among the most prominent measures are shock series based on readings of Federal Open Market Committee (FOMC) minutes ([Romer and Romer; 2004](#); [Coibion; 2012](#); [Miranda-Agrippino and Rey; 2018](#)) and factors based on changes in high frequency future prices around FOMC meetings ([Faust et al.; 2004](#); [Gertler and Karadi; 2015](#); [Barakchian and Crowe; 2013](#); [Nakamura and Steinsson; 2018](#)). However, it is a very difficult task to construct convincing exogenous instruments for monetary policy shocks. With respect to [Romer and Romer \(2004\)](#), the authors themselves claim that their series is only ‘relatively free of endogenous and anticipatory movement’. To ensure against remaining endogeneity they exclude the possibility of a contemporaneous response of the macroeconomic variables to the narrative series. The exogeneity of instruments based on high frequency future data is also questionable. [Ramey \(2016\)](#), for example, finds that the main instrument of [Gertler and Karadi \(2015\)](#) suffers from a nonzero mean, significant autocorrelation and predictability by Greenbook forecasts.¹⁷

Our methodology provides a simple framework to combine identification from sign restrictions and such proxy variables. We illustrate that problems arising if either of the methods is used individually are mitigated to some extent. We start with an identification scheme based on sign restrictions similar in spirit to [Uhlig \(2005\)](#). To further narrow down

¹⁷Greenbook forecasts are those published by the central bank in their FOMC minutes and therefore assumed to be in the information set of the central bank.

the set of models we restrict the covariance of the implied structural shock and a narrative proxy series, thereby discarding all models that imply only a loose relation between the SVAR shock and the narrative series. This sharpens inference of the set identified model, while at the same time avoiding the potentially wrong assumption of exogeneity of the proxy series.

For our explorations, we use a monthly VAR for key US macroeconomic variables, i.e. we use a VAR for $y_t = (\text{ip}_t, \text{cpi}_t, \text{nbr}_t, \text{EBP}_t, R_t^s)'$, where ip_t is the log of industrial production, cpi_t the log of the consumer price index, nbr_t the log of non-borrowed reserves, EBP_t the ‘Excess Bond Premium’, a measure of credit market tightness developed by [Gilchrist and Zakrajšek \(2012\)](#), and R_t^s the federal funds rate. Given the availability of the Excess Bond Premium series and the recent period at the zero lower bound of interest rates, we use a sample period from 1973M07 until 2007M12.¹⁸ We include $p = 6$ lags to account for sufficient dynamics of the time series vector. With respect to the narrative series, we use $m_t = \text{rr}_t$, the Romer and Romer (R&R) narrative shock series updated by [Wieland and Yang \(2016\)](#). We focus specifically on the effects of monetary policy shocks on credit markets, which is an aspect difficult to analyze within a recursive identification scheme. Consider the following set of identifying restrictions in our structural analysis, where ε_t^{mp} is the monetary policy shock:

- R1: Sign restrictions on the contemporaneous impulse responses.

$$\frac{\partial E(\text{ip}_t | \Omega_t)}{\partial \varepsilon_t^{mp}} \leq 0, \quad \frac{\partial E(\text{cpi}_t | \Omega_t)}{\partial \varepsilon_t^{mp}} \leq 0, \quad \frac{\partial E(\text{nbr}_t | \Omega_t)}{\partial \varepsilon_t^{mp}} \leq 0, \quad \frac{\partial E(R_t^s | \Omega_t)}{\partial \varepsilon_t^{mp}} \geq 0.$$

R1 imposes that a contractionary monetary policy shock does not have (contemporaneous) positive effects on output, prices and non-borrowed reserves, and no negative effect on the federal funds rate.

- R2: $\text{Corr}(\varepsilon_t^{mp}, \text{rr}_t) > 0$, imposing that ε_t^{mp} is positively correlated with rr_t .
- R3: $\text{Corr}(\varepsilon_t^{mp}, \text{rr}_t) > \text{Corr}(\varepsilon_{jt}, \text{rr}_t)$ for all $j \neq mp$, imposing that the correlation between ε_t^{mp} and rr_t is the largest among all shocks in the SVAR.
- R4: $\text{Corr}(\varepsilon_t^{mp}, \text{rr}_t)^2 > \text{Corr}(\varepsilon_{jt}, \text{rr}_t)^2$ for all $j \neq mp$, imposing that ε_t^{mp} contributes most to the variance of rr_t among all shocks in the SVAR. Note that if $\varepsilon_t^{mp} = \varepsilon_{1t}$, this is equivalent to imposing that $\psi_{i1} > \psi_{ij}, j = 2, \dots, n$ as denoted in Section 1.2.3.
- R5: $E(\varepsilon_{jt}, \text{rr}_t) = 0$ for all $j \neq mp$, which corresponds to identification via instrumental variables.

¹⁸An exact description of data sources and corresponding time series plots are provided in Appendix 1.B.2

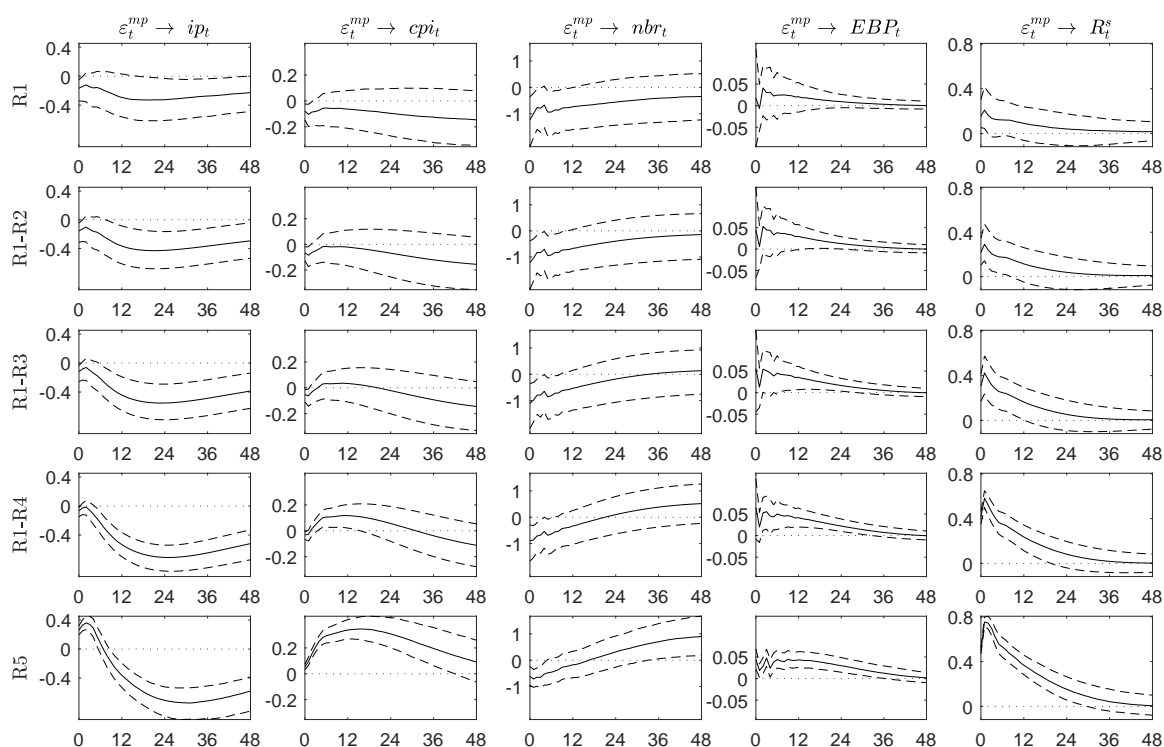


Figure 1.2: Impulse responses in the monetary policy SVAR obtained by using different identifying restrictions. Posterior median (solid line) and 68% posterior credibility sets (dotted lines). Sample period: 1973M07-2007M12.

Restrictions R1 and R5 correspond to identification via sign restrictions and instrumental variables, respectively. Restrictions R2 to R4 will be used in combination with the sign restrictions in order to narrow down the set of admissible models without imposing exogeneity. Note that R4 is typically more restrictive than R3, while R3 is more restrictive than R2. Note that in order to implement R3 and R4, ε_{jt} for $j \neq mp$ need to be identified. Since we do not attempt to identify these shocks economically, we do this in a statistical sense and let them be randomly identified through the rotations generated under the uniform prior on Q .

Figure 1.2 provides the impulse responses obtained from using five different identification schemes for a horizon up to four years. The first row corresponds to scheme R1, that is identifying the policy shock solely with sign restrictions. In line with the prior about the impulse response function, unexpected tightening is associated with a decrease in industrial production, prices, liquidity and an increase in the federal funds rate. However, with exception of output, these effects are almost never significant. With respect to the effects of ε_t^{mp} on the credit market (EBP_t), nothing can be said since the zero line is contained in the 68% posterior credibility set at any point of time. The last row of Figure 1.2 shows the responses from identification scheme R5, which corresponds to a Bayesian proxy SVAR based on pure instrumental variable identification. Some puzzling results emerge in the

impulse response functions. First, output increases significantly for about 6 months in response to a policy tightening, which is certainly at odds with macroeconomic theory. Furthermore, consumer prices rise sharply and remain significant for up to 3 years. Such a ‘price puzzle’ is frequently found in recursively identified VARs and often attributed to an omitted variable bias (Christiano et al.; 1999). Sometimes, including commodity prices in the VAR can mitigate the puzzle but unfortunately this is not the case for our specification. Ramey (2016) finds similar responses based on a proxy SVAR with the R&R shocks as instruments. She finds that including more information in the VAR does not solve the puzzles and argues that they are likely to be caused by endogeneity of the instrument. This would certainly invalidate the use of the R&R shock as an instrumental variable, however, does not pose a problem in our framework.

Rows two, three and four correspond to combining sign restrictions with information of the R&R series, however, without the need of its exogeneity. Each model adds an additional restriction and therefore is based on tighter constraints by construction, always narrowing down the set of admissible models to a somewhat larger extent. We find that adding a simple sign restriction on the correlation (R2) does not change the impulse response functions substantially. Minor effects are found in the persistence of the interest rates and the significance of the response of industrial production. Unfortunately, still nothing can be said on the effects on credit market tightness, which remains insignificant. This implies that the set of monetary policy shocks of the sign restriction only model are already positively correlated with the R&R shock which is why adding R2 does not sharpen inference. However, additionally including restriction R3 makes a considerable difference (row 3 of Figure 1.2). The effects on industrial production and the interest rate are more pronounced, while for prices and liquidity no significant effects are found. With respect to credit markets, the response is positive and a significant tightening is found after six months. As expected, adding restrictions R4 (in addition to R2 and R3) further discards some models, retaining only those where the monetary policy shock additionally contributes most to the variance of the R&R series. The impulse responses in row 4 of Figure 1.2 indicate that the effects on industrial production and interest rates have similar magnitudes as in the point identified proxy SVAR (R5), with the main difference that the former variable is not subject to the increase in the first months. Also the price puzzle is found to be less pronounced and the increase is insignificant in most of the periods. Interestingly, the response of the excess bond premium shows a significant tightening of the credit market similar in magnitude to the proxy SVAR.

In Figure 1.5 of Appendix 1.C, we provide additional plots of impulse response functions that arise if only R2, R2-R3 and R2-R4 are used to identify the model. We find that R2 and R2-R3 in isolation yield too little information to draw meaningful economic conclusions.

Further including R4 tightens the identified set by a large extent, however, the same puzzling positive short run response on output as well as a pronounced price puzzle arise as in point-identification by IV (R5). Therefore, we believe that the combination with sign restrictions (R1) might be more sensible in this scenario since economic theory allows to exclude this patterns a priori.

Summarizing our results, we illustrate that combining sign restrictions with restrictions based on external narrative series can sharpen inference substantially and lead to economically meaningful results. In this context, we also propose a set of restrictions which differ in their strength but are all automatic in a sense that no threshold value (on e.g. correlations) has to be chosen by the researcher. From an empirical point of view, we find that based on our identification scheme, there is evidence for a significant tightening of credit markets in response to a monetary policy shock. This highlights the importance of a credit channel in the transmission mechanism. Such evidence supports the finding of [Gertler and Karadi \(2015\)](#) and implies that theoretical models should pay special attention to this feature.

1.4 Conclusion

For the identification of structural shocks within SVAR models, we suggest to combine sign restrictions with the information in time series that act as proxy variables for the underlying structural shocks. We propose an econometric framework that incorporates the information on the proxy variables by augmenting the SVAR with equations that relate the proxy variables to the structural shocks.

Our econometric framework is fairly general in the following sense: First, the framework allows to simultaneously identify different shocks using either sign restrictions (set-identification) or an external instrument approach (point-identification), always ensuring that all structural shocks are orthogonal. Second, the setup also allows to (set) identify a single shock by combining sign restrictions and the information of external proxy variables, without assuming their exogeneity. Compared to a pure sign restriction approach, the additional information from the external proxy series may help to narrow down the set of admissible models and leads to sharper results, in e.g. impulse response analysis. We essentially discard models that imply structural shocks that have no close relation to the external proxy time series. We measure this relation e.g. by the variance contribution of the shock. Third, the setup also nests the pure sign restriction approach and the pure external instrument variable case. Estimation and inference is done in a full Bayesian setup, which accounts for both, model and estimation uncertainty. An additional advantage of the Bayesian setup is that the inference framework requires no modifications to handle the case of weak instruments and allows for straightforward handling of missing data.

We illustrate the usefulness of our method in two empirical applications. In the first application, we use a standard oil market model from the literature and identify simultaneously an oil supply shock using an external proxy variable as an instrument, and two demand shocks by using sign restrictions. Employing this identification leads to impulse responses that imply much more reasonable oil supply elasticities than a pure sign restriction model. In the second application, we analyze the effects of monetary policy shocks focusing on the credit market. We find that using sign restrictions alone is not informative enough with wide error bands around the response of the excess bond premium. Once we combine the sign restrictions with external information coming from a proxy variable of monetary policy shocks, we find a significant tightening of credit markets.

Overall, our paper suggests that combining sign restrictions and external proxy variables for structural shock identification is a promising way to sharpen results from SVAR models.

References

- Angelini, G. and Fanelli, L. (2018). Identification and estimation issues in structural vector autoregressions with external instruments, *Working Papers 1122*, Dipartimento Scienze Economiche, Università di Bologna.
- Antolín-Díaz, J. and Rubio-Ramírez, J. F. (2018). Narrative sign restrictions for SVARs, *American Economic Review* **108**(10): 2802–29.
- Arias, J. E., Caldara, D. and Rubio-Ramírez, J. F. (2019). The systematic component of monetary policy in SVARs: an agnostic identification procedure, *Journal of Monetary Economics* **101**: 1–13.
- Arias, J. E., Rubio-Ramírez, J. F. and Waggoner, D. F. (2018). Inference based on structural vector autoregressions identified with sign and zero restrictions: Theory and applications, *Econometrica* **86**(2): 685–720.
- Barakchian, S. M. and Crowe, C. (2013). Monetary policy matters: Evidence from new shocks data, *Journal of Monetary Economics* **60**(8): 950–966.
- Baumeister, C. and Hamilton, J. D. (2015). Sign restrictions, structural vector autoregressions, and useful prior information, *Econometrica* **83**(5): 1963–1999.
- Baumeister, C. and Peersman, G. (2013). The role of time-varying price elasticities in accounting for volatility changes in the crude oil market, *Journal of Applied Econometrics* **28**(7): 1087–1109.
- Ben Zeev, N. (2018). What can we learn about news shocks from the late 1990s and early 2000s boom-bust period?, *Journal of Economic Dynamics and Control* **87**: 94–105.
- Bernanke, B. S., Boivin, J. and Elias, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach, *The Quarterly Journal of Economics* **120**(1): 387–422.
- Bertsche, D. and Braun, R. (2018). Identification of structural vector autoregressions by stochastic volatility, *Working Paper Series of the Department of Economics, University of Konstanz 2018-03*, Department of Economics, University of Konstanz.
- Caldara, D. and Herbst, E. (forthcoming). Monetary policy, real activity, and credit spreads: Evidence from Bayesian proxy SVARs, *American Economic Journal: Macroeconomics* .
- Canova, F. and De Nicoló, G. (2002). Monetary disturbances matter for business fluctuations in the G-7, *Journal of Monetary Economics* **49**: 1131–1159.

- Christiano, L. J., Eichenbaum, M. and Evans, C. L. (1999). Monetary policy shocks: What have we learned and to what end?, *Handbook of Macroeconomics*, Vol. 1, Elsevier, chapter 2, pp. 65–148.
- Christiano, L. J., Eichenbaum, M. and Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy, *Journal of Political Economy* **113**(1): 1–45.
- Coibion, O. (2012). Are the effects of monetary policy shocks big or small?, *American Economic Journal: Macroeconomics* **4**(2): 1–32.
- Conley, T. G., Hansen, C. B. and Rossi, P. E. (2012). Plausibly exogenous, *Review of Economics and Statistics* **94**(1): 260–272.
- Drautzburg, T. (2016). A narrative approach to a fiscal DSGE model, *Working Paper 16-11*, Federal Reserve Board Philadelphia.
- Faust, J. (1998). The robustness of identified VAR conclusions about money, *Carnegie-Rochester Conference Series on Public Policy* **49**: 207–244.
- Faust, J., Swanson, E. T. and Wright, J. H. (2004). Identifying VARs based on high frequency futures data, *Journal of Monetary Economics* **51**(6): 1107–1131.
- Fry, R. and Pagan, A. (2011). Sign restrictions in structural vector autoregressions: a critical review, *Journal of Economic Literature* **49**: 938–960.
- Gertler, M. and Karadi, P. (2015). Monetary policy surprises, credit costs, and economic activity, *American Economic Journal-Macroeconomics* **7**(1): 44–76.
- Geweke, J. (2005). *Contemporary Bayesian econometrics and statistics*, Vol. 537, John Wiley & Sons.
- Giacomini, R. and Kitagawa, T. (2015). Robust inference about partially identified SVARs, *Manuscript, University College London* .
- Gilchrist, S. and Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations, *American Economic Review* **102**(4): 1692–1720.
- Hamilton, J. D. (2003). What is an oil shock?, *Journal of econometrics* **113**(2): 363–398.
- Kilian, L. (2008a). The economic effects of energy price shocks, *Journal of Economic Literature* **46**(4): 871–909.
- Kilian, L. (2008b). Exogenous oil supply shocks: How big are they and how much do they matter for the U.S. economy?, *The Review of Economics and Statistics* **90**(2): 216–240.

- Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market, *American Economic Review* **99**(3): 1053–1069.
- Kilian, L. and Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*, Themes in Modern Econometrics, Cambridge University Press.
- Kilian, L. and Murphy, D. P. (2012). Why agnostic sign restrictions are not enough: Understanding the dynamics of oil market VAR models, *Journal of the European Economic Association* **10**(5): 1166–1188.
- Ludvigson, S. C., Ma, S. and Ng, S. (2017). Shock restricted structural vector-autoregressions, *Working Paper 23225*, National Bureau of Economic Research.
- Lütkepohl, H. and Netšunajev, A. (2017). Structural vector autoregressions with heteroskedasticity: A review of different volatility models, *Econometrics and Statistics* **1**: 2–18.
- Mertens, K. and Ravn, M. O. (2012). Empirical evidence on the aggregate effects of anticipated and unanticipated US tax policy shocks, *American Economic Journal-Economic Policy* **4**(2): 145–181.
- Mertens, K. and Ravn, M. O. (2013). The dynamic effects of personal and corporate income tax changes in the United States, *American Economic Review* **103**(4): 1212–1247.
- Mertens, K. and Ravn, M. O. (2014). A reconciliation of SVAR and narrative estimates of tax multipliers, *Journal of Monetary Economics* **68**: 1–19.
- Miranda-Agrippino, S. and Rey, H. (2018). US monetary policy and the global financial cycle, *Working Paper 21722*, National Bureau of Economic Research.
- Montiel-Olea, J. L., Stock, J. H. and Watson, M. W. (2015). Inference in SVARs identified with external instruments, *Harvard Manuscript* .
- Montiel-Olea, J. L., Stock, J. H. and Watson, M. W. (2016). Uniform inference in SVARs identified with external instruments, *Harvard Manuscript* .
- Nakamura, E. and Steinsson, J. (2018). High frequency identification of monetary non-neutrality: The information effect, *Quarterly Journal of Economics* (forthcoming).
- Nevo, A. and Rosen, A. M. (2012). Identification with imperfect instruments, *Review of Economics and Statistics* **94**(3): 659–671.
- Peersman, G. and Van Robays, I. (2012). Cross-country differences in the effects of oil shocks, *Energy Economics* **34**(5): 1532–1547.

- Ramey, V. (2016). Macroeconomic shocks and their propagation, *Handbook of Macroeconomics*, Vol. 2, Elsevier, chapter 2, pp. 71–162.
- Romer, C. D. and Romer, D. H. (1989). Does monetary policy matter - a new test in the spirit of Friedman and Schwartz, *NBER Macroeconomics Annual* **4**: 121–170.
- Romer, C. D. and Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications, *American Economic Review* **94**(4): 1055–1084.
- Rubio-Ramírez, J., Waggoner, D. and Zha, T. (2010). Structural vector autoregressions: Theory of identification and algorithms for inference, *The Review of Economic Studies* **77**(2): 665–696.
- Sims, C. A. (1980). Macroeconomics and reality, *Econometrica* **48**(1): 1–48.
- Stock, J. H. and Watson, M. W. (2012). Disentangling the channels of the 2007-09 recession, *Brookings Papers on Economic Activity* **43**(1): 81–156.
- Stock, J. H. and Watson, M. W. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics, in J. B. Taylor and H. Uhlig (eds), *Handbook of Macroeconomics*, Vol. 2, Elsevier, chapter 8, pp. 415–525.
- Uhlig, H. (2005). What are the effects of monetary policy on output? Results from an agnostic identification procedure, *Journal of Monetary Economics* **52**: 381–419.
- Uhrin, G. B. and Herwartz, H. (2016). Monetary policy shocks, set-identifying restrictions, and asset prices: A benchmarking approach for analyzing set-identified models, *Technical report*, University of Goettingen, Department of Economics.
- Wieland, J. F. and Yang, M.-J. (2016). Financial dampening, *Working Paper 22141*, National Bureau of Economic Research.

Appendix 1.A Gibbs sampler for missing data

Assume that some of the observations m_s are missing at random where s denotes the period of missing data. Instead of simply discarding all values of y_s , we advocate to add an additional step in the Gibbs sampler which imputes the missing values. For this purpose, we need to derive the conditional distribution $p(m_s|Y, M, \Sigma, A)$. Recall the joint likelihood of $\{y_t, m_t\}$:

$$y_t = Ax_t + u_t, \quad \begin{pmatrix} u_t \\ m_t \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \Sigma_u & \Sigma_{mu} \\ \Sigma_{mu'} & \Sigma_m \end{pmatrix}\right).$$

Using standard results of multivariate statistics, the conditional distribution of m_t is given as $p(m_t|Y, M, \Sigma, A) \sim \mathcal{N}(\bar{m}_t, \bar{V}_t)$ where

$$\begin{aligned} \bar{m}_t &= \Sigma_{mu'} \Sigma_u^{-1} (y_t - Ax_t), \\ \bar{V}_t &= \Sigma_m - \Sigma_{mu'} \Sigma_u^{-1} \Sigma_{mu}. \end{aligned}$$

Therefore, a modified Gibbs sampler involves drawing iteratively from the following blocks to generate draws of the reduced form parameters:

1. Draw the autoregressive coefficients from $\alpha|\tilde{y}, \Sigma \sim \mathcal{N}(\bar{\alpha}, \bar{V})$ as in Section 1.2.4.
2. Draw the inverse of the covariance matrix $\Sigma^{-1}|\tilde{y}, \alpha \sim \mathcal{W}((\Psi + S)^{-1}, d + T)$ as in Section 1.2.4.
3. Impute all the missing values m_s of the proxy variables by drawing each of them from a normal distribution $m_s|\tilde{y}, \Sigma, A \sim \mathcal{N}(\bar{m}_s, \bar{V}_s)$ where

$$\begin{aligned} \bar{m}_s &= \Sigma_{mu'} \Sigma_u^{-1} (y_s - Ax_s), \\ \bar{V}_s &= \Sigma_m - \Sigma_{mu'} \Sigma_u^{-1} \Sigma_{mu}. \end{aligned}$$

Appendix 1.B Data

1.B.1 Data for the oil price shock example

The data for the oil market VAR have been taken from the web appendix S2 in [Kilian and Murphy \(2012\)](#). A quarterly time series for the ‘exogenous’ oil price shock is available on Lutz Kilian’s homepage. Since a corresponding time series on the monthly frequency is not readily available, we have constructed it from oil production data of the EIA following exactly the approach described in [Kilian \(2008b\)](#). A time series plot of the variables is shown in Figure 1.3.

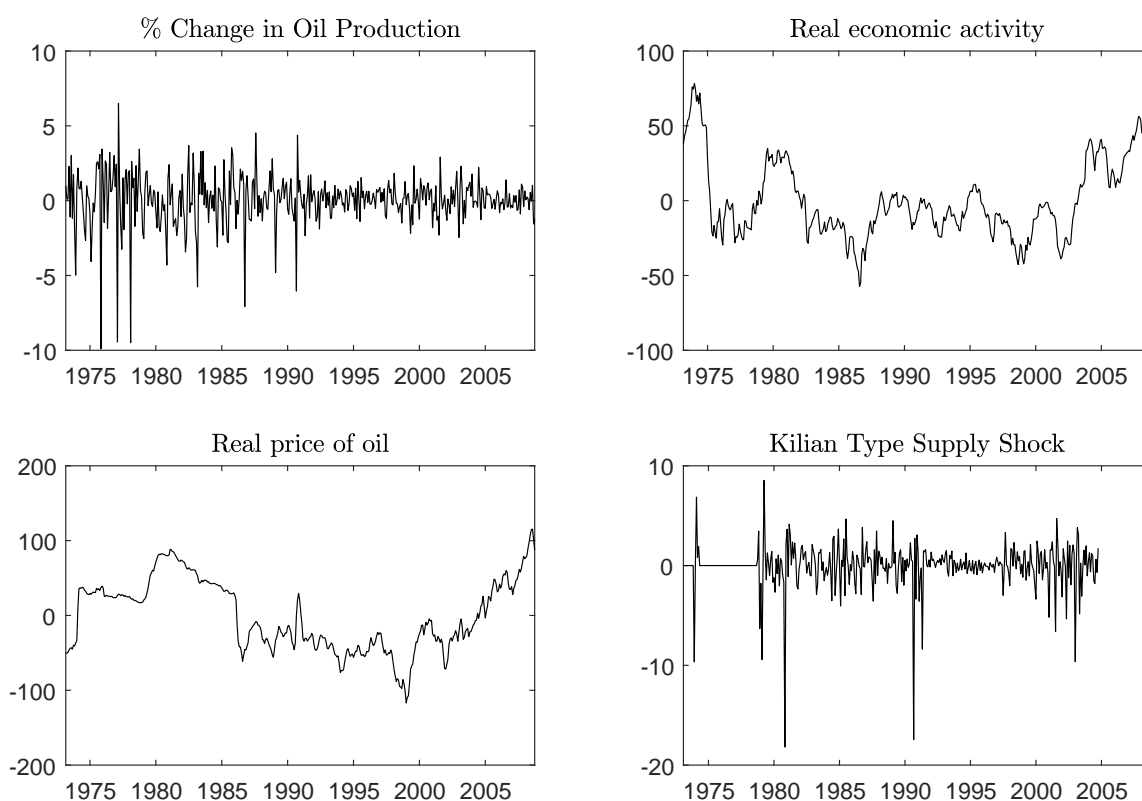


Figure 1.3: Time series plots of oil market variables. Sample: 1973M01-2008M09.

1.B.2 Data for the monetary policy shock example

The time series of the monetary policy VAR were obtained from the following sources. Industrial production, the consumer price index and the federal funds rate were obtained from FRED with series id INDPRO, CPIAUCSL and FEDFUNDS, respectively. The data of the Excess Bond Premium and the Romer and Romer shock were obtained from Valerie Ramey’s homepage and are part of the replication files of her recent chapter in the Handbook

of Macroeconomics (Ramey; 2016). A time series plot for each of the series is given in Figure 1.4.

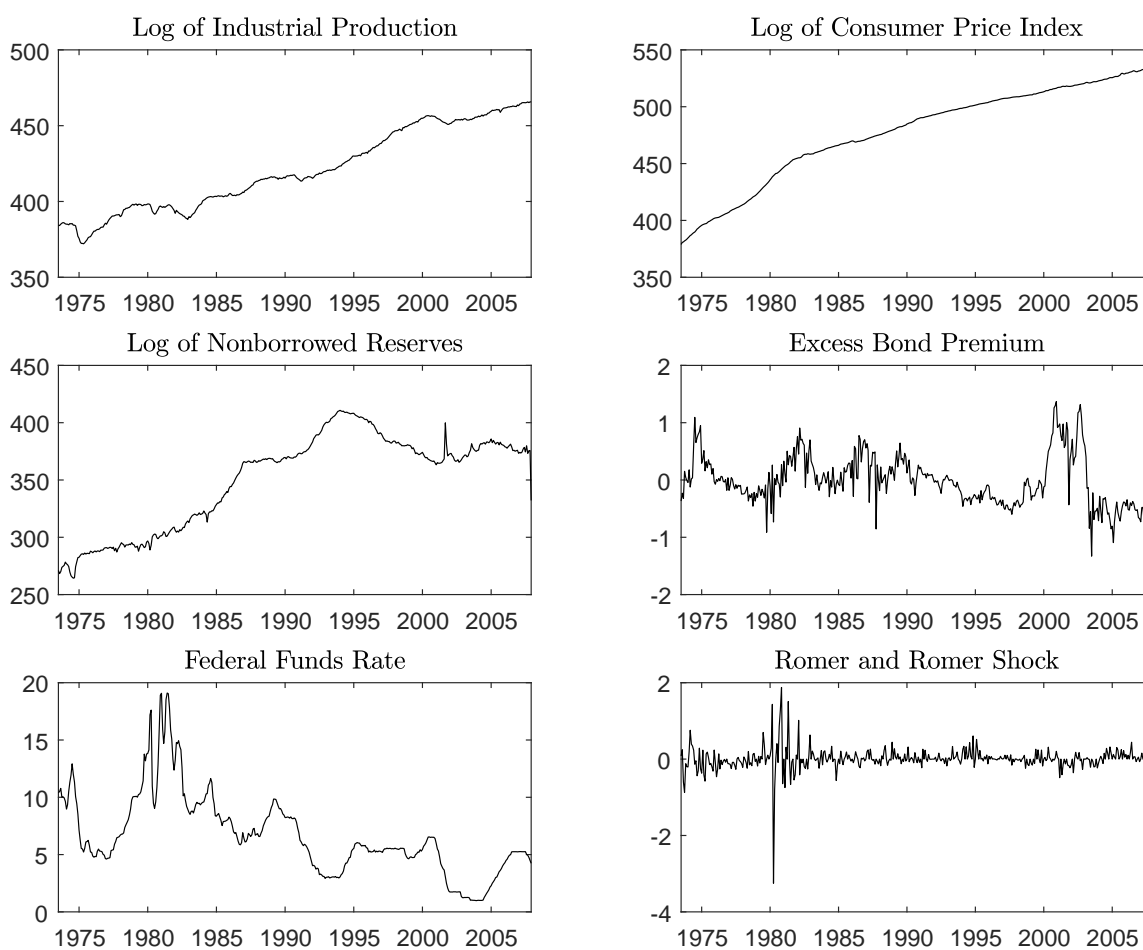


Figure 1.4: Time series plots of monetary policy variables. Sample: 1973M07-2007M12.

Appendix 1.C Additional figures

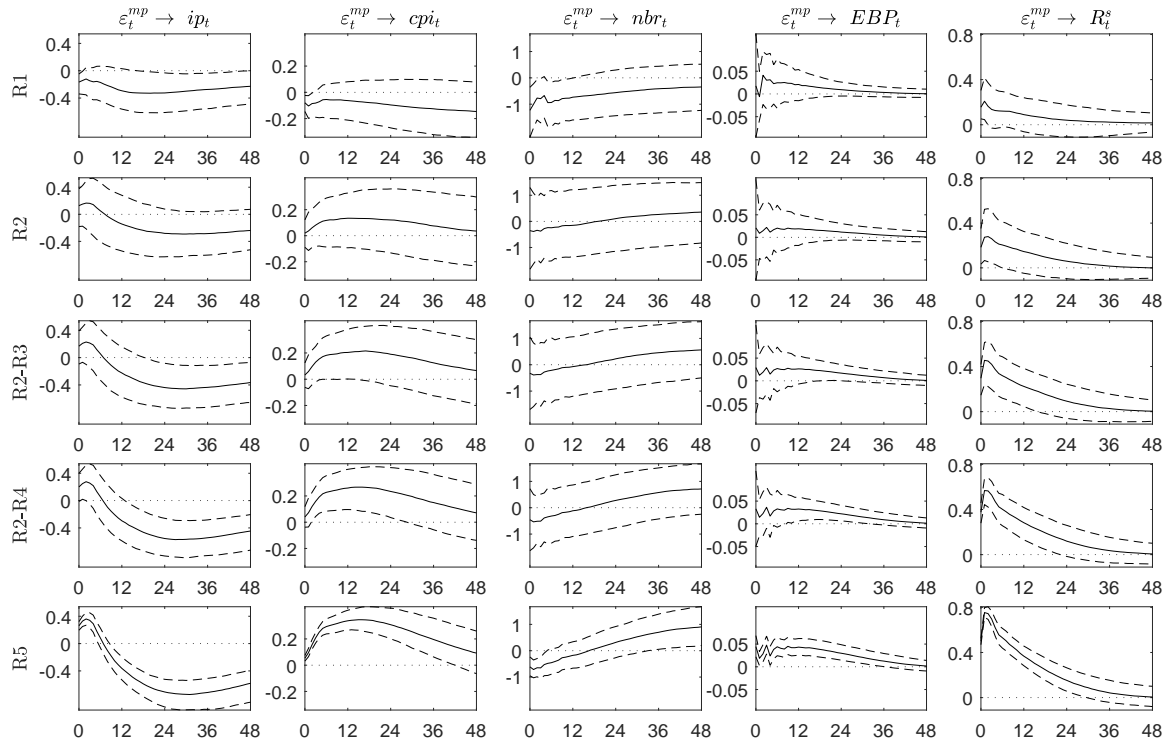


Figure 1.5: Impulse responses in the monetary policy SVAR obtained by using different identifying restrictions. The rows (top to bottom) of the figure show results from using the restriction sets R1, R2, R2+R3, R2+R3+R4, and R5, respectively (see Section 1.3.2 for details). Posterior median (solid line) and 68% posterior credibility sets (dotted lines). Sample period: 1973M07-2007M12.

Chapter **2**

Identification of Structural Vector
Autoregressions by Stochastic Volatility

2.1 Introduction

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

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

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

Closely related to our paper is the work of [Carreiro et al. \(2019\)](#) who also exploit a SV model to identify uncertainty shocks in SVARs. The model of [Carreiro et al. \(2019\)](#) allows for affects of volatility on the conditional mean, and their inference is conducted within a Bayesian framework. In turn, in this paper we focus on the identification of a classical SVAR model and estimate the model with Maximum Likelihood.

Using a stochastic volatility model for the variance of structural shocks is an attractive specification for many reasons. First, SV models enjoy increasing popularity in theoretical and empirical macroeconomics. For example, [Justiniano and Primiceri \(2008\)](#) and [Fernández-](#)

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

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

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

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

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

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

The paper is structured as follows. Section 2.2 introduces the SVAR model with stochastic volatility and discusses under which conditions the structural parameters are identified. Section 2.3 considers Maximum Likelihood estimation and reviews a procedure to test for identification. In Section 2.4, we present simulation evidence while in Section 2.5 we apply the proposed model to study the interdependence between US monetary policy and stock markets. Section 2.6 concludes.

2.2 Identification of SVARs by stochastic volatility

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

$$y_t = \nu + \sum_{j=1}^p A_j y_{t-j} + u_t, \quad (2.2.1)$$

$$u_t = BV_t^{\frac{1}{2}} \eta_t, \quad (2.2.2)$$

where $\eta_t \sim (0, I_K)$ is assumed to be a white noise error term. Equation (2.2.1) corresponds to a standard reduced form VAR(p) model for y_t , capturing common dynamics across the time series data by a linear specification. Here, A_j for $j = 1, \dots, p$ are $K \times K$ matrices of autoregressive coefficients and ν is a $K \times 1$ vector of intercepts. Since we only consider stable time series throughout the paper, we assume:

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

Equation (2.2.2) models the structural part and is set up as a B -model in the terminology of Lütkepohl (2005). The reduced form error terms u_t are decomposed into a linear function of K structural shocks $\varepsilon_t = V_t^{\frac{1}{2}} \eta_t$, with B a $K \times K$ invertible contemporaneous impact matrix and $V_t^{\frac{1}{2}}$ a stochastic diagonal matrix with strictly positive elements capturing potential heteroskedasticity and/or non-normality in each structural shock. This specification yields a time-varying covariance matrix of the reduced form errors u_t given as $\Sigma_t = E(u_t u_t') = BV_t B'$. Throughout the paper, we assume that there are $r \leq K$ heteroskedastic shocks which are ordered such that they appear first in the vector ε_t . To model the time varying second moment of these shocks, we specify an independent Gaussian AR(1) log stochastic volatility model for each of the r heteroskedastic components:

$$V_t = \begin{bmatrix} \text{diag}(\exp([h_{1t}, \dots, h_{rt}]')) & 0 \\ 0 & I_{K-r} \end{bmatrix}, \quad (2.2.3)$$

$$h_{it} = \mu_i + \phi_i(h_{i,t-1} - \mu_i) + \sqrt{s_i} \omega_{it}, \quad \text{for } i = 1, \dots, r, \quad (2.2.4)$$

where $\omega_{it} \sim \mathcal{N}(0, 1)$ and $E(\varepsilon_t' \omega_t) = 0$ for $\omega_t = [\omega_{1t}, \dots, \omega_{rt}]'$. Furthermore, the initial states are assumed to be initialized from the unconditional distribution $h_{i1} \sim \mathcal{N}(\mu_i, s_i / (1 - \phi_i^2))$. Note that the proposed model for equation (2.2.2) is very similar to the Generalized Orthogonal GARCH (GO-GARCH) model of Van der Weide (2002) and Lanne and Saikkonen (2007), with the major difference in the specification (2.2.3)-(2.2.4) of V_t . While for the GO-GARCH the first r diagonal components are modeled by deterministic GARCH(1,1) processes,

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

$$\gamma_i(\tau) = \text{Cov}(\varepsilon_{it}^2, \varepsilon_{i,t+\tau}^2) = \exp(2\mu_i + \sigma_{h_i}^2)(\exp(\sigma_{h_i}^2 \phi_i^\tau) - 1) \quad (2.2.5)$$

$$\kappa_i = \frac{E(\varepsilon_{it}^4)}{(\varepsilon_{it}^2)^2} = E(\eta_{it}^4) \exp(\sigma_{h_i}^2), \quad (2.2.6)$$

$$E(\varepsilon_{it}^2) = E(\exp(h_{it})\eta_{it}^2) = E(\exp(h_{it}))E(\eta_{it}^2) = E(\exp(h_{it})) = \exp\left(\mu_i + \frac{1}{2}\sigma_{h_i}^2\right), \quad (2.2.7)$$

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

The model is able to capture the main stylized facts of structural shocks that are typically encountered in empirical SVAR analysis. First, heteroskedasticity can be modeled by setting $\phi_i > 0$. The respective autocovariance function in the second moment of ε_{it} is given by equation (2.2.5), displaying an exponential decay ϕ_i . This autocovariance function has been found to be very flexible enabling to capture a large variety of heteroskedasticity patterns, an argument that we can confirm based on our simulation evidence. Second, the model can capture heavy tailed errors and the respective kurtosis function κ_i can be decomposed into a part that is due to the kurtosis of the standardized structural shocks η_{it} and a component which inflates the value depending on the underlying SV parameters. That is, given a conditional Gaussian error distribution in ε_{it} , excess kurtosis kicks in as soon as the SV process is nontrivial, that is $s_i > 0$. This means that even if the shock is homoskedastic ($\phi_i = 0$), the model is still able to capture heavy tails under conditional Gaussianity. In this particular case, the structural error would be independent and identically distributed following a mixture of log-normal and Gaussian errors.² We argue that this is a key advantage with respect to a model from the GARCH family, which are generally unable to generate homoskedastic shocks featuring excess kurtosis given the assumption of conditionally Gaussianity. Finally, equation (2.2.7) gives the unconditional scale of the structural shocks as a function of the underlying SV parameters.

In the following, we will use equations (2.2.5)-(2.2.7) to discuss identification in detail. First, note that the structural shocks are latent variables and a unique scaling must be obtained. For this purpose, we follow the widely used normalization of setting the scale to $E(\varepsilon_t \varepsilon_t') = I_K$. Using equation (2.2.7), this can be achieved by restricting the mean of the AR(1) processes

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

to $\mu_i = -0.5s_i/(1 - \phi_i^2)$. It follows that the structural parameters in B are related to the unconditional reduced form covariance matrix by:

$$E(u_t u_t') = \Sigma_u = BB'. \quad (2.2.8)$$

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

Due to the symmetry of the covariance matrix, identification in the SV-SVAR model cannot be discussed based on equation (2.2.8) solely. For that purpose, we rely on Lewis (2018) who treats identification by time varying volatility in a more general context requiring no specific functional forms. In particular, identification can be analyzed based on the lag τ autocovariance in the squared reduced form residuals $\xi_t = \text{vech}(u_t u_t')$. This function takes the following form (Lewis; 2018):

$$\text{Cov}(\xi_t, \xi_{t+\tau}) = L_K(B \otimes B)G_K M_\tau G_K'(B \otimes B)'L_K' \quad (2.2.9)$$

where L_K is the elimination matrix such that $\text{vech}(A) = L_K \text{vec}(A)$, G_K is a selection matrix with zeros and ones such that $\text{vec}(D) = G_K d$ for $D = \text{diag}(d)$ and $M_\tau = \text{diag}(\gamma_1(\tau), \dots, \gamma_r(\tau), 0_{K-r})$.³ Note that one autocovariance has $\sum_{i=1}^5 \binom{i+K-3}{K-2}$ unique elements ($K \geq 2$), while the structural model contains K^2 in B and r autocovariances in $\gamma_i(\tau)$, implicitly parameterized nonlinearly by the underlying SV processes. Lewis (2018) proves general identification of the elements in M_τ and B under the restriction that the diagonal of B is fixed at unity. In order to account for the standard deviation normalization implied by (2.2.8), a modification is necessary. In Proposition 1, we summarize identification of B for any $r \leq K$ for our setting.

Proposition 1. *Let $B = [B_1, B_2]$ with $B_1 \in \mathbb{R}^{K \times r}$, $B_2 \in \mathbb{R}^{K \times (K-r)}$ and. Assume the stable SV-SVAR model presented above with $|\phi_i| < 1$, $\phi_i \neq 0$ and $0 < s_i < \infty$ for $i = 1, \dots, r$, implying that equations (2.2.8) and (2.2.9) hold. Then, matrix B_1 is unique up to permutation and sign switches.*

Proof. See Appendix 2.A.2.

In fact, it is not necessary that $r = K$ shocks are heteroskedastic in order that the impact matrix is identified. The orthogonality constraints implied by equation (2.2.8) yield enough structure to fully identify the model in case of $r = K - 1$, which is summarized in Corollary 1.

³See also Appendix 2.A.1 for a derivation of this function for the SV-SVAR model.

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

Proof. See Appendix 2.A.3.

The presented results are broadly in line with those provided by Lewis (2018). However, our results deviate in the sense that identification is given also under $r = K - 1$ heteroskedastic shocks, based on the additional information provided by (2.2.8). Furthermore, the simple structure assumed for the SV-SVAR allows for a much simpler proof.

At this point we highlight that identification of the model can also be discussed based on non-Gaussianity implied by the SV model. If one is willing to assume mutual independence in ε_{it} , the SV-SVAR model as discussed in this paper is covered by the general framework of Lanne et al. (2017). Specifically, the structural parameters in B are identified up to permutation and sign if the structural shocks are strictly stationary with finite second moments, mutually independent and with at most one Gaussian component. For the SV-SVAR model, this means that in order to achieve strict stationarity and finite second moments, we need $s_i < \infty$ and $|\phi_i| < 1 \forall i$ as discussed above. Furthermore, under conditionally Gaussian errors, at most one structural shock can display a degenerate SV process with $s_i = 0$, implying a Gaussian marginal. Analogous results regarding to partial identification are available in Maxand (forthcoming). As in proposition 2.A.2, the structural parameters associated with the non-Gaussian shocks are locally identified up to permutation and sign-changes.

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

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

$$A_h h_i = \mu_i, i = 1, \dots, r, \quad (2.2.10)$$

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

implying that $E(u_1 u_1') = BB'$ which is typically used to identify the scaling in Markov Switching SVAR models (Lanne et al.; 2010; Herwartz and Lütkepohl; 2014). However, this would require that we leave μ_i unrestricted implying an additional parameter to estimate, which is why we prefer restriction (2.2.10).

2.3 Maximum Likelihood estimation

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

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

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

2.3.1 Evaluation of the likelihood

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

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

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

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

$$p(\varepsilon_i|\theta) = \int p(\varepsilon_i|\theta, h_i)p^c(h_i|\theta)dh_i. \quad (2.3.1)$$

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

To evaluate this integral, we use an importance sampling estimator. Therefore, let $q(h_i)$ be a proposal distribution from which independent random draws $h_i^{(1)}, \dots, h_i^{(R)}$ can be generated, and further let $q(h_i)$ dominate $p(\varepsilon_i|\theta, h_i)p^c(h_i|\theta)$. An unbiased importance sampling estimator of the integral in equation (2.3.1) is:

$$\widehat{p(\varepsilon_i|\theta)} = \frac{1}{R} \sum_{j=1}^R \frac{p(\varepsilon_i|\theta, h_i^{(j)})p^c(h_i^{(j)}|\theta)}{q(h_i^{(j)})}. \quad (2.3.2)$$

Plugging (2.3.2) into the SV-SVAR log-likelihood yields an IS estimator of the SV-SVAR log-likelihood function:

$$\widehat{\mathcal{L}(\theta)} = -T \log |B| + \sum_{i=1}^r \log \widehat{p(\varepsilon_i|\theta)} + \sum_{i=r+1}^K \log p(\varepsilon_i|\theta). \quad (2.3.3)$$

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

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

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

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

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

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

$$\log p(\varepsilon_{it} | \theta, h_{it}) \approx \log p(\varepsilon_{it} | \theta, \tilde{h}_{it}^{(0)}) + b_{it} h_{it} - \frac{1}{2} c_{it} h_{it}^2, \quad (2.3.4)$$

where b_{it} and c_{it} depend on $\tilde{h}_{it}^{(0)}$. Based on the linearized kernel, an approximate smoothing distribution $\pi_G(h_i | \theta, \varepsilon_i)$ takes the form of a Normal distribution with precision matrix $\bar{Q}_i = Q_i + C_i$ and mean $\bar{\delta}_i = \bar{Q}_i^{-1} (b_i + Q_i \delta_i)$, where $C_i = \text{diag}([c_{i1}, \dots, c_{iT}]')$ and $b_i = [b_{i1}, \dots, b_{iT}]'$. The T -dimensional density has a tridiagonal precision matrix which allows for fast generation of random samples and likelihood evaluation. The approximation is evaluated at the mode of the smoothing distribution obtained by a Newton-Raphson method that typically converges in few iterations. Details on the Newton-Raphson method and on explicit expressions for b_{it} and c_{it} are given in Appendix 2.B.1.

As discussed in Section 2.2, our prior density for h_i is subject to the normalizing constraint $A_h h_i = \mu_i$. Therefore, the IS density $\pi_G(h_i | \theta, \varepsilon_i)$ needs a slight modification to account for this linear constraint. In particular, an application of Bayes' theorem yields a constraint density $\pi_G^c(h_i | \theta, \varepsilon_i)$ which is also Gaussian but has mean and covariance:

$$\bar{\delta}_i^c = \bar{\delta}_i - \bar{Q}_i^{-1} A_h' (A_h \bar{Q}_i^{-1} A_h')^{-1} (A_h \bar{\delta}_i - \mu_i), \quad (2.3.5)$$

$$\text{Cov}(h_i | \theta, \varepsilon_i, A_h h_i = \mu_i) = \bar{Q}_i^{-1} - \bar{Q}_i^{-1} A_h' (A_h \bar{Q}_i^{-1} A_h')^{-1} A_h \bar{Q}_i^{-1}. \quad (2.3.6)$$

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

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

$$\pi_G^c(h_i|\theta, \varepsilon_i) = \frac{\pi_G(h_i|\theta, \varepsilon_i)\pi(A_h h_i|h_i)}{\pi_1(A_h h_i)}, \quad (2.3.7)$$

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

$$p^c(h_i|\theta) = \frac{p(h_i|\theta)\pi(A_h h_i|h_i)}{\pi_2(A_h h_i)}, \quad (2.3.8)$$

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

Finally, we recommend to assess the quality of the estimator (2.3.3) by reporting its standard error which can be computed e.g. by the batch means method. Furthermore, for the validity of the standard error and \sqrt{R} -convergence of the IS estimator, the variance of the importance weights has to exist. Since for the high-dimensional integral (2.3.1) this is not clear a-priori, we advise to test for the existence of the variance using e.g. the test of [Koopman et al. \(2009\)](#). However, for sample sizes typically used in macroeconomics we do not expect this to be a serious issue.

2.3.2 EM Algorithm

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

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

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

$$\mathcal{L}(\theta) = \log \int p(y|\theta, h)p(h|\theta)dh \quad (2.3.9)$$

$$= \log \int \frac{p(y|\theta, h)p(h|\theta)}{\tilde{p}(h)} \tilde{p}(h)dh \quad (2.3.10)$$

$$\geq \int \log \left(\frac{p(y|\theta, h)p(h|\theta)}{\tilde{p}(h)} \right) \tilde{p}(h)dh \quad (2.3.11)$$

$$= \int \log(p(y|\theta, h)p(h|\theta)) \tilde{p}(h)dh - \int \log(\tilde{p}(h)) \tilde{p}(h)dh \quad (2.3.12)$$

$$=: F(\tilde{p}, \theta). \quad (2.3.13)$$

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

$$\text{E-step: } \tilde{p}^{(l)} = \arg \max_{\tilde{p}} F(\tilde{p}, \theta^{(l-1)}), \quad (2.3.14)$$

$$\text{M-step: } \theta^{(l)} = \arg \max_{\theta} F(\tilde{p}^{(l)}, \theta). \quad (2.3.15)$$

Under mild regularity conditions the EM algorithm converges reliably towards a local optimum.⁴ It is easy to show that the E-step in (2.3.14) is given by setting $\tilde{p}^{(l)}$ equal to the smoothing distribution $p(h|\theta^{(l-1)}, y)$. This can be seen by noting that for this choice, equation (2.3.11) holds with equality which means that the lower bound $F(\tilde{p}, \theta)$ exactly equals the log-likelihood $\mathcal{L}(\theta)$. Furthermore, the M-step in equation (2.3.15) is given by maximizing the criterion function:

$$Q(\theta; \theta^{(l-1)}) = \int \log(p(y|\theta, h)p(h|\theta)) \tilde{p}^{(l)}(h)dh \quad (2.3.16)$$

$$= E_{\theta^{(l-1)}}(\mathcal{L}_c(\theta)), \quad (2.3.17)$$

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

For the SV-SVAR model, the complete data log-likelihood is rather simple and we refer to [Appendix 2.B.3](#) for an explicit expression. It follows that for a given choice of $\tilde{p}^{(l)}$, computing the M-Step is straightforward. However, since the smoothing distribution in SV models is generally not tractable, we cannot simply set $\tilde{p}^{(l)} = p(h|\theta^{(l-1)}, y)$. Instead, we develop two algorithms which approximate this density to a different extent, one based on an

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

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

Analytical approximation

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

$$\tilde{p}^{(l)}(h) = \prod_{i=1}^r \pi_G^c(h_i|\theta^{(l-1)}, \varepsilon_i), \quad (2.3.18)$$

which is the Gaussian approximation of the smoothing distribution that we already introduced as importance density. This E-step corresponds to maximizing $F(\tilde{p}, \theta^{(l-1)})$ with respect to \tilde{p} considering only the family of Gaussian distributions. To motivate this approach, we follow the arguments of [Neal and Hinton \(1998\)](#) who argue that it is not necessary to work with the exact smoothing distributions in the EM algorithm to get monotonic increases in the log-likelihood function $\mathcal{L}(\theta)$. In fact, it can be shown that $F(\tilde{p}, \theta) = \mathcal{L}(\theta) - D_{KL}(\tilde{p}(h)||p(h|y, \theta))$ where $D_{KL}(\cdot||\cdot)$ is the Kullback - Leibler (KL) divergence measure. Therefore, if the Gaussian approximation is close to the smoothing density in a KL sense, iteratively optimizing $F(\tilde{p}, \theta)$ yields convergence to a point very close to the corresponding local maximum of $\mathcal{L}(\theta)$. In the following, we refer to this algorithm as EM-1 and provide details in [Appendix 2.B.3](#).

Monte Carlo approximation

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

$$\log(\varepsilon_{it}^2) = h_{it} + \log(\eta_{it}^2), \quad (2.3.19)$$

$$h_{it} = \mu_i + \phi_i(h_{i,t-1} - \mu_i) + \sqrt{s_i}\omega_{it}, \quad (2.3.20)$$

where $\eta_{it} \sim N(0, 1)$ and $\omega_{it} \sim N(0, 1)$. [Kim et al. \(1998\)](#) propose to closely approximate the log- χ^2 error distribution in (2.3.19) by a mixture of seven normals. In particular, they specify:

$$p(\log(\eta_{it}^2)|z_{it} = k) \sim \mathcal{N}(\log(\varepsilon_{it}^2); m_k, v_k^2), \quad (2.3.21)$$

$$p(z_{it} = k) = p_k, \quad (2.3.22)$$

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

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

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

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

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

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

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

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

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

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

2.3.3 Properties of the estimator

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

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

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

$$T^{1/2}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}(\theta)^{-1}), \quad (2.3.23)$$

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

$$\widehat{\mathcal{I}}(\hat{\theta}) = T^{-1} \mathcal{J}(\hat{\theta}) \quad (2.3.24)$$

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

To compute estimator (2.3.24) in algorithm EM-1, note that we can evaluate an approximate log-likelihood in closed form based on the Gaussian approximation which we rely on in the E-step. In particular, based on Bayes' Theorem:

$$\log p(\varepsilon_i|\theta) \approx \log p(\varepsilon_i|\theta, h_i) + \log p^c(h_i|\theta) - \log \pi_G^c(h_i|\theta, \varepsilon_i), \quad (2.3.25)$$

which can be evaluated for any h_i . For convenience, the r likelihoods for the heteroskedastic structural shocks are evaluated at the mean $h_i = \bar{\delta}_i^c$, such that the exponential term in $\pi_G^c(h_i|\theta, \varepsilon_i)$ drops out. Therefore, based on (2.3.25) an approximate complete log-likelihood is given as:

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

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

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

$$\mathcal{J}_2(\hat{\theta}) = \mathbb{E} \left[\mathcal{J}_c(\hat{\theta})|y \right] - \text{Cov}(S_c(\hat{\theta})|y), \quad (2.3.26)$$

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

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

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

where the second approximation holds since $E(S_c(\hat{\theta})|y) = 0$. The integrals required to compute the expected values are with respect to the tractable Gaussian distributions $p(h|\hat{\theta}, z^{(j)}, y)$. The derivatives necessary to apply the Louis Method are available in closed form and given in Appendix 2.B.4.

Identification of the SVAR model is ultimately useful to conduct structural analysis. Since Impulse Response Functions (IRFs) are likely to be the most widely used tool for that purpose, we outline in Appendix 2.B.5 how to conduct inference on these quantities within our model. In particular, we describe a Delta Method approach to quantify uncertainty of the identified IRFs.

2.3.4 Testing for identification

For valid likelihood inference on the structural parameters including the impact matrix B , the model must be identified. As highlighted in Section 2.2, at most one component of ε_t is allowed to be homoskedastic if the model is to be identified solely by heteroskedasticity. To determine the number of heteroskedastic shocks in a given application, we recommend to follow a procedure considered by Lanne and Saikkonen (2007) and Lütkepohl and Milunovich (2016) within SVAR-GARCH models. The idea is to conduct the following sequence of tests:

$$H_0 : r = r_0 \quad \text{vs} \quad H_1 : r > r_0, \quad (2.3.27)$$

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

The testing problem given in (2.3.27) is nonstandard since parts of the parameter space differ between null and alternative hypothesis. Therefore, Lanne and Saikkonen (2007) suggest test statistics which require estimation under H_0 only. In particular, suppose that r_0 is the true number of heteroskedastic errors, and separate the structural shocks $\varepsilon_t = B^{-1}u_t = [\varepsilon'_{1t}, \varepsilon'_{2t}]'$ into a heteroskedastic part $\varepsilon_{1t} \in \mathbb{R}^{r_0}$ and homoskedastic innovations $\varepsilon_{2t} \in \mathbb{R}^{K-r_0}$. Note that if the null is true ($r = r_0$), $\varepsilon_{2t} \sim (0, I_{K-r_0})$ is white noise. To test for

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

$$\xi_t = \varepsilon'_{2t} \varepsilon_{2t} - T^{-1} \sum_{t=1}^T \varepsilon'_{2t} \varepsilon_{2t}, \quad (2.3.28)$$

$$\vartheta_t = \text{vech}(\varepsilon_{2t} \varepsilon'_{2t}) - T^{-1} \sum_{t=1}^T \text{vech}(\varepsilon_{2t} \varepsilon'_{2t}), \quad (2.3.29)$$

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

$$Q_1(H) = T \sum_{h=1}^H \left(\frac{\tilde{\gamma}(h)}{\tilde{\gamma}(0)} \right)^2, \quad (2.3.30)$$

$$Q_2(H) = T \sum_{h=1}^H \text{tr} [\tilde{\Gamma}(h)' \tilde{\Gamma}(0)^{-1} \tilde{\Gamma}(h) \tilde{\Gamma}(0)^{-1}], \quad (2.3.31)$$

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

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

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

Proof. See Appendix [2.A.4](#).

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

2.4 Monte Carlo study

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

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

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

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

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

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

For the diagonal matrix $\Lambda_t = \text{diag}([\lambda_{1t}, \lambda_{2t}]')$, the following DGPs are specified:

1. **BP-SVAR:** The BP-SVAR is subject to a one time change in the variance. We set $\Lambda_t = I_2$ for $t = 1, \dots, T/2$ and $\Lambda_t = \text{diag}([2, 7]')$ for $t = T/2 + 1, \dots, T$.
2. **MS(2)-SVAR:** The specified MS-SVAR involves a switching variance with the same regimes than the BP-SVAR. We specify the transition probability matrix:

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

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

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

3. **GARCH-SVAR:** For this specification, the diagonal elements of Λ_t follow univariate GARCH(1,1) processes with unit unconditional variance:

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

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

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

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

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

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

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

$$\text{MSE}(\theta_{jk})_h = \frac{1}{M} \sum_{m=1}^M \left(\sum_{i=0}^h \left(\hat{\theta}_{jk,i}(m) - \theta_{jk,i} \right)^2 \right). \quad (2.4.1)$$

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

The results of the simulation study are provided in [Table 2.1](#). For improved readability, we report relative MSEs in comparison to the correctly specified model. Overall, we find that the SV-SVAR model performs very well regardless of the true DGP or the sample size

Table 2.1: Cumulated MSEs at horizon $h = 5$.

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

Note: MSEs of impulse response functions calculated as in (2.4.1) and displayed relative to true model MSEs.

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

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

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

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

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

2.5 Interdependence between monetary policy and stock markets

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

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

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

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

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

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

2.5.1 Model and identifying constraints

The VAR model of Bjørnland and Leitemo (2009) is based on the following variables: $y_t = (q_t, \pi_t, c_t, \Delta s_t, r_t)'$, where q_t is a linearly detrended index of log industrial production, π_t the annualized inflation rate based on consumer prices, c_t the annualized change in log commodity prices as measured by the World Bank, Δs_t S&P500 real stock returns and r_t the federal funds rate. For detailed description of the data sources, transformations and time series plots see Appendix 2.C. As in Lütkepohl and Netšunajev (2017a), we use an extended sample period including data from 1970M1 until 2007M6, summing up to a total of 450 observations. To make our results comparable, we also choose $p = 3$ lags which is supported by the AIC applied within a linear VAR model.

⁸Yet another branch of the literature relies on sign restrictions of the impulse response functions (Faust; 1998; Canova and De Nicoló; 2002; Uhlig; 2005) or on a combination of sign restrictions and information in proxy variables (Braun and Brüggemann; 2017).

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

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

$$B = \begin{bmatrix} * & 0 & 0 & 0 & 0 \\ * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix} \quad \text{and} \quad \Xi_\infty = \begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & 0 \\ * & * & * & * & * \end{bmatrix}, \quad (2.5.1)$$

where $\Xi_\infty = (I_K - A_1 - \dots - A_p)^{-1}B$ is the long-run impact matrix of the structural shocks on y_t . Note that an asterisk means that the corresponding entry in B and Ξ_∞ is left unrestricted. The last columns of B and Ξ_∞ correspond to the reaction of y_t to a monetary policy shock. Economic activity, consumer- and commodity prices are only allowed to respond with a delay of one month to a monetary policy shock, while stock markets are allowed to react contemporaneously. However, in the long run, a monetary policy shock is assumed to have a zero effect on the stock market. The fourth column of B corresponds to a stock price shock which is constrained to have no contemporaneous impact on activity and prices while the central bank is allowed to adjust the interest rates within the same period. The remaining shocks do not have an economic interpretation. To identify the model, [Bjørnland and Leitemo \(2009\)](#) simply disentangle these shocks by imposing a recursivity assumption. As outlined before, restrictions (2.5.1) are overidentifying in heteroskedastic SVAR models and can be tested against the data. In line with [Lütkepohl and Netšunajev \(2017a\)](#), the following set of restrictions is tested:

R1: Both, B and Ξ_∞ restricted as in (2.5.1).

R2: Only the last two columns of B and Ξ_∞ are restricted as in (2.5.1).

R3: Only B is restricted as in (2.5.1).

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

$$\beta_{21} = (\Sigma_{zu_1}^{-1} \Sigma_{zu_2}') \beta_{11}. \quad (2.5.2)$$

where $\beta_1 = [\beta_{11}, \beta'_{21}]'$ with β_{11} scalar and $\beta_{21} \in \mathbb{R}^{K-1}$. Furthermore, $\Sigma_{zu'} = \text{Cov}(z, u') = [\Sigma_{zu'_1}, \Sigma_{zu'_2}]$ with $\Sigma_{zu'_1}$ scalar and $\Sigma'_{zu'_2} \in \mathbb{R}^{K-1}$. In practice, elements of $\Sigma_{zu'}$ are estimated by the corresponding sample moments.¹⁰ To identify a monetary policy shock, we use the narrative series constructed by [Romer and Romer \(2004\)](#) (RR henceforth) and [Gertler and Karadi \(2015\)](#) (GK henceforth). We test the following Proxy SVAR restrictions that arise when the first column of B is identified via either RR's or GK's instrument:

R4rr: IV moment restrictions (2.5.2) based on the RR shock.

R4gk: IV moment restrictions (2.5.2) based on the GK shock.

We use the RR series extended by [Wieland and Yang \(2016\)](#) which is available for the whole sample. The GK shock is only available for a subsample starting in 1990M1. We use their baseline series which is constructed based on the three months ahead monthly fed funds futures.¹¹ Time series plots of both series are available in Appendix 2.C.

2.5.2 Statistical analysis

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

Table 2.2 reports log-likelihood values, Akaike information criteria (AIC) and Bayesian information criteria (BIC) for a linear VAR and all heteroskedastic models. First of all, we highlight that there is only a small gain in terms of likelihood value of the SV model using the Monte Carlo based algorithm (EM-2) compared to the deterministic approximation (EM-1). To assess the Monte Carlo error of the estimates, we also report approximate 95%-confidence intervals based on an application of the batch means method and $R = 100,000$ draws of the importance density.¹² Comparing the different models, our results suggest that including time-variation in the second moment is strongly supported by both information criteria. Moreover, among the heteroskedastic models we find that particularly models designed for financial variables are favored, that is the GARCH model and the SV model. This may be not surprising given that stock market returns are included in the system.

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

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

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

Table 2.2: Model selection by information criteria.

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

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

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

In accordance with [Lütkepohl and Netšunajev \(2017a\)](#), we also consider standardized residuals as an additional model checking device. Figure 2.1 provides a plot for the standardized residuals of all models computed as $\hat{u}_{it} / \hat{\sigma}_{ii,t}$ where $\hat{\sigma}_{ii,t}^2$ is the i -th diagonal entry of the estimated VAR covariance matrix $\hat{\Sigma}_t$. These plots clearly suggest that none of the other methods is fully satisfactory in yielding standardized residuals that seem to be homoskedastic and approximately normally distributed. However, for the SV-SVAR model, standardized residuals seem well behaved with no apparent heteroskedasticity and virtually no outliers. To confirm this impression, we provide complementary test results in Appendix 2.C.1 concerned with remaining heteroskedasticity and non-normality in standardized structural shocks. We find that only for the shocks of the SV-SVAR model, there is no evidence against both normality and homoskedasticity. To conclude, statistical analysis suggests that the proposed SV-SVAR is the most adequate for this application and we continue our analysis based on this model.

In order to test restrictions R1-R4 as overidentifying, it is necessary to count with enough heteroskedastic shocks ($r \geq K - 1$) to fully identify the impact matrix B . As described in Section 2.3.4, we apply a sequence of tests with $H_0 : r = r_0$ against $H_1 : r > r_0$ for $r_0 = 0, 1, \dots, K - 1$. The results are reported in Table 2.3. We find strong evidence that $r = K$ in our model, implying that the model can be fully identified by heteroskedasticity.

We continue our analysis and test the economically motivated restrictions R1-R4 as overidentifying. In Table 2.4 we provide Likelihood Ratio (LR) test statistics for the

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

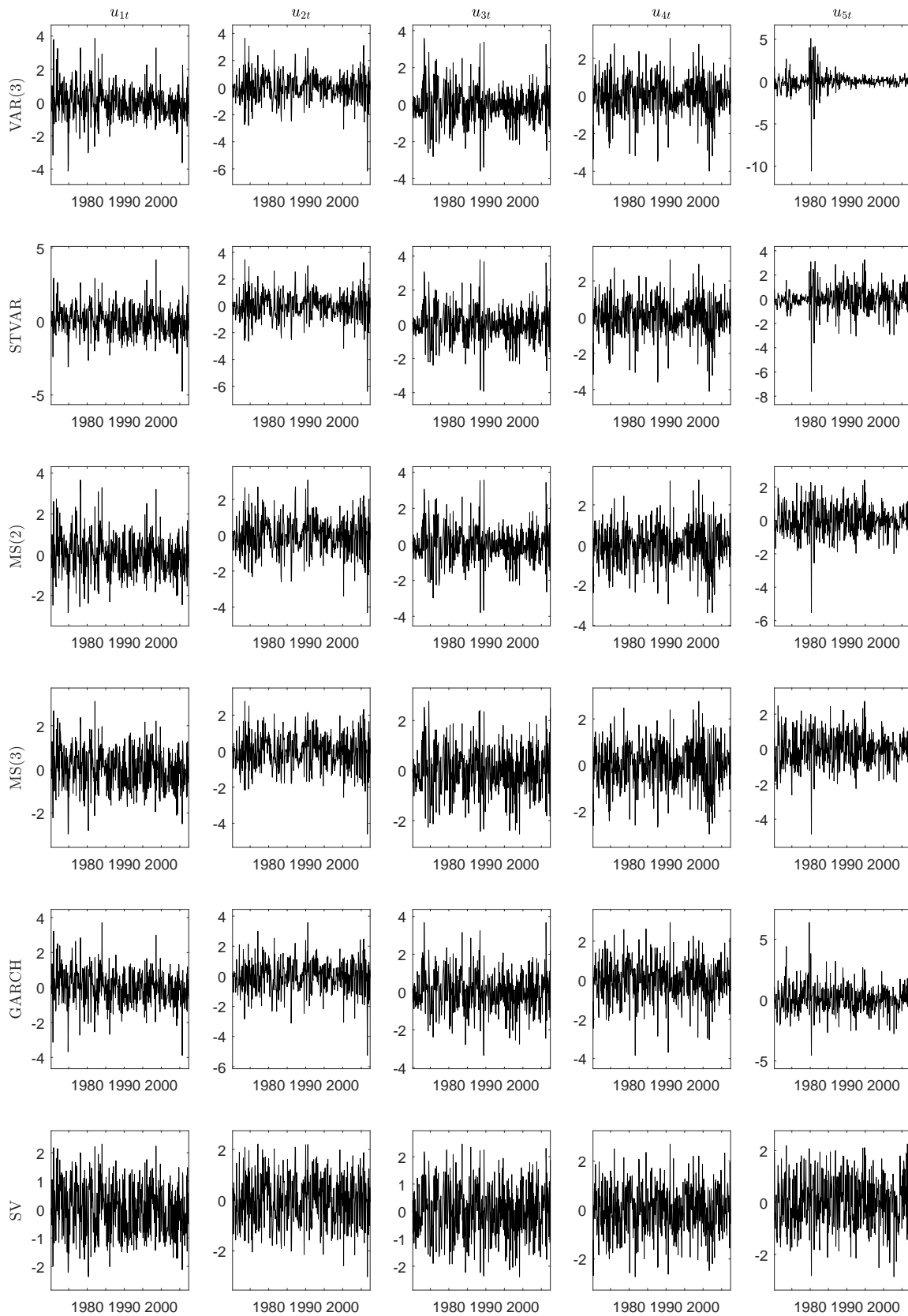


Figure 2.1: Standardized residuals of linear, ST-, MS(2)-, MS(3)-, GARCH- and SV-SVAR model.

Table 2.3: Tests of identification.

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

Note: Sequence of tests to check the number of heteroskedastic shocks in the system as introduced in Section 2.3.4 (Lanne and Saikkonen; 2007).

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

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

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

Table 2.4: Tests for overidentifying restrictions (EM-1).

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

Note: For details about overidentifying restrictions see Section 2.5.1. Likelihood ratio test statistics are computed as $2(\ln L_{H_1} - \ln L_{H_0})$ and are approximately χ^2 -distributed under H_0 . Right columns report approximate 95%-confidence intervals for the p -value resulting from an application of the batch means method to the LR test statistic.

With respect to the Proxy SVAR restrictions, we find that identifying a monetary policy shock with the shock series of [Gertler and Karadi \(2015\)](#) is strongly rejected by the data with a likelihood ratio test statistic exceeding 250. In turn, identification via the narrative series of [Romer and Romer \(2004\)](#) cannot be rejected at any conventional significance level. To further understand these results, we compute sample correlations of the instruments z with $\hat{\varepsilon}$, the estimated structural shocks of the unconstrained SV-SVAR model. For GK, we find $\text{Corr}(z^{GK}, \hat{\varepsilon}) = (0.039, -0.067, 0.050, -0.242, 0.419)$, while for RR, $\text{Corr}(z^{RR}, \hat{\varepsilon}) = (0.042, 0.005, 0.031, -0.021, 0.453)$. While both shocks are subject to a strong correlation with one of the statistically identified shocks, the instrument of GK is highly correlated with at least one additional shock. This clearly violates the exogeneity condition on the instrument. Thereby, our results support the argument of [Ramey \(2016\)](#) who questions the exogeneity of the GK instrument finding that it is autocorrelated and predictable by Greenbook variables. In turn, for the RR shock we find that there is little correlation with the remaining structural residuals of the SVAR. This clearly explains why identification via the RR shock is not rejected. Since the Proxy SVAR restrictions based on RR cannot be rejected, we can interpret the last shock of the unconstrained model as a monetary policy shock for which $\text{Corr}(z^{RR}, \hat{\varepsilon}_5) = 0.45$. In Figure 2.2 we plot impulse response functions (IRFs) up to 72 months (6 years) of the system variables in response to a monetary policy shock. Besides point estimates, we provide 68% asymptotic confidence intervals. Again, we note that there is qualitatively no difference in using EM-1 or EM-2 to compute the estimates and corresponding standard errors.¹⁵ The IRFs and their asymptotic confidence intervals coincide for all variables at all horizons. In line with the IRFs computed by [Lütkepohl and](#)

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

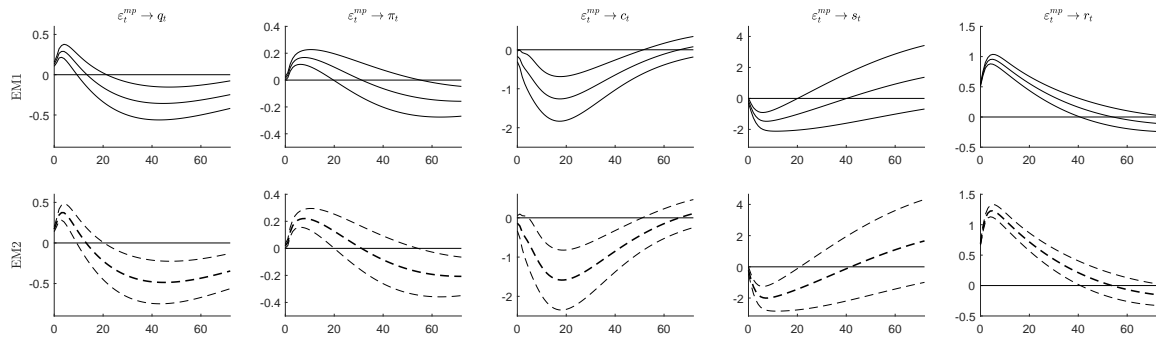


Figure 2.2: Impulse responses monetary policy shock up to a horizon of 72 months with 68% confidence bounds. Figures compare estimates based on EM-1 (solid line) and EM-2 (dashed line) with corresponding asymptotic confidence intervals.

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

2.6 Conclusion

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

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

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

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

References

- Bernanke, B. S., Boivin, J. and Eliasziw, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach, *The Quarterly Journal of Economics* **120**(1): 387–422.
- Bernanke, B. S. and Mihov, I. (1998). Measuring monetary policy, *Quarterly Journal of Economics* **113**(3): 869–902.
- Bjørnland, H. C. and Leitemo, K. (2009). Identifying the interdependence between US monetary policy and the stock market, *Journal of Monetary Economics* **56**(2): 275–282.
- Blanchard, O. J. and Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances, *The American Economic Review* **79**(4): 655–673.
- Braun, R. and Brüggemann, R. (2017). Identification of SVAR models by combining sign restrictions with external instruments, *Working Paper 16*, Department of Economics, University of Konstanz.
- Brüggemann, R., Jentsch, C. and Trenkler, C. (2016). Inference in VARs with conditional heteroskedasticity of unknown form, *Journal of Econometrics* **191**(1): 69–85.
- Caffo, B. S., Jank, W. and Jones, G. L. (2005). Ascent-based Monte Carlo expectation-maximization, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(2): 235–251.
- Canova, F. and De Nicoló, G. (2002). Monetary disturbances matter for business fluctuations in the G-7, *Journal of Monetary Economics* **49**: 1131–1159.
- Cappé, O., Moulines, E. and Ryden, T. (2005). *Inference in Hidden Markov Models*, Springer.
- Carreiro, A., Clark, T. E. and Marcellino, M. G. (2019). The identifying information in vector autoregressions with time-varying volatilities: An application to endogenous uncertainty, *Working Paper*.
- Chan, J. C. C. (2017). The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling, *Journal of Business & Economic Statistics* **35**(1): 17–28.
- Chan, J. C. C. and Grant, A. L. (2016). On the observed-data deviance information criterion for volatility modeling, *Journal of Financial Econometrics* **14**(4): 772–802.

- Chan, J. C. C. and Jeliazkov, I. (2009). Efficient simulation and integrated likelihood estimation in state space models, *International Journal of Mathematical Modelling and Numerical Optimisation* **1**(1-2): 101–120.
- Christiano, L. J., Eichenbaum, M. and Evans, C. L. (1999). Monetary policy shocks: What have we learned and to what end?, *Handbook of Macroeconomics*, Vol. 1, Elsevier, chapter 2, pp. 65–148.
- Christiano, L. J., Eichenbaum, M. and Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy, *Journal of Political Economy* **113**(1): 1–45.
- Coibion, O. (2012). Are the effects of monetary policy shocks big or small?, *American Economic Journal: Macroeconomics* **4**(2): 1–32.
- Danielsson, J. and Richard, J. F. (1993). Accelerated Gaussian importance sampler with application to dynamic latent variable models, *Journal of Applied Econometrics* **8**: 153–173.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **39**: 1–38.
- Durbin, J. and Koopman, S. J. (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models, *Biometrika* **84**(3): 669–684.
- Durbin, J. and Koopman, S. J. (2000). Time series analysis of non-gaussian observations based on state space models from both classical and Bayesian perspectives, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62**(1): 3–56.
- Faust, J. (1998). The robustness of identified VAR conclusions about money, *Carnegie-Rochester Conference Series on Public Policy* **49**: 207–244.
- Faust, J., Swanson, E. T. and Wright, J. H. (2004). Identifying VARs based on high frequency futures data, *Journal of Monetary Economics* **51**(6): 1107–1131.
- Fernández-Villaverde, J. and Rubio-Ramírez, J. F. (2007). Estimating macroeconomic models: A likelihood approach, *Review of Economic Studies* **74**(4): 1059–1087.
- Gertler, M. and Karadi, P. (2015). Monetary policy surprises, credit costs, and economic activity, *American Economic Journal-Macroeconomics* **7**(1): 44–76.
- Gourieroux, C., Monfort, A. and Renne, J. P. (2017). Statistical inference for independent component analysis: Application to structural VAR models, *Journal of Econometrics* **196**(1): 111–126.

- Harvey, A., Ruiz, E. and Shephard, N. (1994). Multivariate stochastic variance models, *Review of Economic Studies* **61**(2): 247–264.
- Herwartz, H. and Lütkepohl, H. (2014). Structural vector autoregressions with Markov switching: Combining conventional with statistical identification of shocks, *Journal of Econometrics* **183**(1): 104–116.
- Jacquier, E., Polson, N. G. and Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models, *Journal of Business & Economic Statistics* **12**(4): 371–389.
- Justiniano, A. and Primiceri, G. E. (2008). The time-varying volatility of macroeconomic fluctuations, *American Economic Review* **98**(3): 604–641.
- Kilian, L. and Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*, Themes in Modern Econometrics, Cambridge University Press.
- Kim, S., Shephard, N. and Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models, *Review of Economic Studies* **65**(3): 361–393.
- Koop, G. and Korobilis, D. (2010). Bayesian multivariate time series methods for empirical macroeconomics, *Foundations and Trends in Econometrics* **3**(4): 267–358.
- Koopman, S. J., Shephard, N. and Creal, D. (2009). Testing the assumptions behind importance sampling, *Journal of Econometrics* **149**(1): 2–11.
- Lanne, M., Lütkepohl, H. and Maciejowska, K. (2010). Structural vector autoregressions with Markov switching, *Journal of Economic Dynamics & Control* **34**(2): 121–131.
- Lanne, M., Meitz, M. and Saikkonen, P. (2017). Identification and estimation of non-Gaussian structural vector autoregressions, *Journal of Econometrics* **196**(2): 288–304.
- Lanne, M. and Saikkonen, P. (2007). A multivariate generalized orthogonal factor GARCH model, *Journal of Business & Economic Statistics* **25**(1): 61–75.
- Lewis, D. J. (2018). Identifying shocks via time-varying volatility, *Federal Reserve Bank of New York Staff Report* (871).
- Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm, *Journal of the Royal Statistical Society Series B-Methodological* **44**(2): 226–233.
- Lütkepohl, H. (2005). *New introduction to multiple time series analysis*, Springer Science & Business Media.

- Lütkepohl, H. and Milunovich, G. (2016). Testing for identification in SVAR-GARCH models, *Journal of Economic Dynamics and Control* **73**: 241–258.
- Lütkepohl, H. and Netšunajev, A. (2017a). Structural vector autoregressions with heteroskedasticity: A review of different volatility models, *Econometrics and Statistics* **1**: 2–18.
- Lütkepohl, H. and Netšunajev, A. (2017b). Structural vector autoregressions with smooth transition in variances, *Journal of Economic Dynamics & Control* **84**: 43–57.
- Lütkepohl, H. and Schlaak, T. (2018). Choosing between different time-varying volatility models for structural vector autoregressive analysis, *Oxford Bulletin of Economics and Statistics* (forthcoming).
- Mahieu, R. J. and Schotman, P. C. (1998). An empirical application of stochastic volatility models, *Journal of Applied Econometrics* **13**(4): 333–359.
- Maxand, S. (forthcoming). Identification of independent structural shocks in the presence of multiple Gaussian components, *Econometrics and Statistics* .
- McLachlan, G. and Krishnan, T. (2007). *The EM algorithm and extensions*, Vol. 382, John Wiley & Sons.
- Melino, A. and Turnbull, S. M. (1990). Pricing foreign-currency options with stochastic volatility, *Journal of Econometrics* **45**(1-2): 239–265.
- Mertens, K. and Ravn, M. O. (2013). The dynamic effects of personal and corporate income tax changes in the United States, *American Economic Review* **103**(4): 1212–1247.
- Montiel-Olea, José, L., Stock, J. H. and Watson, M. W. (2016). Uniform inference in SVARs identified with external instruments, *Harvard Manuscript* .
- Nakamura, E. and Steinsson, J. (2018). High frequency identification of monetary non-neutrality: The information effect, *Quarterly Journal of Economics* (forthcoming).
- Neal, R. M. and Hinton, G. E. (1998). A view of the EM algorithm that justifies incremental, sparse, and other variants, *Learning in graphical models*, Springer, pp. 355–368.
- Normandin, M. and Phaneuf, L. (2004). Monetary policy shocks: Testing identification conditions under time-varying conditional volatility, *Journal of Monetary Economics* **51**(6): 1217–1243.
- Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy, *Review of Economic Studies* **72**(3): 821–852.

- Ramey, V. (2016). Macroeconomic shocks and their propagation, *Handbook of Macroeconomics*, Vol. 2, Elsevier, chapter 2, pp. 71–162.
- Rigobon, R. (2003). Identification through heteroskedasticity, *Review of Economics and Statistics* **85**(4): 777–792.
- Romer, C. D. and Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications, *American Economic Review* **94**(4): 1055–1084.
- Roweis, S. and Ghahramani, Z. (2001). Learning nonlinear dynamical systems using the Expectation–Maximization algorithm, *Kalman filtering and neural networks* **6**: 175–220.
- Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, *Journal of the Royal Statistical Society Series B-Statistical Methodology* **71**(2): 319–392.
- Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models, *Journal of Econometrics* **63**(1): 289–306.
- Sims, C. A. (1980). Macroeconomics and reality, *Econometrica* **48**(1): 1–48.
- Stock, J. H. and Watson, M. W. (2012). Disentangling the channels of the 2007–09 recession, *Brookings Papers on Economic Activity* **43**(1): 81–156.
- Uhlig, H. (2005). What are the effects of monetary policy on output? Results from an agnostic identification procedure, *Journal of Monetary Economics* **52**: 381–419.
- Van der Weide, R. (2002). GO-GARCH: A multivariate generalized orthogonal GARCH model, *Journal of Applied Econometrics* **17**(5): 549–564.
- Wieland, J. F. and Yang, M.-J. (2016). Financial dampening, *Working Paper 22141*, National Bureau of Economic Research.
- Wright, J. H. (2012). What does monetary policy do to long-term interest rates at the zero lower bound?, *The Economic Journal* **122**(564): 447–466.

Appendix 2.A Derivations and proofs

To ensure identification of impact matrix B in model (2.2.1)-(2.2.4) we show that under sufficient heterogeneity in the second moments of the structural shocks, i.e. $r \geq K - 1$, there is no B^* different from B except for column permutations and sign changes which yields an observationally equivalent model with the same time-varying second moment properties in reduced form errors u_t for all $t = 1, \dots, T$. Furthermore, for $r < K - 1$, we show which parameters in impact matrix B are identified and which are not. This also includes one possible identification scheme for this scenario. We start with the derivation of the autocovariance function of the second moments of reduced form residuals u .

2.A.1 Autocovariance function of the second moments

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

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

The entries of this expression are only non-zero if both $i = j = k = l$ and $i \leq r$ hold for $i, j, k, l \in \{1, \dots, K\}$ due to the structure of the SV-SVAR model (2.2.1)-(2.2.4). Thus, it is

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

with G_K being a selection matrix and M_τ as defined in Section 2.2. Briefly recap that we define (Lewis; 2018):

$$\xi_t = \text{vech}(u_t u_t') = L_K \text{vec}(u_t u_t').$$

Consequently, the autocovariance function in ξ reads:

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

2.A.2 Proof of proposition 1

Proof. Suppose $\tilde{B} = BQ$ and $\tilde{\varepsilon}_t = Q^{-1}\varepsilon_t$ with $Q = \begin{pmatrix} Q_1 & Q_3 \\ Q_2 & Q_4 \end{pmatrix}$, where $Q_1 \in \mathbb{R}^{r \times r}$, $Q_2, Q_3 \in \mathbb{R}^{(K-r) \times r}$ and $Q_4 \in \mathbb{R}^{(K-r) \times (K-r)}$ define an observationally equivalent model satisfying (2.2.8) and (2.2.9). Due to (2.2.8), it is:

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

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

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

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

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

This yields the following conditions:

$$\forall i = 1, \dots, r : \sum_{l=1}^r q_{li}^4 \gamma_l(\tau) = \tilde{\gamma}_i(\tau) \neq 0, \quad (2.A.1)$$

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

Because of (2.A.2), it is $\sum_{l=1}^r q_{l\bullet}' \underbrace{q_{li}^2 q_{lj}}_{=: \lambda_{lij}} = 0$ for all $i, j \in \{1, \dots, K\}$ with $i \neq j$. As it is an

orthogonal matrix, row vectors $q_{l\bullet}$ of matrix Q are linearly independent such that $\lambda_{lij} = 0$ for all $l \in \{1, \dots, r\}, i, j \in \{1, \dots, K\} : i \neq j$. Consequently, considering the first r rows of Q , i.e. matrix $[Q_1, Q_3]$, only one element per row can be different from zero. Due to the orthogonality, this element has to be ± 1 .

Because of (2.A.1), in each column of $r \times r$ matrix Q_1 at least one element has to be non-zero. Following the previous argument, these r non-zero entries correspond to the $r \pm 1$ entries

in $[Q_1, Q_3]$. This directly implies that Q_3 is a zero matrix and Q_1 has exactly one element different from zero per row and column which is ± 1 . Thus, Q_1 can be decomposed in DP where D is a diagonal matrix with ± 1 entries and P is a permutation matrix.

In addition, orthogonality of Q yields that Q_2 has to be a zero matrix. Finally, Q_4 has to be a $(K - r) \times (K - r)$ orthogonal matrix to satisfy $QQ' = I_K$. Therefore, block B_1 is unique up to permutation and sign changes.

□

2.A.3 Proof of corollary 1

Using Proposition 1 shows that an observationally equivalent model with the same autocovariance function in the second moment of the reduced form errors can be obtained by

$\tilde{B} = BQ$ if and only if Q has the structure $\begin{pmatrix} Q_1 & 0 \\ 0 & Q_4 \end{pmatrix}$, $Q_1 = DP$ with D a diagonal matrix with ± 1 entries on the diagonal, P a permutation matrix and $Q_4 \in \mathbb{R}^{(K-r) \times (K-r)}$ any orthogonal matrix. Thus, the decomposition $B = [B_1, B_2]$ with $B_1 \in \mathbb{R}^{K \times r}$ and $B_2 \in \mathbb{R}^{K \times (K-r)}$ yields uniqueness of B_1 apart from multiplication of its columns by -1 and permutation. Moreover, in case that $r = K - 1$, column vector B_2 is also unique up to multiplication with -1 :

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

2.A.4 Proof of corollary 2

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

Appendix 2.B Estimation

2.B.1 Importance density

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

$$\log p(h_i|\theta, \varepsilon_i) \propto \log p(\varepsilon_i|\theta, h_i) + \log p(h_i). \quad (2.B.1)$$

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

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

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

Since the measurement equation is nonlinear in h_i , the normalizing constant of the smoothing distribution in equation (2.B.1) is not known. An approximate distribution, however, can be obtained by a second order Taylor approximation of the measurement equation (2.B.3). The corresponding partial derivatives are given as:

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

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

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

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

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

Consequently, the approximate smoothing density is:

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

The restricted density $\pi_G^c(h_i|\theta, \varepsilon_i)$ is constructed as outlined in Section 2.3. Note that $\pi_G^c(h_i|\theta, \varepsilon_i)$ yields a good approximation only if $\tilde{h}_i^{(0)}$ is chosen appropriately. In the following, we sketch how the Newton Raphson method is used to evaluate the IS density at the mode of the smoothing distribution (2.B.1).

2.B.2 Newton Raphson method

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

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

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

2.B.3 EM algorithm

To fix notation, define the following quantities:

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

Using this, VAR equation (2.2.1) can be compactly written as:

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

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

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

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

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

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

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

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

which correspond to persistent heteroskedasticity with initial kurtosis of about 3.7 for the estimated structural shocks $\hat{\varepsilon}_i, i = 1, \dots, r$.

Note that in order to satisfy linear restriction (2.2.10) we set for $i = 1, \dots, r$ and $l \geq 1$:

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

EM-1

Because of $\hat{\varepsilon}_i^{(l-1)} = \hat{B}^{(l-1)}(y_t - \hat{A}^{(l-1)}x_t)$, it is equivalent to condition the approximate smoothing densities π_G^c and their moments to $(\theta^{(l-1)}, \hat{\varepsilon}_i^{(l-1)})$ or $(\theta^{(l-1)}, y)$, respectively.

Based on starting values $\theta^{(0)} = \left[\left(\hat{\alpha}^{(0)} \right)', \text{vec} \left(\hat{B}^{(0)} \right)', \left(\hat{\phi}^{(0)} \right)', \left(\hat{s}^{(0)} \right)' \right]'$, the EM algorithm iteratively cycles through the following steps for $l \geq 1$:

1. E-step: For $i = 1, \dots, r$, evaluate the moments of the approximate smoothing densities, mean $\bar{\delta}_i^c$ and variance $\bar{Q}_i^{-1} - \bar{Q}_i^{-1} A_h' (A_h \bar{Q}_i^{-1} A_h')^{-1} A_h \bar{Q}_i^{-1}$, as described in Appendix 2.B.1. Thereby, directly inverting \bar{Q}_i is unnecessary costly since we only need its diagonal elements representing the marginal variances $\text{Var}(h_{it} | \theta^{(l-1)}, y)$ and the entries of the first off-diagonal corresponding to $\text{Cov}(h_{it}, h_{i,t-1} | \theta^{(l-1)}, y)$. Similar to the Kalman smoother recursions, they can be obtained without computing the whole inverse using sparse matrix routines based on Takahashi's equations (Rue et al.; 2009). An efficient implementation in Matlab is available at the MathWorks File Exchange (see *sparseinv* by Tim Davis).
2. M-step: Conditional on the approximate smoothing density of log-variances h_i ($i = 1, \dots, r$), we update parameters of both state and measurement equation of the SV-SVAR model.

- (a) Update ϕ_i and s_i for $i = 1, \dots, r$:

Conditional on the moments of the approximate smoothing density we maximize the expected value of the complete data log-likelihood (2.B.5) with respect to the state equation parameters. Therefore, define

$$\frac{\partial^{a_1+a_2} \mathcal{L}_c}{\partial^{a_1} \phi \partial^{a_2} s} = \left[\frac{\partial^{a_1+a_2} \mathcal{L}_c}{\partial^{a_1} \phi_1 \partial^{a_2} s_1}, \dots, \frac{\partial^{a_1+a_2} \mathcal{L}_c}{\partial^{a_1} \phi_r \partial^{a_2} s_r} \right]$$

for $a_1, a_2 \in \{0, 1, 2\}$ with $a_1 + a_2 \leq 2$, $\nabla G(\phi, s) = \mathbf{E} \left[\frac{\partial \mathcal{L}_c}{\partial \phi}, \frac{\partial \mathcal{L}_c}{\partial s} \right]$ and

$$H(\phi, s) = \mathbf{E} \left(\begin{array}{cc} \text{diag} \left(\frac{\partial^2 \mathcal{L}_c}{\partial \phi^2} \right) & \text{diag} \left(\frac{\partial^2 \mathcal{L}_c}{\partial \phi \partial s} \right) \\ \text{diag} \left(\frac{\partial^2 \mathcal{L}_c}{\partial \phi \partial s} \right) & \text{diag} \left(\frac{\partial^2 \mathcal{L}_c}{\partial s^2} \right) \end{array} \right).$$

The detailed expressions for first and second derivatives of the complete data log-likelihood are printed in 2.B.4. Then, set $\hat{\phi}_k = \hat{\phi}^{(l-1)}$ and $\hat{s}_k = \hat{s}^{(l-1)}$ and update parameters using Newton-Raphson, i.e. set

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

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

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

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

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

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

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

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

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

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

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

EM-2

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

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

$$Q(\theta, \theta^{(l-1)}) \approx \frac{1}{R} \sum_{j=1}^R \mathbb{E}_{\theta^{(l-1)}}^{(j)} [\mathcal{L}(\theta)], \quad (2.B.6)$$

where the expectations are taken with respect to the tractable distribution $p(h | \theta^{(l-1)}, z^{(j)}, y)$. To generate random draws of z , we rely on the methodology of

Kim et al. (1998). For each of the heteroskedastic shocks ($i = 1, \dots, r$), this involves iteratively drawing from the following conditional distributions:

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

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

with $y_{it}^* = \log\left[\left(\hat{\varepsilon}_{it}^{(l-1)}\right)^2\right]$, $\hat{\varepsilon}_t^{(l-1)} = \left(\hat{B}^{(l-1)}\right)^{-1}\left(y_t - \hat{A}^{(l-1)}x_t\right)$ and $\phi(\cdot; m_k, v_k^2)$ indicating the pdf of a normal distribution with mean m_k and variance v_k^2 . Mixture parameters p_k 's, m_k 's and v_k 's are tabulated in Table 2.5.

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

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

and

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

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

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

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

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

Note that the corrected moments are those used to compute the Monte Carlo expected complete data log-likelihood from equation (2.B.6). As in EM-1, we only compute the diagonal and first off-diagonal of the covariance matrix Σ_{ij} using the same sparse matrix routines.

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

- (a) Update ϕ_i and s_i for $i = 1, \dots, r$: Conditional on the mixture indicators z , the expected value of the complete data log-likelihood (2.B.5) is maximized. To do so, define $\nabla G_R(\phi, s) = \frac{1}{R} \sum_{j=1}^R \mathbb{E} \left(\frac{\partial \mathcal{L}_c}{\partial \phi}, \frac{\partial \mathcal{L}_c}{\partial s} \middle| z^{(j)} \right)$ containing the first and

$$H_R(\phi, s) = \frac{1}{R} \sum_{j=1}^R \mathbb{E} \left(\begin{array}{cc} \text{diag} \left(\frac{\partial^2 \mathcal{L}_c}{\partial \phi^2} \right) & \text{diag} \left(\frac{\partial^2 \mathcal{L}_c}{\partial \phi \partial s} \right) \\ \text{diag} \left(\frac{\partial^2 \mathcal{L}_c}{\partial \phi \partial s} \right) & \text{diag} \left(\frac{\partial^2 \mathcal{L}_c}{\partial s^2} \right) \end{array} \middle| z^{(j)} \right)$$

including the second derivatives. The detailed expressions are printed in Section 2.B.4. All expectations of functions of the log-variances are uniquely determined by the sampled mixture indicators. Then, set $\hat{\phi}_k = \hat{\phi}^{(l-1)}$ and $\hat{s}_k = \hat{s}^{(l-1)}$ and update parameters using Newton-Raphson, i.e. set

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

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

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

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

where everything is as in EM-1 but:

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

- (c) Update B as in EM-1.

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

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

Table 2.5: Mixture approximation $\log \chi_{(1)}^2$ distribution (Kim et al.; 1998).

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

Note: Seven Normal Mixture components to approximate a $\log(\chi_{(1)}^2)$ distribution adjusted by its mean -1.2704 .

2.B.4 Derivatives complete data log-likelihood

The respective derivatives of the complete data log-likelihood (2.B.5) are given in the following. Let $\bar{h}_{it} = h_{it} - \mu_i$ for $i = 1, \dots, r$ and $t = 1, \dots, T$. First and second derivatives with respect to state equation parameters ϕ_i and s_i are given as follows:

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

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

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

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

2.B.5 Inference on structural impulse response functions

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

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

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

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

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

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

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

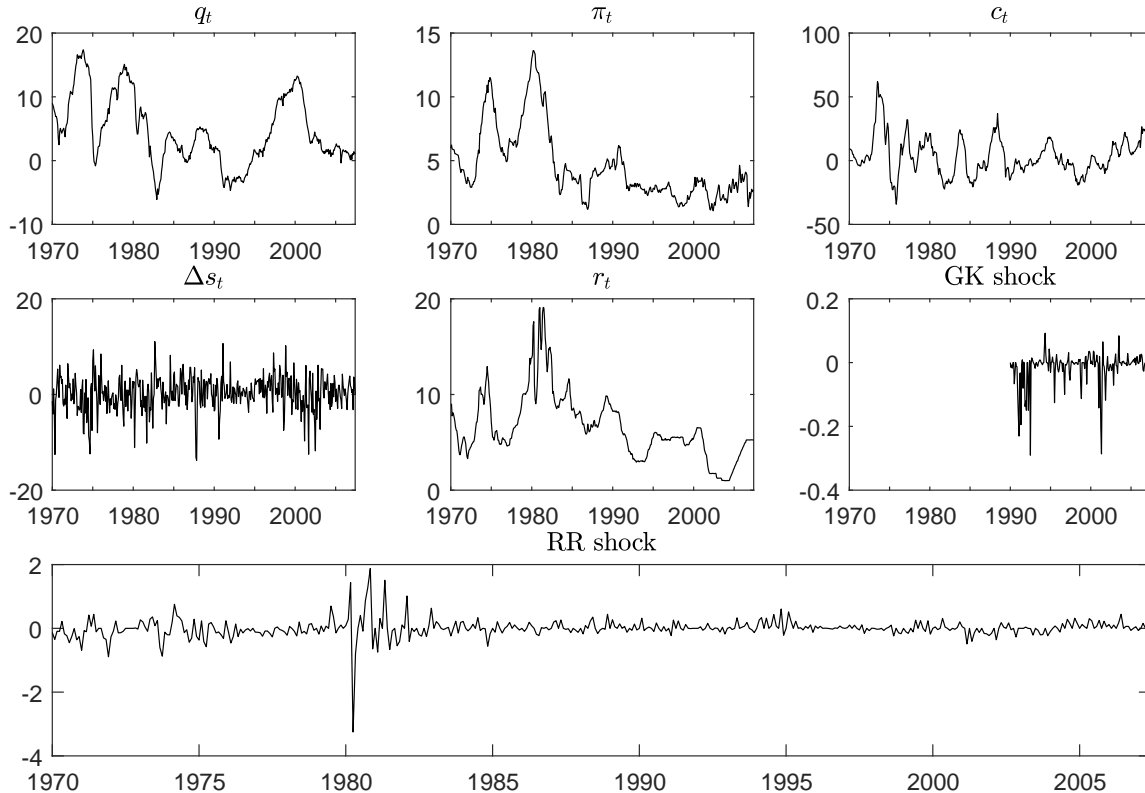


Figure 2.3: Time series plots.

2.C.1 Complementary results

Table 2.6: Tests on standardized structural shocks.

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

Note: Multivariate Jarque-Bera (MJB) test conducted as in (Lütkepohl; 2005, p. 181). Test statistics Q_1 and Q_2 as discussed in Section 2.3.4, applied to estimated standardized structural shocks $\hat{\varepsilon}_t / \exp(\hat{h}_t/2)$.

Table 2.7: Tests for overidentifying restrictions (EM-2).

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

Note: For details about overidentifying restrictions see Section 2.5.1. Likelihood ratio test statistics are computed as $2(\ln L_{H_1} - \ln L_{H_0})$ and are approximately χ^2 -distributed under H_0 . Right columns report an approximate 95%-confidence interval for the *p*-value resulting from an application of the batch means method to the LR test statistic.

Chapter **3**

The Importance of Supply and Demand for
Oil Prices: Evidence from a SVAR
Identified by non-Gaussianity

3.1 Introduction

There is compelling evidence that oil price shocks are important drivers of the business cycle (Kilian; 2008a; Hamilton; 2009). To estimate their effects, much of the early literature treats oil prices as exogenous to the business cycle, arguing that oil prices were largely driven by wars and political events (Shapiro and Watson; 1988; Blanchard and Galí; 2010).¹ However, at the latest since Kilian (2009), the focus has shifted towards further disentangling oil price fluctuations into supply and demand shocks with possibly very different implications for the economy. A sound understanding of the driving forces is important for the design of optimal monetary and fiscal policy or market regulation.

In this paper, I provide novel evidence on the relative importance of supply and demand shocks for oil price fluctuations. Following Kilian (2009), a structural vector autoregressive (SVAR) model is set up for the global oil market, which is identified exploiting non-Gaussianity in the structural shocks (Lanne et al.; 2017; Gouriéroux et al.; 2017). The identification approach is further sharpened by combining the statistical approach with conventional restrictions typically imposed on oil price elasticities. In order to incorporate this information in a coherent way, a new Bayesian non-Gaussian SVAR model is developed.

With this model, I find that oil supply shocks have contributed very little to fluctuations in oil prices since 1985. In terms of the relative contribution of supply shocks to the long term forecast error variance of oil prices, the model arrives at point estimates between 1% and 13% percent, depending on the exact prior specification. Therefore, my analysis contributes to the ongoing debate about the importance of supply for fluctuations in oil prices. One branch of the literature, including Kilian (2009), Kilian and Murphy (2012, 2014) and Antolín-Díaz and Rubio-Ramírez (2018) finds that oil prices are mainly driven by demand and that supply shocks account for less than 10% of the long term variability in oil prices. In contrast, Baumeister and Hamilton (forthcoming) and Caldara et al. (forthcoming) report very different results for the importance of supply shocks, estimating substantially larger contributions. All of these papers make use of prior information on oil price elasticities to disentangle supply and demand shocks, most of them in combination with sign restrictions.

As highlighted in Caldara et al. (forthcoming), the disagreement in this literature can be largely attributed to minor differences in the identifying information used on oil price elasticities. The reason is that oil production and prices are fairly uncorrelated, rendering a large class of models identified by solely sign restrictions equally consistent with the data. In this context, minor differences in the prior for oil price elasticities can translate into very different implications for the drivers of oil prices. In this paper, no such prior

¹Examples include the Arab oil embargo after the October war (1973-1974), the Iranian Revolution (1978-1979), the Iran-Iraq war late 1980s, the collapse of the OPEC 1986, the Gulf war 1990, the Venezuelan Oil strike (2002) or the outbreak of the Iraq war in 2003.

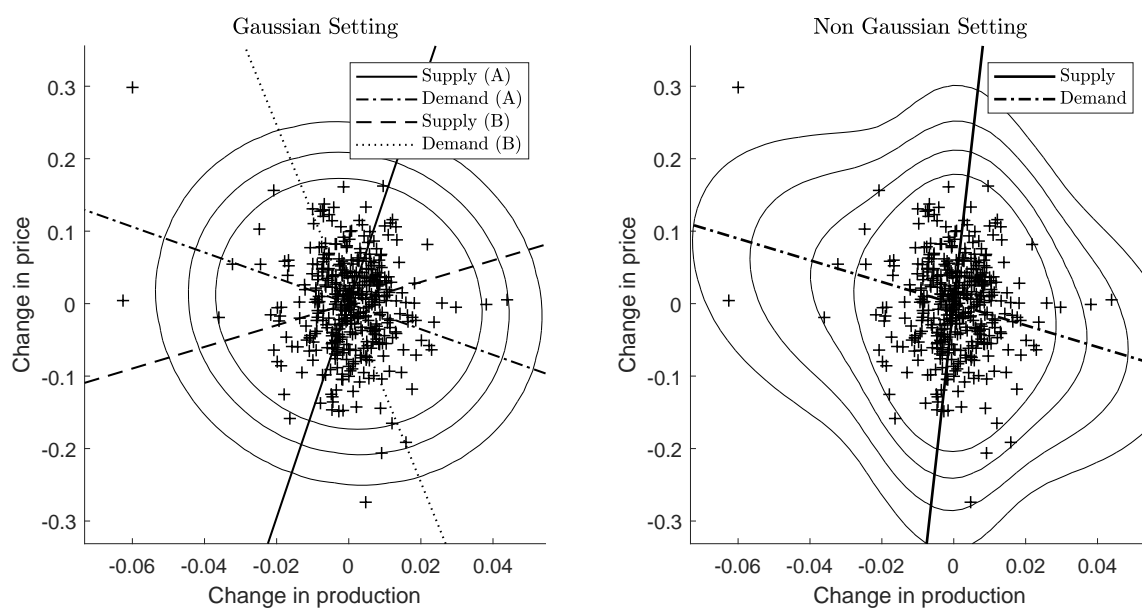


Figure 3.1: Identifying oil demand and supply curves by non-Gaussianity.

information is needed to achieve identification, given that identification is primarily based on non-Gaussianity.

To illustrate the differences in identification, consider Figure 3.1, which contains scatter plots for the unpredicted changes in oil production and oil prices.² In a standard SVAR model featuring Gaussian errors (left), many different models are in accordance with the observed correlation structure. For example, consider two arbitrary chosen supply and demand schedules (A and B), both satisfying common sense sign restrictions on the respective slopes.³ While both models are observationally equivalent, they imply very different dynamics for the oil market. In model A, supply is fairly inelastic and demand is elastic. Consequently, in such a model oil production would be mainly driven by supply shocks while oil prices would be largely caused by demand shocks. In turn, for model B, the supply is more elastic and demand is fairly inelastic, implying the exact opposite for the driving forces of oil prices. Incorporating prior knowledge on elasticities ultimately boils down to picking a range of models from the class of observationally equivalent models, thereby shaping the answer about the drivers of oil price fluctuations a priori.

On the right of Figure 3.1, I illustrate how non-Gaussianity provides an alternative way to discriminate among observationally equivalent models. The solid lines correspond to contour lines of the estimated joint density, implied by the non-Gaussian model developed in this paper. As visualized in the Figure, the model captures the irregular distribution to

²For this Figure, the unexpected changes were computed as residuals of a linear VAR with $p = 24$ lags. The residuals span the period from January 1985 to October 2018.

³Many economists would agree that the slope of the demand curve is negative, and that of the supply curve is positive.

come up with a unique supply and demand schedule consistent with the data. Hereby, the non-Gaussian shape makes certain shifts of the supply and demand curve more likely than others, this way working as a probabilistic instrument (see also [Rigobon \(2003\)](#) for a similar interpretation in SVARs identified by heteroskedasticity). For the data considered in the scatter plot, the statistical identification scheme estimates a very steep supply curve and rather flat demand curve. This result is in line with the main message of the paper: supply shocks contribute very little to fluctuations of oil prices.

A potential drawback of a purely statistically identified oil market model is that implausible large oil price elasticities are included in the posterior distribution. Therefore, I suggest to combine identification via non-Gaussianity with the conventional approach of restricting oil price elasticities. Note that in comparison to a classical SVAR, including prior information on elasticities has less drastic consequences on identification in the non-Gaussian framework. In particular, any additional information is overidentifying and thereby, the data has a chance to substantially revise prior beliefs on structural parameters. I illustrate this approach incorporating recent instrumental variable (IV) estimates of [Caldara et al. \(forthcoming\)](#) into the non-Gaussian model through a prior distribution. I find that inference on structural parameters is sharpened substantially once this information is incorporated. At the same time, the prior distribution is updated by fairly large amounts, pushing the posterior into parameter regions that imply a more inelastic supply and a more elastic demand, in line with the identifying information contained in non-Gaussianity.

In order to coherently combine statistical identification with prior information available for structural parameters, I develop a new Bayesian framework for non-Gaussian SVAR models. In particular, I exploit nonparametric density estimation for simultaneous identification of structural parameters and their underlying distributions. Univariate Dirichlet process mixture models (DPMM) ([Escobar and West; 1995](#)) are specified for each of the latent shocks in the SVAR, allowing to accurately model very general forms of density functions.⁴ Because structural shocks are unobserved, I make use of a constrained Dirichlet process prior ([Yang et al.; 2010](#)) to fix the mean and variance of the structural shocks at zero and unity, respectively. By exploiting a nonparametric framework, I avoid the need to specify a likelihood function in the Bayesian framework. Taking the wrong stance about the underlying form of non-Gaussianity can have severe consequences if parameters are to be identified by the distributional assumption. Furthermore, assessing identification is a straightforward task in this framework. Given that for each component the posterior distribution of the predictive density is readily available, a simple comparison with the kernel of a standard normal distribution suffices to check if there is enough non-Gaussianity in the data to use it for identification.

⁴Much alike kernel density estimators, DPMM are the workhorse model in Bayesian statistics to model unknown density functions nonparametrically.

With this model, I complement a list of econometric techniques, which have been proposed to identify shocks by non-Gaussianity over the last years. Among the frequentist approaches, [Lanne et al. \(2017\)](#) discuss Maximum Likelihood (ML) estimation and use a t -distribution in their empirical application involving monetary policy shocks. In turn, [Gourieroux et al. \(2017\)](#) consider pseudo-ML inference for independent component analysis (ICA) in general and apply it to SVAR analysis. More recently, GMM estimation is considered by [Lanne and Luoto \(2018\)](#) and [Herwartz \(2018\)](#) uses non-parametric dependence measures to disentangle non-Gaussian shocks. To the best of my knowledge, the only Bayesian approach in this literature is that of [Lanne and Luoto \(2016\)](#), who use t -distributed error terms to capture deviations from normality.

Compared to this literature, there are potential benefits from adopting the framework developed in this paper. First of all, the procedure is completely automatic and there is no need to take a stance about the form of non-Gaussianity underlying the structural shocks. This contrasts to those approaches which require to select either a likelihood function, a contrast function, or a set of moments prior estimation. Furthermore, as noted by [Boscolo et al. \(2004\)](#) in the context of ICA, only a flexible model which is able to learn the source distributions can yield consistent estimates in very general settings.⁵ On the other hand, the increase in robustness from a nonparametric approach comes at the cost of some loss in efficiency, if compared to an accurately specified parametric model. However, finding a reasonably accurate likelihood can be a difficult task. Furthermore, as noted in [Boscolo et al. \(2004\)](#), ICA approaches based on nonparametric density estimates do not necessarily require large sample sizes to achieve a more accurate source separation than less flexible methods.

Furthermore, some benefits are related to the Bayesian perspective adopted in this paper. As noted in [Lanne and Luoto \(2016\)](#), a Bayesian approach allows for a probabilistic labeling of the statistical shocks as economically meaningful shocks. This is an important tool given that the statistical identification approach may yield shocks, which carry no economic meaning at all. In the context of the oil market model considered in this paper, an economic labeling of the statistically identified shocks is obtained by computing posterior probabilities to satisfy conventional sign restrictions ([Kilian and Murphy; 2012](#)) and narrative sign restrictions ([Antolín-Díaz and Rubio-Ramírez; 2018](#)). Another benefit from adopting the Bayesian framework is the possibility to include prior information. This can be used to combine statistical with economic identification as suggested in the present paper, but also to regularize reduced form autoregressive parameters by Minnesota priors ([Doan et al.; 1984](#)) or priors learned on training samples.

This paper is not the first in considering a statistical identification approach to disentangle oil price fluctuations. [Herwartz and Plödt \(2016\)](#) and [Lütkepohl and Netšunajev \(2014\)](#)

⁵Note that in the ICA literature, the term “sources” is used for the unknown latent variables, and correspond to structural shocks in SVARs.

consider identification by non-Gaussianity and heteroskedasticity, respectively, for a similar SVAR model of the global oil market. However, both approaches rely on frequentist methods which do not allow for the incorporation of prior information about oil price elasticities. Furthermore, they do only consider a “most likely” labeling. In contrast, the probabilistic approach considered in this paper accounts for all the uncertainty with respect to these labels. Finally, given evidence by [Baumeister and Peersman \(2013\)](#) of a structural break in the oil market in the mid 80’s, I only consider data after 1985 while the samples considered in [Herwartz and Plödt \(2016\)](#) and [Lütkepohl and Netšunajev \(2014\)](#) go back to 1974.

The structure of the paper is as follows. In Section 3.2, the SVAR model with nonparametric density estimators is introduced and posterior inference by Markov chain Monte Carlo (MCMC) is discussed. Moreover, a set of tools helpful to assess identification in a given dataset is provided and illustrated on simulated data. In Section 3.3, the empirical analysis is conducted starting with the model specification, testing for identification, labeling statistical shocks with economic meaning and the incorporation of prior information about oil price elasticities. Finally, the model is used to conduct structural analysis. Section 3.4 concludes.

3.2 Methodology

In this section, I describe the SVAR model endowed with unknown density functions used to identify structural shocks by non-Gaussianity. The core of the model is a linear SVAR(p) specification for the conditional mean of a K -dimensional time series vector y_t :

$$y_t = \nu + \sum_{j=1}^p A_j y_{t-j} + B \varepsilon_t, \quad (3.2.1)$$

$$= \mu_y + \sum_{j=1}^{\infty} \Theta_j \varepsilon_{t-j}, \quad (3.2.2)$$

where $\varepsilon_t \sim (0, I_K)$ are serially uncorrelated, strictly stationary structural shocks with unconditional mean $E(\varepsilon_t) = 0$ and covariance $E(\varepsilon_t \varepsilon_t') = I_K$. Throughout the paper, invertibility is assumed, that is:

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

Therefore, the SVAR(p) has a MA(∞) representation as in (3.2.2) where $\Theta_j = \Phi_j B$, $\Phi_0 = I_K$, $\Phi_j = \sum_{i=1}^j \Phi_{j-i} A_i$ for $j \in \mathbb{N}$ with $A_i = 0$ for $i > p$. The ik -th entry of matrix Θ_j contains the impulse response, capturing the dynamic effect of structural shock k on the i -th variable in y_t , j periods after the shock has occurred.

The contemporaneous impact matrix B contains all structural parameters of the model and without further assumptions, B is only identified up to orthogonal rotations. To see this, note that the covariance matrix of the reduced form errors $u_t = B\varepsilon_t$ is given by $\Sigma_u = E(u_t u_t') = BB'$. An observationally equivalent model with the same first and second moments can be obtained by simply post-multiplying the impact matrix with an orthogonal matrix $\tilde{B} = BQ$ satisfying $QQ' = I_K$ and $Q^{-1} = Q'$.⁶ In classical SVAR analysis, economically motivated identifying restrictions are introduced on B in order to achieve identification of the model. Among the most popular are the imposition of short- and long run restrictions on the effect of structural shocks (Bernanke; 1986; Blanchard and Quah; 1989), sign restrictions (e.g. Uhlig (2005)) or restrictions implied by external instruments (Stock and Watson; 2012; Mertens and Ravn; 2013).

In this paper, identification of the shocks is achieved by distributional assumptions requiring that the components of ε_t are mutually independent and that at most one shock follows a Gaussian marginal distribution.⁷ This is sufficient to reduce the identification problem in B to that of sign switches and column permutations (see Lanne et al. (2017) and Gouriou et al. (2017) for details). To be precise, in a non-Gaussian scenario with independent shocks, all observationally equivalent models are of the form $\tilde{B} = BPD$ where P is a K -dimensional permutation matrix and D a diagonal matrix with elements ± 1 . That is, once a permutation and sign pattern is fixed, the model is identified.

In case of more than one Gaussian shock, the model is only partially identified (Maxand; forthcoming). Without loss of generality, let the first $r < K - 1$ components in ε_t be non-Gaussian and the others Gaussian. Further, divide $B = [B_1, B_2]$ where $\dim(B_1) = K \times r$ and $\dim(B_2) = K \times (K - r)$. Then, only B_1 is identified up to permutation and sign changes, corresponding to all structural parameters that capture the contemporaneous impact of the non-Gaussian shocks.

3.2.1 Constrained Dirichlet process mixture models

In the present paper, Bayesian nonparametric methods are used to explicitly model the non-Gaussian distributions of the structural shocks ε_t . In particular, independent univariate Dirichlet process mixture models (DPMM) are specified for each component in ε_t . Much alike the kernel density estimators, DPMMs are very popular tools in Bayesian statistics to model distributions of very general forms. Given that DPMMs are developed for observed data, I further adjust the density estimators to satisfy the identifying constraints for location and scale of the latent shocks, that is $E(\varepsilon_{it}) = 0$ and $E(\varepsilon_{it}^2) = 1$ ($i = 1, \dots, K, t = 1, \dots, T$).

⁶In particular, $E(\tilde{B}\varepsilon_t\varepsilon_t'\tilde{B}') = \tilde{B}\tilde{B}' = BQQ'B' = BB' = \Sigma_u$.

⁷Note that assuming mutual independence is stronger than assuming contemporaneous uncorrelatedness.

For the ease of exposition, assume that the structural shocks were observed, and therefore, one would not need the mean and variance constraints on their distribution. In the remainder of the paper, denote such “unconstrained” shocks by ε_{it}^* for $i = 1, \dots, K$. A DPMM is given by the following hierarchical model:

$$\begin{aligned}\varepsilon_{it}^* | \theta_{it}^* &\sim p(\varepsilon_{it}^* | \theta_{it}^*), \\ \theta_{it}^* &\sim G_i^*, \\ G_i^* &\sim \text{DP}(G_{i0}, \alpha_i),\end{aligned}$$

where $p(\varepsilon_{it}^* | \theta_{it}^*)$ is a probability distribution parameterized by θ_{it}^* and can be thought of a conditional likelihood at time t . G_i^* is the corresponding prior distribution for θ_{it}^* and has the additional feature of being random itself, following a Dirichlet process (DP) $G_i^* \sim \text{DP}(G_{i0}, \alpha_i)$ (Ferguson; 1973). A DP is uniquely characterized by a base distribution G_{i0} and a concentration parameter α_i . Throughout the paper, the Gaussian DPMM of Escobar and West (1995) is adopted implying that $p(\varepsilon_{it}^* | \theta_{it}^*) \sim \mathcal{N}(\mu_{it}^*, \sigma_{it}^{*2})$ and therefore $\theta_{it}^* = (\mu_{it}^*, \sigma_{it}^*)'$. As a base distribution G_{i0} , the conjugate normal inverse gamma distribution is chosen.⁸

In the following, two instructive representations of the DPMM are discussed to attain a better understanding of the Bayesian model casted on the structural shocks. The first is known as the Pólya Urn representation (Blackwell and MacQueen; 1973) and gives an intuitive way to generate samples of θ_{it}^* from its prior, by integrating out G_i^* . In particular, for $t = 1, 2, \dots, T$:

$$\theta_{it}^* | \theta_{i1}^*, \dots, \theta_{it-1}^* \sim \frac{1}{t-1+\alpha_i} \sum_{l=1}^{t-1} \delta_{\theta_{il}^*} + \frac{\alpha_i}{t-1+\alpha_i} G_{i0},$$

where $\delta_{(\cdot)}$ is the Dirac measure. In words, with (equal) probability $\frac{1}{t-1+\alpha_i}$, θ_{it}^* is drawn from one of the $t-1$ previously drawn parameters, and with probability $\frac{\alpha_i}{t-1+\alpha_i}$, θ_{it}^* is drawn from the base distribution (G_{i0}). The Pólya Urn scheme illustrates the main properties of the DPM prior of θ_{it}^* . First, the realizations are almost surely discrete. Second, there is a “richer get richer” property implied by this prior which leads to a clustering of θ_{it}^* . Therefore, the model for ε_{it}^* can be interpreted as a flexible mixture model where the number of components is random and increasing in T . The strength of clustering is governed by the concentration parameter α_i and lower values are associated with less mixture components for a given T .⁹

⁸A normal inverse Gamma distribution is given by $(\mu, \sigma^2) \sim \text{NIG}(a, b, m, s) \sim p(\mu | \sigma^2) p(\sigma^2)$ where $p(\sigma^2) \sim i\mathcal{G}(a, b)$ and $\mu | \sigma^2 \sim \mathcal{N}(m, \sigma^2 s)$.

⁹Given that α_i governs the complexity of the density estimator, it is often treated as random endowed it with a hyperprior. See Appendix 3.A.1 for details.

An alternative instructive way to think about the DP prior can be obtained by what is known as its stick breaking representation (Sethuraman; 1994). Here, G_i^* is not marginalized out, but rather given an explicit form which is constructed as follows:

$$\begin{aligned}\theta_{i1}^*, \theta_{i2}^*, \dots &\sim G_{i0}, \\ V_{i1}, V_{i2}, \dots &\sim \text{Beta}(1, \alpha_i), \\ \pi_{ik} &:= V_{ik} \prod_{j=1}^{k-1} (1 - V_{ij}), \\ G_i^* &= \sum_{k=1}^{\infty} \pi_{ik} \delta_{\theta_{ik}^*}.\end{aligned}$$

Here, π_{ik} 's are mixing probabilities obtained iteratively based on the random beta variables V_{ik} and θ_{ik}^* denoting the parameters associated with the k -th mixture component rather than the t -th observation. The stick breaking representation of G_i^* clarifies why the DPMM implies an infinite mixture model for ε_{it}^* . Furthermore, given the Gaussian setting adopted in this paper, the first two moments of the distribution of $\varepsilon_{it}^* | G_i^*$ can be easily calculated using standard formulas for Gaussian mixture models:

$$\begin{aligned}\text{E}(\varepsilon_{it}^* | G_i^*) &= \mu_i^* = \sum_{k=1}^{\infty} \pi_{ik} \mu_{ik}^*, \\ \text{Var}(\varepsilon_{it}^* | G_i^*) &= \sigma_i^{*2} = \sum_{k=1}^{\infty} \pi_{ik} ((\mu_{ik}^* - \mu_i^*)^2 + \sigma_{ik}^{*2}).\end{aligned}$$

With this explicit expression of the first two moments of the unconstrained shocks ε_{it}^* , it is straightforward to define a constrained DP prior G_i , such that the DPMM generates shocks ε_{it} with zero mean and unit variance. In particular, let $G_i | G_i^* = \sum_{k=1}^{\infty} \pi_{ik} \delta_{\theta_{ik}}$ where $\mu_{ik} = \sigma_i^{*-1} (\mu_{ik}^* - \mu_i^*)$, $\sigma_{ik}^2 = \sigma_i^{*-2} \sigma_{ik}^{*2}$, and $\theta_{ik} = (\mu_{ik}, \sigma_{ik}^2)'$. Then, the constrained DPMM considered in this paper reads:

$$\begin{aligned}\varepsilon_{it} | \theta_{it} &\sim p(\varepsilon_{it} | \theta_{it}), \\ \theta_{it} &\sim G_i | G_i^*, \\ G_i^* &\sim \text{DP}(G_{i0}, \alpha_i).\end{aligned}$$

It is straightforward to show that for the constrained DPMM, indeed it is $\text{E}(\varepsilon_{it} | G_i) = 0$ and $\text{Var}(\varepsilon_{it} | G_i) = 1$ as required for the identification of the latent shocks.

3.2.2 Bayesian inference

In the following, I outline how inference can be conducted for the resulting SVAR model. Assuming the availability of p fix presample values y_0, \dots, y_{-p+1} , the model reads in compact notation:

$$y_t = \nu + Ax_t + B\varepsilon_t, \quad (3.2.3)$$

$$\varepsilon_{it} | \theta_{it} \sim \mathcal{N}(\mu_{it}, \sigma_{it}^2), \quad (3.2.4)$$

$$\theta_{it} \sim G_i, \quad (3.2.5)$$

for $i = 1, \dots, K$, $t = 1, \dots, T$, $x_t = [y'_{t-1}, y'_{t-2}, \dots, y'_{t-p}]'$, $A = [A_1, A_2, \dots, A_p]$ and G_i distributed according to a constrained DP as considered in Section 3.2.1. Given that the mixture parameters underlying G_i are subject to a complex nonlinear constraint, direct inference on θ_{it} would be very challenging. However, following the ideas in Yang et al. (2010), it is straightforward to conduct inference for an underidentified parameter expanded (PX) model that is based on the unconstrained DP prior G_i^* . Their strategy, which I adopt in this paper, involves to design an inference algorithms for a PX model, and ex-post map the draws back to the inferential model described in equations (3.2.3)-(3.2.5).

The PX representation can be obtained by rewriting the model in terms of unconstrained shocks ε_t^* . In particular, define $\varepsilon_t = (\Sigma^*)^{-\frac{1}{2}}(\varepsilon_t^* - \mu^*)$, where $\mu^* = [\mu_1^*, \dots, \mu_K^*]'$ and $\Sigma^* = \text{diag}([\sigma_1^{*2}, \dots, \sigma_K^{*2}])$. Then, an observationally equivalent model to equations (3.2.3)-(3.2.5) is given by:

$$y_t = \Gamma X_t + B^* \varepsilon_t^*, \quad (3.2.6)$$

$$\varepsilon_{it}^* | \theta_{it}^* \sim \mathcal{N}(\mu_{it}^*, \sigma_{it}^{*2}), \quad (3.2.7)$$

$$\theta_{it}^* \sim G_i^*, \quad (3.2.8)$$

where $\Gamma = [\nu^*, A]$, $X_t = [1, x_t']'$, $\nu^* = \nu - B(\Sigma^*)^{-\frac{1}{2}}\mu^*$ and $B^* = B(\Sigma^*)^{-\frac{1}{2}}$. MCMC inference for the PX model given by equations (3.2.6)-(3.2.8) is relatively easy. Given that θ_{it}^* is unconstrained, standard algorithms in the literature developed for DPMMs can be used as a building block. After generating a large amount of draws from the (underidentified) “working parameters” of the PX model, the output is simply transformed back in order to conduct inference on the (identified) parameters of the original model, which are of ultimate

interest. In particular, for the l -th draw of the MCMC output, the auxiliary parameters are mapped back to the original SVAR parameters by:

$$\begin{aligned} B^{(l)} &= (B^*)^{(l)} \left((\Sigma^*)^{\frac{1}{2}} \right)^{(l)}, \\ v^{(l)} &= (v^*)^{(l)} + (B^*)^{(l)} (\mu^*)^{(l)}. \end{aligned}$$

Furthermore, inference for the corresponding parameters underlying G_i can be based on the transformations:

$$\begin{aligned} \mu_{ik}^{(l)} &= (\sigma_i^{*(l)})^{-1} (\mu_{ik}^{*(l)} - \mu_i^{*(l)}), \\ \sigma_{ik}^{2(l)} &= (\sigma_i^{*(l)})^{-2} (\sigma_{ik}^{*(l)})^2. \end{aligned}$$

Note that since MCMC inference is performed in the PX model, it is of computational convenience to specify prior distributions directly on the PX parameters Γ and B^* , rather than on the SVAR parameters of equation (3.2.3).¹⁰ For most of the applications, the fact that prior distributions are specified for the PX model should not be problematic. First, consider the autoregressive parameters contained in A . Given that these parameters are the same in the inferential and PX model, prior information e.g. in form of a Minnesota prior or a prior based on training samples, can be set up in the usual way. With respect to B , most of the prior information used in the literature is scale free and therefore, can be directly set on B^* . For example, sign restrictions can be imposed by restricting the posterior of B^* yielding equivalent results. Furthermore, prior information on elasticities, as considered in the present paper, are just ratios of elements in a certain column of B and therefore, can also be specified directly on B^* .

In Appendix 3.A.1, a detailed MCMC algorithm is developed that cycles through the conditional distributions of the PX model parameters. The algorithm uses the conditional sampler of [Ishwaran and James \(2001\)](#) as a building block for the DPMMs. This sampler is based on a truncation of the DP prior for a large number of clusters N , imposed by setting $V_{iN} = 1$.¹¹

Recall that within the non-Gaussian model, the impact matrix B is only identified up to sign switches and column permutations. Therefore, the posterior distribution is multimodal around $K! \times 2^K$ possible permutations and sign configurations of B .¹² This can be highly

¹⁰The reason is that otherwise, the conditional independence structure of the PX model parameters is lost by tying together SVAR (Γ, B^*) with DP prior parameters (those underlying μ^* and Σ^*). In turn, this complicates the design of efficient MCMC inference.

¹¹By setting $V_{iN} = 1$, it is guaranteed that $\sum_{k=1}^N \pi_{ik} = 1$ and hence, the truncated DP prior $G_i^* = \sum_{k=1}^N \pi_{ik} \delta_{\theta_{ik}^*}$ is a well defined distribution.

¹²In particular, there are $K!$ ways to permute the columns and 2^K ways to order the signs. See also [Woźniak and Droumaguet \(2015\)](#) for a detailed discussion.

problematic for Monte Carlo based inference as considered in this paper. In particular, if the modes are not isolated sufficiently, the Markov Chain eventually explores multiple permutations, and ergodic averages of the output can become meaningless tools to draw inference on elements in B . To avoid this problem known as “label switching”, one approach is to formulate a dogmatic prior that isolates one mode, see e.g. the approach taken in [Lanne and Luoto \(2016\)](#). However, as discussed in [Jasra et al. \(2005\)](#), this solution is undesirable given that it implicitly restricts the posterior in potentially odd ways. An alternative way adopted in this paper is to solve the label switching problem ex-post, that is only after running an unconstrained MCMC sampler. In case that label-switching occurred in the Markov Chains, I use an algorithm based on k -means clustering to partition the MCMC output according to the different modes it explores. Once a partition is obtained by the algorithm, the draws are re-permuted to represent a unique mode. The reader interested in the details of the algorithm is referred to [Appendix 3.A.2](#). Note that for most of the computations conducted in this paper, the posterior modes of B were isolated sufficiently and label switching was of no concern. However, particularly in case of running very large Markov Chains the MCMC eventually explored multiple modes. In that case, I used the algorithm discussed in the appendix to repermute the output before proceeding with inference.

3.2.3 Assessing identification

A major advantage of the present framework is that the structural parameters in B are estimated jointly with the density of each structural shock. Therefore, deviations from Gaussianity at any point of the distribution are exploited for identification. Furthermore, the availability of density estimates allows for an easy assessment of the identifying conditions in a given dataset. Recall that for all structural parameters to be identified, at most one structural shock is allowed to have a Gaussian marginal distribution. In case of more Gaussian components, only those columns of B are identified which correspond to the non-Gaussian shocks.

Let φ denote the vector of parameters in the model and $p(\varphi|Y)$ its posterior distribution. The first strategy to statistically verify identification for a given dataset is to compute the posterior distribution of the predictive densities $p(\varepsilon_{i,T+1}|Y) = \int p(\varepsilon_{i,T+1}|Y, \varphi)p(\varphi|Y)d\varphi$, and compare it with density values of a standard normal distribution. If at least $K - 1$ of the components show substantial deviations from Gaussianity, this constitutes statistical evidence for identification. For a given MCMC output of length M , a Monte Carlo estimator of the predictive density is given by:

$$p(\varepsilon_{i,T+1}|Y) \approx M^{-1} \sum_{l=1}^M p\left(\varepsilon_{i,T+1}|Y, \varphi^{(l)}\right) = M^{-1} \sum_{l=1}^M \sum_{k=1}^N \pi_{ik}^{(l)} \phi\left(\varepsilon_{i,T+1}; \mu_{ik}^{(l)}; \sigma_{ik}^2{}^{(l)}\right),$$

where $\phi(\cdot; \mu, \sigma)$ denotes the density of a normal distribution with mean μ and variance σ^2 . The Monte Carlo estimator also allows for a straightforward assessment of the uncertainty of the density estimator by means of posterior quantiles. Corresponding Monte Carlo estimates can be obtained by the empirical quantiles of the draws $p(\varepsilon_{i,T+1}|Y, \varphi^{(l)})$, $l = 1, \dots, M$.

An alternative possibility to assess identification is to compute posterior moments of the predictive density and compare these to moments of a standard normal distribution. Monte Carlo estimates of skewness and kurtosis of the predictive density are given as (Frühwirth-Schnatter; 2006):

$$\begin{aligned} E(\varepsilon_{i,T+1}^3|Y) &\approx M^{-1} \sum_{l=1}^M E(\varepsilon_{i,T+1}^3|Y, \varphi^{(l)}), \\ &= M^{-1} \sum_{l=1}^M \sum_{k=1}^N \pi_{ik}^{(l)} (\mu_{ik}^{3(l)} + 3\mu_{ik}^{(l)} \sigma_{ik}^{2(l)}), \\ E(\varepsilon_{i,T+1}^4|Y) &\approx M^{-1} \sum_{l=1}^M E(\varepsilon_{i,T+1}^4|Y, \varphi^{(l)}), \\ &= M^{-1} \sum_{l=1}^M \sum_{k=1}^N \pi_{ik}^{(l)} (\mu_{ik}^{4(l)} + 6\mu_{ik}^{2(l)} \sigma_{ik}^{2(l)} + 3\sigma_{ik}^{4(l)}), \end{aligned}$$

where again, the Monte Carlo framework allows for a straightforward assessment of the uncertainty of these estimates by means of posterior quantiles. Corresponding Monte Carlo estimates can be obtained by the empirical quantiles of the draws $E(\varepsilon_{i,T+1}^3|Y, \varphi^{(l)})$ and $E(\varepsilon_{i,T+1}^4|Y, \varphi^{(l)})$ for $l = 1, \dots, M$.

Finally, it can also be instructive to inspect the MCMC output for each column of the structural impact matrix B . If there are less than $K - 1$ non-Gaussian shocks, columns in B corresponding to Gaussian shocks are not identified and therefore, the MCMC output will be ill-behaved and subject to high autocorrelation. In turn, those columns corresponding to non-Gaussian shocks are point-identified which should translate into stationary draws and more reasonable autocorrelation patterns.

In the following, I illustrate the above mentioned tools for simulated data from a trivariate model. To show the different behavior under non-identification and identification, two scenarios are considered. In the first scenario, there is just one Gaussian shock, implying that the model is identified. In particular, I set $\varepsilon_{1t} \sim \log \mathcal{N}(0, 0.25)$, $\varepsilon_{2t} \sim t_5$, and $\varepsilon_{3t} \sim \mathcal{N}(0, 1)$, and standardize the shocks ex-post to have zero mean and unit variance. In the second scenario, ε_{1t} follows the same non-Gaussian distribution as in the first scenario, while the remaining shocks follow a standard normal distribution, that is $[\varepsilon_{2t}, \varepsilon_{3t}]' \sim \mathcal{N}(0, I_2)$. Here, there is a lack of identification, and only the first column of B is identified, representing the structural parameters corresponding to the impact of the only non-Gaussian shock in the

model. For both scenarios, shocks of length $T = 1000$ observations are simulated. The shocks are then fed into the following stable SVAR(2), which is based on a model considered in [Herwartz \(2018\)](#) for a simulation study:¹³

$$B = \begin{pmatrix} 2 & 0.5 & -0.5 \\ 1 & 2 & 0 \\ 0 & 1 & 1 \end{pmatrix}, A_1 = \begin{pmatrix} 1.24 & -0.09 & -0.16 \\ 0.13 & 0.94 & -0.06 \\ 0.24 & 0.30 & 1.03 \end{pmatrix}, A_2 = \begin{pmatrix} -0.37 & 0.04 & 0.08 \\ -0.06 & -0.22 & 0.03 \\ -0.12 & -0.15 & -0.26 \end{pmatrix}$$

In [Figure 3.2](#), estimated predictive densities are plotted together with 90% highest posterior density (HPD) credible intervals for the output of both datasets. In addition, MCMC chains of length 5000 are plotted for the structural parameters in B . The first two columns of the figure correspond to estimation output of the identified model (scenario 1). In line with the data generating process, estimates of the predictive densities indicate that there is considerable evidence of non-Gaussianity in two of the three shocks. As expected, the resulting draws of the structural parameters are well behaved, showing a reasonable autocorrelation pattern fluctuating around a constant mean. The right two columns of the figure correspond to the output of the non-identified model with two Gaussian components (scenario 2). For two of the three shocks, predictive density indicate no evidence against Gaussianity since the 90% HPD credible intervals include values implied by a standard normal. Furthermore, the Markov chains of the structural parameters are ill-behaved, that is subject to high autocorrelation and without a clear mean, underlining the lack in identification. However, in line with partial identification results, those parameters in B measuring the contemporaneous impact of the non-Gaussian component are identified. Finally, note that plots of Markov chains are also suitable to detect if label switching is an issue for the MCMC output of identified columns. If the answer is positive, the output must be re-permuted and the identification analysis repeated. For the output in [Figure 3.2](#), there is no label-switching visible.

To complement the illustration, [Table 3.1](#) provides posterior quantiles of skewness (s_α) and kurtosis (κ_α) of the structural shock, as implied by the Gaussian mixture distribution. For scenario 1, the posterior estimates successfully reveal that the first two shocks are non-Gaussian, given that the HPD credible intervals of skewness and kurtosis are very different from values implied by a standard normal distribution (0 and 3 respectively). As expected, for the last shock there is no evidence against Gaussianity. The method also gives the correct information in scenario 2. In line with the data generating process, only ε_{1t} is detected as non-Gaussian with considerable skewness and excess kurtosis as indicated by HPD credible intervals. In turn, for ε_{2t} and ε_{3t} there is no evidence against Gaussianity, indicating a lack of identification. Overall, looking at posterior estimates for higher moments

¹³The values of B were slightly adjusted for illustrative purposes.

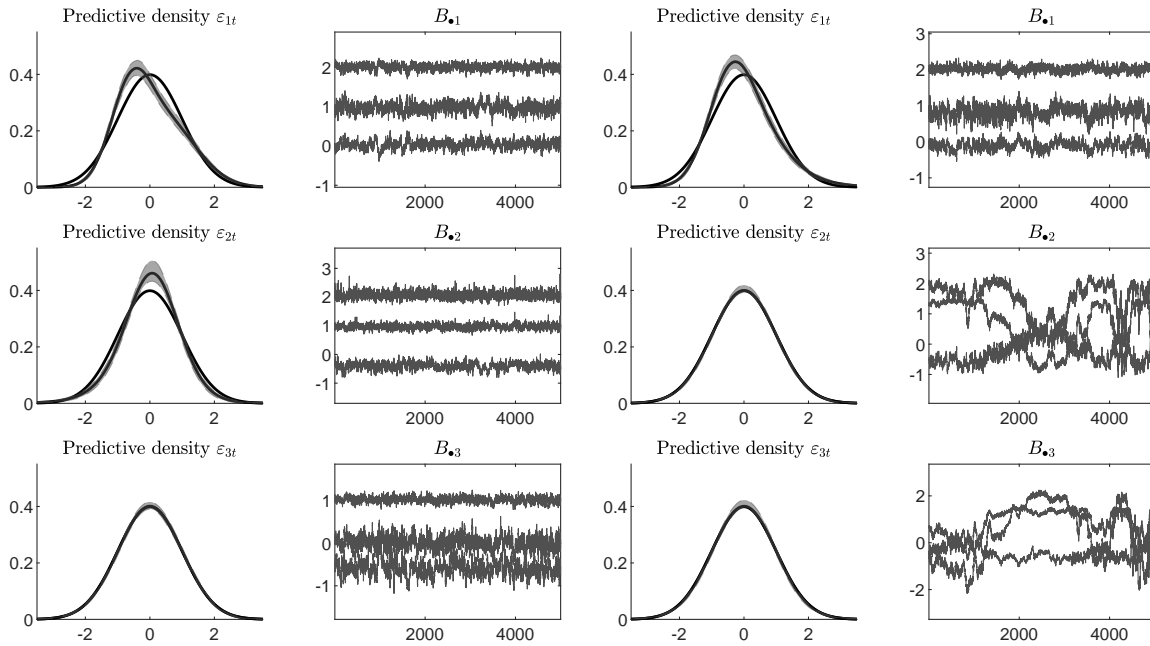


Figure 3.2: Illustration of identification analysis for an identified (left two columns) and an unidentified model (right two columns). Column 1 and 3 provide plots of predictive densities together with 90% HPD credible intervals, set in contrast to the density of a standard normal distribution. Column 2 and 4 provide plots of the Markov Chains corresponding to elements in B .

can provide an alternative tool to detect non-Gaussianity in structural shocks and thereby help to assess identification in a given dataset.

3.3 Importance of supply and demand channels for oil price fluctuations

In this section, I apply the methodology to identify supply and demand shocks driving the global oil market. The structure of the empirical analysis is as follows. In Section 3.3.1, I introduce the exact specification of the SVAR model, including data sources, sample period and priors. Furthermore, I check if there is enough non-Gaussianity in the shocks to identify the model by statistical means. In Section 3.3.2, I compute posterior probabilities to satisfy conventional sign restrictions (Kilian and Murphy; 2012) as well as narrative restrictions (Antolín-Díaz and Rubio-Ramírez; 2018) in order to attach economic meaning to the statistical shocks. Once a labeling is available, I incorporate additional prior information on supply and demand elasticities (Section 3.3.3) before I present a range of empirical implications (Section 3.3.4), including impulse response functions, historical decompositions and forecast error variance decompositions.

Table 3.1: Illustration of identification analysis based on higher moments.

	Identified Model						Unidentified Model					
	$s_{0.05}$	$s_{0.5}$	$s_{0.95}$	$\kappa_{0.05}$	$\kappa_{0.5}$	$\kappa_{0.95}$	$s_{0.05}$	$s_{0.5}$	$s_{0.95}$	$\kappa_{0.05}$	$\kappa_{0.5}$	$\kappa_{0.95}$
ε_{1t}	0.65	0.8	0.98	3.56	4.03	4.79	0.55	0.72	0.93	3.69	4.22	5.14
ε_{2t}	-0.3	0.28	1.16	6.44	8.7	15.24	-0.13	0	0.12	2.97	3.03	3.51
ε_{3t}	-0.08	0	0.16	2.92	3.04	3.5	-0.13	0.03	0.18	2.89	3.08	3.58

The table gives posterior α quantiles for kurtosis κ_α and skewness s_α implied by each structural shock's predictive density $p(\varepsilon_{it}|Y), i = 1, 2, 3$.

3.3.1 Model specification and identification analysis

The SVAR considered in this paper mainly follows the specification proposed in [Kilian \(2009\)](#) for the global crude oil market. The model is based on $p = 24$ lags and includes three monthly variables: $y_t = [\Delta\text{prod}_t, \text{rea}_t, \text{rpo}_t]'$. Here, Δprod_t indicates the percentage change in world crude oil production, rea_t is a measure of global economic activity and rpo_t is the log of the real oil price.¹⁴ To proxy global economic activity, the world industrial production index of [Baumeister and Hamilton \(forthcoming\)](#) is used. Following the recommendations in [Hamilton \(2018\)](#), the cyclical component is extracted using the two-year change in logarithms. A plot of all time series data included in the analysis is provided in Appendix 3.B. Given strong evidence of a structural break in the oil market during the 80's ([Baumeister and Peersman; 2013](#)), I use data from January 1985 to July 2018 for inference purposes. Furthermore, observations from 1983 to 1985 serve as presample values.¹⁵ Instead of completely discarding the data available prior 1983, I use it as a training sample for the prior distribution of the autoregressive parameters. In particular, a normal prior is specified as $\text{vec}(A) \sim \mathcal{N}(\text{vec}(\hat{A}), \text{Cov}(\text{vec}(\hat{A})))$, where the corresponding moments are calibrated using ordinary least squares estimates for data from 1974M1 to 1982M12. Until further notice, the remainder of the prior distributions are chosen to be uninformative, specifically $\text{vec}(v^*) \sim \mathcal{N}(0_{K \times 1}, 1000I_K)$ and $\text{vec}(B^*) \sim \mathcal{N}(0_{K^2 \times 1}, 1000I_{K^2})$. The hyperparameters for the K individual DPMMs are set to $a = 3, b = 0.5, m = 0, s = 2, a_\alpha = 2$ and $b_\alpha = 2$.¹⁶

Before any meaningful economic analysis can be conducted with the non-Gaussian SVAR, it is necessary to assess if the structural parameters are identified in the specified model. Recall that for identification of the parameters in B , one needs that at least two of the three

¹⁴World oil production has been downloaded from the U.S. Energy Information Administration (EIA). The real price of oil is constructed on the basis of U.S. Refiner Acquisition Cost of Crude Oil, available from the EIA, which is further deflated by the U.S. Consumer Price Index (CPI) available at the Federal Reserve Economic Database (FRED).

¹⁵Reestimating the model for a larger sample that starts in 1974 yields substantially different results in terms of estimated impact matrix B supporting the argument of a structural break.

¹⁶Reestimating the model for other values for these hyperparameters yields very similar results.

Table 3.2: Identification analysis for the oil market SVAR based on higher moments of the structural shocks.

	$s_{0.05}$	$s_{0.5}$	$s_{0.95}$	$\kappa_{0.05}$	$\kappa_{0.5}$	$\kappa_{0.95}$
ε_{1t}	-1.60	-0.50	0.41	6.64	9.39	15.90
ε_{2t}	-0.90	0.07	1.20	5.13	8.28	16.02
ε_{3t}	-0.68	-0.17	0.36	3.09	4.06	6.47

The table gives posterior α quantiles for kurtosis κ_α and skewness s_α implied by each structural shock's predictive density $p(\varepsilon_{it}|Y), i = 1, 2, 3$.

shocks follow non-Gaussian marginal distributions. As described in Section 3.2, this can be easily checked by comparing density values of the predictive distributions with those of a standard normal distribution. In Figure 3.3, the corresponding estimates together with 90% HPD credible intervals are plotted for each structural shock's predictive density. For the model specified above, there is considerable evidence that at least two out of the three structural shocks are non-Gaussian. Both, the first and second shock, show considerable heavier concentration of probability mass near zero than implied by a standard normal. For the third shock, the evidence is less concise given that the standard normal density is contained in the 90% HPD credible intervals of the predictive density. An equivalent conclusion can be obtained by looking at posterior skewness and kurtosis of the predictive densities (Table 3.2). First, with respect to skewness, none of the shocks shows significant deviations from normality. However, for two structural shocks there is strong evidence of excess kurtosis. Given that this points toward identification of the model, it is of no surprise that the posterior output of the structural parameters is well behaved in line with full identification. In the second row of Figure 3.3, the Markov Chains for the corresponding columns in B are plotted separately and a quick visual inspection confirms that the parameters are likely to be identified. Furthermore, the plots also indicate an absence of label switching issues, which implies that the modes are well separated and the output can be processed safely for inference purposes.

3.3.2 Identifying supply and demand components in the non-Gaussian model

Without further analysis, the structural shocks from the non-Gaussian SVAR are only statistically identified and cannot be interpreted. In the following exercise, an economic labeling is attached to these shocks. To this end, I compute posterior probabilities that

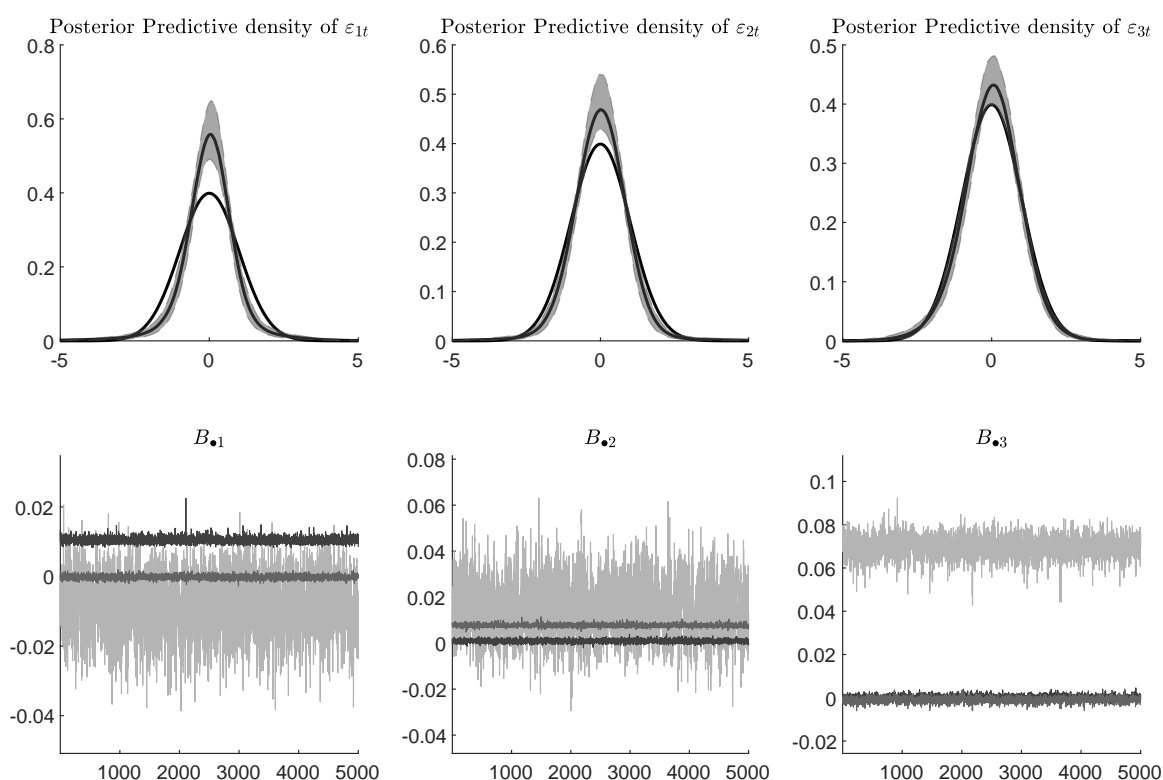


Figure 3.3: Identification analysis for the oil market SVAR. Top: Posterior estimates of each structural shocks predictive density with 90% HPD credible intervals (shaded areas). Bottom: MCMC output for the corresponding structural parameters in B .

statistical shocks satisfy economically motivated restrictions.¹⁷ The restrictions considered in this paper are the sign restrictions used by [Kilian and Murphy \(2012\)](#) as well as the narrative restrictions of [Antolín-Díaz and Rubio-Ramírez \(2018\)](#). The idea is that if a certain statistical shock displays a high probability to satisfy the restrictions considered in these papers, the corresponding economic label can be adopted. Both papers identify the following three structural drivers of the global oil market: a supply shock (ε_t^s), an aggregate demand (AD) driven shock (ε_t^{ad}) and a oil market specific demand (OD) shock (ε_t^{od}), capturing precautionary and speculative components.

[Kilian and Murphy \(2012\)](#) consider the following sign restrictions (SR) on the impact effects of these shocks. First, a negative supply shock is restricted to decrease production, while both price and global activity must increase. Second, the AD shock is associated with an increase in all three variables. Finally, the oil market specific demand shock must increase production, decrease activity and increase the oil price. In [Table 3.3](#), the corresponding posterior probabilities to satisfy these sign restrictions are tabulated. In the first panel of the

¹⁷In the Bayesian framework, these probabilities are given as the relative frequencies the posterior simulation algorithm visits parameter regions where the restrictions are satisfied.

Table 3.3: Posterior probabilities to satisfy sign restrictions of [Kilian and Murphy \(2012\)](#).

	jointly						shock by shock		
	permutation						$p(\varepsilon_{1t} = \cdot Y)$	$p(\varepsilon_{2t} = \cdot Y)$	$p(\varepsilon_{3t} = \cdot Y)$
ε_t^s	3	3	2	2	1	1	0.31	0.09	0.42
ε_t^{ad}	2	1	3	1	3	2	0.05	0.83	0.14
ε_t^{od}	1	2	1	3	2	3	0.14	0.01	0.38
$p(\text{SR} Y)$	0.12	0.00	0.00	0.00	0.01	0.14			

Left: posterior probabilities $p(\text{SR}|Y)$ that a certain permutation of shocks yields a system which satisfies the sign restrictions of [Kilian and Murphy \(2012\)](#). Right: posterior probabilities $p(\varepsilon_{it} = \cdot | Y)$ that the i th shock qualifies as supply shock (ε_t^s), aggregate demand shock (ε_t^{ad}), or oil market specific shock (ε_t^{od}).

Table, the probabilities are given that a certain labeling of the statistical shocks jointly satisfy all sign restrictions. Recall that there exist $K!$ ways of ordering the structural shocks, and therefore, each permutation must be considered separately. For example, the first column gives the joint probability that ε_{3t} satisfies the restrictions of a supply shocks, ε_{2t} satisfies the restrictions of an AD shock and ε_{1t} those of an OD shock. This analysis yields that there are two permutations with nonzero probabilities, yielding 12% and 14% respectively. Both values are low and cannot be used convincingly for labeling purposes. In the second panel of Table 3.3, the individual posterior probabilities are given for each shock. That means for each of the statistically identified shocks (ε_{1t} , ε_{2t} and ε_{3t}), the posterior probabilities of qualifying as a supply shock ε_t^s , an AD shock ε_t^{ad} or an oil market specific shock ε_t^{od} , are computed separately. The first component (ε_{1t}) satisfies the sign restrictions of a supply shock with 31% probability, that of an AD shock with 5% and an other demand shock with probability 14%. While these results point into the direction of a supply shock, there seems to be too much uncertainty about its impact effect to label the shock convincingly as such. However, for the second shock this analysis is more meaningful. With 83% percent probability ε_{2t} satisfies the SRs of an AD shock which allows to adopt its economic meaning. For the last shock, there is again too much uncertainty about its impact effect, which allows for no ultimate statement whether it qualifies as supply or demand shock. Summing up, it seems difficult to properly disentangle the first and last shock into supply or demand components using impact sign restrictions.

Therefore, the exercise is repeated for a set of “narrative restrictions” suggested by [Antolín-Díaz and Rubio-Ramírez \(2018\)](#) for the global oil market model. Their idea is to identify SVARs by imposing restrictions on the economic implications the model has for a list of historical events. In particular, identifying restrictions are imposed on the sign of structural shocks at certain dates and on corresponding historical decompositions (HD). HDs decompose the movements in y_t into its primitive driving forces $\varepsilon_\tau, \tau = 1, \dots, t$. They are

defined as $HD_{it}^{(j)} = \sum_{\tau=0}^{t-1} \Theta_{ij,\tau} \varepsilon_{t-\tau}$ and give the cumulative effect of the j -th structural shock on the i -th variable at time t in the system.¹⁸

For each structural shock, posterior probabilities to satisfy the following narrative restrictions are computed.¹⁹ The first set (NR-1) restricts the sign of the oil supply shock to be negative at the following dates: the start of the Persian Gulf War (1990M8), the Venezuelan oil strike (2002M12), the start of the Iraq War (2003M3) and the start of the Libyan Civil War (2011M2). The second set of restrictions (NR-2) considers the contribution of the supply shock for the HD of global oil production during the same dates. Specifically, the absolute value of the contribution of an oil supply shock must be larger than the absolute value of the contribution made by any other shock in the model. Finally, the last restriction (NR-3) considers the historical decomposition in the real price of oil during the outbreak of the Persian Gulf War in August 1990. According to [Antolín-Díaz and Rubio-Ramírez \(2018\)](#), economists agree that AD shocks were not responsible for the increase in oil prices. Therefore, the last restriction imposes that the absolute value of the contribution of ε_t^{ad} for the movement of rpo_t in August 1990 must be smaller than the absolute contribution made by any other shock in the model.

In Table 3.4, the posterior probabilities are tabulated for each statistical shock. First, consider the restrictions on the sign of the shock necessary to qualify as supply component. A good candidate seems to be ε_1 , given that this shock satisfies all but one restrictions with 100% probability. With a probability of just 11%, the problematic date is 2003M3 which is why the joint probability (baseline) to satisfy NS-1 is also very low. Therefore, I also consider an alternative scenario where the supply shock becomes negative just in 2003M4. This might not be totally unreasonable given that the outbreak of the war was on March 20. Once the restriction is moved from March to April, the probability raises to 1 and ε_1 satisfies all restrictions jointly with 100% probability. Therefore, ε_1 may be interpreted as an oil supply shock. This analysis is complemented by comparing the estimated ε_1 with two prominent supply shocks in the literature: the exogenous oil supply shock of [Kilian \(2008b\)](#) and the supply shock of [Baumeister and Hamilton \(forthcoming\)](#).²⁰ First of all, a visual comparison is made in Figure 3.4, which demonstrates clear similarities between both shocks and ε_{1t} . Computing correlations with the posterior median estimates of the statistical shocks yields $\hat{\rho}_{\varepsilon^{Kilian08}, \varepsilon} = [0.296, -0.026, 0.069]$ and $\hat{\rho}_{\varepsilon^{BH18}, \varepsilon} = [0.879, -0.096, -0.273]$, which confirms

¹⁸HDs are simply obtained from truncating the MA(∞) representation $y_t = \sum_{\tau=0}^{t-1} \Theta_{\tau} \varepsilon_{t-\tau} + \sum_{\tau=t}^{\infty} \varepsilon_{t-\tau} \approx \sum_{\tau=0}^{t-1} \Theta_{\tau} \varepsilon_{t-\tau}$, considering only the shocks observed in-sample $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T$. See also [Kilian and Lütkepohl \(2017\)](#).

¹⁹Given that the sample starts in 1985, no narrative restrictions suggested by [Antolín-Díaz and Rubio-Ramírez \(2018\)](#) prior to these dates are taken into account.

²⁰The comparison is based on a monthly version of [Kilian \(2008b\)](#) as recomputed in [Braun and Brüggemann \(2017\)](#). The oil supply shock of [Baumeister and Hamilton \(forthcoming\)](#) was downloaded on 13/03/2019 from Christiane Baumeister's homepage <https://sites.google.com/site/cjsbaumeister/research>.

that the first shock strongly qualifies as a supply shock.²¹ Furthermore, note that neither the supply shock of Kilian (2008b) nor that of Baumeister and Hamilton (forthcoming) are negative in March 2003, which indicates that violating NR1 at this particular date might not be too problematic. Now, consider the second set of narrative restrictions which are also related to the supply shock. With 73% probability, ε_1 satisfies these restrictions which underlines the finding that ε_1 can be labeled as supply shock. Finally, consider the last narrative restriction, imposing that the AD shock is the least important driver of the oil price surge in August 1990. With 66% probability, ε_1 satisfies this restriction followed by ε_2 with 25% and ε_3 with 9% probability. Therefore, there is too much uncertainty in the historical decomposition at this particular date, and one cannot convincingly use this restriction for labeling.

To sum up the main results of the labeling exercise, there is considerable evidence that ε_1 qualifies as an oil supply shock and ε_2 as an aggregate demand shock. For ε_3 , the labeling exercise is based on sign- and narrative restrictions is not too informative. However, it might not be totally unreasonable to label the last component as the oil market specific demand shock given that with 40% probability it satisfies the corresponding sign restrictions.

²¹Here, $\hat{\rho}_{z,\varepsilon} = T^{-1} \sum_{t=1}^T z_t \varepsilon_t'$.

Table 3.4: Posterior probabilities to satisfy narrative restrictions of [Antolín-Díaz and Rubio-Ramírez \(2018\)](#).

	restrictions	events (t)					jointly	
		1990M8	2002M12	2003M3	2003M4	2011M2	baseline	alternative
NR-1	$p(\varepsilon_{1t} < 0 Y)$	1	1	0.11	1	1	0.11	1
	$p(\varepsilon_{2t} < 0 Y)$	0.49	0.89	0.50	0.42	0.99	0.17	0.15
	$p(\varepsilon_{3t} < 0 Y)$	0	0	1	1	0.09	0	0
NR-2	$p\left(\text{HD}_{1t}^{(1)} > \text{HD}_{1t}^{(s)} , s \neq 1 Y\right)$	1	1	0.8	-	0.89	0.73	-
	$p\left(\text{HD}_{1t}^{(1)} > \text{HD}_{1t}^{(s)} , s \neq 2 Y\right)$	0	0	0.06	-	0.07	0	-
	$p\left(\text{HD}_{1t}^{(1)} > \text{HD}_{1t}^{(s)} , s \neq 3 Y\right)$	0	0	0.13	-	0.04	0	-
NR-3	$p\left(\text{HD}_{3t}^{(1)} < \text{HD}_{3t}^{(s)} , s \neq 1 Y\right)$	0.66	-	-	-	-	-	-
	$p\left(\text{HD}_{3t}^{(2)} < \text{HD}_{3t}^{(s)} , s \neq 2 Y\right)$	0.25	-	-	-	-	-	-
	$p\left(\text{HD}_{3t}^{(3)} < \text{HD}_{3t}^{(s)} , s \neq 3 Y\right)$	0.09	-	-	-	-	-	-

Posterior probabilities $p(\cdot|Y)$ to satisfy narrative sign restrictions are tabulated for each structural shock ε_{it} , $i = 1, 2, 3$. When evaluated jointly, the “baseline” set of dates refers to 1990M8, 2002M12, 2003M3 and 2011M2, while “alternative” is based on 1990M8, 2002M12, 2003M4 and 2011M2.

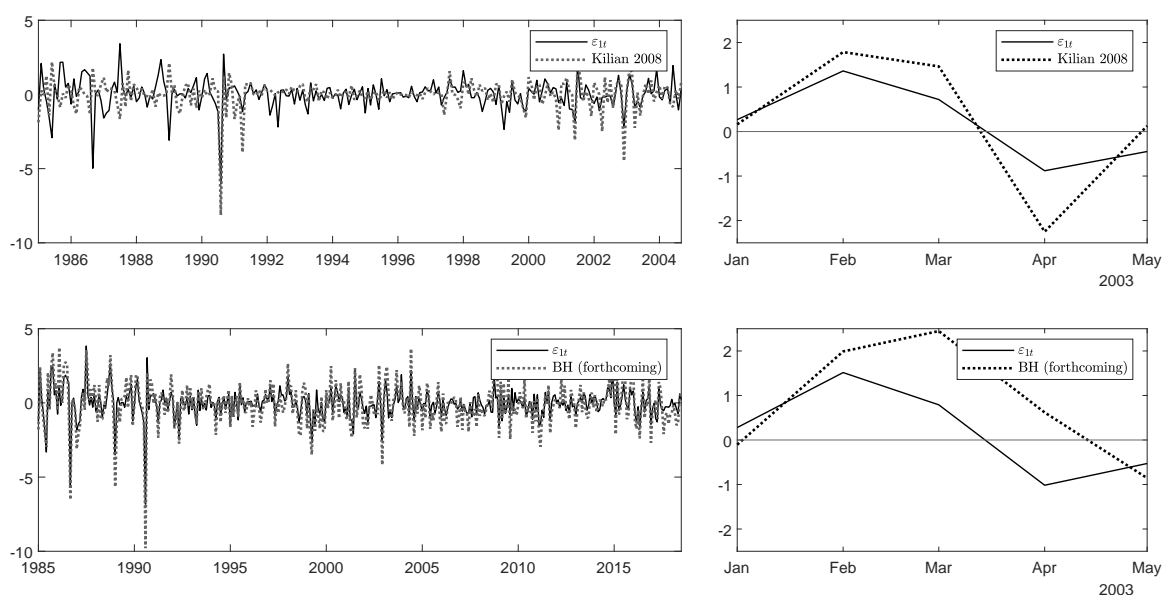


Figure 3.4: Comparison of the first statistical shock with oil supply from the literature. The posterior median estimate of ε_{1t} is compared to the shock of Kilian (2008b) (top) and the shock of Baumeister and Hamilton (forthcoming) (bottom).

3.3.3 Incorporating prior knowledge about oil price elasticities

In the following, I will demonstrate how the statistical identification approach can be improved by including prior information on oil price elasticities. There are two reasons why this can yield more reasonable results. First of all, it can help with the labeling, since the incorporation of identifying information on oil price elasticities goes along with attaching economic meaning to the shocks. This is particularly valuable in the light of the large uncertainty associated with the label of the third statistical shock (ε_3). Second, the identification can be sharpened by ruling out implausible large oil price elasticities from the posterior distribution of the structural parameters, which ultimately increases the economic plausibility of the results.

In the following, assume that ε_1 is a supply shock, and that ε_2 and ε_3 are both demand shocks. Then, three short run oil price elasticities are implicitly given as ratios of elements in B . First, $\eta_1 = \frac{\Delta\% \text{ quantity}}{\Delta\% \text{ price}} = \frac{B_{12}}{B_{32}}$ and $\eta_2 = \frac{\Delta\% \text{ quantity}}{\Delta\% \text{ price}} = \frac{B_{13}}{B_{33}}$ can be interpreted as oil supply elasticities. Both measure the change in production responding to an oil price increase, originally caused by a shift in the demand curve. Second, $\eta_3 = \frac{\Delta\% \text{ quantity}}{\Delta\% \text{ price}} = \frac{B_{11}}{B_{31}}$ can be interpreted as an oil demand elasticity. It measures the same ratio, however, triggered by a shift in the supply curve. For the non-Gaussian SVAR identified solely by statistical means, identification is not sharp enough to rule out elasticities that are considered implausible from an economic point of view. In particular, 5%, 50% and 95% quantiles of the posterior distribution of η_1 are given by $[-0.21, 0.05, 0.37]$, of η_2 by $[-0.02, 0.01, 0.03]$ and of η_3

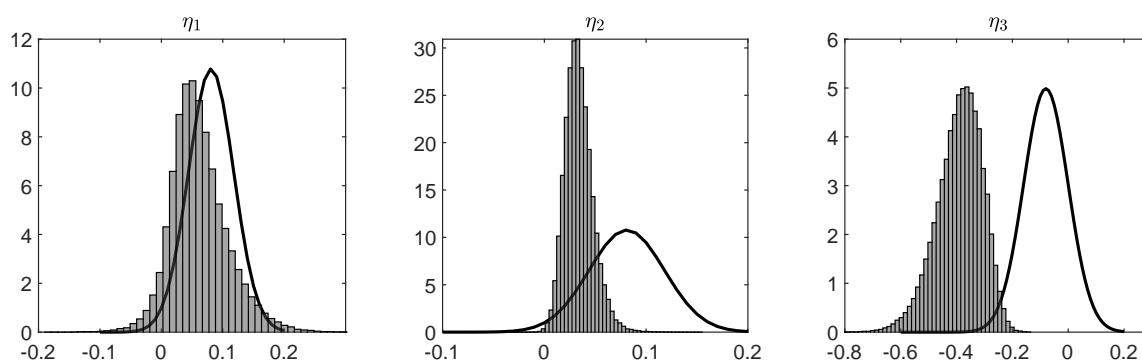


Figure 3.5: Oil supply elasticities: prior distribution (solid line) vs posterior distribution (histogram).

by $[-7.43, -0.89, 6.90]$. While the median estimates are sensible, the credible intervals of η_1 and η_3 are extremely wide and include unreasonably large values in both sign and magnitude. Given the wide availability of microeconomic estimates for these elasticities, it can be desirable to incorporate the knowledge into the non-Gaussian SVAR and achieve sharper identification. This strategy is considered in the following.

First, consider the supply elasticities (η_1 and η_2). While exact magnitudes remain controversial for the oil market, there is a certain consensus that the supply side is rather inelastic. In particular, [Kilian and Murphy \(2012\)](#) impose the dogmatic belief that $\eta_{1/2} \in (0, 0.0258)$, while [Baumeister and Hamilton \(forthcoming\)](#) impose a (positively) truncated t -prior with mode at 0.1, scale 0.2 and 3 degrees of freedom. Somewhere in between, [Caldara et al. \(forthcoming\)](#) exploit country level IV estimates to come up with a range of estimates depending on the construction of their instrument. In their preferred specification, they arrive at an estimate of $\hat{\eta}_{1/2} = 0.08(0.037)$ for the supply elasticity and $\hat{\eta}_3 = -0.08(0.079)$ for the demand elasticity, where the values in brackets give the estimated standard errors.²² In the following, I will continue incorporating the values of [Caldara et al. \(forthcoming\)](#) as prior information into the model. Note that in contrast to a Gaussian model, the structural parameters are identified and therefore, the data has the possibility to substantially revise this prior information. With respect to the exact form, normal distributions are used, that is the priors are $\eta_1 \sim N(0.08, 0.037^2)$, $\eta_2 \sim N(0.08, 0.037^2)$, and $\eta_3 \sim N(-0.08, 0.079^2)$ respectively.²³

In Figure 3.5, a comparison of prior and posterior distributions for the oil price elasticities is provided. First, note that all prior distributions are revised by the identifying information

²²[Kilian and Murphy \(2014\)](#) argue that in a model without oil inventories, η_3 cannot be interpreted properly as the demand elasticity. However, [Caldara et al. \(forthcoming\)](#) find that changes in inventories are of secondary importance and including them does not change their findings.

²³One can handle this prior in the PX model in a standard manner. To see this, note that e.g. for η_1 one has $\eta_1 = \frac{B_{12}}{B_{32}} = \frac{B_{12}^*}{B_{32}^*}$ since the scale cancels out. Therefore, priors on elasticities can be handled as priors on B^* .

provided by non-Gaussianity, although to a different extent. With respect to the supply elasticities, both posterior means are below values implied by the prior distribution, revising the parameters towards a model with a steeper supply schedule. However, for η_1 , prior and posterior are quite similar indicating that there is little objection by the data against the elasticities considered by the prior. In turn, the posterior of η_2 is more narrow and concentrated at substantially lower values than implied by the prior, indicating a substantial revision by the data. Finally, the prior of η_3 is strongly revised into the direction of a more elastic demand schedule. Overall, the posterior regions for oil price elasticities are more reasonable than for the model using only non-Gaussianity for identification. To see this, recall the 5%, 50% and 95% quantiles of the posterior distribution in the purely statistical identification: $[-0.21, 0.05, 0.37]$ for η_1 , $[-0.02, 0.01, 0.03]$ for η_2 and $[-7.43, -0.89, 6.90]$ for η_3 . In contrast, for the combined identification approach, these values are given by $[0, 0.057, 0.15]$ for η_1 , $[0.01, 0.03, 0.06]$ for η_2 and $[-0.5, -0.38, -0.27]$ for η_3 .

The second benefit of combining statistical with economic identification are more informative results for the labeling exercise. In Table 3.5, I provide posterior probabilities to satisfy the sign restrictions of Kilian and Murphy (2012), recomputed for the combined identification approach. Note that incorporating prior on oil price elasticities goes along with a labeling as either supply or demand shock. In the numbers, this translates into zero probability that ε_{1t} satisfies the restrictions of a demand shock, and zero probabilities that ε_{2t} or ε_{3t} satisfy the restrictions of supply shocks. Interestingly, one can also use the posterior probabilities to disentangle ε_{2t} and ε_{3t} into AD and OD shocks, despite the fact that the prior is not informative about the exact nature of the demand shocks. In particular, ε_{2t} clearly qualifies as an aggregate demand shock given that it satisfies the respective restrictions with 95% probability. Furthermore, ε_{3t} can be labeled as OD shock given that the probability of satisfying the respective restrictions went up to 89%. Overall, the probability to jointly satisfy all sign restrictions increased substantially from 14% to 57%, which underlines the benefits of the combined identification approach.

3.3.4 Structural analysis

In the remainder of this section, I will discuss the main implications of the SVAR model in terms of impulse response functions, historical decompositions and forecast error variance decompositions. Throughout the analysis, I will constantly compare the results from a purely statistical identification approach with those obtained from a combined identification.

In Figure 3.6, I provide impulse responses up to 6 years to structural shocks of size one standard deviation. A negative supply shock (ε_1) is associated with a persistent decrease in the level of oil production, a decline in global activity and an increase in the oil price. The magnitudes of these effects depend substantially on the inclusion of the prior knowledge on

Table 3.5: Posterior probabilities to satisfy sign restrictions of Kilian and Murphy (2012), combined identification approach.

	jointly						shock by shock		
	permutation						$p(\varepsilon_{1t} = \cdot Y)$	$p(\varepsilon_{2t} = \cdot Y)$	$p(\varepsilon_{3t} = \cdot Y)$
ε_t^s	3	3	2	2	1	1	0.65	0	0
ε_t^{ad}	2	1	3	1	3	2	0	0.95	0.11
ε_t^{od}	1	2	1	3	2	3	0	0.03	0.89
$p(\text{SR} Y)$	0	0	0	0	0.02	0.57			

Left: posterior probabilities $p(\text{SR}|Y)$ that a certain permutation of shocks yields a system which satisfies the sign restrictions of Kilian and Murphy (2012). Right: posterior probabilities $p(\varepsilon_{it} = \cdot | Y)$ that the i th shock qualifies as supply shock (ε_t^s), aggregate demand shock (ε_t^{ad}), or oil market specific shock (ε_t^{od}).

oil price elasticities. Specifically, the response of the oil price is much more pronounced and 90% HPD credible intervals are only positive for more than two years in the combined identification approach. With respect to the AD shock (ε_2), which is normalized to increase global economic activity, oil prices are estimated to increase persistently with 90% HPD credible intervals positive at all horizons. Furthermore, global oil production is estimated to increase significantly for a couple of months, though there is more uncertainty than for the other IRFs. For the effects of ε_2 , there seems to be no additional information from the prior on oil price elasticities, as the estimates largely coincide for both identification approaches. In turn, for the last shock there are substantial differences. A positive OD shock (ε_3), normalized to increase the real oil price, is estimated to increase crude oil production substantially for about two years, but only if the combined identification approach is considered. The response of economic activity coincides for both approaches, and is estimated to be negative with a lag of about 2 years. From an economic point of view, the directions of the IRFs are reasonable and in line with the related literature. The exact magnitudes, however, depend substantially on the inclusion of the prior information. In particular, the impulse responses of ε_1 and ε_3 are affected by the inclusion of prior information. This is totally sensible given that the additional prior distribution is particularly informative for the posterior of η_1 and η_3 , which effectively are ratios of impact impulse responses for ε_3 and ε_1 (see Section 3.3.3).

In Figure 3.7, historical decompositions are plotted to analyze important historical oil price fluctuations through the lens of the structural model. Each row of the graph shows the unexpected movement in rpo_t (dashed lines) decomposed into the contributions made by each shock. The first row corresponds to the unexpected rise of the oil price during the outbreak of the gulf war in August 1990. Based on statistical identification only, almost the entire increase in the oil price is attributed to precautionary demand shocks (ε_3), a finding

that is broadly in line with [Kilian and Murphy \(2012\)](#). Interestingly, once prior information is incorporated the picture changes somewhat and more weight is put on the supply shock (ε_1), in particular during the first months. A significant effect of oil supply shocks is also found by [Baumeister and Hamilton \(forthcoming\)](#) and [Caldara et al. \(forthcoming\)](#). The next episode under consideration is the surge of oil prices in 2007–2008. Most of it can be explained by AD shocks (ε_2), while precautionary demand (ε_3) and oil supply shocks (ε_1) explain less of the fluctuation. The exact extent depends again on the identification approach, with slightly more weight on supply shocks if the prior information on elasticities is included. The results are in line with other studies analyzing the causes of the 2007–2008 oil price surge ([Hamilton; 2009](#); [Kilian and Murphy; 2012](#)). The next event considered is the drop of the oil price during the global financial crisis. Here, a combination of AD (ε_2) and oil market specific demand components (ε_3) are found to be the main driver, and both identification approaches largely coincide in this finding. Finally, the oil price decrease in 2014–2015 seems to be driven by a combination of supply (ε_1) and oil market specific demand components (ε_2). Again, in the combined identification approach, more weight is put on the supply side than in a model identified by solely statistical means.

Finally, [Figure 3.8](#) reports the results obtained from forecast error variance decompositions (FEVD).²⁴ Posterior median estimates of the FEVDs are plotted together with 90% HPD credible intervals for horizons up to 6 years. With respect to global crude oil production, the most important driver are found to be supply shocks (ε_1). Depending on the inclusion of additional prior information, median estimates are 75% or 85% for the contribution of supply shocks to the long term variability (6 years) of oil production. On the short run, the contributions of ε_1 are found to be even higher, reflecting the steep supply curve estimated in a non-Gaussian setting. The additional prior information has some effect though, shifting weight from the supply to the OD shock (ε_3). Real economic activity is found to be mostly driven by AD shocks (ε_2). In terms of FEVDs, these are estimated to contribute between 70% and 95% percent depending on the horizon. Oil supply shocks (ε_1) account for around 10% of the long term variation in global activity, while OD shocks (ε_3) account for a maximum of 20%, depending on the exact horizon. Similar to what has been found in the impulse response analysis, there is very little additional information from the prior distribution on elasticities.

The FEVDs of oil price fluctuations uncovers very interesting effects of the additional prior information. In a purely statistical identification approach, most of the contributions are made by the oil market specific demand shock (ε_3) (70%–90%), followed by the AD shock (ε_2) (5%–30%) and very little is attached to supply shocks (ε_1) (2–3%). Once the IV estimates of [Caldara et al. \(forthcoming\)](#) are incorporated through the prior distribution, the

²⁴See [Kilian and Lütkepohl \(2017\)](#) for a textbook treatment of FEVDs.

picture changes. The share of variability accounted for by supply shocks (ε_1) is substantially larger, contributing between 12% and 20% depending on the horizon. At the same time, less weight is given to the oil market specific shock (ε_3). To understand the role of the prior for the magnitude of these estimates, it is necessary to compare them with values implied by a model that does not rely on non-Gaussianity. Therefore, I re-estimated the SVAR using Gaussian errors and identified the structural parameters by the Minimum Distance approach used by [Caldara et al. \(forthcoming\)](#).²⁵ Using their identification approach, I obtain posterior median estimates of 60% on impact and 40% after 6 years, in terms of the contribution of the supply shock (ε_1) to the forecast error variance of the oil price. These numbers highlight the amount of information available in non-Gaussianity. The prior for elasticities is revised substantially towards lower contributions of supply shocks for fluctuations in oil prices, reflecting more inelastic supply and elastic demand curves.

Given that including prior information on oil price elasticities has substantial implications for the causes of oil prices, I further conduct a prior sensitivity analysis. The goal is to understand the implications of different prior values for the median estimate of the (long term) contribution of supply shocks for fluctuations in oil prices. In [Table 3.6](#), each row displays different values specified for mean and variance for the normal priors of the oil price elasticities ($\mu_{-\eta_i}, \sigma_{\eta_i}, i = 1, 2, 3$), the corresponding posterior moments ($\bar{\mu}_{\eta_i}, \bar{\sigma}_{\eta_i}, i = 1, 2, 3$) as well as the estimates for the contribution of ε_1 for fluctuations in rpo_t .

The first row corresponds to the estimates obtained by a merely statistical identification approach, arriving at a point estimate of 3%. However, unreasonably large oil price elasticities are part of the posterior distribution. Very extreme values can be ruled out by using a weakly informative prior distribution, centered at zero with variance 2 (second row). The median estimates are barely affected, and supply shocks still account for only 3% of the variation in oil prices. However, particularly the variance of the demand elasticity remains high and the posterior distribution still includes unreasonably large values from an economic point of view.

The 3rd and 4th row correspond to prior specifications which aim to mimic the very low values of $\eta_{1/2}$ used by [Kilian and Murphy \(2012\)](#) to identify the model (KM). Two versions are considered, the first is based on a prior mean of 0.02 with a standard deviation of 0.05, while the second is based on twice the dispersion (0.1). Both yield very similar results, which are negligible contribution of supply shocks for oil price fluctuations. However, also for this specification, unreasonably large values for the demand elasticity are part of the posterior distribution.

²⁵Specifically, they identify the model minimizing a weighted squared distance between model-implied elasticities and the corresponding IV estimates. The weighting matrix is set as diagonal with its elements equal to the IV standard errors.

In the last four specifications I consider different priors based on IV estimates calculated in [Caldara et al. \(forthcoming\)](#) (CCI). The first two specifications are based on the values obtained by using their “narrow” instrument, yielding estimates of $\hat{\eta}_{1/2} = 0.056(0.019)$ and $\hat{\eta}_3 = -0.031(0.037)$, where the values in brackets indicate the standard errors. Both specification are based on setting the prior mean equal to the IV estimates, but differ in the prior dispersion considered. In the first, I consider a standard deviation equal to the IV standard errors ($\underline{\sigma}_{\eta_{1/2}} = 0.019, \underline{\sigma}_{\eta_3} = 0.037$) while for the second, I consider twice the dispersion ($\underline{\sigma}_{\eta_{1/2}} = 0.038, \underline{\sigma}_{\eta_3} = 0.074$). Finally, the last two specifications are based on the estimates obtained by their “broad” instrument, which are $\hat{\eta}_{1/2} = 0.081(0.037)$ and $\hat{\eta}_3 = -0.08(0.079)$. The two specification considered in the sensitivity analysis differ only in the prior dispersion. Again, one specification is based on setting the prior standard deviation equal to the IV standard errors, and therefore corresponds to the values considered in the empirical analysis. The second is based on setting the standard deviation to twice these values. With respect to the posteriors of elasticities, note that all specifications based on CCI yield plausible values, which is sensible given that they are highly informative about both supply and demand elasticities. Furthermore, the estimates of the demand elasticity are considerable revised toward lower values, implying a more elastic demand schedule. Regarding the estimated contribution of supply shock to the forecast error variance of oil prices, the values based on the first specification yield the largest estimate (21%). Once the prior dispersion is doubled, this estimate goes down to 13%. For the broad instrument, which was already considered in the previous empirical analysis, the estimate is 11%. In turn, if the prior dispersion is doubled, only 3% of the fluctuations can be contributed to the oil supply shock at a 6 year horizon. Overall, the sensitivity analysis underlines the main finding that oil supply shocks are playing a minor role for oil price fluctuations. With the exception of one very tight prior specification, the broad range of estimates lies between 1% and 13%.

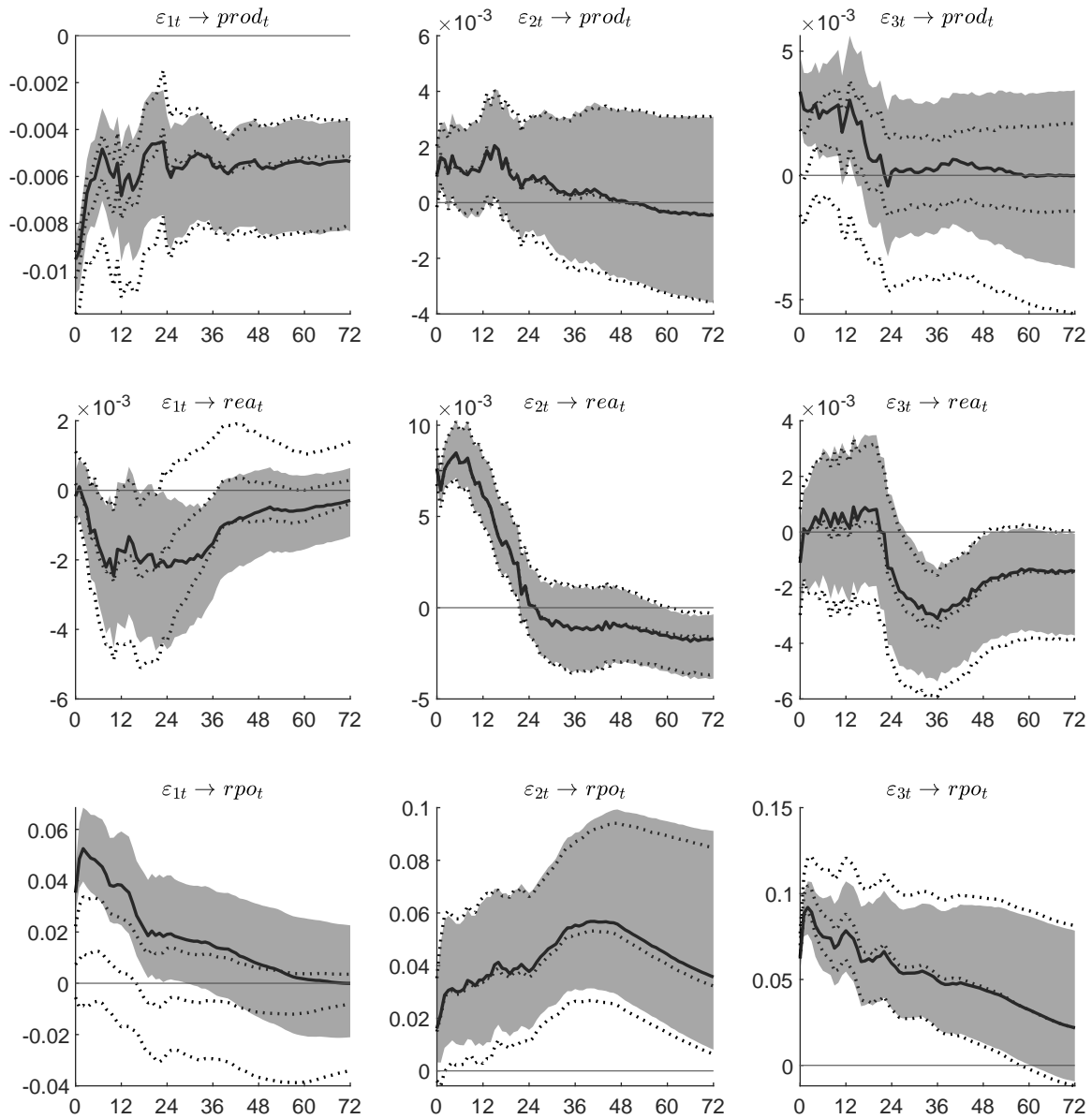


Figure 3.6: Posterior median estimates of IRFs together with 90% HPD credible intervals. IRFs based on a combined identification approach are indicated by a solid line with shaded areas. IRFs identified by solely non-Gaussianity are indicated by dotted lines. ε_1 is labeled as a supply shock, ε_2 as an aggregate demand shock and ε_3 as an oil market specific demand shock.

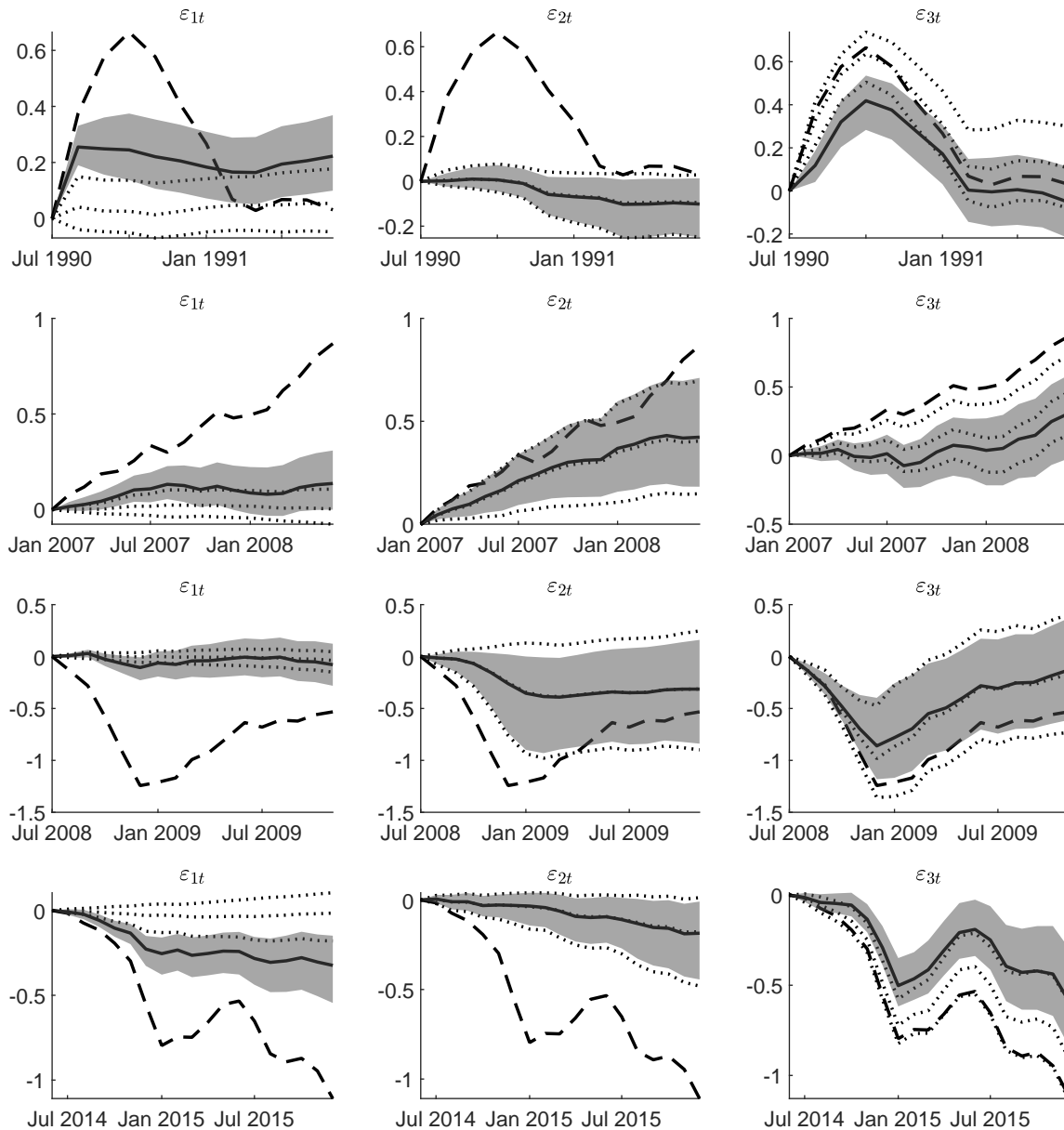


Figure 3.7: Posterior median estimates of historical decompositions of the (log) real oil price at key events together with 90% HPD credible intervals. The dashed line indicates the unexpected movement of rpo_t . HDs based on identification by non-Gaussianity and elasticity priors are given by the solid line with shaded areas. HDs identified by solely non-Gaussianity are indicated by dotted lines. ε_1 is labeled as a supply shock, ε_2 as an aggregate demand shock and ε_3 an oil market specific demand shock.

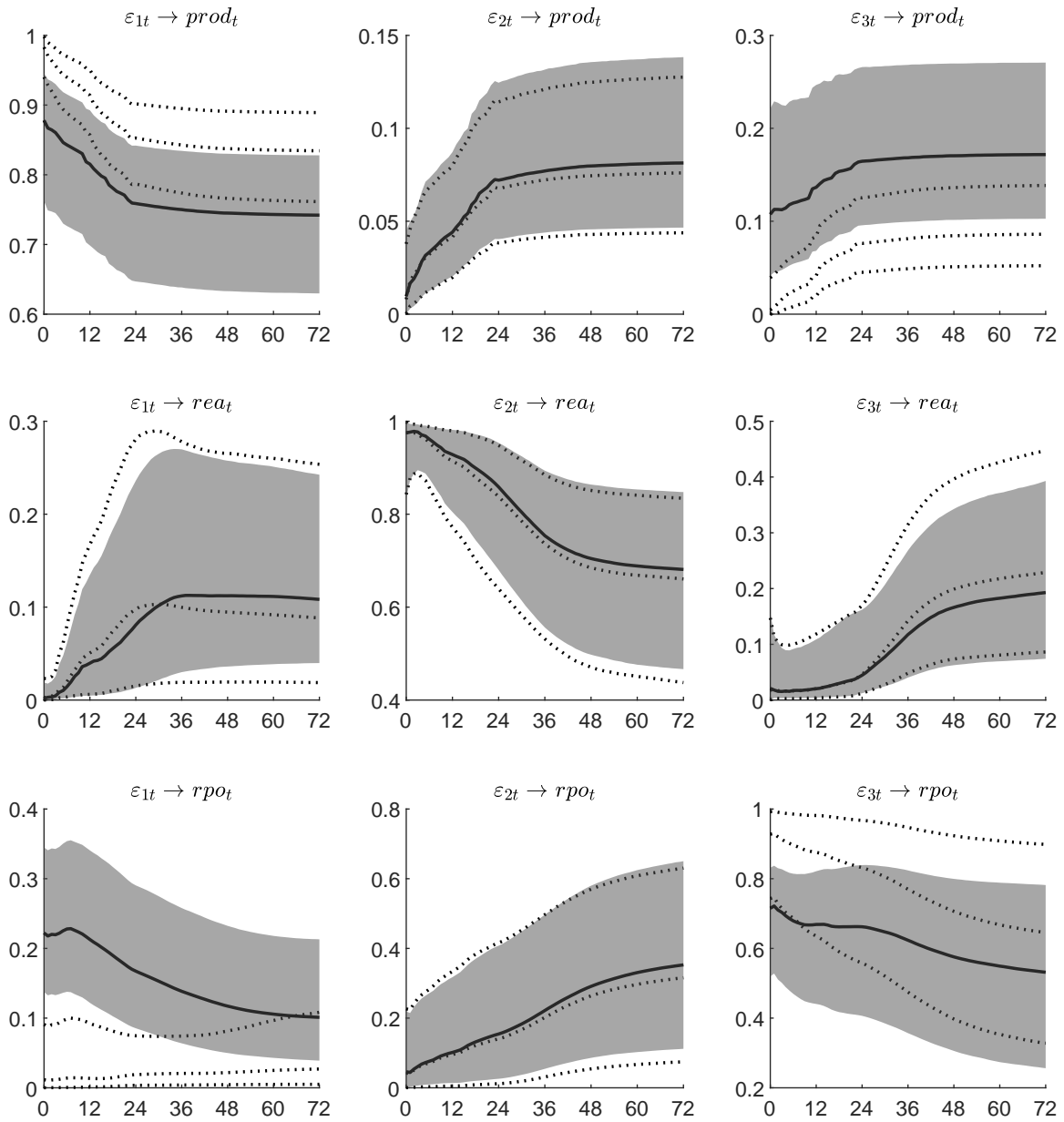


Figure 3.8: Posterior median estimates of forecast error variance decompositions together with 90% HPD credible intervals. FEVDs based on a combined identification approach are indicated by a solid line with shaded areas. FEVDs identified by solely non-Gaussianity are indicated by dotted lines. ϵ_1 is labeled as a supply shock, ϵ_2 as an aggregate demand shock and ϵ_3 an oil market specific demand shock.

Table 3.6: Prior sensitivity analysis.

	$\underline{\mu}_{\eta_1}$	$\underline{\sigma}_{\eta_1}$	$\underline{\mu}_{\eta_2}$	$\underline{\sigma}_{\eta_2}$	$\underline{\mu}_{\eta_3}$	$\underline{\sigma}_{\eta_3}$	$\bar{\mu}_{\eta_1}$	$\bar{\sigma}_{\eta_1}$	$\bar{\mu}_{\eta_2}$	$\bar{\sigma}_{\eta_2}$	$\bar{\mu}_{\eta_3}$	$\bar{\sigma}_{\eta_3}$	FEVD $_{\varepsilon_{1t} \rightarrow \text{rpo}_t}$
	-	-	-	-	-	-	0.02	3.67	0.00	0.01	-1.73	69.97	0.03
	0	2	0	2	0	2	(0.05)	(0.08)	(0.00)	(0.02)	(-0.98)	(1.37)	
	0.02	0.05	-	0.02	0.05	-	0.04	0.34	0.01	0.02	-0.19	1.51	0.03
KM	0.02	0.05	-	0.02	0.05	-	0.04	0.03	0.00	0.01	-1.07	72.30	0.01
	(0.04)	(0.04)	(0)	(0.02)	(-0.91)	(1.26)	0.05	0.05	0.00	0.01	-2.38	128.00	0.02
CCI	(0.04)	(0.05)	(0)	(0.02)	(-0.99)	(1.38)	0.05	0.02	0.08	0.01	-0.19	0.02	0.21
	0.05	0.02	0.05	0.02	-0.03	0.04	0.05	0.03	0.06	0.02	-0.26	0.04	0.13
	0.05	0.04	0.05	0.04	-0.03	0.07	0.05	0.03	0.06	0.02	-0.27	0.04	0.11
	0.08	0.04	0.08	0.04	-0.08	0.08	0.07	0.03	0.06	0.02	-0.27	0.04	0.11
	0.08	0.07	0.08	0.07	-0.08	0.16	0.06	0.05	0.03	0.01	-0.38	0.08	0.03

Prior sensitivity analysis for the share of the long term (6 years) variance of the real oil price, accounted for by supply shocks (FEVD $_{\varepsilon_{1t} \rightarrow \text{rpo}_t}$). $\eta_1 = B_{12}/B_{32}$ and $\eta_2 = B_{13}/B_{33}$ denote the oil price elasticities, while $\eta_3 = B_{11}/B_{31}$ denote the demand elasticity. All priors are based on normal distributions. Prior means and variances are denoted as $\underline{\mu}_{(\cdot)}$ and $\underline{\sigma}_{(\cdot)}$, for posterior the notation $\bar{\mu}_{(\cdot)}$ and $\bar{\sigma}_{(\cdot)}$ is used. When additional values are provided in brackets, they correspond to median and interquartile range respectively.

3.4 Conclusion

In this paper, new evidence is obtained on the relative importance of supply and demand components for oil price fluctuations. To identify a SVAR model of the global crude oil market, non-Gaussianity is exploited in combination with prior distributions on oil price elasticities. The latter can substantially sharpen inference by ruling out unreasonably large values a priori. Furthermore, it helps to achieve an economically meaningful label of the structural shocks. Given that the additional information is overidentifying in a non-Gaussian setting, the data has a chance to speak against the prior. Therefore, incorporating this information has less drastic consequences for identification than in a Gaussian setting. The empirical results indicate that in terms of the contribution to the forecast error variance of oil prices, supply shocks have been minor drivers post 1985. Depending on the exact specification of the prior, most of the point estimates are located between 1% and 13%. Furthermore, even tightly specified priors are updated substantially in direction of a more inelastic supply and elastic demand curve, underlining the minor role of supply shocks for oil prices. Consequently, the empirical findings are broadly in line with papers that impose very low oil supply elasticity for identification purposes ([Kilian and Murphy; 2012, 2014](#); [Antolín-Díaz and Rubio-Ramírez; 2018](#)).

This paper also proposes a new Bayesian SVAR model, which gives a coherent tool to combine identification by non-Gaussianity with economic identification. Structural shocks are assumed to be independent and follow unknown distributions, normalized to have zero mean and unit variance. The distributions are modeled using constrained Dirichlet process mixture models. The resulting SVAR is capable to simultaneously identify the structural parameters with the unknown source densities, and requires no stance about the form of non-Gaussianity in the data prior estimation. Assessing identification in a dataset is straightforward and can be done by simply comparing posterior predictive densities with those implied by standard normal distributions. Besides the possibility to incorporate external identifying information, another benefit of the Bayesian framework is the possibility to assess uncertainty with respect to the economic labeling of the shocks, as well as the possibility to use prior information to regularize VAR reduced form parameters.

References

- Antolín-Díaz, J. and Rubio-Ramírez, J. F. (2018). Narrative sign restrictions for SVARs, *American Economic Review* **108**(10): 2802–29.
- Baumeister, C. J. and Hamilton, J. D. (forthcoming). Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks, *American Economic Review* .
- Baumeister, C. and Peersman, G. (2013). The role of time-varying price elasticities in accounting for volatility changes in the crude oil market, *Journal of Applied Econometrics* **28**(7): 1087–1109.
- Bernanke, B. S. (1986). Alternative explanations of the money-income correlation, *Carnegie-Rochester Conference Series on Public Policy*, Vol. 25, Elsevier, pp. 49–99.
- Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via Pólya urn schemes, *The Annals of Statistics* **1**(2): 353–355.
- Blanchard, O. J. and Galí, J. (2010). *The Macroeconomic Effects of Oil Price Shocks: Why are the 2000s so different from the 1970s?*, University of Chicago Press.
- Blanchard, O. J. and Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances, *The American Economic Review* **79**(4): 655–673.
- Boscolo, R., Pan, H. and Roychowdhury, V. P. (2004). Independent component analysis based on nonparametric density estimation, *IEEE Transactions on Neural Networks* **15**(1): 55–65.
- Braun, R. and Brüggemann, R. (2017). Identification of SVAR models by combining sign restrictions with external instruments, *Working Paper 16*, Department of Economics, University of Konstanz.
- Caldara, D., Cavallo, M. and Iacoviello, M. (forthcoming). Oil price elasticities and oil price fluctuations, *Journal of Monetary Economics* .
- Chan, J. C. C. (2017). The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling, *Journal of Business & Economic Statistics* **35**(1): 17–28.
- Chib, S. and Greenberg, E. (1995). Understanding the Metropolis–Hastings algorithm, *The American Statistician* **49**(4): 327–335.

- Chib, S. and Jeliazkov, I. (2005). Accept–reject Metropolis–Hastings sampling and marginal likelihood estimation, *Statistica Neerlandica* **59**(1): 30–44.
- Doan, T., Litterman, R. and Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions, *Econometric Reviews* **3**(1): 1–100.
- Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures, *Journal of the American Statistical Association* **90**(430): 577–588.
- Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems, *The Annals of Statistics* **1**(2): 209–230.
- Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*, Springer Science & Business Media.
- Gourieroux, C., Monfort, A. and Renne, J. P. (2017). Statistical inference for independent component analysis: Application to structural VAR models, *Journal of Econometrics* **196**(1): 111–126.
- Hamilton, J. D. (2009). Causes and consequences of the oil shock of 2007-08, *Brookings Papers on Economic Activity* **40**(1): 215–283.
- Hamilton, J. D. (2018). Measuring global economic activity, *Working Paper*, University of California at San Diego.
- Herwartz, H. (2018). Hodges–Lehmann detection of structural shocks: An analysis of macroeconomic dynamics in the euro area, *Oxford Bulletin of Economics and Statistics* .
- Herwartz, H. and Plödt, M. (2016). The macroeconomic effects of oil price shocks: Evidence from a statistical identification approach, *Journal of International Money and Finance* **61**: 30–44.
- Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors, *Journal of the American Statistical Association* **96**(453): 161–173.
- Jasra, A., Holmes, C. C. and Stephens, D. A. (2005). Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling, *Statistical Science* pp. 50–67.
- Kilian, L. (2008a). The economic effects of energy price shocks, *Journal of Economic Literature* **46**(4): 871–909.
- Kilian, L. (2008b). Exogenous oil supply shocks: How big are they and how much do they matter for the U.S. economy?, *The Review of Economics and Statistics* **90**(2): 216–240.

- Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market, *American Economic Review* **99**(3): 1053–1069.
- Kilian, L. and Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*, Themes in Modern Econometrics, Cambridge University Press.
- Kilian, L. and Murphy, D. P. (2012). Why agnostic sign restrictions are not enough: Understanding the dynamics of oil market VAR models, *Journal of the European Economic Association* **10**(5): 1166–1188.
- Kilian, L. and Murphy, D. P. (2014). The role of inventories and speculative trading in the global market for crude oil, *Journal of Applied Econometrics* **29**(3): 454–478.
- Koop, G. and Korobilis, D. (2010). Bayesian multivariate time series methods for empirical macroeconomics, *Foundations and Trends in Econometrics* **3**(4): 267–358.
- Lanne, M. and Luoto, J. (2016). Data-driven inference on sign restrictions in Bayesian structural vector autoregression, *Technical report*, Department of Economics and Business Economics, Aarhus University.
- Lanne, M. and Luoto, J. (2018). GMM estimation of non-Gaussian structural vector autoregression, *Technical report*, HECER, Helsinki Center of Economic Research.
- Lanne, M., Meitz, M. and Saikkonen, P. (2017). Identification and estimation of non-Gaussian structural vector autoregressions, *Journal of Econometrics* **196**(2): 288–304.
- Lütkepohl, H. and Netšunajev, A. (2014). Disentangling demand and supply shocks in the crude oil market: How to check sign restrictions in structural VARs, *Journal of Applied Econometrics* **29**(3): 479–496.
- Maxand, S. (forthcoming). Identification of independent structural shocks in the presence of multiple Gaussian components, *Econometrics and Statistics* .
- Mertens, K. and Ravn, M. O. (2013). The dynamic effects of personal and corporate income tax changes in the United States, *American Economic Review* **103**(4): 1212–1247.
- Rigobon, R. (2003). Identification through heteroskedasticity, *Review of Economics and Statistics* **85**(4): 777–792.
- Sethuraman, J. (1994). A constructive definition of Dirichlet priors, *Statistica Sinica* pp. 639–650.
- Shapiro, M. D. and Watson, M. W. (1988). Sources of business cycle fluctuations, *NBER Macroeconomics Annual* **3**: 111–148.

- Stock, J. H. and Watson, M. W. (2012). Disentangling the channels of the 2007-09 recession, *Brookings Papers on Economic Activity* **43**(1): 81–156.
- Tibshirani, R., Walther, G. and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63**(2): 411–423.
- Tierney, L. (1994). Markov chains for exploring posterior distributions, *The Annals of Statistics* **22**(4): 1701–1728.
- Uhlig, H. (2005). What are the effects of monetary policy on output? Results from an agnostic identification procedure, *Journal of Monetary Economics* **52**: 381–419.
- Woźniak, T. and Droumaguet, M. (2015). Assessing monetary policy models: Bayesian inference for heteroskedastic structural VARs, *Working Paper*, University of Melbourne.
- Yang, M., Dunson, D. B. and Baird, D. (2010). Semiparametric Bayes hierarchical models with mean and variance constraints, *Computational Statistics & Data Analysis* **54**(9): 2172–2186.

Appendix 3.A Bayesian inference

3.A.1 Markov chain Monte Carlo algorithm

In the following, MCMC inference is discussed for a SVAR model with the structural shocks that follow independent constrained Dirichlet Process mixture models (DPMM). Adopting the sampler of [Yang et al. \(2010\)](#), the algorithm is based on a truncated version of the stick breaking process ([Ishwaran and James; 2001](#)). In particular, for a reasonably large N , a close approximation of the DPMMs can be obtained by setting $V_{iN} = 1$. Based on this truncation, the model is summarized once more for $i = 1, \dots, K$, $t = 1, \dots, T$ and $k = 1, \dots, N$:

$$\{\mu_{i1}^*, \sigma_{i1}^{*2}\}, \{\mu_{i2}^*, \sigma_{i2}^{*2}\}, \dots, \{\mu_{iN}^*, \sigma_{iN}^{*2}\} \sim \text{NIG}(a, b, m, s), \quad (3.A.1)$$

$$\alpha_i \sim \mathcal{G}(a_\alpha, b_\alpha), \quad (3.A.2)$$

$$V_{i1}, V_{i2}, \dots, V_{i,N-1} \sim \text{Beta}(1, \alpha_i), \quad (3.A.3)$$

$$\pi_{ik} = V_{ik} \prod_{j=1}^{k-1} (1 - V_{ij}), \quad (3.A.4)$$

$$\mu_i^* = \sum_{j=1}^N \pi_{ij} \mu_{ij}^*, \quad (3.A.5)$$

$$\sigma_i^{*2} = \sum_{j=1}^N \pi_{ij} \left((\mu_{ij}^* - \mu_i^*)^2 + \sigma_{ij}^{*2} \right), \quad (3.A.6)$$

$$\mu_{ik} = (\mu_{ik}^* - \mu_i^*) / \sigma_{ik}^*, \quad (3.A.7)$$

$$\sigma_{ik} = \sigma_{ik}^* / \sigma_i^*, \quad (3.A.8)$$

$$z_{it} \sim \text{Multinomial}(\pi_i), \quad (3.A.9)$$

$$\varepsilon_{it} | z_{it} \sim \mathcal{N}(\varepsilon_{it} | \mu_{i,z_{it}}, \sigma_{i,z_{it}}^2), \quad (3.A.10)$$

$$y_t = \nu + Ax_t + B\varepsilon_t, \quad (3.A.11)$$

where $x_t = [y'_{t-1}, y'_{t-2}, \dots, y'_{t-p}]'$ and $A = [A_1, A_2, \dots, A_p]$. As indicated in Section 3.2, α_i is treated as a random variable endowed with a Gamma prior, such that the complexities of the K density estimators are automatically inferred from the data ([Escobar and West; 1995](#)).

As discussed in Section 3.2, the MCMC algorithm is constructed for a parameter extended (PX) model, obtained by reformulating equations (3.A.10)-(3.A.11) to:

$$\varepsilon_{it}^* | z_{it} \sim \mathcal{N}(\varepsilon_{it}^* | \mu_{i,z_{it}}^*, \sigma_{i,z_{it}}^{*2}), \quad (3.A.12)$$

$$y_t = \Gamma X_t + B^* \varepsilon_t^*, \quad (3.A.13)$$

where $\mu^* = [\mu_1^*, \dots, \mu_K^*]'$, $\Sigma^* = \text{diag}(\sigma_1^{*2}, \dots, \sigma_K^{*2})$, $\Gamma = [v^*, A]$, $X_t = [1, x_t']'$, $v^* = v - B(\Sigma^*)^{-\frac{1}{2}}\mu^*$ and $B^* = B(\Sigma^*)^{-\frac{1}{2}}$. The following prior distributions are considered for the parameters in (3.A.13). For $\gamma = \text{vec}(\Gamma)$, a conditionally conjugate normal prior is convenient, that is $\gamma \sim \mathcal{N}(\underline{\mu}_\gamma, \underline{V}_\gamma)$.²⁶ For the elements in $b^* = \text{vec}(B^*)$, no such priors exists. Therefore, there is no additional computational cost of considering non-standard prior distribution. If no information is available on B , a reasonable choice could be an improper prior $p(b^*) \propto 1$ or a weakly informative prior $b^* \sim \mathcal{N}(\underline{\mu}_{b^*}, \underline{V}_{b^*})$ for very large values in \underline{V}_{b^*} . If prior information on B is available, for example in terms of sign restrictions or prior information about elasticities, they can also be incorporated in $p(b^*)$.

Let the set of PX parameters including latent variables be $\varphi = [v^*, A, B^*, \theta^*, V, Z, \alpha]$, where θ^* is the collection of cluster parameters, V the Beta random variables, Z the assignment indicators and α the set of concentration parameters. Based on arbitrary initial values, the following MCMC algorithm eventually generates draws $\varphi^{(l)}, l = 1, 2, \dots$ from the posterior distribution of $p(\varphi|Y)$ by cycling through the conditionals of each member in φ .

1. In the first block, consider drawing the parameters underlying the K (unconstrained) DPMMs. Following the conditional sampler of [Ishwaran and James \(2001\)](#) gives:

- (a) For $i = 1, \dots, K$ and $k = 1, \dots, N$, draw the cluster moments from their normal inverse gamma posteriors:

$$p\left(\sigma_{ik}^{*2} | Y, \varphi_{-\sigma_{ik}^{*2}}\right) \sim \begin{cases} i\mathcal{G}(a, b) & \text{if } z_{it} \neq k \quad \text{f.a. } t, \\ i\mathcal{G}(\bar{a}_{ik}, \bar{b}_{ik}) & \text{else,} \end{cases}$$

$$p\left(\mu_{ik}^* | Y, \sigma_{ik}^{*2}, \varphi_{-\mu_{ik}^*}\right) \sim \begin{cases} \mathcal{N}(m, \sigma_{ik}^{*2}s) & \text{if } z_{it} \neq k \quad \text{f.a. } t, \\ \mathcal{N}(\bar{m}_{ik}, \bar{V}_{ik}) & \text{else,} \end{cases}$$

²⁶Popular choices for $\underline{\mu}_\gamma$ and \underline{V}_γ are to set a fully non-informative prior ($\underline{\mu}_\gamma = 0_{K(Kp+1)}$ and $\underline{V}_\gamma^{-1} = c_\gamma I_{K(Kp+1)}$ with c_γ very close to zero), a Minnesota type of prior ([Koop and Korobilis; 2010; Doan et al.; 1984](#)) or to use a training sample to learn prior moments as used in Section 3.3.

where the posterior moments are defined by:

$$\begin{aligned}\bar{a}_{ik} &= a + \frac{T_{ik}}{2}, & \text{with } T_{ik} &= \sum_{t=1}^T \mathbb{1}\{z_{it} = k\}, \\ \bar{b}_{ik} &= b + 0.5 \left(\frac{m^2}{s} + \sum_{t:z_{it}=k} \varepsilon_{it}^{*2} - \frac{\bar{m}_{ik}^2}{\bar{V}_{ik}} \right), \\ \bar{V}_{ik} &= \left(\frac{1}{s} + T_{ik} \right)^{-1}, \\ \bar{m}_{ik} &= \bar{V}_{ik} \left(\frac{m}{s} + \sum_{t:z_{it}=k} \varepsilon_{it}^* \right).\end{aligned}$$

- (b) For $i = 1, \dots, K$ and $t = 1, \dots, T$, sample from the distribution of the mixture indicators $p(z_{it}|Y, \varphi_{-Z})$. For each component and point in time, the probability to be in cluster $k \in \{1, \dots, N\}$ is proportional to:

$$P(z_{it} = k|Y, \varphi_{-Z}) = \frac{\pi_{ik} \phi(\varepsilon_{it}^*; \mu_{ik}^*, \sigma_{ik}^{*2})}{\sum_{j=1}^N \pi_{ij} \phi(\varepsilon_{it}^*; \mu_{ij}^*, \sigma_{ij}^{*2})},$$

where $\phi(\cdot; \mu, \sigma^2)$ represents the value of the probability function of the normal distribution with mean μ and variance σ^2 .

- (c) Draw the beta random variables for $i = 1, \dots, K$ and $k = 1, \dots, N$:

$$p(V_{ik}|Y, \varphi_{-V}) \sim \text{Beta} \left(1 + T_{ik}, \alpha_i + \sum_{j=k+1}^N T_{ij} \right), \quad \text{for } k = 1, \dots, N-1,$$

$$V_{iN} = 1.$$

- (d) Sample the hyperparameters α_i ($i = 1, \dots, K$) as outlined in [Escobar and West \(1995\)](#). Therefore, let $n_i = \sum_{j=1}^N \mathbb{1}\{T_{ij} \neq 0\}$ be the number of active clusters in component $i = 1, \dots, K$. Then, first draw an auxiliary variable ξ_i and conditional on ξ_i , the concentration parameters α_i for $i = 1, \dots, K$:

$$\begin{aligned}p(\xi_i|\alpha_i) &\sim \text{Beta}(\alpha_i + 1, T), \\ p(\alpha_i|Y, \varphi_{-\alpha}, \xi_i) &\sim \pi_{\xi_i} \mathcal{G}(a_\alpha + n_i, b_\alpha - \log(\xi_i)) \\ &\quad + (1 - \pi_{\xi_i}) \mathcal{G}(a_\alpha + n_i - 1, b_\alpha - \log(\xi_i)),\end{aligned}$$

where π_{ξ_i} is defined as:

$$\frac{\pi_{\xi_i}}{1 - \pi_{\xi_i}} = \frac{a_\alpha + n_i - 1}{T(b_\alpha - \log(\xi_i))}.$$

2. Now consider sampling from the conditional distribution of the (PX) SVAR parameters. Denote $\mu_t^* = [\mu_{1,z_{1t}}^*, \dots, \mu_{K,z_{Kt}}^*]'$ and $\Sigma_t^* = \text{diag}([\sigma_{1,z_{1t}}^{*2}, \dots, \sigma_{K,z_{Kt}}^{*2}])$ for $t = 1, \dots, T$. Then, conditional on the current state of the Markov Chain, the model can be rewritten as:

$$y_t = (X_t' \otimes I_K) \gamma + B^* \varepsilon_t^*, \quad \varepsilon_t^* \sim \mathcal{N}(\mu_t^*, \Sigma_t^*).$$

which allows for an easy derivation of the conditional distributions.

- (a) With respect to the intercept and autoregressive parameters, the conditional distribution is a normal distribution $p(\gamma|Y, \varphi_{-\gamma}) \sim \mathcal{N}(\bar{\mu}_\gamma, \bar{V}_\gamma)$ where

$$\bar{V}_\gamma = \left(V_\gamma^{-1} + \sum_{t=1}^T (X_t \otimes I_K) (B^* \Sigma_t^* [B^*]')^{-1} (X_t' \otimes I_K) \right)^{-1},$$

$$\bar{\mu}_\gamma = \bar{V}_\gamma \left(V_\gamma^{-1} \underline{\mu}_\gamma + \sum_{t=1}^T (X_t \otimes I_K) (B^* \Sigma_t^* [B^*]')^{-1} \tilde{y}_t \right),$$

and $\tilde{y}_t = y_t - B^* \mu_t^*$.

- (b) With respect to the impact matrix, the conditional distribution is of unknown form. Therefore, an accept-reject Metropolis-Hastings (ARMH) algorithm is considered (Tierney; 1994; Chib and Greenberg; 1995). Let the target density denote $p(b^*|Y, \varphi_{-b^*})$ and define:

$$\bar{b}^* = \underset{\underline{b}^*}{\text{argmax}} \underbrace{p(\underline{b}^*) \prod_{t=1}^T \phi(y_t; \Gamma X_t + \underline{B}^* \mu_t^*, \underline{B}^* \Sigma_t^* [\underline{B}^*]')}_{=: p(Y|\underline{b}^*)}, \quad (3.A.14)$$

where $p(\underline{b}^*)$ is the prior density and $\phi(y_t|\mu_\phi, \Sigma_\phi)$ denotes a normal density at y_t with mean μ_ϕ and variance Σ_ϕ . Let $\mathcal{I}_{\bar{b}^*}$ be the inverse Hessian of $p(\cdot)p(Y|\cdot)$ evaluated at its maximum \bar{b}^* . Then, the algorithm continues as follows (Chib and Jeliazkov; 2005):

i. *Accept-reject step:*

Generate $\tilde{b}^* \sim \mathcal{N}(\bar{b}^*, \mathcal{I}_{\bar{b}^*}) =: h(\bar{b}^*|Y)$ and accept it with probability:

$$\alpha_{\text{AR}}(\tilde{b}^*|Y) = \min \left\{ 1, \frac{p(Y|\tilde{b}^*) p(\tilde{b}^*)}{c_{\bar{b}^*} h(\tilde{b}^*|Y)} \right\}.$$

which is repeated until a draw is accepted. In this paper, it is $c_{\bar{b}^*} = 3p(Y|\bar{b}^*) p(\bar{b}^*)$ following the implementation in Chan (2017), which for the computations conducted in this paper yielded acceptance rates of around 98% for the ARMH algorithm.

ii. *Metropolis-Hastings step:*

Let $\mathcal{D}(\bar{b}^*) = \{\underline{b}^* : p(Y|\underline{b}^*) p(\underline{b}^*) \leq c_{\bar{b}^*} h(\underline{b}^*|Y)\}$ and $\mathcal{D}^C(\bar{b}^*)$ its complement. Then, distinguish between the following cases where $b^{*(l-1)}$ denotes the current state of the Markov chain b^* :

A. If $b^{*(l-1)} \in \mathcal{D}(\bar{b}^*)$, set $\alpha_{\text{MH}}(\tilde{b}^*|Y) = 1$.

B. If $b^{*(l-1)} \in \mathcal{D}^C(\bar{b}^*)$ and $\tilde{b}^* \in \mathcal{D}(\bar{b}^*)$, set

$$\alpha_{\text{MH}}(\tilde{b}^*|Y) = \frac{c_{\bar{b}^*} h(\bar{b}^*|Y)}{p(Y|\bar{b}^*) p(\bar{b}^*)}.$$

C. If $b^{*(l-1)}, \tilde{b}^* \in \mathcal{D}^C(\bar{b}^*)$, set

$$\alpha_{\text{MH}}(\tilde{b}^*|Y) = \min \left\{ 1, \frac{p(Y|\tilde{b}^*) p(\tilde{b}^*) h(\tilde{b}^*|Y)}{p(Y|\bar{b}^*) p(\bar{b}^*) h(\bar{b}^*|Y)} \right\}.$$

Finally, set the next state of the Markov chain $b^{*(l)}$ to:

$$b^{*(l)} = \begin{cases} \tilde{b}^* & \text{with probability } \alpha_{\text{MH}}(\tilde{b}^*|Y), \\ b^{*(l-1)} & \text{with probability } 1 - \alpha_{\text{MH}}(\tilde{b}^*|Y). \end{cases}$$

3.A.2 Dealing with label switching

As described in Section 3.2, the model is only identified up to $2^K \times K!$ different permutations and sign configurations, which eventually get explored by the MCMC algorithm. Therefore, an algorithm based on k -means clustering is presented in the following, suitable to re-permute the output before it is used for inference. In the following, denote M the length of the Markov Chain and let $X_b = [\text{vec}(B^{(1)}), \text{vec}(B^{(2)}), \dots, \text{vec}(B^{(M)})]$ be the $K^2 \times M$ matrix of raw MCMC output containing the vectorized elements of the contemporaneous impact matrix B .

The following algorithm re-permutes X_b to \tilde{X}_b such that \tilde{X}_b contains draws of $\text{vec}(B)$ that correspond to a unique permutation and sign-configuration.

1. Using “data” X_b , choose a conservative number of clusters $k_{ub} \leq 2^K \times K!$ for a k -means algorithm, e.g. using the Gap statistic (Tibshirani et al.; 2001).
2. Run the k -means algorithm for k_{ub} on the “data” X_b to find the partition $S = \{S_1, \dots, S_{k_{ub}}\}$ that minimizes the within cluster sum of squares:

$$S = \underset{\hat{S}}{\text{argmin}} \sum_{i=1}^{k_{ub}} \left\| X_b(\hat{S}_i) - \mu(\hat{S}_i) \right\|^2,$$

where each cluster centroid $\mu(S_i) = \frac{1}{|S_i|} \sum_{l \in S_i} \text{vec}(B^{(l)})$ corresponds to the posterior mean estimate of $\text{vec}(B)$ based on the draws associated with S_i . Furthermore, save the cluster assignments $\zeta_l \in (1, \dots, k_{ub})$ for $l = 1, \dots, M$ providing the assignment of the l -th draw to the i -th component S_i .

3. Denote \hat{B}_i as the corresponding $K \times K$ matrix based on the centroid $\mu(S_i)$, \mathcal{P} the set of K -dimensional permutation matrices and \mathcal{D} the set of K -dimensional diagonal matrices with each diagonal entry being either +1 or -1. For $i = 1, \dots, k_{ub}$:

$$(P_i, D_i) = \underset{\hat{P}_i \in \mathcal{P}, \hat{D}_i \in \mathcal{D}}{\text{argmax}} \text{tr}(\hat{B}_i \hat{P}_i \hat{D}_i).$$

In words, find a permutation and sign configuration that maximizes the trace of the centroid for each cluster. For the SVAR model, this implies that the shocks will be ordered such the i th shocks has the largest (positive) impact on the i th variable ($i = 1, \dots, K$).

4. Re-permute the MCMC output based on these permutation and sign configurations, that is set:

$$\tilde{X}_b = [\text{vec}(B^{(1)} P_{\zeta_1} D_{\zeta_1}), \text{vec}(B^{(2)} P_{\zeta_2} D_{\zeta_2}), \dots, \text{vec}(B^{(M)} P_{\zeta_M} D_{\zeta_M})].$$

In case that inference is conducted on the posterior predictive densities of the structural shocks, the same permutation is used to reorder the output for the parameters underlying the corresponding DPMMs.

Appendix 3.B Time series plots

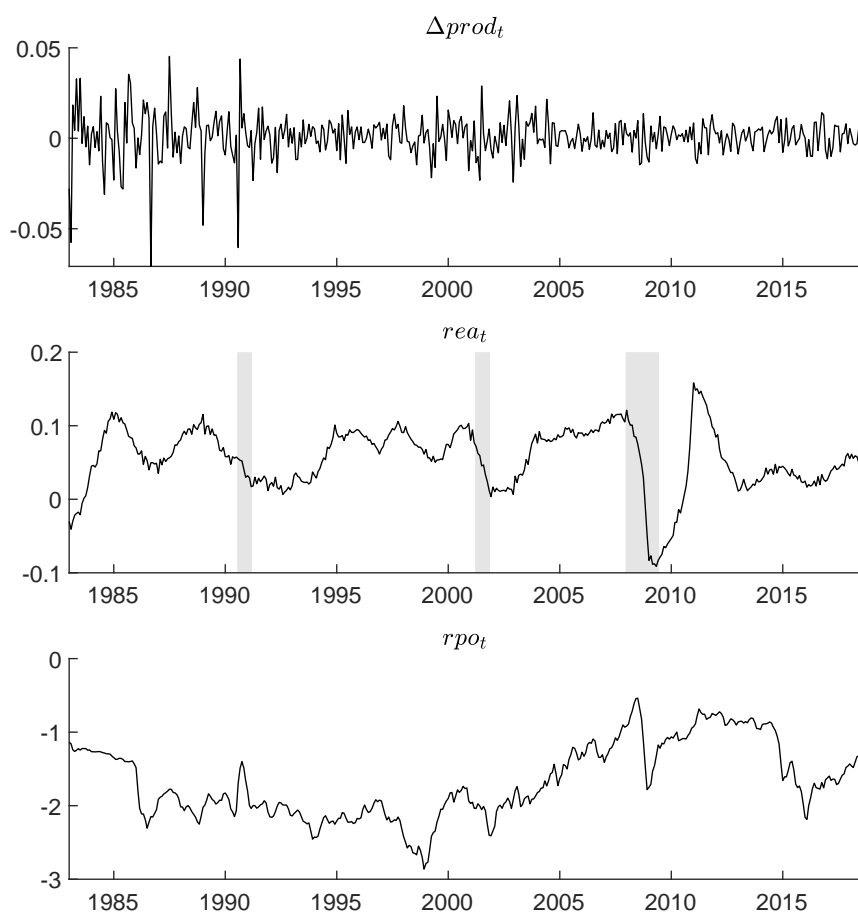


Figure 3.9: Time series plots for change in global crude oil production ($\Delta prod_t$), real economic activity (rea_t) and the (log of) real price of oil (rpo_t). Shaded areas indicate periods that overlap with NBER recession dates.

Complete References

- Angelini, G. and Fanelli, L. (2018). Identification and estimation issues in structural vector autoregressions with external instruments, *Working Papers 1122*, Dipartimento Scienze Economiche, Università di Bologna.
- Antolín-Díaz, J. and Rubio-Ramírez, J. F. (2018). Narrative sign restrictions for SVARs, *American Economic Review* **108**(10): 2802–29.
- Arias, J. E., Caldara, D. and Rubio-Ramírez, J. F. (2019). The systematic component of monetary policy in SVARs: an agnostic identification procedure, *Journal of Monetary Economics* **101**: 1–13.
- Arias, J. E., Rubio-Ramírez, J. F. and Waggoner, D. F. (2018). Inference based on structural vector autoregressions identified with sign and zero restrictions: Theory and applications, *Econometrica* **86**(2): 685–720.
- Barakchian, S. M. and Crowe, C. (2013). Monetary policy matters: Evidence from new shocks data, *Journal of Monetary Economics* **60**(8): 950–966.
- Baumeister, C. and Hamilton, J. D. (2015). Sign restrictions, structural vector autoregressions, and useful prior information, *Econometrica* **83**(5): 1963–1999.
- Baumeister, C. J. and Hamilton, J. D. (forthcoming). Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks, *American Economic Review* .
- Baumeister, C. and Peersman, G. (2013). The role of time-varying price elasticities in accounting for volatility changes in the crude oil market, *Journal of Applied Econometrics* **28**(7): 1087–1109.
- Ben Zeev, N. (2018). What can we learn about news shocks from the late 1990s and early 2000s boom-bust period?, *Journal of Economic Dynamics and Control* **87**: 94–105.
- Bernanke, B. S. (1986). Alternative explanations of the money-income correlation, *Carnegie-Rochester Conference Series on Public Policy*, Vol. 25, Elsevier, pp. 49–99.
- Bernanke, B. S., Boivin, J. and Elias, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach, *The Quarterly Journal of Economics* **120**(1): 387–422.
- Bernanke, B. S. and Mihov, I. (1998). Measuring monetary policy, *Quarterly Journal of Economics* **113**(3): 869–902.

- Bertsche, D. and Braun, R. (2018). Identification of structural vector autoregressions by stochastic volatility, *Working Paper Series of the Department of Economics, University of Konstanz 2018-03*, Department of Economics, University of Konstanz.
- Bjørnland, H. C. and Leitemo, K. (2009). Identifying the interdependence between US monetary policy and the stock market, *Journal of Monetary Economics* **56**(2): 275–282.
- Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via Pólya urn schemes, *The Annals of Statistics* **1**(2): 353–355.
- Blanchard, O. J. and Galí, J. (2010). *The Macroeconomic Effects of Oil Price Shocks: Why are the 2000s so different from the 1970s?*, University of Chicago Press.
- Blanchard, O. J. and Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances, *The American Economic Review* **79**(4): 655–673.
- Boscolo, R., Pan, H. and Roychowdhury, V. P. (2004). Independent component analysis based on nonparametric density estimation, *IEEE Transactions on Neural Networks* **15**(1): 55–65.
- Braun, R. and Brüggemann, R. (2017). Identification of SVAR models by combining sign restrictions with external instruments, *Working Paper 16*, Department of Economics, University of Konstanz.
- Brüggemann, R., Jentsch, C. and Trenkler, C. (2016). Inference in VARs with conditional heteroskedasticity of unknown form, *Journal of Econometrics* **191**(1): 69–85.
- Caffo, B. S., Jank, W. and Jones, G. L. (2005). Ascent-based Monte Carlo expectation-maximization, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(2): 235–251.
- Caldara, D., Cavallo, M. and Iacoviello, M. (forthcoming). Oil price elasticities and oil price fluctuations, *Journal of Monetary Economics* .
- Caldara, D. and Herbst, E. (forthcoming). Monetary policy, real activity, and credit spreads: Evidence from Bayesian proxy SVARs, *American Economic Journal: Macroeconomics* .
- Canova, F. and De Nicoló, G. (2002). Monetary disturbances matter for business fluctuations in the G-7, *Journal of Monetary Economics* **49**: 1131–1159.
- Cappé, O., Moulines, E. and Ryden, T. (2005). *Inference in Hidden Markov Models*, Springer.

- Carreiro, A., Clark, T. E. and Marcellino, M. G. (2019). The identifying information in vector autoregressions with time-varying volatilities: An application to endogenous uncertainty, *Working Paper* .
- Chan, J. C. C. (2017). The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling, *Journal of Business & Economic Statistics* **35**(1): 17–28.
- Chan, J. C. C. and Grant, A. L. (2016). On the observed-data deviance information criterion for volatility modeling, *Journal of Financial Econometrics* **14**(4): 772–802.
- Chan, J. C. C. and Jeliazkov, I. (2009). Efficient simulation and integrated likelihood estimation in state space models, *International Journal of Mathematical Modelling and Numerical Optimisation* **1**(1-2): 101–120.
- Chib, S. and Greenberg, E. (1995). Understanding the Metropolis–Hastings algorithm, *The American Statistician* **49**(4): 327–335.
- Chib, S. and Jeliazkov, I. (2005). Accept–reject Metropolis–Hastings sampling and marginal likelihood estimation, *Statistica Neerlandica* **59**(1): 30–44.
- Christiano, L. J., Eichenbaum, M. and Evans, C. L. (1999). Monetary policy shocks: What have we learned and to what end?, *Handbook of Macroeconomics*, Vol. 1, Elsevier, chapter 2, pp. 65–148.
- Christiano, L. J., Eichenbaum, M. and Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy, *Journal of Political Economy* **113**(1): 1–45.
- Coibion, O. (2012). Are the effects of monetary policy shocks big or small?, *American Economic Journal: Macroeconomics* **4**(2): 1–32.
- Conley, T. G., Hansen, C. B. and Rossi, P. E. (2012). Plausibly exogenous, *Review of Economics and Statistics* **94**(1): 260–272.
- Danielsson, J. and Richard, J. F. (1993). Accelerated Gaussian importance sampler with application to dynamic latent variable models, *Journal of Applied Econometrics* **8**: 153–173.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **39**: 1–38.

- Doan, T., Litterman, R. and Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions, *Econometric Reviews* **3**(1): 1–100.
- Drautzburg, T. (2016). A narrative approach to a fiscal DSGE model, *Working Paper 16-11*, Federal Reserve Board Philadelphia.
- Durbin, J. and Koopman, S. J. (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models, *Biometrika* **84**(3): 669–684.
- Durbin, J. and Koopman, S. J. (2000). Time series analysis of non-gaussian observations based on state space models from both classical and Bayesian perspectives, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62**(1): 3–56.
- Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures, *Journal of the American Statistical Association* **90**(430): 577–588.
- Faust, J. (1998). The robustness of identified VAR conclusions about money, *Carnegie-Rochester Conference Series on Public Policy* **49**: 207–244.
- Faust, J., Swanson, E. T. and Wright, J. H. (2004). Identifying VARs based on high frequency futures data, *Journal of Monetary Economics* **51**(6): 1107–1131.
- Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems, *The Annals of Statistics* **1**(2): 209–230.
- Fernández-Villaverde, J. and Rubio-Ramírez, J. F. (2007). Estimating macroeconomic models: A likelihood approach, *Review of Economic Studies* **74**(4): 1059–1087.
- Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*, Springer Science & Business Media.
- Fry, R. and Pagan, A. (2011). Sign restrictions in structural vector autoregressions: a critical review, *Journal of Economic Literature* **49**: 938–960.
- Gertler, M. and Karadi, P. (2015). Monetary policy surprises, credit costs, and economic activity, *American Economic Journal-Macroeconomics* **7**(1): 44–76.
- Geweke, J. (2005). *Contemporary Bayesian econometrics and statistics*, Vol. 537, John Wiley & Sons.
- Giacomini, R. and Kitagawa, T. (2015). Robust inference about partially identified SVARs, *Manuscript, University College London* .

- Gilchrist, S. and Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations, *American Economic Review* **102**(4): 1692–1720.
- Gourieroux, C., Monfort, A. and Renne, J. P. (2017). Statistical inference for independent component analysis: Application to structural VAR models, *Journal of Econometrics* **196**(1): 111–126.
- Hamilton, J. D. (2003). What is an oil shock?, *Journal of econometrics* **113**(2): 363–398.
- Hamilton, J. D. (2009). Causes and consequences of the oil shock of 2007-08, *Brookings Papers on Economic Activity* **40**(1): 215–283.
- Hamilton, J. D. (2018). Measuring global economic activity, *Working Paper*, University of California at San Diego.
- Harvey, A., Ruiz, E. and Shephard, N. (1994). Multivariate stochastic variance models, *Review of Economic Studies* **61**(2): 247–264.
- Herwartz, H. (2018). Hodges–Lehmann detection of structural shocks: An analysis of macroeconomic dynamics in the euro area, *Oxford Bulletin of Economics and Statistics* .
- Herwartz, H. and Lütkepohl, H. (2014). Structural vector autoregressions with Markov switching: Combining conventional with statistical identification of shocks, *Journal of Econometrics* **183**(1): 104–116.
- Herwartz, H. and Plödt, M. (2016). The macroeconomic effects of oil price shocks: Evidence from a statistical identification approach, *Journal of International Money and Finance* **61**: 30–44.
- Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors, *Journal of the American Statistical Association* **96**(453): 161–173.
- Jacquier, E., Polson, N. G. and Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models, *Journal of Business & Economic Statistics* **12**(4): 371–389.
- Jasra, A., Holmes, C. C. and Stephens, D. A. (2005). Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling, *Statistical Science* pp. 50–67.
- Justiniano, A. and Primiceri, G. E. (2008). The time-varying volatility of macroeconomic fluctuations, *American Economic Review* **98**(3): 604–641.
- Kilian, L. (2008a). The economic effects of energy price shocks, *Journal of Economic Literature* **46**(4): 871–909.

- Kilian, L. (2008b). Exogenous oil supply shocks: How big are they and how much do they matter for the U.S. economy?, *The Review of Economics and Statistics* **90**(2): 216–240.
- Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market, *American Economic Review* **99**(3): 1053–1069.
- Kilian, L. and Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*, Themes in Modern Econometrics, Cambridge University Press.
- Kilian, L. and Murphy, D. P. (2012). Why agnostic sign restrictions are not enough: Understanding the dynamics of oil market VAR models, *Journal of the European Economic Association* **10**(5): 1166–1188.
- Kilian, L. and Murphy, D. P. (2014). The role of inventories and speculative trading in the global market for crude oil, *Journal of Applied Econometrics* **29**(3): 454–478.
- Kim, S., Shephard, N. and Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models, *Review of Economic Studies* **65**(3): 361–393.
- Koop, G. and Korobilis, D. (2010). Bayesian multivariate time series methods for empirical macroeconomics, *Foundations and Trends in Econometrics* **3**(4): 267–358.
- Koopman, S. J., Shephard, N. and Creal, D. (2009). Testing the assumptions behind importance sampling, *Journal of Econometrics* **149**(1): 2–11.
- Lanne, M. and Luoto, J. (2016). Data-driven inference on sign restrictions in Bayesian structural vector autoregression, *Technical report*, Department of Economics and Business Economics, Aarhus University.
- Lanne, M. and Luoto, J. (2018). GMM estimation of non-Gaussian structural vector autoregression, *Technical report*, HECER, Helsinki Center of Economic Research.
- Lanne, M., Lütkepohl, H. and Maciejowska, K. (2010). Structural vector autoregressions with Markov switching, *Journal of Economic Dynamics & Control* **34**(2): 121–131.
- Lanne, M., Meitz, M. and Saikkonen, P. (2017). Identification and estimation of non-Gaussian structural vector autoregressions, *Journal of Econometrics* **196**(2): 288–304.
- Lanne, M. and Saikkonen, P. (2007). A multivariate generalized orthogonal factor GARCH model, *Journal of Business & Economic Statistics* **25**(1): 61–75.
- Lewis, D. J. (2018). Identifying shocks via time-varying volatility, *Federal Reserve Bank of New York Staff Report* (871).

- Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm, *Journal of the Royal Statistical Society Series B-Methodological* **44**(2): 226–233.
- Ludvigson, S. C., Ma, S. and Ng, S. (2017). Shock restricted structural vector-autoregressions, *Working Paper 23225*, National Bureau of Economic Research.
- Lütkepohl, H. (2005). *New introduction to multiple time series analysis*, Springer Science & Business Media.
- Lütkepohl, H. and Milunovich, G. (2016). Testing for identification in SVAR-GARCH models, *Journal of Economic Dynamics and Control* **73**: 241–258.
- Lütkepohl, H. and Netšunajev, A. (2014). Disentangling demand and supply shocks in the crude oil market: How to check sign restrictions in structural VARs, *Journal of Applied Econometrics* **29**(3): 479–496.
- Lütkepohl, H. and Netšunajev, A. (2017a). Structural vector autoregressions with heteroskedasticity: A review of different volatility models, *Econometrics and Statistics* **1**: 2–18.
- Lütkepohl, H. and Netšunajev, A. (2017b). Structural vector autoregressions with smooth transition in variances, *Journal of Economic Dynamics & Control* **84**: 43–57.
- Lütkepohl, H. and Schlaak, T. (2018). Choosing between different time-varying volatility models for structural vector autoregressive analysis, *Oxford Bulletin of Economics and Statistics* (forthcoming).
- Mahieu, R. J. and Schotman, P. C. (1998). An empirical application of stochastic volatility models, *Journal of Applied Econometrics* **13**(4): 333–359.
- Maxand, S. (forthcoming). Identification of independent structural shocks in the presence of multiple Gaussian components, *Econometrics and Statistics* .
- McLachlan, G. and Krishnan, T. (2007). *The EM algorithm and extensions*, Vol. 382, John Wiley & Sons.
- Melino, A. and Turnbull, S. M. (1990). Pricing foreign-currency options with stochastic volatility, *Journal of Econometrics* **45**(1-2): 239–265.
- Mertens, K. and Ravn, M. O. (2012). Empirical evidence on the aggregate effects of anticipated and unanticipated US tax policy shocks, *American Economic Journal-Economic Policy* **4**(2): 145–181.

- Mertens, K. and Ravn, M. O. (2013). The dynamic effects of personal and corporate income tax changes in the United States, *American Economic Review* **103**(4): 1212–1247.
- Mertens, K. and Ravn, M. O. (2014). A reconciliation of SVAR and narrative estimates of tax multipliers, *Journal of Monetary Economics* **68**: 1–19.
- Miranda-Agrippino, S. and Rey, H. (2018). US monetary policy and the global financial cycle, *Working Paper 21722*, National Bureau of Economic Research.
- Montiel-Olea, J. L., Stock, J. H. and Watson, M. W. (2015). Inference in SVARs identified with external instruments, *Harvard Manuscript* .
- Montiel-Olea, J. L., Stock, J. H. and Watson, M. W. (2016a). Uniform inference in SVARs identified with external instruments, *Harvard Manuscript* .
- Montiel-Olea, José, L., Stock, J. H. and Watson, M. W. (2016b). Uniform inference in SVARs identified with external instruments, *Harvard Manuscript* .
- Nakamura, E. and Steinsson, J. (2018). High frequency identification of monetary non-neutrality: The information effect, *Quarterly Journal of Economics* (forthcoming).
- Neal, R. M. and Hinton, G. E. (1998). A view of the EM algorithm that justifies incremental, sparse, and other variants, *Learning in graphical models*, Springer, pp. 355–368.
- Nevo, A. and Rosen, A. M. (2012). Identification with imperfect instruments, *Review of Economics and Statistics* **94**(3): 659–671.
- Normandin, M. and Phaneuf, L. (2004). Monetary policy shocks: Testing identification conditions under time-varying conditional volatility, *Journal of Monetary Economics* **51**(6): 1217–1243.
- Peersman, G. and Van Robays, I. (2012). Cross-country differences in the effects of oil shocks, *Energy Economics* **34**(5): 1532–1547.
- Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy, *Review of Economic Studies* **72**(3): 821–852.
- Ramey, V. (2016). Macroeconomic shocks and their propagation, *Handbook of Macroeconomics*, Vol. 2, Elsevier, chapter 2, pp. 71–162.
- Rigobon, R. (2003). Identification through heteroskedasticity, *Review of Economics and Statistics* **85**(4): 777–792.

- Romer, C. D. and Romer, D. H. (1989). Does monetary policy matter - a new test in the spirit of Friedman and Schwartz, *NBER Macroeconomics Annual* **4**: 121–170.
- Romer, C. D. and Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications, *American Economic Review* **94**(4): 1055–1084.
- Roweis, S. and Ghahramani, Z. (2001). Learning nonlinear dynamical systems using the Expectation–Maximization algorithm, *Kalman filtering and neural networks* **6**: 175–220.
- Rubio-Ramírez, J., Waggoner, D. and Zha, T. (2010). Structural vector autoregressions: Theory of identification and algorithms for inference, *The Review of Economic Studies* **77**(2): 665–696.
- Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, *Journal of the Royal Statistical Society Series B-Statistical Methodology* **71**(2): 319–392.
- Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models, *Journal of Econometrics* **63**(1): 289–306.
- Sethuraman, J. (1994). A constructive definition of Dirichlet priors, *Statistica Sinica* pp. 639–650.
- Shapiro, M. D. and Watson, M. W. (1988). Sources of business cycle fluctuations, *NBER Macroeconomics Annual* **3**: 111–148.
- Sims, C. A. (1980). Macroeconomics and reality, *Econometrica* **48**(1): 1–48.
- Stock, J. H. and Watson, M. W. (2012). Disentangling the channels of the 2007-09 recession, *Brookings Papers on Economic Activity* **43**(1): 81–156.
- Stock, J. H. and Watson, M. W. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics, in J. B. Taylor and H. Uhlig (eds), *Handbook of Macroeconomics*, Vol. 2, Elsevier, chapter 8, pp. 415–525.
- Tibshirani, R., Walther, G. and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63**(2): 411–423.
- Tierney, L. (1994). Markov chains for exploring posterior distributions, *The Annals of Statistics* **22**(4): 1701–1728.
- Uhlig, H. (2005). What are the effects of monetary policy on output? Results from an agnostic identification procedure, *Journal of Monetary Economics* **52**: 381–419.

- Uhrin, G. B. and Herwartz, H. (2016). Monetary policy shocks, set-identifying restrictions, and asset prices: A benchmarking approach for analyzing set-identified models, *Technical report*, University of Goettingen, Department of Economics.
- Van der Weide, R. (2002). GO-GARCH: A multivariate generalized orthogonal GARCH model, *Journal of Applied Econometrics* **17**(5): 549–564.
- Wieland, J. F. and Yang, M.-J. (2016). Financial dampening, *Working Paper 22141*, National Bureau of Economic Research.
- Woźniak, T. and Droumaguet, M. (2015). Assessing monetary policy models: Bayesian inference for heteroskedastic structural VARs, *Working Paper*, University of Melbourne.
- Wright, J. H. (2012). What does monetary policy do to long-term interest rates at the zero lower bound?, *The Economic Journal* **122**(564): 447–466.
- Yang, M., Dunson, D. B. and Baird, D. (2010). Semiparametric Bayes hierarchical models with mean and variance constraints, *Computational Statistics & Data Analysis* **54**(9): 2172–2186.

Eigenabgrenzung

Das erste Kapitel habe ich zusammen mit Ralf Brüggemann verfasst (Universität Konstanz). Meine individuelle Leistung bei der Erstellung dieses Kapitel beträgt 60%.

Das zweite Kapitel ist in Zusammenarbeit mit Dominik Bertsche entstanden, der ebenfalls Doktorand an der Universität Konstanz ist. Meine individuelle Leistung bei der Erstellung dieses Kapitel beträgt 50%.

Ich versichere hiermit, dass ich Kapitel 3 ohne Hilfe Dritter verfasst habe.