Causal effects in macroeconomics through higher moments

by

Jaakko Nelimarkka

M. Sc.

Academic dissertation to be presented for public examination,
by due permission of the Faculty of Social Sciences
of the University of Helsinki,
in the Lecture Hall of Economicum, Arkadiankatu 7,
on Friday, 18 January 2019 at 12 noon.
Research Reports
Publications of the Helsinki Center of Economic Research, No. 2019:1
Dissertationes Oeconomicae

JAAKKO NELIMARKKA

CAUSAL EFFECTS IN MACROECONOMICS THROUGH HIGHER MOMENTS

ISBN: 978-952-10-8748-6 (print)
ISBN: 978-952-10-8749-3 (online)
ISSN: 2323-9786 (print)
ISSN: 2323-9794 (online)

Unigrafia Oy
Helsinki 2019
Abstract

This dissertation examines causal effects in macroeconomics by identifying structural shocks using higher and time-varying moments of data. In particular, the use of non-normality provides more conditions from data to draw inference on the macroeconomic shocks and their propagation. In this way, I complement the conventional structural vector autoregressive analysis with techniques that facilitate the recovery of exogenous variation based on additional features of data.

In the first essay, I empirically analyse monetary policy transmission. The surprise announcements by the central bank allow to identify the causal effects of monetary policy. However, this identification strategy becomes problematic when a combination of shocks moves the market surprises. I introduce a flexible structural vector autoregression to identify several types of monetary policy actions triggering the market surprises, based on a novel combination of high-frequency proxies and higher moments of data. I estimate three distinct shocks from the surprise component of monetary policy, a conventional, a long-run monetary policy and an information shock. By a policy measured in the long-run shock, the central bank is able to influence the economy similarly to conventional monetary policy but with instruments other than the short-run interest rate.

The second essay studies the effects of government spending under anticipation of fiscal policy. When economic agents foresee future fiscal policies, measuring the causal effects of government spending is confronted by econometric challenges. The essay explores the propagation of government spending shocks using a noncausal model that allows for anticipation of exogenous fiscal policy changes. Overcoming the issue of insufficient information, the shock is extracted from an anticipated error term by using institutional information about the conduct of fiscal policy. In the U.S. economy, the shock increases investment, employment, and wages one and a half years prior to
its arrival, and consumption eventually rises. The estimated fiscal multiplier is above but close to unity. Importantly, neglecting the anticipation leads to underestimation of the multiplier.

In the last essay, I provide evidence on the effects of news shocks under insufficient information. News shocks about future productivity can be correctly inferred from a conventional vector autoregressive model only if information contained in the observables is rich enough. The methodology of the essay is able to measure the anticipation of permanent changes in total factor productivity independent of available information. By means of a noncausal model, economic shocks are recovered from both past and future variation, which solves the problem of insufficient information per se. Consequently, the model produces impulse responses to the anticipated structural shocks. In the U.S. economy, news about improving total factor productivity moves investment and stock prices, but the measured impact effects are modest. The estimated news shock gradually diffuses to productivity and generates smooth reactions of forward-looking variables.
Acknowledgments

This dissertation project started as an intention to understand better how the empirical and theoretical macroeconomic models are connected. In this respect, the product is not particularly far from the plan. However, I could not expect in 2013 that one type of distribution and a prefix “non” could play such a crucial role for all the pitfalls and advances during the following five years.

Several people have helped me through this research project. In particular, I am grateful to my thesis supervisor, Professor Markku Lanne. Since our first meeting in 2012, he has given me enormous support, guidance and valuable advice. The projects that seemed rather endless have turned into feasible ones after visiting his office. His reading of my early and unpolished drafts has considerably improved the quality, structure and language of the chapters of the thesis.

I am also thankful to Professor Luca Fanelli and Assistant Professor Lars Winkelmann who agreed to act as pre-examiners of this thesis. The constructive comments I received were extremely helpful and raised many important issues. I would also like to thank Associate Professor Luca Sala for agreeing to act as my opponent.

At different stages of the research, I have received feedback from the academic community in Helsinki. Professor Antti Ripatti has encouraged me in the research and provided me valuable comments. In spite of his scarce time resources, he has extensively helped me in administrative matters and given me professional advice. I have also had fruitful talks with Associate Professor Jani Luoto about computational details, estimation issues and other methodological matters. Eleonora Granzier’s thorough discussions on the raw drafts of my papers helped me to sharpen the focus of the research and to better justify the use of rather unconventional methods. The useful discussions with Docent Henri Nyberg highlighted, among other things, many desirable and
undesirable aspects of noncausal models. The members of the Research Group in Financial and Macroeconomics and the participants in several Econometric workshops of the FDPE have also given useful feedback on my research.

During my doctoral studies, I had an opportunity to visit the Department of Economics at Cornell University. I thank Karel Mertens for sponsoring this visit and Assistant Professor Kristoffer Nimark for the opportunity to attend the Macro Lunch seminars where I also had a chance to present my on-going work.

I presented and discussed the research contained in this dissertation in various conferences and occasions. The suggestions from Jacopo Cimadomo, Adam Gulan, Tomi Kortela, Annika Kuusela, Karel Mertens, Kristoffer Nimark, Pentti Saikkonen and Lauri Vilmi, amongst others, have given new perspectives and improved the current work.

I would also like to thank for financial support received from various sources. I am thankful for the four-year-salaried position in the Doctoral Programme of Economics at the University of Helsinki. In addition, financial support from the Yrjö Jahnsson Foundation, the OP Group Research Foundation, the Research Foundation of the Savings Banks and the Academy of Finland is gratefully acknowledged.

Pursuing doctoral studies would not have been that pleasant without the support from the fellow students of the HECER (now Helsinki GSE). The coffee and lunch breaks considerably diversified my social life during the thesis work. I was also honoured to sit with Juho Nyholm in the same office. His constant willingness to help with relevant and irrelevant issues has been both extremely valuable and entertaining.

Finally, I am grateful to my parents who have supported me from the beginning on to pursue studies in economics. My siblings Jufa and Jussi have played an important role in providing activities other than economic research. Most importantly, I wish to thank my dear wife Aino Silvo for her constant support, love and patience that made the realisation of this project possible. This dissertation is dedicated to our beloved daughter Saara who has brought us so much joy and happiness during the final phase of this project.

Helsinki, December 2018

Jaakko Nelimarkka
## Contents

Abstract i

1 Introduction 1
  1.1 Background ................................................. 1
  1.2 Identification of dynamic causal effects in macroeconomics 2
    1.2.1 The VARMA class of models ......................... 3
    1.2.2 Structural vector autoregression ..................... 5
    1.2.3 Dynamic identification problem ..................... 6
    1.2.4 Static identification problem ........................ 9
    1.2.5 Non-normality in macroeconomic variables .......... 13
  1.3 Summary of the essays ................................... 14
    1.3.1 Chapter 2: Identification of monetary policy shocks through proxies and non-normality 16
    1.3.2 Chapter 3: The effects of government spending under anticipation: the noncausal VAR approach 17
    1.3.3 Chapter 4: Evidence on news shocks under information deficiency ............................ 18

2 Identification of monetary policy shocks through proxies and non-normality 23
  2.1 Introduction .............................................. 23
  2.2 Empirical strategy ....................................... 27
    2.2.1 Monetary SVAR model ................................. 27
    2.2.2 Identification of monetary policy shocks through proxy variables ............................ 29
    2.2.3 Estimation ............................................. 34
  2.3 Results .................................................. 35
2.3.1 Data and estimation ........................................... 36
2.3.2 Proxies and monetary policy shocks ....................... 38
2.3.3 The effects of monetary policy shocks ...................... 40
2.3.4 Discussion .................................................. 43
2.4 Conclusions .................................................... 48
Appendix 2.A Derivation of the Gibbs sampler ................... 53

3 The effects of government spending under anticipation: the noncausal VAR approach 57
3.1 Introduction .................................................... 57
3.2 Theory ........................................................ 61
  3.2.1 Identification of government spending shocks under antici-
      pation .......................................................... 61
  3.2.2 Analytical example: a model of fiscal foresight ............ 66
  3.2.3 Estimation .................................................. 69
3.3 The impact of government spending in the U.S. economy .... 70
  3.3.1 Data and estimation ........................................ 71
  3.3.2 Impulse responses to a government spending shock ........ 73
  3.3.3 The size of the fiscal multiplier .......................... 78
  3.3.4 Relation to government spending shock measures .......... 80
3.4 Conclusions .................................................... 87
Appendix 3.A Estimation of the noncausal VAR(r,s) ............. 91
Appendix 3.B Convergence of the posterior sampler .............. 93
Appendix 3.C Further empirical results ........................... 93
Appendix 3.D Alternative identification scheme .................. 97

4 Evidence on news shocks under information deficiency 101
4.1 Introduction .................................................... 101
4.2 Theory ........................................................ 104
  4.2.1 Nonfundamentalness and noncausality ..................... 104
  4.2.2 Noncausal VAR ............................................. 107
  4.2.3 Noncausality in a stylised model with news shocks ....... 110
  4.2.4 Identification of news shocks ............................ 112
4.3 Monte Carlo simulation ........................................ 115
  4.3.1 New Keynesian model with news shocks .................. 116
  4.3.2 Simulation design ......................................... 117
  4.3.3 Nonfundamentalness and model selection ................ 118
  4.3.4 Impulse responses ......................................... 119
4.4 News shocks and U.S. business cycles ........................ 123
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.4.1</td>
<td>Data</td>
<td>124</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Estimation and identification</td>
<td>125</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Results from the baseline model</td>
<td>127</td>
</tr>
<tr>
<td>4.4.4</td>
<td>Reactions of further forward-looking variables</td>
<td>131</td>
</tr>
<tr>
<td>4.5</td>
<td>Conclusions</td>
<td>133</td>
</tr>
<tr>
<td>Appendix 4.A</td>
<td>Noncausal representation of a noninvertible VARMA model</td>
<td>138</td>
</tr>
<tr>
<td>Appendix 4.B</td>
<td>Maximum likelihood estimation of the noncausal VAR model</td>
<td>139</td>
</tr>
<tr>
<td>Appendix 4.C</td>
<td>Details on the stylised example</td>
<td>139</td>
</tr>
<tr>
<td>4.C.1</td>
<td>Derivation of the MA coefficients</td>
<td>139</td>
</tr>
<tr>
<td>4.C.2</td>
<td>Noncausal representation with a more general technology process</td>
<td>141</td>
</tr>
<tr>
<td>Appendix 4.D</td>
<td>Solution to the identification problem</td>
<td>143</td>
</tr>
<tr>
<td>Appendix 4.E</td>
<td>Monte Carlo simulations</td>
<td>144</td>
</tr>
<tr>
<td>4.E.1</td>
<td>NK model of Section 4.3</td>
<td>144</td>
</tr>
<tr>
<td>4.E.2</td>
<td>Further simulation results</td>
<td>145</td>
</tr>
<tr>
<td>Appendix 4.F</td>
<td>Details on the empirical part</td>
<td>148</td>
</tr>
<tr>
<td>4.F.1</td>
<td>Bayesian estimation of the noncausal VAR</td>
<td>148</td>
</tr>
<tr>
<td>4.F.2</td>
<td>On the convergence of the posterior sampling</td>
<td>150</td>
</tr>
<tr>
<td>4.F.3</td>
<td>Impulse responses from the additional specifications</td>
<td>153</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 Background

Knowledge of causal relations in macroeconomics is essential for the understanding of economic mechanisms and business cycle fluctuations. However, the recovery of causal effects in the interconnected world requires versatile methods, rigorous analysis and ambitious scientific enterprise. In particular, a macroeconometric method needs to isolate the movements important for the question of interest from the continuously evolving dynamic system which constantly reacts to new information about economic fundamentals. To this endogeneity problem, the line of attack in empirical macroeconomics is to use econometric and theoretical models that impose certain structure to the underlying economy.

Modern macroeconometric thinking traces back to Lucas (1976) who heavily criticised the way how simultaneous equations models were used at that time for policy analysis. According to the critique, coefficients estimated from correlations of the observed variables cannot directly be used to evaluate the impact of macroeconomic policies. In an economy with constantly optimising economic agents, a change in the coefficients regarding the policy directly implies a shift in the interaction within the economic system such that those correlations are no more valid. To correctly measure the effects, a change in a policy or fundamentals must come outside the economic system.

Since the Lucas’ critique, macroeconomic research has moved towards the study of macroeconomic shocks. These shocks induce unexpected fluctuations of the economy, as the economic agents adjust to new fundamentals in a system where the premises such as policy and behavioural rules remain time-invariant. Apart from explaining the economic fluctuations, the shocks allow to examine the dynamic effects of the economy to unexpected events. Hence,
as such, the shocks provide a framework to compare, evaluate and measure the impact of different macroeconomic mechanism and policies.

In the early 1980s, macroeconomics evolved in a way to take the Lucas’ critique into consideration. Starting from Kydland and Prescott (1982), theoretical macroeconomics introduced the use dynamic stochastic general equilibrium (DSGE) models which inferred macroeconomic fluctuation to stem from responses of economic agents to unexpected shocks on fundamentals. Simultaneously, Sims (1980) proposed the use of structural vector autoregressive (SVAR) models for policy evaluation in a reduced-form framework, where no particular theoretical structure is imposed on dynamic relationships of variables. In particular, the introduced SVAR methodology infers the causal effects not from the coefficients of the model but, rather, from exogenous changes extracted from an unpredictable component through structural assumptions. Importantly, the new macroeconomic techniques survived the Lucas’ critique by studying the dynamic path of stochastic shocks.

This dissertation introduces and uses new techniques for the study of macroeconomic effects in the SVAR framework. In the three essays, I will show that time-varying and higher moments of data can be exploited through non-Gaussianity to recover the dynamic effects of monetary policy, government spending and productivity on the economy. Specifically, the techniques refine the SVAR methodology when theoretical structure or observables are unable to provide sufficiently information to identify the economic shocks of interest.

1.2 Identification of dynamic causal effects in macroeconomics

The focus of dynamic macroeconomics is on the effects of macroeconomic shocks. The shocks are mutually uncorrelated and represent unexpected variation that moves the economic variables as they induce economic agents to reoptimise their behaviour. By estimating the effects of the shocks in a reduced form, i.e. imposing as few assumptions as possible to the underlying economy, it is possible to test implications of economic models and discriminate between competing paradigms. Fundamentally, the dynamic paths due to the shocks give information about the causal relationships in the economy.

The estimation of the responses to the macroeconomic shocks of interest is, however, subject to the issue that economic agents are forward-looking and
1.2 Identification of dynamic causal effects in macroeconomics

constantly reoptimise their behaviour. As a result, endogeneity is inherently present in any empirical question, which poses a difficulty for the isolation of exogenous shocks. Hence, an identification technique needs to take into account, first, how the economic agents form expectations about the future and, second, how to extract exogenous and unanticipated variation in the environment with general equilibrium effects.

To formalise the above requirement, let an $n$-dimensional column vector $y_t$ include the economic variables of interest, and let $E[y_t|I_{t-1}]$ be the conditional expectation of the economic agents about $y_t$ on the information set of economic agents, $I_{t-1}$, at $t-1$. Generally, to find the effects of macroeconomic shocks on the economy, the model needs to define expectations $E[y_t|I_{t-1}]$, identify macroeconomic shocks from the forecast error $y_t - E[y_t|I_{t-1}]$ and, under these premises, derive the reactions of the variables included in $y_t$.

1.2.1 The VARMA class of models

A general framework for the analysis of macroeconomic dynamics is a vector autoregressive moving average (VARMA) representation of the $n$ variables in $y_t$:

$$y_t = \sum_{i=1}^{p} A_i y_{t-i} + \sum_{i=0}^{q} B_i \varepsilon_{t-i}, \quad (1.1)$$

where $(n \times n)$ matrices $\{A_i\}_{i=1}^{p}$ contain the autoregressive coefficients of the lags of $y_t$ and $(n \times n)$ matrices $\{B_i\}_{i=0}^{q}$ are the moving average terms related to the $n$ mutually uncorrelated structural shocks in vector $\varepsilon_t$.\footnote{For the sake of illustration, constant terms are ignored from the representation.} In general, the VARMA representation (1.1) infers that the dynamics of economic variables are caught by their $p$ lags and by the current and past $q$ structural shocks. In addition to this reduced-form interpretation, the VARMA representation encompasses any solution of a linearised dynamic stochastic general equilibrium model (DSGE) model up to a truncation of the lags of $y_t$ (see, e.g. Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson, 2007. The latter aspect implies that finding an adequate VARMA model is equivalent to correctly capturing dynamics generated by the underlying economy, where forward-looking agents form rational expectations about the future states and optimally choose their actions.
The VARMA representation (1.1) is more compactly written as

\[ A(L)y_t = B(L)\varepsilon_t \]  (1.2)

with an \((n \times n)\) matrix polynomial \(A(L) = I_n - A_1 L - \ldots - A_p L^p\), \(B(L) = B_0 + B_1 L + \ldots + B_q L^q\) and \(L\) the usual lag operator. When the autoregressive polynomial \(A(L)\) is stable, i.e. its roots are outside the unit circle, the model (1.2) has a moving average (MA) representation

\[ y_t = A(L)^{-1}B(L) = \sum_{j=0}^{\infty} \Theta_j \varepsilon_{t-j} = \Theta(L)\varepsilon_t, \]  (1.3)

with coefficients of \(\{\Theta_j\}_{j=0}^{\infty}\) obtained recursively from identity \(A(z)\Theta(z) = B(z)\). The movements of variables in \(y_t\) are thus explained by the history of the structural shocks. Accordingly, the economic fluctuations stem from the responses of the economy to exogenous and unanticipated disturbances \(\varepsilon_t\) – the interpretation tracing back to Slutzky (1937). In particular, the reactions of variables \(y_t\) to an unexpected macroeconomic shock \(\varepsilon_{i,t}\), the \(i\)th element of vector \(\varepsilon_t\), can be deduced from the impulse response function

\[ \frac{\partial y_{t+j}}{\partial \varepsilon_{i,t}} = [\Theta_j]_{i}, \; j = 0, 1, \ldots, \]  (1.4)

where \([\Theta_j]_{i}\) is the \(i\)th column of matrix \(\Theta_j\). The MA coefficients thus reveal the dynamic path of the economy in response to a shock \(\varepsilon_{i,t}\).\(^2\)

The MA representation (1.3) is the building block of causal inference in macroeconomics. First, the produced impulse responses show how the economy moves as a result of a shock hitting the system. When the character of these shocks is known, tracking the dynamic effects helps to understand the importance of different economic mechanisms. Second, through the lens of the MA representation, business cycle fluctuations originate from variation in the macroeconomic shocks. Moreover, the representation is able to measure the relative importance of these shocks for different time horizons and unveil the sources of business cycle fluctuations. Third, provided the existence of a shock that reflects exogenous variation in the variable of interest, the respec-

\(^2\)Even if the polynomial \(A(L)\) does not satisfy the stability condition and the MA representation (1.3) does not exist, the impulse responses coincide with those obtained from the same recursion as above.
tive impulse responses depict the causal effects of this specific variable on the economy. For instance, it is possible to measure the implications of productivity increases or examine the effectiveness of monetary or fiscal policy, if there is a shock – or, equivalently, an instrument – that induces a direct change in the variable of interest.

1.2.2 Structural vector autoregression

Given that a large body of macroeconomic models, such as all linearised rational expectations models, is covered by the VARMA and MA representations (1.2) and (1.3), macroeconomic effects can be estimated in a reduced form and by imposing as few assumptions on the economic structure as possible. Prominently, structural vector autoregressions (SVAR) are able to measure – under certain assumptions – the dynamic effects of macroeconomic shocks generated by the underlying VARMA model. If the matrix polynomial $B(L)$, or equivalently $\Theta(L)$, is invertible in the past, i.e. it has no roots inside the unit circle, the VARMA model (1.2) has a vector autoregressive (VAR) representation

$$B_0B(L)^{-1}A(L)y_t = B_0\varepsilon_t, \quad (1.5)$$

where $B_0B(L)^{-1}A(L)$ is an $(n \times n)$ square summable matrix polynomial with possibly infinite powers of $L$.

Truncating the lags of $y_t$, the empirical counterpart of the representation (1.5) is a causal VAR model

$$C(L)y_t = u_t, \quad C(L) = I_n - C_1L - \ldots - C_pL^p, \quad (1.6)$$

where $\{C_i\}_{i=1}^p$ are $(n \times n)$ matrices and the residual $u_t = B_0\varepsilon_t$ is the reduced-form error term. As a result, the structural shocks $\varepsilon_t$ are contained in the forecast error of the model as $u_t = y_t - E[y_t|y_{t-1}, y_{t-2}, \ldots] = B_0\varepsilon_t$ and can be recovered after imposing identifying restrictions on matrix $B_0$. Moreover, the model has an MA representation

$$y_t = C(L)^{-1}u_t = \sum_{j=0}^{\infty} D_jB_0\varepsilon_{t-j}, \quad (1.7)$$

which coincides with the true representation (1.3) provided the validity of the invertibility condition and the identifying restrictions. Hence, as long as the underlying economy is assumed to be of the linear VARMA form (1.1), the
impulse responses of the structural VAR (SVAR) model (1.6) track the true propagation of the macroeconomic shocks.

The SVAR methodology faces, however, two identification problems (Hansen and Sargent, 1980, 1991; Lütkepohl, 2014). First, if the invertibility of the MA polynomial \( B(L) \) fails to hold, no mapping between the underlying VARMA model (1.2) and the VAR representation (1.5) exists. The second identification problem concerns finding a correct static rotation to map the reduced-form error term \( u_t \) to the true structural shocks, depending on the validity of the identifying restrictions imposed on the impact matrix \( B_0 \). Under a failure of these conditions, the SVAR model is unable to extract from the data the true structural shocks and impulse responses.\(^3\)

### 1.2.3 Dynamic identification problem

The first of the aforementioned identification problems is dynamic: invertibility of the VARMA model (1.2) is needed for filtering out an error term \( u_t \) consisting of the structural shocks \( \varepsilon_t \) through the current and lagged values of \( y_t \). However, invertibility is far from what equilibrium conditions of macroeconomic models usually imply for a small set of observables included in \( y_t \). In particular, noninvertibility may arise from the slow diffusion of technology shocks (Lippi and Reichlin, 1994a), from the ability of economic agents to predict future fiscal policies (Ramey, 2011; Leeper, Walker, and Yang, 2013) or from the news on productivity (Forni, Gambetti, and Sala, 2014; Beaudry and Portier, 2014). Above all, when economic agents are able to foresee certain shocks of the economy before they strongly influence the variables in \( y_t \), the underlying economy is likely to imply noninvertibility of the MA polynomial \( B(L) \) (Leeper et al., 2013; Beaudry and Portier, 2014). Fundamentally, the problem reduces to the fact that forward-looking agents have a larger information set than the econometrician possesses. That is, the conditional expectation \( E[y_t | y_{t-1}, \ldots] \) derived from the VAR model no more represents the forecast made by the economic agents.

Under this nonfundamentalness problem, induced by the above noninvertibility, the history of observables \( y_t \) does not contain enough information to recover the linear combination of the structural shocks \( B_0 \varepsilon_t \). In that case, fitting a VAR model produces an error term that is a linear combination of the current and lagged values of \( \varepsilon_t \) (Lippi and Reichlin, 1994b; Fernández-Villaverde

\(^3\)An alternative approach to recover the structural shocks of interest is the use of local projections (Jordà, 2005). On the identification of causal relationships in macroeconomics, see also Nakamura and Steinsson (2018b).
1.2 Identification of dynamic causal effects in macroeconomics

et al., 2007). This implies that no static impact matrix $B_0$ exists to retrieve the structural shocks. Furthermore, the implicitly invertible MA representation (1.7) does not correspond to the underlying noninvertible MA representation (1.3). The impulse responses produced from the structural VAR model may then be severely distorted.

Ultimately, the dynamic identification concerns information: the variables need to be reactive enough with respect to the current shocks. Hence, a straightforward approach to tackle nonfundamentalness is to augment the model with variables that are forward-looking enough to capture the expectations of the public (Leeper et al., 2013). In the fiscal policy, this strategy has been proceeded by Ramey (2011), Fisher and Peters (2010) and Leeper et al. (2012) by variables based on news paper sources or stock price that reflect expectations of economic agents. Information can also be added by extracting factors from large datasets, as proceeded by Forni et al. (2014) in studying news shocks on productivity or by Bernanke et al. (2005) to estimate the effects of monetary policy. Alternatively, the structural shocks can be obtained by imposing dynamic theoretical restrictions on the nonfundamental error term (Lippi and Reichlin, 1994b; Mertens and Ravn, 2010; Forni, Gambetti, Lippi, and Sala, 2017), or by directly estimating a theoretical model (Schmitt-Grohé and Uribe, 2012). Imposing such dynamic structure departs from the reduced-form analysis by shrinking the set of economic processes.

Compared to the above approaches, in Chapters 3 and 4, I propose an approach to recover the structural shocks under insufficient information while still imposing few assumptions on the dynamic structure. In particular, I solve the nonfundamentalness problem with a noncausal model that remains valid both under invertibility and noninvertibility. The model includes, in addition to the lagged values of $y_t$, future terms of observables, by which it is possible to filter out an error term that consists of anticipated structural shocks. Specifically, under noninvertibility of the VARMA model (1.2) to the past, the MA polynomial $B(L)$ can be inverted to the past and future such that

$$
\sum_{j=-\infty}^{\infty} F_j y_{t-j} = u_{t-l},
$$

(1.8)

where $l$ is the number of roots of the MA polynomial located within the unit circle. The above noncausal representation is, however, generic and infeasible to estimate. To make the representation operational, I use the noncausal VAR model of Lanne and Saikkonen (2013) to recover a potentially anticipated
reduced-form error term $\tilde{u}_t$:

$$\Pi(L)\Phi(L^{-1})y_t = \tilde{u}_t,$$

(1.9)

where $\Pi(L) = I_n - \Pi_1 L - \ldots - \Pi L_1$ and $\Phi(L) = I_n - \Phi_1 L^{-1} - \ldots - \Phi_s L^{-s}$ are the lag and lead polynomials. The use of future values circumvents the need for adding information to the model, as the model (1.9) remains valid both under nonfundamentalness and fundamentalness. In other words, the model allows for the misalignment between the information sets of economic agents and econometrician. In particular, the expectation $E[\tilde{u}_t | y_{t-1}, \ldots]$ is nonzero and can be forecast by the past values of $y_t$ (Lanne and Saikkonen, 2013; Lanne and Luoto, 2016).

Now, the anticipated error term, $\tilde{u}_t$ is a linear combination of the underlying but potentially lagged structural shocks contained in an $n$-dimensional vector $\bar{\varepsilon}_t$. The latter can then be obtained as $\tilde{u}_t = \bar{B} \varepsilon_t$ after imposing identifying restrictions on the static rotation matrix $\bar{B}$. The propagation of these anticipated shocks are determined by the two-sided MA representation of the model, obtained by inverting the stable polynomials $\Pi(L)$ and $\Phi(L^{-1})$:

$$y_t = \sum_{j=-\infty}^{\infty} \Psi_j \bar{B} \varepsilon_{t-j},$$

(1.10)

where $\sum_{j=-\infty}^{\infty} \Psi_j = \Phi(L^{-1})^{-1} \Pi(L)^{-1}$. This implies that the shocks obtained from the error term affect the variables before $t$ such that they are of anticipated nature. Instead, under fundamentalness, the lead terms $\{\Phi_j\}_{j=1}^s$ become redundant and the representation (1.10) reduces to the conventional one-sided MA representation (1.7). Hence, noncausality facilitates the recovery of the structural shocks as anticipated when the underlying model implies nonfundamentalness. Then, the true MA representation (1.3) is derived as the two-sided representation (1.10), where the original structural shocks are contained in $\bar{\varepsilon}_t$ as time-shifted.

However, to distinguish between noninvertible and invertible – correspondingly between noncausal and causal – representations, more conditions are needed from data for identification. Specifically, the causal VAR (1.6) and the noncausal VAR (1.9) are observationally equivalent when the structural shocks are Gaussian.\(^4\) This non-identifiability stems from the Gaussian distribution,
1.2 Identification of dynamic causal effects in macroeconomics

where the first and second moments determine the shape of the distribution. Thus a nonfundamental error term produced by a causal VAR (1.6) may well be uncorrelated over time, as Gaussianity does not distinguish between correlation and statistical independence. Under non-Gaussianity, in turn, the causal and noncausal models can be distinguished as higher moments of data are needed to produce an error term $\bar{u}_t$ that is not only uncorrelated but also independent over time.

Following Lanne and Saikkonen (2013) and Lanne and Luoto (2016), I identify and estimate the noncausal VAR by using a simple deviation from non-normality. The normally distributed structural shocks are augmented with a stochastic volatility factor as follows:

$$\bar{\varepsilon}_t = \omega^{-1/2}_t \varepsilon_t,$$

where $\varepsilon_t \sim N(0, I_n)$ are the mutually uncorrelated structural shocks and $\omega^{-1/2}_t$ is a volatility term such that $\lambda \omega_t$ is $\chi^2_\lambda$ distributed. As a whole, $\bar{\varepsilon}_t$ is multivariate t-distributed. When $\lambda \to \infty$, the distribution resembles Gaussian distribution. On the other hand, low values of $\lambda$ imply that the distribution of the structural shocks has fatter tails than normality would imply. In particular, such non-normality is usually present in the economic time series (Fagiolo, Napoletano, and Roventini, 2008). As $\lambda$ can be estimated from data, it is eventually an empirical question whether the noncausal model can be identified and used to tackle the nonfundamentalness problem.

1.2.4 Static identification problem

Although the SVAR model survives the dynamic identification problem presented above, a static identification problem remains to be solved, i.e. finding a mapping between the structural shocks and the reduced-form error term:

$$u_t = B_0 \varepsilon_t.$$  

In the conventional SVAR literature, the correct static rotation $B$ is inferred through covariance restrictions

$$\Sigma = E[u_t u_t'] = E[B_0 \varepsilon_t \varepsilon_t' B_0'] = B_0 B_0',$$

representation (1.3) directly. Their inference may be feasible only with strong prior information about the underlying economic mechanism, which enables to distinguish between invertible and noninvertible representations.
where $\Sigma$ is a positive definite $n$-dimensional covariance matrix of the reduced-form errors $u_t$ and can be estimated from data. The relation (1.13) is, however, able to provide only $n(n + 1)/2$ conditions for the $n$ unknown parameters in $B_0$. Hence, the data moments have to be complemented by prior restrictions based on external information to derive the impact matrix $B_0$ and produce the impulse responses from the MA representation (1.7).

The SVAR literature has developed various techniques to find $B_0$, as recently reviewed by Ramey (2016), Stock and Watson (2016) and Kilian and Lütkepohl (2017). I concentrate here on identification strategies that have been used to recover the class of structural shocks considered in the three subsequent chapters. First, the monetary policy shocks are deviations of the central bank from the systematic policy rule. Although these shocks, per se, determine only a small fraction of the movements of interest rates and economic activity, they can be used to derive the causal effects of monetary policy on the economy. Second, government spending shocks are exogenous changes in fiscal policy and are orthogonal to the state of the economy. Due to this exogeneity, the implied impulse responses show the propagation of an exogenous increase in the fiscal policy, which allows to measure the effectiveness of government spending on the economy. Third, news shocks on technology (Beaudry and Portier, 2006) have emerged as a factor that can explain how changes in productivity generate business cycle fluctuations. As these shocks induce immediate reactions in the forward-looking variables but diffuse to total factor productivity with a delay, the news shocks have been proposed to be an important source of short-run fluctuations.

The most common identification strategy imposes restrictions on the contemporaneous relations of the variables in the SVAR model. These restrictions assume that certain variables respond to the structural shocks or policy variables only with a lag. In other words, the restrictions exclude feedback from a current change in the economy to a set of variables. In the monetary SVAR literature, Christiano, Eichenbaum, and Evans (1999) use a recursive identification scheme according to which monetary policy adjusts within a period to real economic activity and prices whereas the latter react to monetary policy only with a lag. In the fiscal policy literature, Blanchard and Perotti (2002) use a timing restriction that government spending does not respond to the current circumstances in the economy. Hence, a government spending shock is a deviation from the endogenous path of spending. Correspondingly, Beaudry and Portier (2006) identify a news shock as a shock that affects total factor productivity only with a lag.

The above short-run restrictions are, however, not always justifiable from
the viewpoint of economic theory. Notably, the recursiveness assumption of Christiano et al. (1999) is at odds with the modern macroeconomic models that imply an immediate reaction of prices and output to monetary policy as well as with the fact that monetary policy often reacts to the current changes in the financial market. The short-run restrictions are also vague to model misspecification. In particular, when the underlying model implies nonfundamentalness and the reduced-form error is a linear combination of the current and past shocks, the identifying assumptions are starkly violated and even more incorrectly rotate the nonfundamental errors.

Apart from the above misspecification issues, it is often difficult to find suitable short-run restrictions as the economy continuously responds to the most recent shocks. One prominent technique, introduced by Uhlig (2005), uses sign restrictions on the variables of the model to discriminate between different shocks of the economy. In the context of monetary policy, the shock is then identified according to the signs of the responses of output, inflation and interest rate to an unexpected shock. Similarly, fiscal shocks can be identified through information about their effects on spending, taxes and consumption (Mountford and Uhlig, 2009). However, the approach provides only set identification, giving a spectrum of potential impact matrices that satisfy the restrictions. Moreover, the sign restrictions are often infeasible to use when the objective is to study implications of different paradigms which usually concern the sign of the impulse responses.

The identification can also be based on the medium- and long-run effects of the shocks. To identify a news shock on productivity, Beaudry and Portier (2006) apply the technique of Blanchard and Quah (1989) to find a shock that has a permanent effect on technology. Such long-run restrictions with infinite horizon have, however, weak small sample properties. These problems may be avoided when the horizon of interest is finite. The news shock literature has recently combined a medium-run identification scheme introduced by Uhlig (2004) and extended by Francis, Owyang, Roush, and DiCecio (2014) with the short-run restrictions. Proposed by Barsky and Sims (2011), the news shock explains the most of the forecast error variance in total factor productivity among the shocks that have a delayed effect on the variable.

The external information needed for the static identification can also be based on the use of variables informative about the structural shocks of interest. These types of variables are collected from external sources such as news papers, financial market data or administrative records and measure directly the shocks of interest. Hence, the identification is based on the timing and relevance of the variables with respect to the latent shocks. In the proxy SVAR
methodology (Stock and Watson, 2012; Mertens and Ravn, 2013), these variables are externally used for the identification of the structural shocks from the forecast error (1.12).

The empirical measures of structural shocks have widely used both in the empirical monetary and fiscal policy literature. Barakchian and Crowe (2013), Gertler and Karadi (2015) and Nakamura and Steinsson (2018a), amongst other, extract a monetary policy shock from movements in the futures and interest rates within a narrow window around the announcements by the central bank. In other words, the nonsystematic component of monetary policy is extracted from the surprise actions and language in the statements. Resolving both dynamic and static identification problems, Ramey (2011) uses a narrative series constructed from newspaper sources to measure the expected government expenditures to identify a spending shock.

The techniques discussed in this subsection can – to some extent – be used in the noncausal framework (1.9) to identify anticipated structural shocks. The use of these restrictions are, however, constrained by the fact that the timing of anticipated shocks can less precisely be defined a priori. In Chapters 3 and 4, after resolving the dynamic identification problem with the noncausal VAR (1.9), I complete the static identification with the use of short-run and medium-run restrictions. In particular, I identify the government spending shocks with the same exclusion restrictions as Blanchard and Perotti (2002) but allow now the shock to be anticipated by the economic agents. Similarly, the news shocks under potential noninvertibility are isolated from the anticipated error term with an assumption that the shock explains the most variation of total factor productivity in the medium run.

The above approaches complement the covariance restrictions (1.13) with assumptions on the timing, shape, magnitude of the structural shocks. The need for this external information originates from considering only the first and second moments in the identification. Nonetheless, it is possible to learn more from data as soon as higher or time-varying moments are used in the estimation. Starting from Rigobon (2003), growing literature uses non-normal features of data to identify structural shocks. The use of such features increases the conditions obtained from data such that the shocks can be identified without imposing theoretical restrictions on the impact matrix \( B \). More conditions can be obtained by unconditional (Rigobon, 2003) or conditional (Sentana and Fiorentini, 2001) heteroskedasticity, Markov-switching

\[ \text{The impulse responses can then be derived either in the SVAR framework or directly using local projections of Jordà (2005).} \]
1.2 Identification of dynamic causal effects in macroeconomics

heteroskedasticity (Lanne, Lütkepohl, and Maciejowska, 2010) or by any non-normal distribution (Lanne, Meitz, and Saikkonen, 2017). Then, the structural shocks are discriminated by using information about their statistical independence or orthogonality conditional on time-varying moments.

In particular, I use the statistical identification approach in Chapter 2 to extract a set of monetary policy shocks from proxy variables. I proceed by assuming that the $i$th element of the shock vector $\varepsilon_t$ is distributed as

$$\varepsilon_{i,t} = h_{i,t}^{-1/2} \varepsilon_{i,t}^*, \quad i = 1, \ldots, n,$$

(1.14)

where $\lambda_i h_{i,t} \sim \chi^2_{\lambda_i}$ and $\varepsilon_{i,t}^* \sim N(0, 1)$. Conditionally, the structural shocks are accompanied by a stochastic volatility factor $h_{i,t}^{-1/2}$, which implies that conditions relating to moments $E[(\varepsilon_{i,t})^k(\varepsilon_{j,t})^l]$ for $k + l \geq 3$, $i \neq j$ and $i, j = 1, \ldots, n$ are no more determined by the first and second moments only. Unconditionally, the structural shocks are Student’s t-distributed. The assumption is similar to the multivariate t-distribution in (1.11) but, now, the structural shocks are required to be cross-sectionally independent in addition to the time independence. Consequently, it is possible to distinguish multiple structural shocks by the statistical properties of proxy variables. Hence, in place of discriminating between a number of structural shocks using zero restrictions that are difficult to justify, I combine the proxy-based identification with the knowledge about the statistical properties of data, facilitated by the assumption of non-normality (1.14). By the distributional assumption, the model can efficiently be estimated by Bayesian methods.

1.2.5 Non-normality in macroeconomic variables

Naturally, non-Gaussianity is a useful property to exploit only if the higher moments of data exhibit patterns that cannot be justified by normally distributed shocks. In effect, distributions of growth rates of output have been documented to be heavy-tailed (Fagiolo et al., 2008; Ascari et al., 2015), i.e. they have excess kurtosis that translates to large deviations from the mean. Moreover, the estimated VAR models often produce residuals that do not survive tests for normality (See Kilian and Lütkepohl 2017, Ch. 14, and the references therein). In a similar vein, Cúrdia, del Negro, and Greenwald (2014) and Chib and Ramamurthy (2014) show that augmenting a typical DSGE model with t-distributed innovations improves its performance in the low-frequency data, suggesting the importance of the excess kurtosis and the existence of
rare large shocks.

For motivation, Figure 1.1 plots first-hand evidence on non-normality in data from the post-war U.S. economy. In particular, I plot distributions of several macroeconomic indicators used in the analyses of the subsequent chapters. The solid lines depict the kernel density estimates of the variables, whereas the dashed-dotted lines show the corresponding fitted Gaussian distribution. Panel (a) plots a set of variables at monthly frequency and Panel (b) at quarterly frequency. All nonstationary variables are expressed in growth rates. Panel (a) reveals that none of the monthly variables aligns with the corresponding normal distribution. Rather, probability mass is excessively concentrated on the middle and on the tails of the distributions. A similar observation concerns the lower-frequency quarterly data: non-normality does not vanish due to the time aggregation.

Strikingly, the evidence suggests that non-normality is not a concern in the high-frequency financial data only but prevails in more time aggregated monthly and quarterly macroeconomic data. In principle, this phenomenon could be a result of the non-linear economic structure with heavy-tailed distributions emerging from the underlying endogenous decision rules. In that case, a VARMA model necessarily captures dynamics only partially, leading to the observed non-normal innovation terms. However, according to Ascari et al. (2015), the non-linearity of DSGE models is insufficient to generate the observed heavy-tailed distributions. Instead, the distributions of economic variables may well be generated by a linear combination of non-normal structural shocks. In this perspective, non-Gaussianity need not be considered a symptom of misspecification but an inherent feature of data, which additionally provides a useful source of information for the static and dynamic identification of causal effects.

1.3 Summary of the essays

In this section, I briefly summarise the content of the three self-contained essays of this thesis. The ultimate aim of the essays is to recover macroeconomic shocks to study their propagation in the economy. For this purpose, non-normality is used to overcome the lack of identification. Chapter 2 concentrates on the static identification problem by exploiting both high-frequency proxy variables and non-normality to measure the causal effects of monetary policy. In Chapters 3 and 4, the interest is on the dynamic identification problem to find a fundamental error term by additionally exploiting the fu-
Figure 1.1: Distributions of selected monthly and quarterly U.S. macroeconomic variables

Kernel density estimates (solid lines) and estimates of the corresponding normal distribution (dot-dashed lines) reported. Growth rates ($\Delta\%$) of industrial production, consumer price index (CPI) and the quarterly variables have been annualised. GDP is the real per-capita gross domestic product and TFP the capacity-utilisation adjusted total factor productivity. For details on the quarterly variables, see Subsections 3.3.1 and 4.4.1. Data span the time period 1948–2016 for all except the 2 and 10-year yields (1961:6–2016). Data sources: National accounts variables (National Income and Product Accounts, Bureau of Economic Analysis); the utilisation-adjusted TFP (Fernald, 2012); Industrial production, Unemployment rate and CPI (FRED, Federal Reserve Bank of St. Louis); the interest rates (Gürkaynak et al., 2007); S&P index (Yahoo! Finance). All variables except the interest rates and the S&P index are seasonally adjusted.
ture variation of macroeconomic indicators. Recovering the macroeconomic shocks as anticipated, these two essays analyse anticipated fiscal policy and news on productivity.

1.3.1 Chapter 2: Identification of monetary policy shocks through proxies and non-normality

To identify the causal effects of monetary policy, it is necessary to find variation from the non-systematic component of monetary policy. However, as the actions of the central bank mainly originate from its endogenous response to the economy, empirical strategies using different identification schemes do not produce analogous results about the propagation of monetary policy shocks.

In the recent literature, surprise announcements by the central bank are observed in movements of financial variables within a narrow time window around the release of monetary policy statements. These high-frequency proxies reflect deviations from the endogenous policy rule and allow to measure the effects of monetary policy. However, the latter identification strategy becomes problematic when the market surprises are triggered by more than one shock. In that case, the identified shock is a mixture of underlying shocks and may induce reactions that are far from the true impulse responses. As monetary policy is to a large part communication in multiple dimensions, statements of the central bank highly likely induce surprises that cannot be condensed into one factor only.

I introduce a flexible structural vector autoregression with a novel identification scheme to find several types of unexpected monetary policy actions. Specifically, I exploit comovement of high-frequency proxies and macroeconomic variables in combination with higher moments of data. I proceed by introducing time-varying volatility to the proxy variables, by which I am able to distinguish various channels of monetary policy transmission without imposing prior restrictions. The approach combines two appealing identification strategies. First, the proxy variables accurately measure monetary policy surprises induced by the statements. Second, non-normality is used to distinguish between the relevant shocks.

I estimate three types of policy surprises that drive the reactions to the monetary policy statements in the U.S. economy. First, a conventional monetary policy shock relates to surprise changes in the short-run rate. Second, the statements induce, to a smaller extent, the arrival of a shock that reveals new information to the public about the state of the economy through the
endogenous response of monetary policy. The third shock affects the risk and expectations of economic agents – orthogonal to the present short-run rate change – originating from communication of the central bank regarding its future actions. Importantly, by this policy akin to forward guidance, the central bank is able to influence the economy similarly to the conventional interest rate change.

1.3.2 Chapter 3: The effects of government spending under anticipation: the noncausal VAR approach

Fiscal policy usually includes an implementation lag during which the economy may adjust to the new conditions. As a result, new policies are observed by economic agents before the shock arrives in an empirical model. This implies that the shocks identified from the VAR model may in fact be anticipated by the public. The conventional VAR models are thus at high risk to mismeasure the effects of government spending.

I explore the propagation of government spending shocks using a noncausal model that allows for anticipation of exogenous fiscal policy changes. In particular, I identify a government spending shock with the exclusion restrictions of Blanchard and Perotti (2002) according to which the economic policy does not respond to the current state of the economy. However, unlike the previous literature, I allow the identified government spending shock to be anticipated such that the issue of insufficient information becomes innocuous. The shock is extracted from an anticipated error term, and the impulse responses can be derived from the two-sided MA representation of the model. Importantly, the proposed method remains valid regardless of the nonfundamentalness issue.

In the U.S. economy, the shock increases investment, employment, and wages one and a half years prior to its arrival, and consumption eventually rises. Ex-post, the shock captures movements in defence spending, in line with the previous literature. The estimated fiscal multiplier is above but close to unity. Disregarding the anticipation and using a conventional VAR model lead to underestimation of the multiplier. I also show that variables gauging the expectations of the public can in part anticipate the future policy shock identified in the noncausal VAR framework. The findings suggest that government spending is mildly expansionary, supporting the existence of Keynesian mechanisms.
1.3.3 Chapter 4: Evidence on news shocks under information deficiency

News shocks about future productivity generate fluctuation of the economy before radical changes can be observed in total factor productivity. However, the news shocks can correctly be inferred from a conventional VAR model only if information contained in observables is rich enough. In particular, an underlying theoretical model with news shocks is likely to imply a nonfundamentalness problem for the observables, where the shock cannot be retrieved from the past only.

The essay provides evidence on the anticipation of permanent changes in total factor productivity independent of the available information. By means of the noncausal VAR model, economic shocks are recovered from both the past and future variation, which solves the problem of insufficient information per se. Consequently, the model is able to show the impulse responses to the anticipated structural shocks. Methodologically, the essay contributes by introducing an identification technique which facilitates the conduct of structural analysis with the noncausal VAR model. I show both with a stylised example as well as with simulations from a DSGE model that the approach and the proposed identification strategy perform well in the recovery of news shocks.

In the U.S. economy, news about improving total factor productivity moves investment and stock prices, but the measured impact effects are modest. The estimated news shock gradually diffuses to productivity and generates smooth reactions of forward-looking variables. Hence, the news shocks are unlikely to generate significant short-run fluctuations of the economy. The measured strong reactions of the previous literature may be a consequence of information deficiency and the ignored contemporaneous effects of the shock on total factor productivity.

References


References


2 Identification of monetary policy shocks through proxies and non-normality

2.1 Introduction

A fundamental issue in macroeconomics is how monetary policy propagates to the real economic activity and prices. For the estimation of the dynamic causal effects of monetary policy, an identification strategy requires exogenous variation stemming from surprise actions of the central bank. That is, by finding non-systematic movements in monetary policy – deviations from its systematic, endogenous decision rule – it is possible to measure the direct effects of money and interest rates on the macroeconomy. Despite vast research, the literature has reached no consensus about the transmission of monetary policy or about how the non-systematic variation is measured.

This essay contributes to the monetary policy literature by estimating the effects of multiple shocks driving the reactions of the public to monetary policy announcements. The approach combines identification based on variables that reflect the surprise movements of the central bank with the use of statistical properties of data. As a result, multiple monetary policy shocks revealed in the statements of the central bank can flexibly be recovered. I find that the statements of the Federal Reserve are connected with three distinct shocks that have asymmetric effects on the interest rates, on credit spreads and on the economy. The existence of multiple shocks is able to give insight into various
dimensions of monetary policy communication.

Since Sims (1980), vector autoregressive (VAR) models have been used to isolate monetary policy shocks, identified by structural assumptions based on prior information. In the structural VAR (SVAR) literature, identification is traditionally achieved by short-run recursive (Christiano, Eichenbaum, and Evans, 1999) or sign (Uhlig, 2005) restrictions. These restriction are, however, not always justifiable from the view point of economic theory or external information. In particular, they may unrealistically exclude simultaneous feedback of the economy from monetary policy or limit the reactions of the central bank to current financial conditions. Moreover, extracting the nonsystematic part from the policy rate – as usually proceeded – may be problematic as soon as the central bank communicates its actions in advance. In particular, under forward-looking monetary policy, the actions of the central bank tend to be anticipated such that changes in the target rate are unlikely to contain monetary policy shocks (Barakchian and Crowe, 2013).

To address these problems, exogenous variation in monetary policy has been identified through construction of a variable informative about surprise actions of the central bank. Romer and Romer (2004) identify monetary policy shocks as deviations of the target rate change from the one implied by the internal forecasts of the Federal Reserve. On the other hand, the high-frequency identification approach interprets that the market reactions within a narrow window around the announcements by the central bank reflect the surprise component of monetary policy. To this line of research belong, amongst others, Kuttner (2001), Faust, Swanson, and Wright (2004), Gürkaynak, Sack, and Swanson (2005), Cochrane and Piazzesi (2002) Barakchian and Crowe (2013), and Gertler and Karadi (2015). As an immediate advantage, the strategy extends the analysis of monetary policy communication beyond a conventional interest rate change.

Using these proxies of the underlying structural shocks, the impulse responses to a monetary policy shock can be derived using local projections (Jordà, 2005) or a proxy SVAR model (Stock and Watson, 2012; Mertens and Ravn, 2013). While the former approach provides inference that is flexible with respect to model misspecification, a SVAR model controls efficiently for remaining endogeneity and produces precise estimates even if the proxy is a noisy measure of the latent shock. Applying the proxy SVAR, Gertler and Karadi (2015) recover a monetary policy shock from the high-frequency move-

---

1See Ramey (2016) for a recent survey of identification strategies and empirical literature.
2See also Stock and Watson (2018).
ments of Federal funds futures around the meetings of the Federal Open Market Committee (FOMC).

However, identifying one monetary policy shock from the reactions of financial variables may constitute a problem as soon as the central bank communication relates to more than one instrument. Specifically, apart from the interest rate target, the language of monetary policy statements may signal its future actions (Gürkaynak, Sack, and Swanson, 2005; Campbell, Evans, Fisher, and Justiniano, 2012), contain information about the state of the economy (Nakamura and Steinsson, 2018) or affect the term structure of interest rates in a way inconsistent with standard macroeconomic models (Hanson and Stein, 2015). In other words, a proxy variable used to identify one shock may in reality be moved by a number of policy shocks. Violating the exclusion restriction, the proxy is no more uncorrelated with the remaining structural shocks of the economy but, rather, the identified shock is a compound of several types of policies and announcements.

In this essay, I estimate the effects of monetary policy using a wide range of high-frequency variables to identify multiple sources how monetary policy announcements and actions influence the economy. In particular, I exploit both information contained in the variables informative about the underlying shocks and statistical properties of the data to recover a set of shocks that drive the monetary policy announcements. Consequently, the dynamic effects of different types of monetary policies can simultaneously be derived regardless of the strength and number of the proxy variables or of prior knowledge about the propagation of the shocks.

Conventionally, parameters of the SVAR model are obtained from least squares estimation or by assuming normality of the structural shocks. In spite of being computationally convenient and tractable, the identification of structural shocks then requires further restrictions based on external information. In the proxy SVAR, this information is incorporated through a proxy variable. However, if the proxies are correlated with more than one shock, the latter need to be distinguished by prior knowledge (Mertens and Ravn, 2013; Arias, Rubio-Ramirez, and Waggoner, 2018), but these restrictions may be difficult to justify. In this essay, instead, learning from data is extended to higher than second moments in order to obtain additional information for the identification. Specifically, I proceed by assuming idiosyncratic time-varying volatility in the structural shocks, which allows me to statistically identify a set of shocks that both move the proxy variables on impact and produce economically meaningful impulse responses. Compared to the proxy SVAR, the approach is now able to identify multiple structural shocks based on a number of proxies with-
out prior assumptions about their short-run effects.

Exploiting the non-normality and proxy variables, I estimate the SVAR model involving idiosyncratic volatility with Bayesian methods. To draw posterior parameters, I develop a computationally efficient algorithm that is particularly suitable for large-scale models. The Gibbs sampler is similar to Lanne and Luoto (2016) but uses the technique of Waggoner and Zha (2003) to draw the coefficients of contemporaneous relations. The latter step circumvents numerical problems involved with a large number of parameters, and the estimation routine can efficiently be applied to any overidentified SVAR model both under Gaussianity and non-Gaussianity.\(^3\)

This essay is also linked to a growing body of literature on statistical identification of structural shocks.\(^4\) By assuming a non-normal distribution, data can bring additional conditions for the identification of the structural shocks through unconditional (Rigobon, 2003) and conditional (Sentana and Fiorentini, 2001) heteroskedasticity, a mixture of normal distributions (Lanne and Lütkepohl, 2010) or through any non-Gaussian distribution (Lanne, Meitz, and Saikkonen, 2017; Gouriéroux, Monfort, and Renne, 2017). Alternatively, non-normality can be exploited through higher-moment conditions directly (Guay and Normandin, 2018; Lanne and Luoto, 2018). To establish full identifiability of the SVAR model, this essay applies the former strategy. Accordingly, structural shocks are required to be not only mutually uncorrelated but also independent, and the higher-moment conditions are implicitly taken into account. The statistically identified shocks do not, however, contain economic interpretation per se, as they are extracted from statistical properties of data only. The labelling of the shocks can thus be facilitated by the use of multiple proxy variables informative about the structural shocks of interest.

Similar to this essay, Caldara and Herbst (2018) and Arias et al. (2018), amongst others, rely on the Bayesian inference to estimate the proxy SVAR model.\(^5\) Whereas Caldara and Herbst (2018) use reduced-form parameterisation and restrict attention to one shock using a single proxy, the non-normality allows to draw structural parameters of the model directly. Moreover, the lack of identification in a multiple-shock-setting, as raised by Arias et al. (2018), is absent under non-Gaussianity. Thanks to the overidentifiability, the inference under non-normality is not conditional on exogeneity assumptions or quality

\(^{3}\)The model used in this chapter is also similar to Brunnermeier, Palia, Sastry, and Sims (2017) who apply statistical identification with the use of t-distributed shocks and heteroskedasticity.

\(^{4}\)For recent review, see Kilian and Lütkepohl (2017), Ch. 13.

\(^{5}\)In addition, Antolín-Díaz and Rubio-Ramírez (2018) combine sign restrictions and proxy-based identification.
of the proxy.

In the empirical part of the chapter, I identify various monetary policy shocks driving the reactions to the release of Federal Reserve’s monetary policy statements, observed by changes in the future rates around the FOMC meetings. I am able to identify three shocks associated with the monetary policy statements. First, a conventional monetary policy shock affects the short-run rate. The second shock I identify is related to the part of the statement that reflects the endogenous component of monetary policy but is unexpected to the economic agents. In other words, the shock induces information effects similar to Campbell et al. (2012) and Nakamura and Steinsson (2018). The information shock plays, however, a less central role for the overall surprise movements. The third identified shock has no effects on the short-run rate but affects the risk premium. In particular, the shock mirrors language in the monetary policy statements regarding the future policy actions, being closely related to the path shock of Gürkaynak et al. (2005). Such a forward-guidance-type policy reflected in the shock has macroeconomic effects similar to the conventional interest rate change. Finally, the existence of these three distinct shocks potentially explains why the proxy variable used by Gertler and Karadi (2015) is sensitive to the model specification, the issue raised by Ramey (2016).

The remainder of the essay is organised as follows. The next section outlines the empirical strategy. Empirical results are shown in Section 2.3. The last section concludes.

2.2 Empirical strategy

This section presents the structural vector autoregression that uses proxies combined with statistical properties of data to identify monetary policy shocks. First, I review the monetary SVAR framework. Second, I present the identification based on the proxy variables and non-normality of data. Last, the estimation of the model is outlined.

2.2.1 Monetary SVAR model

Here, I present the standard structural VAR framework to study the causal effects of monetary policy. Assume, ignoring deterministic terms, that \( n \) observables in vector \( y_t \) have a \( \text{VAR}(p) \) representation with respect to \( n \) mutually
Identification of monetary policy shocks through proxies and non-normality

uncorrelated structural shocks in vector $\varepsilon_t$ with a unit variance,

$$y_t = \sum_{i=1}^{p} A_i y_{t-i} + B_y \varepsilon_t$$

(2.1)

where $B_y$ and $A_i$, $i = 1, \ldots, p$ are ($n \times n$) matrices. The dynamic effects of variables can then be analysed through impulse responses to the macroeconomic shocks, derived from the moving average representation of the model. However, without further assumptions, the SVAR model (2.1) is not identified since the covariance restrictions of the reduced-form residuals $u_t = B_y \varepsilon_t$,

$$E[B_y \varepsilon_t \varepsilon_t'] = E[u_t u_t'] = \Sigma,$$

(2.2)

do not provide enough conditions to obtain a unique set of $n^2$ elements of matrix $B_y$.

To achieve identifiability and recover a monetary policy shock, the standard monetary SVAR literature proceeds by theoretical restrictions on the elements of $B_y$. Prominently, under the recursiveness assumption of Christiano et al. (1999), the policy rate is allowed to adjust to contemporaneous macroeconomic conditions, whereas the real variables and prices react to current monetary policy with a lag. Following Uhlig (2005), the monetary policy shock can be set-identified through sign restrictions on the elements of matrix $B_y$. Alternatively, $y_t$ may include a measure of the monetary policy shock, which is ordered either as the first variable in $y_t$ with lower triangular $B_y$ or after the real variables and prices when imposing the recursiveness assumption.\(^\text{6}\)

Provided those identifying assumptions are valid, the above strategies can measure the causal effects of monetary policy. However, the recursive, zero and sign restrictions are subject to theoretical assumptions that may be questioned. The recursive identification presupposes lagged adjustment of output and inflation to monetary policy and no contemporaneous reaction of the central bank to financial variables. On the other hand, it may be difficult to come up with suitable sign restrictions such that the identified monetary policy surprises can credibly be distinguished from the other shocks of the economy. Moreover, the latter approach produces identification only up to a set of impact matrices $B_y$ that satisfy the imposed sign restrictions.

\(^{6}\)For instance, this measure can be related to deviations of the intended target rate change from the one implied by the internal forecasts of the central bank (Romer and Romer, 2004) or to changes in private sector’s beliefs about the policy stance based on Federal funds rate future contracts (Barakchian and Crowe, 2013).
In contrast, the restrictions based on the simultaneous relations or the signs of responses can be relaxed as soon as there exists a direct measure of the latent shock of interest. That is, instead of relying on these restrictions, the identification is established through knowledge of the informativeness of a variable with respect to a subset of the underlying structural shocks. The variable can then either be included to \( y_t \) to derive the shock as an orthogonalised innovation or, alternatively, be treated as a proxy variable, as discussed in detail in the following subsection.

### 2.2.2 Identification of monetary policy shocks through proxy variables

Starting from Kuttner (2001) and followed by Faust et al. (2004), Gürkaynak et al. (2005), Gertler and Karadi (2015) and Nakamura and Steinsson (2018), amongst others, the high-frequency identification approach estimates the effects of monetary policy based on price changes in future contracts on Federal Funds and Eurodollar bonds within a narrow time window around the announcements by the central bank. In other words, the immediate reactions to the unexpected part of Federal Reserve’s statement reflect exogenous variation stemming from the non-systematic part of monetary policy.

Let \( f_t^{i+i} \) be the rate implied by an \( i \)-month-ahead interest rate future contract, expiring in month \( t + i \) and settled around a meeting held in month \( t \). Accordingly, a surprise due the monetary policy reads as

\[
\nu_t^{i+i} = f_t^{i+i} - f_{t-\tau}^{i+i} = g(\Omega_t) - E_t[\nu(\Omega_t)],
\]

(2.3)

where \( t - \tau \) is the time point shortly before the monetary policy announcement in month \( t \) has taken place, \( g(\cdot) \) is the monetary policy rule, \( E_t[\cdot] \) the conditional expectation of the public at \( t \) and \( \Omega_t \) denotes the information set of the central bank. The revision \( \nu_t^{i+i} \) in the market expectations then measures the monetary policy shock, the unanticipated deviation of the central bank from its known decision rule \( g(\cdot) \), provided that no other shocks systematically occur around those statements.

The proxy \( \nu_t^{i+i} \) can be used to derive the causal effects of monetary policy either by the local projection framework (Jordà, 2005) or by the structural VAR methodology. Prominently, the proxy SVAR (Stock and Watson, 2012; Mertens and Ravn, 2013) uses \( \nu_t^{i+i} \) as an external instrument to find the matrix \( B_y \) in (2.1). The covariance restrictions of the VAR residuals (2.2) are therein comple-
Identification of monetary policy shocks through proxies and non-normality

mented by moment conditions which require the proxy to be correlated with the structural shock of interest but uncorrelated with the remaining shocks of the economy. As an advantage, the approach is robust to measurement and missing values problems.

Applying the proxy SVAR, Gertler and Karadi (2015) identify a monetary policy shock from changes in the Federal funds rate and Eurodollar futures around the FOMC announcements. As most of the literature, they concentrate on one shock based on a single external instrument, selected among the set of futures by an F-test at the first stage of the identification. The use of a single proxy may be, however, particularly restrictive if the announcements of the central bank are mixtures of different types of shocks, and financial variables react to a variety of surprises concerning monetary policy. The shocks may relate not only to the conventional, unanticipated change of the current policy rate, but also to communication regarding the future path of interest rates, to unconventional measures and to information revealed by the central bank about economic prospects (See, e.g. Gürkaynak et al. 2005, Wright 2012, Nakamura and Steinsson 2018 and Campbell et al. 2012).

To extend the analysis, let the change (2.3) in the future for horizon $i$ be a compound of multiple shocks,

$$v_t^{i+1} = f_t^{i+1} - f_t^{i-1} = \phi_{i,1} e_{1,t} + \ldots + \phi_{i,l} e_{l,t} + s_i' \xi_t, \quad i = 0, \ldots, k - 1$$

(2.4)

where $s_i$ is a column vector of dimension $k$ and $\xi_t$ contains $k$ noise terms. The market reactions in $k$ futures $f_t^{i+1}, i = 0, \ldots, k - 1$, are then driven by $l$ structural shocks $e_t^1 = (e_{1,t}, \ldots, e_{l,t})'$ and $k$ noise terms. Consequently, using a single proxy would catch variation that consists of all policy disturbances occurring at the time of the announcement, violating the identifying assumption of the scheme.\(^7\)

The existence of multiple shocks driving the proxy variables requires thus refinement of the methodology. In principle, the proxy SVAR identification of Mertens and Ravn (2013) is able to extract the $l = k$ shocks in $e_t^1$ based on the same number of proxy variables. However, identifying $e_t^1$, i.e. the $l$ first columns of $B_y$, requires $l(l - 1)/2$ additional restrictions to be imposed on these columns. These restrictions would assume that the shocks differ in their timing how they propagate to the variables of the model, which may be difficult to justify. As the high-frequency proxies are strongly mutually

\(^7\)The proxy SVAR literature uses an F-test to evaluate the relevance of the instrument with respect to the reduced-form error. However, it does not test whether a given proxy is unrelated to the remaining structural shocks of the system.
2.2 Empirical strategy

correlated, they cannot be used in the identification one at a time for the recovery of multiple shocks, either.

In a linear proxy SVAR model, Arias et al. (2018) show that additional zero and sign restrictions are needed to distinguish the $k$ structural shocks of interest. Alternatively, Angelini and Fanelli (2018) propose a model that includes $y_t$ and $m_t$ and derive conditions under which a set of shocks can be identified. However, these conditions are sensitive as they depend on the number of variables, shocks and proxies, $n$, $l$ and $k$, as well as on the covariance structure between the proxies and the underlying shocks. On the other hand, assuming a lower-triangular impact matrix in the augmented SVAR would set on parameters $\{\phi_{i,1}, \ldots, \phi_{i,l}\}_{i=0}^{k-1}$ and $\{s_t\}_{i=0}^{k-1}$ in (2.4) undesirable zero restrictions that would ignore measurement errors and contemporaneous relations between the proxies.

In general, the need for prior assumptions regarding the short-run effects originates from the fact that the residual covariance restrictions (2.2) and the additional moment conditions obtained from the proxy SVAR exploit only the time-invariant second moments of data. This information, however, does not provide full identification of the matrix $B_y$. Therefore, the issue of partial identifiability is present as long as the SVAR model is estimated by least squares or under Gaussian errors.

In turn, once higher moments become relevant to capture features of data, the SVAR model (2.5) can be identified with no prior restrictions on the impact matrix $B_y$. Specifically, having structural shocks with at most one component Gaussian ensures that all elements of the impact matrix can uniquely be identified up to a sign and a permutation of the columns.\footnote{This result is established, amongst others, in Lanne et al. (2017) and Gouriéroux et al. (2017).} The result follows from statistical independence. Under Gaussianity, orthogonality of structural shocks directly implies independence, and learning from data is limited to the covariance restrictions (2.2). Under non-Gaussianity instead, data are able to provide further conditions by taking non-normality such as time-varying volatility, leptokurtosis and skewness into account.

What follows, the non-normality-based identification is complemented with proxy variables to facilitate the empirical relevance and the labelling of the estimated shocks. To implement the strategy, consider the joint dynamics of $y_t$
Identification of monetary policy shocks through proxies and non-normality

and $k$ proxies in $m_t = (v_t^1, \ldots, v_t^{t+k-1})'$,

$$
\begin{bmatrix}
    I_k \\
    0_{n,k}
\end{bmatrix}
\begin{bmatrix}
    m_t \\
    y_t
\end{bmatrix} = \bar{B}
\begin{bmatrix}
    \xi_t \\
    \xi_t
\end{bmatrix},
$$

(2.5)

or

$$
\bar{A}(L)\tilde{y}_t = \bar{B} \bar{\varepsilon}_t,
$$

(2.6)

where $A(L) = I_n - A_1 L - \ldots - A_p L^p$, $L$ the usual lag operator, $\bar{A}(L) = \text{diag}(I_k, A(L))$, $\bar{y}_t = (m_t', y_t')'$ and $\bar{\varepsilon}_t = (\varepsilon_1^t, \varepsilon_2^t, \xi_t') \sim (0, I_{\bar{n}})$ is a vector of dimension $\bar{n} = n + k$ containing the $l$ structural shocks of interest, the remaining $n - l$ shocks in $\varepsilon_2^t$ and $k$ measurement errors in $\xi_t$.

In the augmented SVAR model (2.5), the shocks and errors are linearly related to the proxy and economic variables through a $(\bar{n} \times \bar{n})$ impact matrix $\bar{B}$. Given $\bar{\varepsilon}_t$, the model (2.5) nests the original model (2.1) in its last $n$ equations, the last $n$ rows of matrix $\bar{B}$ being $[B_y \ 0_k]$. Under linearity, the model (2.5) also coincides with the proxy SVAR identification. In particular, let $l = k$ and the first $k$ rows of $\bar{B}$ be $[\Phi \ 0_{k,n-k} \ S]$, where $[\Phi]_{i,j} = \phi_{i,j}$, $i,j = 1, \ldots, k$, and $S$ stacks the row vectors $s'_i$, $i = 1, \ldots, k$. $m_t$ is an external instrument, as both the instrument relevance condition, $E[m_t \varepsilon_1^t] = \Phi$, and the instrument exogeneity condition, $E[m_t \varepsilon_2^t] = 0$, are satisfied.\(^9\)

Under non-normality, restrictions on $\bar{B}$ are no more necessary to identify the structural shocks of interest $\varepsilon_1^t$. Allowing for non-normality gives thus flexibility needed when the economic theory or the proxies do not bring enough information to discriminate between the structural shocks. To this end, I assume that the $i$th element of the vector $\bar{\varepsilon}_t$ is distributed as

$$
\bar{\varepsilon}_{i,t} = h_{i,t}^{-1/2}\eta_{i,t}, \eta_{i,t} \sim N(0,1), (\lambda_i - 2)h_{i,t} \sim \chi^2_{\lambda_i}, i = 1, \ldots, \bar{n}.
$$

(2.7)

Hence, the structural shocks are, unconditionally, independently Student’s $t$-distributed with unit variance and degrees of freedom parameter $\lambda_i > 2$, $i = 1, \ldots, \bar{n}$. Conditional on the stochastic volatility factor $h_{i,t}^{-1/2}$, the shocks are Gaussian. The independent and identically distributed stochastic factors

\(^9\)Compared to the model (2.5), the proxy SVAR allows the measurement errors to enter into equation (2.4) in a nonlinear manner. This situation is present, in particular, when dummy variables or discontinuous series are used as proxies, whereas here the movements of financial variables during the announcement days are continuous. Linearity is also usually assumed in the existing literature.
thus account for non-Gaussian features of data, and their existence provides further conditions to identify the structural shocks.\textsuperscript{10} That is, the higher moments $E[\bar{\varepsilon}_i, t^{q} \bar{\varepsilon}_j, t^{r}]$ for $i, j = 1, \ldots, \tilde{n}$ and $q + r \geq 3$ are no more determined by the first and second moments of data. It is also worth noting that large values of degrees of freedom $\lambda_i$ implies Gaussianity. Therefore, plausibility of the distributional assumption and the implied identifiability are empirical questions.\textsuperscript{11}

To gain intuition on the identification, combine the distributional assumption (2.7) with (2.4) and rewrite the high-frequency change $v_{t}^{i+i}$ due to a monetary policy surprise as

$$v_{t}^{i+i} = \sum_{j=1}^{l} \phi_{i,j} h_{i,t}^{-1/2} \eta_{j,t} + s'_{t} \xi_{t}, i = 0, \ldots, k - 1,$$

where $s'_{t} \xi_{t} = \sum_{j=0}^{k-1} s_{i,j} h_{n+j,t}^{1/2} \eta_{n+j,t}$. Under normality, $h_{i,t} = 1$ for all $j$ and $t$, and prior restrictions are needed either on the effects of $v_{t}^{i+i}$ or $y_{t}$. Under non-normality instead, the magnitude of impulse responses to different shocks, $h_{i,t}^{-1/2}, i = 1, \ldots, \tilde{n}$, varies independently over time. Hence, the public reacts to different unanticipated monetary policy actions under idiosyncratic volatility, by which additional information is available for identification.\textsuperscript{12} Suppose now $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are two structural shocks revealed at the FOMC meeting. When departing from normality, the following two higher-moment conditions, inter alia, become implicitly non-redundant. First, volatilities of the structural shocks are uncorrelated, $	ext{Cov}(\varepsilon_{1,t}^{2}, \varepsilon_{2,t}^{2}) = E[\eta_{1,t}^{2} h_{1,t}^{2} \eta_{2,t}^{2} h_{2,t}^{2}] - 1 = 0$,

\textsuperscript{10}Compared to using principle components from a larger set such as in Barakchian and Crowe (2013) to find the most relevant proxy, the non-normality facilitates the recovery of the shocks uniquely from the joint dynamics of macroeconomic variables and proxies.

\textsuperscript{11}The distribution used here is one distinct approach to ensure identifiability of $B_{y}$. As an advantage, the assumption provides convenient and computationally feasible estimation routines and parsimoniously models time-varying volatility. Statistical identification of structural shocks can also be achieved by unconditional (Rigobon, 2003) and conditional (Sentana and Fiorentini, 2001) heteroskedasticity, mixture-normal distribution (Lanne and Lütkepohl, 2010) or by a combination of heteroskedasticity and Student’s t distribution (Brunnermeier et al., 2017). In the Bayesian estimation, all these approaches would involve significantly more computational burden.

\textsuperscript{12}In addition, there is no persistence in those volatilities across time, i.e. between different meetings of the FOMC. This assumption is justifiable by the fact that the proxy series $v_{t}^{i+i}, i = 0, \ldots, k - 1$ are discontinuous as they only measure the high-frequency reactions at particular point of time.
Identification of monetary policy shocks through proxies and non-normality

which holds only if \(h_{1,t}\) and \(h_{2,t}\) are independent. Second, the shocks differ in terms of their kurtosis, as their fourth moments, \(E[\varepsilon_{i,t}] = (\lambda_i - 2)^{-1}(\lambda_i - 4)^{-1}\), \(i = 1, \ldots, \bar{n}\), depend on the shape of the distribution.

Conveniently, non-Gaussianity provides unique identification for all \(\bar{n}^2\) elements of \(\tilde{\beta}\) such that restrictions in (2.5) become unnecessary. Accordingly, the number of shocks affecting the proxies is unrestricted and can be inferred ex post at the labelling of the shocks. It should, however, be noted that the identification is only up to a sign and permutation of the columns. Following Lanne et al. (2017), I concentrate on a permutation that satisfies

\[
c_{ij} = [\tilde{B}]_{ij}, \quad |c_{ii}| > |c_{ij}| \quad \forall i < j, \quad (2.9)
\]

and

\[
b_{ii} = [\tilde{B}]_{ii} > 0, \quad \forall i = 1, \ldots, \bar{n} \quad (2.10)
\]

where \(\tilde{B} = D\hat{B}\) and \(D\) is a diagonal matrix by which the columns of \(\hat{B}\) have a unit Euclidean norm. In general, the permutation rule (2.9)–(2.10) imposes no prior restrictions on the impact of structural shocks but restricts the analysis on one of the \(\bar{n}!\) permutations of \(\hat{B}\).

As the matrix \(\hat{B}\) is uniquely identified regardless of the number of restrictions, it is the labelling of the shocks that determines their economic importance. Here, the latter is achieved by the proxies \(m_t\) that contain information about surprise actions of the central bank. Therefore, the monetary policy shocks can be identified as structural shocks that both explain the most of the variation in the included proxies and move policy rates on impact.

2.2.3 Estimation

The augmented SVAR model (2.5) is overidentified under non-normality in terms of its likelihood up to a permutation and sign of the impact matrix. Including a large number of proxies and variables of interest implies, however, a large parameter space, which can efficiently be handled by Bayesian methods. Moreover, the Bayesian approach allows to incorporate prior information on the contemporaneous relations of the variables. Next, I outline a Metropolis-within-Gibbs sampler to draw from the posterior distribution of the model parameters. Details of the algorithm are found in Appendix 2.A. The sampler, which builds upon Lanne and Luoto (2016) and Waggoner and Zha (2003), has two distinct features. First, drawing the parameters of the model exploits conditional normality arising from the distributional assump-
2.3 Results

As a special case, when the volatility terms $\{h_{i,t}\}_{T=1}^T$ are fixed and enough identifying restrictions are set, the algorithm estimates an overidentified Gaussian SVAR model as Waggoner and Zha (2003). Second, the rows of $\bar{B}^{-1}$ have a marginal posterior distribution from which it is possible to sample without a computationally demanding Metropolis-Hastings step.

I use standard prior distributions for the parameters of the model. First, free parameters related to the autoregressive terms in $\bar{A}(L)$ and to the contemporaneous relations in $\bar{B}^{-1}$ are assumed Gaussian a priori. Additionally, I restrict $\bar{B}$ to satisfy the permutation rule (2.9)–(2.10), as the same likelihood would be attained by any permutation of the rows of $\bar{B}^{-1}$. Last, for the degrees-of-freedom parameter $\lambda_i, i = 1, \ldots, \bar{n}$, I assume an exponential prior distribution.

Using the prior distribution and conditional likelihood, sampling from the posterior proceeds as follows. First, the posterior of the free parameters in $\bar{A}(L)$, given present draws of $\bar{B}^{-1}$ and $\{h_{i,t}\}_{T=1}^T, i = 1, \ldots, \bar{n}$, is multivariate normal. Second, drawing unrestricted parameters of $\bar{B}^{-1}$ from the posterior distribution involves a nonstandard distribution. Instead of a Metropolis-Hastings step, as in Canova and Pérez (2015), Brunnermeier et al. (2017) and Lanne and Luoto (2016), I follow the strategy of Waggoner and Zha (2003) who draw each row of $\bar{B}^{-1}$ separately given current draws of its remaining rows and other parameters of the model. The approach efficiently tackles multimodality, nonlinearity and tightness of the conditional posterior of $\bar{B}^{-1}$, which often leads to an excessive amount of rejections in Metropolis-Hastings algorithms.\(^{13}\) Now, drawing a row of $\bar{B}^{-1}$ is equivalent to drawing from a number of univariate normal distributions and one special distribution with the use of an orthonormal rotation. Third, the volatility factors $\{h_{i,t}\}_{i=1}^T$ are sampled from $\chi^2$-distributions for $i = 1, \ldots, \bar{n}$, and a Metropolis-Hastings step is used to obtain the posterior draws of the degrees-of-freedom parameters.

2.3 Results

This section studies the effects of monetary policy when the shocks are identified through proxies and non-normality. Compared to the proxy SVAR, the

\(^{13}\) The nonstandard algorithm is due to the likelihood of the SVAR model which does not belong to any known distribution family. For alternative techniques, for instance, Canova and Pérez (2015) and Lanne and Luoto (2016). The method of Waggoner and Zha (2003) remains valid as long as the prior distribution of the rows of $\bar{B}^{-1}$ are, apart from the permutation rule, Gaussian and independent a priori.
Identification of monetary policy shocks through proxies and non-normality

approach requires neither a first-stage test to select the strongest instrument for a structural shock nor additional exclusion restrictions to distinguish between the shocks of interest. Instead, all variables informative about potentially multiple shocks may simultaneously be used in the analysis. Eventually, the impulse responses determine, first, how the proxies react to the shocks and, second, the economic plausibility of the identified shocks. As a result, all surprise monetary actions occurring at the announcements of the Federal Reserve can be jointly identified from the high-frequency proxies and macroeconomic variables.

2.3.1 Data and estimation

I estimate the model using the following U.S. monthly data. My measures of real economic activity and prices are the logs of industrial production and consumer price index (CPI), respectively. I obtain the daily nominal 1-year and 10-year government bond yields from the Gürkaynak, Sack, and Wright (2007) database. The monthly average 1-year rate is used as the policy indicator, and I additionally construct a spread between the two interest rates, aggregated from daily frequency. The latter variable accounts for changes in the term structure. Following Gertler and Karadi (2015), I use the excess bond premium (EBP) of Gilchrist and Zakrajšek (2012) to measure credit conditions. This indicator captures variation in the aggregate credit risk related to the U.S. corporate bonds. To control for signalling effects of monetary policy, I include the log of the S&P 500 index.\footnote{The industrial production and CPI series are downloaded from the FRED data base. The excess bond premium is taken from the website of the Board of Governors of the Federal Reserve System. The stock price index is downloaded from Yahoo! Finance and aggregated to the monthly frequency.}

The analysis uses five proxies obtained from the dataset of Gertler and Karadi (2015) to extract the monetary policy shocks. They include the changes in the following futures in a 30-minute window around the FOMC meetings: the current and 3-month Federal Funds futures and the 6, 9 and 12-month futures on 3-month Eurodollar deposits. The data span the months 1990:1–2012:6.

The model with 12 lags is estimated using the Minnesota prior with overall and relative tightness parameters of 2 and 1, respectively, and a decay parameter set to 1. I assume that the mean of matrix $\vec{B}^{-1}$ is a diagonal matrix and each element has an independent normal distribution with standard deviation $10^3$, implying an uninformative prior distribution. The prior mean
2.3 Results

of the degrees of freedom parameters are set to 10, close to Gaussianity. The results presented are insensitive to less informative priors. Finally, I impose the matrix $\hat{A}(L)$ be as in (2.5).

Due to non-normality, $\bar{B}^{-1}$ could be left without any restrictions. However, the global identification of all structural shocks is beyond the scope of the essay. To shrink the number of parameters and identify only the shocks related to the proxies, I impose the matrix $\bar{B}$, or equivalently $\bar{B}^{-1}$, to be lower-block triangular,

$$
\bar{B} = \begin{bmatrix}
B_{11} & 0_{k,n} \\
B_{21} & B_{22}
\end{bmatrix},
$$

where $B_{11}$ and $B_{21}$ are $(k \times k)$ and $(n \times k)$ matrices with $k^2$ and $kn$ free parameters, respectively, and $B_{22}$ is lower triangular. This restricts that only $k = 5$ of the total $n + k = 11$ shocks affect the proxies on impact with no prior assumptions imposed. On the other hand, the lower-triangular matrix $B_{22}$ concerns the block containing remaining shocks of the economy, and they are recovered only as reduced-form-type errors. The restriction facilitates the computational burden by reducing the number of possible permutations, while allowing for a sufficient number of shocks and noise terms to affect the proxies. The results do not change if the restrictions on $\bar{B}^{-1}$ are relaxed.

As no zero restrictions are imposed on the first $k$ columns of the $\bar{B}$ matrix, their identification hinges upon the non-Gaussian assumption to distinguish between the shocks driving the proxies. To motivate the non-normal assumption, Figure 2.1 plots the kernel density estimates of the standardised proxy variables of the analysis. The density estimates suggest that large probability mass is concentrated on the neighbourhood of the means. In addition, the distributions of the proxies have fatter tails than implied by Gaussianity. Assuming that the proxies are linear combinations of Student’s t-distributed elements, the non-normality observed in Figure 2.1 is parsimoniously accounted

Figure 2.1: Kernel density estimates of the proxy variables
2.3.2 Proxies and monetary policy shocks

I start with the evaluation of how the proxies are affected by the identified shocks. This analysis sheds light on the relevance of the proxies and the character of the shocks. In Panel (a) of Table 2.1, I report the fractions by which the shocks contribute to the variance of each proxy, i.e. the forecast error variance decompositions. The contributions suggest that three shocks, \( \epsilon_{1,t} \), \( \epsilon_{2,t} \) and \( \epsilon_{3,t} \), emerge as driving the proxies the most. Among them, \( \epsilon_{1,t} \) accounts for the most variation in the current, 3-month and 6-month futures, whereas \( \epsilon_{3,t} \) mostly influences the movements in the longer-horizon Eurodollar futures. In contrast, the shock \( \epsilon_{2,t} \) explains approximately 10 percent of the movements at shorter and longer horizons.

To gain more interpretation, it is instructive to examine the impact effects of the shocks on the future rates, reported in Panel (b) of Table 2.1. The proxies react to the shocks \( \epsilon_{1,t} \) and \( \epsilon_{3,t} \) with uniform signs, as all coefficients apart from the current futures with respect to \( \epsilon_{3,t} \) are statistically significantly positive. In contrast, the responses to \( \epsilon_{2,t} \) have ambiguous signs. While the change of the current month future is significantly negative, the future rates at longer horizons increase. Hence, the expectations on the current policy rate are revised downwards in response to the shock, simultaneously with increasing slope of the term structure. Overall, the impulse responses to all three shocks are monotone in horizon: the reactions of \( \epsilon_{1,t} \) decrease in maturity, whereas the opposite is true for the shocks \( \epsilon_{2,t} \) and \( \epsilon_{3,t} \).

The weighting of the shocks reported in Table 2.1, produced from the statistical identification of the SVAR model, is similar to the factorisation employed by Gürkaynak et al. (2005), Campbell et al. (2012) and Barakchian and Crowe (2013). Notably, this literature has extracted “target” and “path” factors from the surprise movements, where the latter factor is orthogonal to the changes in the current Federal Funds futures. Instead of these two factors, the statistical identification recovers three structural shocks based on movements in both proxies and macroeconomic variables. Despite this methodological approach, it is essential to consider the non-normal distributional assumptions for the shocks.

15 In an extreme case, the non-normal densities in Figure 2.1 could have been generated by a single non-normal shock. Hence, the plots give only first-hand information on the distributional assumption. In turn, the low degrees-of-freedom parameters imply that also the underlying shocks are non-normal.

16 An examination based on the F-test of the proxy SVAR is proceeded in Gertler and Karadi (2015).
2.3 Results

<table>
<thead>
<tr>
<th>Panel (a)</th>
<th>Forecast error variance decomposition (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proxy: FF1</td>
<td>$\varepsilon_{1,t}$</td>
</tr>
<tr>
<td>Proxy: FF1</td>
<td>$\varepsilon_{2,t}$</td>
</tr>
<tr>
<td>Proxy: FF3</td>
<td>$\varepsilon_{3,t}$</td>
</tr>
<tr>
<td>Proxy: FF3</td>
<td>$\varepsilon_{4,t}$</td>
</tr>
<tr>
<td>Proxy: ED6</td>
<td>$\varepsilon_{5,t}$</td>
</tr>
<tr>
<td>Proxy: ED6</td>
<td>$\varepsilon_{6,t}$</td>
</tr>
<tr>
<td>Proxy: ED9</td>
<td>$\varepsilon_{7,t}$</td>
</tr>
<tr>
<td>Proxy: ED9</td>
<td>$\varepsilon_{8,t}$</td>
</tr>
<tr>
<td>Proxy: ED12</td>
<td>$\varepsilon_{9,t}$</td>
</tr>
<tr>
<td>Proxy: ED12</td>
<td>$\varepsilon_{10,t}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel (b)</th>
<th>Impact effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proxy: FF1</td>
<td>$\lambda_{1}$</td>
</tr>
<tr>
<td>Proxy: FF1</td>
<td>$\lambda_{2}$</td>
</tr>
</tbody>
</table>

Table 2.1: Forecast error variance decompositions and impact effect of the identified shocks on the proxies.

The table reports the forecast error variance decomposition and the impact effects of the five proxies to the identified shocks. Posterior medians reported and 90 % credible sets shown in parentheses. Proxies: the current (FF1) and 3-month (FF3) Federal funds futures, 6-month (ED6), 9-month (ED9) and 12-month (ED12) futures on Eurodollar deposit contracts. The last row reports the posterior median and credible sets of the degrees-of-freedom parameters.
difference, $\varepsilon_{1,t}$ resembles the target factor by the short-run future changes attaining the highest loadings. Likewise, $\varepsilon_{3,t}$ is potentially connected with the path factor.

In addition, Table 2.1 reports statistics on the two remaining shocks that affect the proxies, $\varepsilon_{4,t}$ and $\varepsilon_{5,t}$, and on the shock distribution. First, by the forecast error variance decomposition and the impact effects, $\varepsilon_{4,t}$ and $\varepsilon_{5,t}$ can be regarded as negligible noise terms: they contribute less to the variance of the proxies and their coefficients are to a large part close to zero. As the relevant variation in the future rates are therefore driven by three out of five shocks, the zeros assumed in the upper-right block of $\bar{B}^{-1}$ are likely non-restrictive. Second, according to Table 2.1, the estimates of the degrees-of-freedom parameters are low, which indicates non-normality of the shocks and supports the validity of the identification scheme.\textsuperscript{17}

2.3.3 The effects of monetary policy shocks

Figure 2.2 plots the periodwise posterior medians of the impulse responses to the three relevant shocks, $\varepsilon_{1,t}$, $\varepsilon_{2,t}$, and $\varepsilon_{3,t}$, and the 90 and 68 percent credible sets. These one standard deviation shocks induce different responses of the interest rates. First, in response to the shock $\varepsilon_{1,t}$ in Panel (a), the 1-year rate increases and the term spread decreases with the same magnitude, implying that the 10-year rate does not move on impact. Second, the shock $\varepsilon_{2,t}$ in Panel (b) prompts a drop in the 1-year rate and a jump of a greater magnitude in the term spread, which together imply an increase of the 10-year rate. Last, the shock $\varepsilon_{3,t}$ in Panel (c) affects, positively but statistically insignificantly, only the term spread.\textsuperscript{18}

All the three shocks have adverse effects on the financial market and the macroeconomy. Credit costs rise due to a positive shift in the excess bond premium, and the stock price index decreases. The shocks also imply a gradual decline of industrial production. This effect is strongest for the shock $\varepsilon_{2,t}$, whereas the index remains positive for the first months after a shock $\varepsilon_{1,t}$ and falls afterwards only temporarily. In turn, the shocks have asymmetric effects on prices. After a shock $\varepsilon_{1,t}$, the price level remains persistently low for the

\textsuperscript{17}It is noteworthy to mention that the degrees-of-freedom parameters need not be distinct for the identification be valid, as opposed to the heteroskedasticity-based strategies, where the shock variances are required to vary. It is the permutation rule (2.9)–(2.10) that eventually discriminates between the non-normal shocks.

\textsuperscript{18}With the shadow rate of Wu and Xia (2016) in the model, the results remain the same. The movements in the 1-year rate and the shadow rate are broadly similar.
2.3 Results

![Impulse response graphs](image)

Figure 2.2: Impulse responses to the three candidate monetary policy shocks
Each panel plots in solid lines the periodwise posterior median impulse responses to a one standard deviation shock. The light and dark grey areas border the 90 and 68 percent periodwise credible sets, respectively.

subsequent periods. The shock $\epsilon_{3,t}$ has a similar deflationary effect, whereas prices remain constant in response to $\epsilon_{2,t}$.

By the impulse responses of Figure 2.2, it is possible to label the three shocks that drive reactions to the monetary policy announcements. First, the shock $\epsilon_{1,t}$, associated predominantly with movements in the short-run Federal funds futures, increases the short-run rate on impact while stock prices react negatively. Given its adverse effects on the macroeconomy, it can be labelled as a conventional contractionary monetary policy shock.

Second, the remaining variation in the short-run futures are explained by the shock $\epsilon_{2,t}$. The shock decreases the 1-year yield and is associated with negatively moving industrial production and non-moving prices, inconsistent with general conclusions about a monetary expansion. Concluding from Table 2.1, the shock also induces positive reactions of the longest-maturity futures
I dentification of monetary policy shocks through proxies and non-normality

Figure 2.3: The shocks driving the proxies over time

Posterior medians of the shocks $\varepsilon_{1,t}$, $\varepsilon_{2,t}$, $\varepsilon_{3,t}$ over time.

despite decreasing the short-run rate on impact. Importantly, the private sector reacts negatively to a reduction in the short-run rate, as the stock price index falls on impact. The decreasing interest rate is thus likely to reflect an unexpected response of the central bank to the unfavourable economic outlook.

Finally, the shock $\varepsilon_{3,t}$ is primarily related to the risk in the economy, observed as rising excess bond premium and declining stock prices under no significant interest rate responses. Nonetheless, given its large contribution to the variation of the futures on the longer horizon and its negative impact on the economic activity and prices, the shock is likely linked to central bank communication.

Figure 2.3 plots the evolution of the three shocks over time. The conventional monetary policy shock $\varepsilon_{1,t}$ as well as the shock $\varepsilon_{2,t}$ gain much of its variation before the zero-lower-bound period. Especially, positive values of
2.3 Results

$\epsilon_{2,t}$ take place, i.e. the shock induces reductions in the short-run rates, at the surge of the financial crisis. Simultaneously, $\epsilon_{1,t}$ induces large surprise falls of the interest rate. None of the two shocks enters the economy, however, after the Federal funds rate reached the zero lower bound in December 2008. Compared to the two latter shocks, $\epsilon_{3,t}$ occurs constantly over time. That is, policy surprises both before and during the zero-lower-bound period contain the component related to the long-run future rates.

2.3.4 Discussion

Due to the existence of multiple shocks, statements of the central bank contain information beyond a conventional short-run interest rate change. Specifically, the monetary policy surprises can be decomposed to three factors: to a contractionary short-run interest rate shock, $\epsilon_{1,t}$, to changes in the short-run rate due to the present and future economic conditions, $\epsilon_{2,t}$, and to the long-run shock $\epsilon_{3,t}$ that affects mainly the risk premium.

As mentioned above, a concept related to the changes in the longer-run future rates, captured by the shocks $\epsilon_{2,t}$ and $\epsilon_{3,t}$, is the path factor, identified from the range of future rate changes around the central bank announcements. Gürkaynak et al. (2005) interpret it to reflect the future interest rate path announced in the statements. Campbell et al. (2012) define two distinct types of communication contained in the factor. First, Odyssean forward guidance commits the FOMC to future actions. Second, language regarding Delphic forward guidance forecasts macroeconomic outcomes, and interest rate hikes may then change expectations of the public about the state of the economy. According to their results, private forecasts about future economic activity are revised upwards after this path shock. In a similar fashion, Nakamura and Steinsson (2018) argue that the FOMC statements affect the beliefs of economic agents about economic fundamentals due to the superior information set of the Federal Reserve.

Monetary policy may also affect corporate credit spreads and long-run rates by more than the expectations hypothesis would imply. Hanson and Stein (2015), Gertler and Karadi (2015) and Caldara and Herbst (2018) highlight that a surprise monetary policy shock is followed by tightening of the credit market – seen as rising credit spreads and increasing long-term rates. Accordingly, monetary policy transmits to the financial market through risk and term premia (Gertler and Karadi, 2015; Hanson and Stein, 2015), potentially due to financial frictions and incomplete credit markets. Furthermore, Caldara and Herbst (2018) emphasise the endogenous response of the cen-
Identification of monetary policy shocks through proxies and non-normality

central bank to credit conditions: ignoring the credit spreads in the model distorts conclusions about the effects of monetary policy. However, these studies employ identification strategies based on a single variable, which may be a compound of the identified shocks $\varepsilon_{1,t}$, $\varepsilon_{2,t}$ and $\varepsilon_{3,t}$.

Instead, the present non-normal identification extracts shocks from a broader range of proxies, combined with variation in macroeconomic aggregates. The weighting of these proxies produces a conventional monetary policy shock $\varepsilon_{1,t}$ that induces effects usually observed in the monetary SVAR literature, reviewed by Ramey (2016). In particular, an interest rate hike identified through the high-frequency future rate changes produces now a significant reduction in prices following a conventional monetary policy shock, as opposed to Gertler and Karadi (2015), Ramey (2016) and Caldara and Herbst (2018) who find only modest price responses. This finding may be due to the fact that the identification is able to purge out of the conventional monetary policy shock factors related to other types of central bank communication, i.e. to the shocks $\varepsilon_{2,t}$ and $\varepsilon_{3,t}$.

In effect, the changes in the short-run futures are not due to the $\varepsilon_{1,t}$ only. Instead, they are affected by the shock $\varepsilon_{2,t}$ that triggers a fall of the 1-year yield and an increase of the longer-run future rates. Despite lowering the interest rate, the shock has adverse effects on the economy, seen as declining stock prices, increasing excess bond premium and falling industrial production. Therefore, its impulse responses unlikely measure the causal effects of conventional monetary policy. Rather, the shock is Delphic, reflecting an endogenous but unanticipated response of monetary policy to adverse economic circumstances. The public thus learns about the state of the economy through statements by the Federal Reserve, similar to the information effect of Nakamura and Steinsson (2018).

Additionally, the third surprise component due to the monetary policy announcements, the long-run shock $\varepsilon_{3,t}$, explains a large fraction of the 6-, 9- and 12-month futures. Its impact on the future rates is similar to the path factor of Gürkaynak et al. (2005). However, no significant response of the long-run interest rate can be seen, as the term spread reacts only mildly positively. Given that the excess bond premium signals investors’ risk appetite and is informative about likelihood of recessions in the medium term (Gilchrist and Zakrajšek, 2012), the shock can be interpreted to influence the expectations of the private sector about risk, economic prospects and future actions similar to forward guidance. Importantly, since its existence is not restricted to the conventional monetary policy period, the shock transmits through a channel other than a policy rate change.
2.3 Results

Table 2.2: Correlation of the identified shocks with different shock measures

<table>
<thead>
<tr>
<th>Series</th>
<th>$\epsilon_{1,t}$</th>
<th>$\epsilon_{2,t}$</th>
<th>$\epsilon_{3,t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>R&amp;R shock</td>
<td>0.26</td>
<td>-0.04</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>[0.25 ; 0.27]</td>
<td>[-0.07 ; -0.02]</td>
<td>[-0.01 ; 0.04]</td>
</tr>
<tr>
<td>N&amp;S FFR shock</td>
<td>0.57</td>
<td>-0.29</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>[0.56 ; 0.58]</td>
<td>[-0.34 ; -0.24]</td>
<td>[0.09 ; 0.19]</td>
</tr>
<tr>
<td>N&amp;S policy news shock</td>
<td>0.52</td>
<td>-0.03</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>[0.51 ; 0.53]</td>
<td>[-0.08 ; 0.03]</td>
<td>[0.28 ; 0.39]</td>
</tr>
</tbody>
</table>

Correlation of the shocks in posterior medians with the monetary policy shock of Romer and Romer (2004) (R&R) as well as with the Federal funds rate (FFR) and the policy news shock of Nakamura and Steinsson (2018) (N&S). Posterior medians reported and 90% credible sets shown in parentheses. The R&R shock is the series based on the updates of Wieland and Yang (2016) and covers the months 1990:1–2007:12, the both N&S series span the months 1995:2–2012:6.

Next, I examine the relation of the identified shocks to three prominent series that have been used to identify the effects of monetary policy. In Table 2.2, I report their correlations with the three identified shocks. The shock series of Romer and Romer (2004), which measures non-systematic interest rate changes based on the records and internal forecasts of the Federal Reserve, is positively correlated with the conventional monetary policy shock $\epsilon_{1,t}$. Moreover, the correlation between the other two shocks, $\epsilon_{2,t}$ and $\epsilon_{3,t}$ is virtually zero. Hence, the estimated conventional monetary policy shock is consistent with the measure that captures non-systematic variation in the policy rule based on external sources.

Furthermore, the three shocks are related to the high frequency measures of Nakamura and Steinsson (2018), the Federal funds rate (FFR) and the policy news shocks. The former series consists of changes in the current futures, whereas the latter shock is closely related to the path factor of Gürkaynak et al. (2005). Specifically, Nakamura and Steinsson (2018) show that the policy news shock induces strong information effects, leading to the public to update its belief about the state of the economy. According to Table 2.2, the both series are strongly correlated with the monetary policy shock $\epsilon_{1,t}$, due to the fact that they weight the short-run futures similarly to $\epsilon_{2,t}$. For the same reason, the correlation of the FFR shock with $\epsilon_{2,t}$ is negative, consistent with the negative impulse response of the current month future rate shown in Table 2.1. In turn, the long-run shock $\epsilon_{3,t}$ correlates with the policy news shock – in contrast to $\epsilon_{2,t}$. In light of this evidence, $\epsilon_{3,t}$ is a particular combination of information and risk premium effects, operating in a dimension orthogonal to the current...
short-run interest rate.

To investigate the nature of the shocks $\varepsilon_{2,t}$ and $\varepsilon_{3,t}$ more concretely, Table 2.3 lists their five largest occurrences since July 1995 when the FOMC started to regularly announce statements.\(^{19}\) The table reports the size of the shock, the relevant FOMC meeting, the decided target rate and its change as well as the change of the S&P 500 index on the meeting day.\(^{20}\) Furthermore, in the last column, I compile from the statements relevant language that potentially characterises the shock and its sign. As this compilation does not rely on any quantitative methodology, the examination of relevant language is rather illustrative.

Broadly, the labelling of $\varepsilon_{2,t}$ is in line with the statistics in Panel (a) of Table 2.3. The shock is followed by a stock price response of opposite sign in all except one meeting.\(^{21}\) In line with the impulse responses, the positive shocks are associated with interest rate cuts. The contrary is, however, not the case for the shock due to the meeting of 18 April 2001, potentially caused by a simultaneous surprise interest rate cut and positive manufacturing data announced on the same day. In terms of relevant language, all five statements emphasise the state of the economy in the medium run, for which the change of the Federal funds rate target is justified.

Similar to $\varepsilon_{2,t}$, the largest values of the long-term shock $\varepsilon_{3,t}$ are associated with stock price reactions of opposite signs on the relevant FOMC meeting day. The shocks are, however, not necessarily linked to changes in the Federal funds rate. The statements rather contain language that concerns the policy horizon or the use of multiple instruments to sustain price stability and maintain economic growth. Moreover, the relevant meeting days coincide with the largest observations of the path factor of Gürkaynak et al. (2005). In this respect, the shock potentially influences the behaviour of economic agents through information about medium-run policy actions and economic outcomes, leading to changes in prices, production and financial risk similarly to the conventional monetary policy shock.

Overall, the approach based on non-normality and proxies gives insight into the identification of monetary policy shocks, where the measured mon-

\[^{19}\]For the shock $\varepsilon_{1,t}$, the corresponding five months are May 2001, February 2008, January 2001, April 2001 and November 2001.

\[^{20}\]The aggregation of the proxy series due to the use of the average monthly interest rate implies that the relevant meeting may be the one of the previous month. See footnote 11 of Gertler and Karadi (2015) for details.

\[^{21}\]In January 2001, Fed announced a surprise interest rate cut, also observed as large value of $\varepsilon_{1,t}$. As a result, stock prices responded on the FOMC day strongly positively.
Panel (a) The largest values of the shock $\varepsilon_{2,t}$

<table>
<thead>
<tr>
<th>Month</th>
<th>Shock</th>
<th>Relevant meeting</th>
<th>FFR target / FFR change (%)</th>
<th>Stock price change (%)</th>
<th>Relevant language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feb 2008</td>
<td>5.2</td>
<td>30 Jan</td>
<td>3.0 / -0.5</td>
<td>-0.48</td>
<td>&quot;...downside risks to growth remain.&quot;</td>
</tr>
<tr>
<td>Jan 2001</td>
<td>4.8</td>
<td>3 Jan</td>
<td>6.0 / -0.5</td>
<td>5.0</td>
<td>&quot;...the risks are weighted mainly toward conditions that may generate economic weakness in the foreseeable future.&quot;</td>
</tr>
<tr>
<td>Oct 2008</td>
<td>2.7</td>
<td>8 Oct</td>
<td>1.5 / -0.5</td>
<td>-1.1</td>
<td>&quot;...the pace of economic activity has slowed markedly in recent months. [...] the intensification of financial market turmoil is likely to exert additional restraint on spending. [...] the Committee believes that the decline in energy and other commodity prices and the weaker prospects for economic activity have reduced the upside risks to inflation.&quot;</td>
</tr>
<tr>
<td>Apr 2001</td>
<td>-2.7</td>
<td>18 April</td>
<td>4.5 / -0.5</td>
<td>3.9</td>
<td>&quot;...Appreciable downside risks to growth remain.&quot;</td>
</tr>
<tr>
<td>Jan 2008</td>
<td>2.5</td>
<td>22 Jan</td>
<td>3.5 / -0.75</td>
<td>-1.1</td>
<td></td>
</tr>
</tbody>
</table>

Panel (b) The largest values of the shock $\varepsilon_{3,t}$

<table>
<thead>
<tr>
<th>Month</th>
<th>Shock</th>
<th>Relevant meeting</th>
<th>FFR target / FFR change (%)</th>
<th>Stock price change (%)</th>
<th>Relevant language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feb 2004</td>
<td>-4.0</td>
<td>28 Jan</td>
<td>1 / 0</td>
<td>-1.4</td>
<td>&quot;...the Committee believes that it can be patient [replacing “considerable period” of the previous statement] in removing its policy accommodation.&quot;</td>
</tr>
<tr>
<td>Jul 1995</td>
<td>-3.6</td>
<td>6 Jul</td>
<td>5.75 / -0.25</td>
<td>1.2</td>
<td>&quot;...the Federal Open Market Committee decided to decrease slightly the degree of pressure on bank reserve positions [...] inflationary pressures have receded enough to accommodate a modest adjustment in monetary conditions.&quot;</td>
</tr>
<tr>
<td>Nov 2003</td>
<td>-2.6</td>
<td>28 Oct</td>
<td>1 / 0</td>
<td>1.5</td>
<td>&quot;...the Committee believes that policy accommodation can be maintained for a considerable period.”</td>
</tr>
<tr>
<td>Oct 1998</td>
<td>-2.5</td>
<td>15 Oct</td>
<td>5 / -0.25</td>
<td>4.2</td>
<td>&quot;...further easing of the stance of monetary policy was judged to be warranted to sustain economic growth in the context of contained inflation.”</td>
</tr>
<tr>
<td>Dec 2008</td>
<td>-2.2</td>
<td>16 Dec</td>
<td>0-0.25 / -1</td>
<td>5.1</td>
<td>&quot;The Federal Reserve will employ all available tools to promote the resumption of sustainable economic growth and to preserve price stability.”</td>
</tr>
</tbody>
</table>

Table 2.3: The five most significant incidents of $\varepsilon_{2,t}$ and $\varepsilon_{3,t}$ since July 1995
Table reports the posterior medians of the shock, the relevant FOMC meeting, the change in the Federal Funds target rate, percentage change of the S&P 500 index on the meeting day and shock-relevant language taken from the statement of the Federal Reserve. Source: Board of Governors of the Federal Reserve System (2018) and Campbell et al. (2012).
etary tightening often implies counterfactually expansionary effects (Ramey, 2016). Here, the monetary policy announcements are decomposed into three shocks that have different effects on the interest rates and the economy. The results show that the short-run futures are driven not only by non-systematic monetary policy component $\varepsilon_{1,t}$ but also, to a smaller extent, by the shock $\varepsilon_{2,t}$ which reveals information about the state of the economy. Neglecting $\varepsilon_{2,t}$ and using a single short-run future as a proxy imply that the surprise interest rate changes have not fully been cleaned from the information effect. It should, however, be noted that the size of the information shock is considerably smaller than suggested by Nakamura and Steinsson (2018).

Moreover, a large portion of variation around the FOMC meeting is contributed by the long-term shock $\varepsilon_{3,t}$. The emergence of the shock is independent of the zero lower bound, i.e. it occurs both under conventional and unconventional monetary policy periods. The shock shows effects in the risk premium but leaves the short-run interest rate intact and is closely related to the path factor of Gürkaynak et al. (2005). The existence of the shock may be due to the ability of the central bank to affect the expectations of the public (Cochrane and Piazzesi, 2002; Gürkaynak et al., 2005) or due to financial frictions (Gertler and Karadi, 2015; Hanson and Stein, 2015). By implementing this type of forward guidance reflected in the shock, the central bank is able to affect production and prices similarly to the interest rate change.

2.4 Conclusions

This chapter investigated the identification of monetary policy shocks in a framework that exploited both variation around the announcements of the central bank and statistical properties of data. The proxies were used to identify movements relevant for the monetary surprises; assuming non-normality of data facilitated the discrimination between multiple shocks driving the announcements. Unlike in the previous studies, the macroeconomic effects of various monetary policy announcements could simultaneously be estimated without a priori knowledge about the propagation of the shocks.

I used a general estimation routine to measure the impact of the monetary shocks identified from various high-frequency reactions to the statements of the Federal Reserve. The results showed that the announcements are driven by three distinct shocks: a conventional interest rate shock, an information shock that reveals central bank’s endogenous but unanticipated reactions to the current state of the economy and a long-term shock. Importantly, the long-
term shock affects the risk and the expectations in the economy by instruments different from the short-run interest rate. Nonetheless, the macroeconomic effects of such a policy surprise and a conventional interest rate change are analogous.

The existence of a monetary policy shock that affects the risk premium before and after the financial crisis of 2008 is beyond the implications usually inferred from theoretical macroeconomic models. Nonetheless, the shock may relate to the ability of the central bank to affect the medium-term expectations by informing the public about the future policy actions, similar to forward guidance. The central bank may also influence the risk premium under incomplete information or financial frictions.

Finally, it is worth emphasising that the SVAR model and framework of this essay are readily available for the identification of other than monetary policy shocks. In particular, as soon as there exist proxy variables that may be noisy measures of the latent shocks and data suggest non-normality, the model is able to recover all relevant shocks and their impulse responses.

References


Identification of monetary policy shocks through proxies and non-normality


Appendix

2.A Derivation of the Gibbs sampler

Here, I present details on the Gibbs sampler outlined in subsection 2.2.3. As a starting point for the derivation of the posterior distribution, multiply the SVAR model (2.5) by the inverse of $\bar{B}$ to represent it equivalently as a system of simultaneous equations, including a $(\bar{n} \times 1)$ constant $a$

$$\bar{B}^{-1}(\bar{y}_t - \bar{A}_1\bar{y}_{t-1} - \ldots - \bar{A}_p\bar{y}_{t-p} - a) = H_t^{-1/2}\eta_t, \ t = 1, \ldots, T$$

(2.12)

with a conditional likelihood function

$$p(y|\bar{A}, \bar{B}^{-1}, H) = (2\pi)^{-T\bar{n}/2}|\det(\bar{B}^{-1})|^{T}\prod_{t=1}^{T}|H_t|^{1/2}\exp\left(-\frac{1}{2}\sum_{t=1}^{T}\bar{u}_t'\bar{B}^{-1}\bar{H}_t\bar{B}^{-1}\bar{u}_t\right),$$

(2.13)

where $y = (y'_1, \ldots, y'_T)'$, $\bar{A}' = [a \bar{A}_1 \cdots \bar{A}_p]'$, $H_t = \text{diag}(h_{1,t}, \ldots, h_{\bar{n},t})$, $H = \text{diag}(h_{1,T}, \ldots, h_{\bar{n},T})$ and $\bar{u}_t = \bar{A}(L)\bar{y}_t - a$.

Subsequently, define column vectors $\alpha = \text{vec}(\bar{A})$ and $b_i$, $i = 1, \ldots, \bar{n}$, where the latter corresponds to the $i$th row of matrix $\bar{B}^{-1}$, i.e. $\bar{B}^{-1} = [b_1 \cdots b_{\bar{n}}]$. Let these vectors be linearly related to vectors $\alpha_r ((\bar{n}^2p + \bar{n} - r_\alpha) \times 1)$ and $b_{r,i} ((\bar{n} - r_{b,i}) \times 1)$ through matrices $R_\alpha ((\bar{n}^2p + \bar{n} - R_\alpha) \times (\bar{n}^2p + \bar{n} - \bar{n}))$ and $R_{b,i} (\bar{n} \times q_i)$, $q_i = \bar{n} - r_{b,i}$:

$$\alpha = R_\alpha\alpha_r$$

(2.14)

$$b_i = R_{b,i}b_{r,i}, \ i = 1, \ldots, \bar{n},$$

(2.15)

where $r_\alpha$ and $r_{b,i}$ are the number of restricted parameters in $\bar{A}$ and in the $i$th row of $\bar{B}^{-1}$, respectively. Accordingly, the vectors $\alpha_r$ and $b_{r,i}$ contain the free parameters of the model and are mapped by matrices $R_\alpha$ and $R_{b,i}$, whose elements are either 1 or 0, to $\bar{A}$ and $\bar{B}^{-1}$.

I set the prior distribution for the parameters as follows. First, as standard in the Bayesian VAR literature, the autoregressive parameters in $\alpha_r$ are,

---

\[ \text{The restrictions are set to the inverse of } \bar{B}. \] However, imposing restrictions of a block-triangular form such as in (2.5) is straightforward.
Identification of monetary policy shocks through proxies and non-normality

A priori, normally distributed as $\alpha_r \sim N(\bar{\alpha}_r, V_{\alpha_r})$. Following Sims and Zha (1998) and Waggoner and Zha (2003), the free parameters in the rows of $\bar{B}^{-1}$ have a prior normal distribution $b_{r,i} \sim N(b_{r,i}, V_{b_{r,i}})I(\bar{B}^{-1})$, where $I(\bar{B}^{-1})$ is an indicator function equal to one if $\bar{B}$ satisfies the permutation rule (2.9)–(2.10). The degrees-of-freedom parameter $\lambda_i$ has exponential prior distribution $\lambda_i \sim \text{Exp}(\bar{\lambda}_i)$, $i = 1, \ldots, \bar{n}$ and truncated to be greater than 2.

Given $y$, $\bar{B}^{-1}$ and $H$ and combining the likelihood (2.13) and the prior, the conditional posterior of $\alpha_r$ reads as

$$p(\alpha_r | y, \bar{B}^{-1}, H) \propto \exp \left( -\frac{1}{2} (\alpha_r - \bar{\alpha}_r)' \bar{V}_{\alpha_r}^{-1} (\alpha_r - \bar{\alpha}_r) \right)$$ \hspace{1cm} (2.16)

where

$$\bar{V}_{\alpha_r}^{-1} = X_r' \Omega^{-1} X_r + V_{\alpha_r}^{-1}$$ \hspace{1cm} (2.17)

$$\bar{\alpha}_r = \bar{V}_{\alpha_r} \left( X_r' \Omega^{-1} y + V_{\alpha_r}^{-1} \bar{\alpha}_r \right)$$ \hspace{1cm} (2.18)

and

$$\Omega^{-1} = (I_T \otimes \bar{B}^{-1'}) H (I_T \otimes \bar{B}^{-1})$$ \hspace{1cm} (2.19)

$$X_r = XR_{\alpha}$$ \hspace{1cm} (2.20)

$$X = (X_1', \ldots X_T')'$$ \hspace{1cm} (2.21)

$$X_t = I_n \otimes (1, y_{t-1}', \ldots, y_{t-p}').$$ \hspace{1cm} (2.22)

Hence, $\alpha_r$ is conditionally drawn from a multivariate normal distribution with mean $\bar{\alpha}_r$ and variance $\bar{V}_{\alpha_r}$.

On the other hand, the $i$th row of $\bar{B}^{-1}$ can be drawn from a distribution conditional on the remaining rows of $\bar{B}^{-1}$, $\bar{A}$ and $H$. The strategy follows Waggoner and Zha (2003) but with a difference that no further restrictions are needed to identify $\bar{B}^{-1}$. The likelihood (2.13) can be written, conditional on $\bar{A}$, $\bar{B}^{-1}_{-i} = (b_1, \ldots, b_{i-1}, 0, b_{i+1}, \ldots, b_\bar{n})'$ and $H$, as

$$p(y | \bar{A}, \bar{B}^{-1}_{-i}, H) \propto \left| \det(\bar{B}^{-1}) \right|^T \exp \left( -\frac{1}{2} b_{r,i}' R_{b,i} \Psi_{b,i} R_{b,i} b_{r,i} \right),$$ \hspace{1cm} (2.23)
where $\Psi_{u,i} = \sum_{t=1}^{T} \tilde{u}_{i} h_{i,t} \tilde{u}_{i}$. Combining with the prior distribution yields

$$p(b_{r,i}|y, \bar{A}, \bar{B}_{-i}^{-1}, H) = |\det(\bar{B}^{-1})|^T \exp \left( -\frac{T}{2} (b_{r,i} - \bar{b}_{r,i})' \bar{V}_{b,i} (b_{r,i} - \bar{b}_{r,i}) \right) I(\bar{B}^{-1}),$$

(2.24)

where $\bar{V}_{b,i}^{-1} = \frac{1}{T} \left( \bar{V}_{b,i}^{-1} + R_{b,i} \Psi_{u,i} R_{b,i} \right)$. As shown by Waggoner and Zha (2003), drawing from (2.24) is equivalent to drawing from a number of univariate normal distributions and one special distribution. In particular, let $w$ be an $(n \times 1)$ vector orthogonal to matrix $\bar{B}_{-i}$ and define a $(q_i \times 1)$ vector $v_1 = C_i R_{b,i} w / ||C_i R_{b,i} w||$, where $C_i C_i' = \bar{V}_{b,i}^{-1}$. Furthermore, by forming an orthonormal basis $(v_1, \ldots, v_{q_i})$, $b_{r,i}$ is equal to

$$b_{r,i} = C_i \sum_{j=1}^{q_i} \beta_j.$$  

(2.25)

As $b_{r,i}$ is a linear function of $\beta_j, j = 1, \ldots, q_i$, drawing $b_{r,i}$ is equivalent to drawing from conditional distribution (See, Waggoner and Zha, 2003; Villani, 2009)

$$p(\beta_1, \ldots, \beta_{q_i}|y, \bar{A}, \bar{B}_{-i}^{-1}, H) \propto |\beta_1|^T \exp \left( -\frac{T}{2} (\beta_1 - \bar{\beta}_1)^2 \right)$$

$$\exp \left( -\frac{T}{2} \sum_{j=2}^{q_i} (\beta_j - \bar{\beta}_j)^2 \right) I(\bar{B}^{-1})$$

(2.26)

with $\bar{\beta}_j = v_j' C_i^{-1} \bar{b}_{r,i}$. According to (2.26), $\beta_2, \ldots, \beta_{q_i}$ are conditionally normally distributed with mean $\bar{\beta}_j$. On the other hand, the kernel for $\beta_1$ is non-standard due to the term $|\beta_1|^T$. Nonetheless, following Villani (2009), the distribution of $\beta_1$ can be approximated by a mixture normal distribution

$$f(\beta_1|y, \bar{A}, \bar{B}_{-i}^{-1}, H) \approx p_1 N(\mu_1, \sigma_1^2) + (1 - p_1) N(\mu_2, \sigma_2^2)$$

(2.27)

with $\mu_1 = \frac{\bar{\beta}_1^2}{2} - \frac{1}{2} \sqrt{\bar{\beta}_1^2 + 4}$, $\mu_2 = \frac{\beta_1^2}{2} + \frac{1}{2} \sqrt{\beta_1^2 + 4}$, $\sigma_1^2 = \frac{1}{T} \frac{\mu_1^2}{1 + \rho_1}$ and $p_1 = \frac{1}{1 + e^{\rho_1 T}}$. The algorithm is run for $i = 1, \ldots, \tilde{n}$ such that for each $i$, $b_{r,i}$ is backed out from $\beta_1, \ldots, \beta_{q_i}$ using (2.25). Additionally, at each iteration, only draws that belong to a permutation (2.9)–(2.10) are accepted.

To draw $H$, the hierarchical prior for $\lambda_i h_{i,t}$ is $\lambda_i^2$ is combined by the likeli-
Identification of monetary policy shocks through proxies and non-normality

hood (2.13) to obtain

\[ p\left(h_{i,t} | y, \bar{A}, \bar{B}^{-1}, \lambda_i\right) \propto h_{i,t}^{\lambda_i+1/2 - 1} \exp \left(-\frac{h_{i,t}(\varepsilon_{i,t}^2 + \lambda_i - 2)}{2}\right), \quad (2.28) \]

i.e. \( h_{i,t} \) can be drawn using \( h_{i,t}(\varepsilon_{i,t}^2 + \lambda_i - 2) \sim \chi^2_{\lambda_i + 1} \) for \( i = 1, \ldots, \bar{n} \) and \( t = 1, \ldots, T \). Last, by the hierarchical structure, \( \lambda_i \) is updated by data only through \( H \). Combining (2.28) with the prior distribution yields

\[ p\left(\lambda_i | y, \{h_{i,t}\}_{t=1}^T\right) \propto 2^{-T/2} \Gamma(\lambda_i/2)^{-T} (\lambda_i - 2)^{\lambda_i/2 - T} \left( \prod_{t=1}^T h_{i,t}^{\lambda_i/2} \right) \exp \left(-\left(\frac{\lambda_i - 2}{2\lambda_i} \frac{T}{\sum_{t=1}^T h_{i,t} + \lambda_i} \lambda_i\right)\right) \quad (2.29) \]

from which, following Lanne and Luoto (2016), I draw using independence-chain Metropolis-Hasting algorithm. As a candidate density, I use the univariate normal distribution with mean and the precision parameter set to the mode and the negative hessian of the log conditional distribution, respectively.
3 The effects of government spending under anticipation: the noncausal VAR approach

3.1 Introduction

What is the impact of government spending on the economy? Despite a large body of literature, disagreement prevails on how an increase in government spending moves consumption, investment and output. To a great extent, the lack of consensus is due to uncertainty about valid methods to identify exogenous fiscal policy changes and to measure the reactions of economic agents. In particular, an identification strategy has to take into consideration that policy measures are usually implemented with a delay by the nature of the political process. The economic agents thus foresee and internalise fiscal policies before they materialise. If agents’ expectations are not accounted for, such fiscal foresight causes an econometric obstacle to measure the causal effects of government spending.

As the most prominent tool, the vector autoregressive (VAR) model derives the macroeconomic effects of fiscal policy by identifying a government spending shock through exclusion, sign or medium-run restrictions, normally based on fiscal rules of the government (Blanchard and Perotti, 2002; Galí, López-Salido, and Vallés, 2007; Mountford and Uhlig, 2009; Ben Zeev and Pappa, 2017). However, as shown by Ramey (2011b), a large part of exogenous changes in the U.S. government spending is related to defence expenditures and is predictable by information held by the public. Consequently,
The effects of government spending under anticipation: the noncausal VAR approach

the VAR model is at risk to identify government spending shocks that are in fact anticipated, and the underlying propagation of a policy change cannot be revealed. Under this nonfundamentalness problem, the VAR model is unable to retrieve the structural shocks from the past observables as the expectations of the public are not properly included in the information set of an econometrician (Hansen and Sargent, 1980, 1991; Lippi and Reichlin, 1994). As shown by Leeper, Walker, and Yang (2013), the fiscal foresight in a rational expectations model inherently leads to the nonfundamentalness problem through a noninvertible moving average (MA) representation for typical observables included to a VAR model.

The literature has tackled the problem of fiscal foresight by exploiting proxies and narrative measures for expectations of economic agents. The government spending shock is then derived as an innovation to the proxy or by local projections following Jordà (2005). Following the narrative approach, Ramey (2011b) constructed a proxy variable from administrative and news sources about the expected exogenous changes in military spending over time, also used in various subsequent studies to control for information. Using stock market data, Fisher and Peters (2010) recovered the spending shock from the excess returns of U.S. military contractors. Auerbach and Gorodnichenko (2012) and Caggiano, Castelnuovo, Colombo, and Nodari (2015) extracted the shock from the revisions in the professional forecasters. However, the validity of all of these approaches hinges upon how well the additional variable catches the information held by the public. Another possibility is to use Blaschke matrices to find the corresponding fundamental representation (see Mertens and Ravn 2010). These theoretical dynamic restrictions may, nevertheless, excessively shrink the set of possible underlying economic processes.

In this essay, I contribute to the fiscal policy literature by estimating the impact of government spending with a noncausal model that implicitly controls for the foresight of economic agents, while being flexible about the underlying economic process. I deviate from the conventional VAR analysis by augmenting the specification with the lead terms of observables, which corresponds to the noncausal VAR model of Lanne and Saikkonen (2013). Without assuming the included variables to align with economic agents’ information set, the future terms resolve the nonfundamentalness problem as the predictable error term of the model may now contain anticipated shocks. The impulse responses to the shocks are then derived from the two-sided MA representation

---

2For recent surveys, see Leeper et al. (2013), Ramey (2016) and Stock and Watson (2018).
of the model, which depends both on the past and future errors.

In contrast to its causal counterpart, the noncausal model is able to recover a shock that may already be internalised by the economic agents. I parsimoniously identify an anticipated spending shock using typical exclusion restrictions imposed on a fiscal policy rule. According to the rule, the government responds to the recent shocks of the economy with a lag. Under fiscal foresight, the model generates impulse responses of forward-looking variables to an anticipated government spending shock. By contrast, when anticipation does not matter, the model reduces to a causal VAR model with the exclusion restrictions following Blanchard and Perotti (2002). However, to distinguish between causal and noncausal specifications, the estimation requires non-Gaussianity as the models are observationally equivalent by their first and second moments. To that end, I assume multivariate $t$-distributed errors, under which the Gaussian structural shocks share a volatility term and normality is nested as limiting case. Consequently, the noncausal model can be estimated by a computationally efficient Gibbs sampler following Lanne and Luoto (2016).

On the methodological side, the essay belongs to the literature on nonfundamentalness in the VAR models, tracing back to Hansen and Sargent (1980), followed by Hansen and Sargent (1991) and Lippi and Reichlin (1994), and more recently discussed in Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson (2007), Lütkepohl (2014), Forni and Gambetti (2014) and Beaudry and Portier (2014). In general, nonfundamentalness could be tested and resolved by adding information in terms of factors extracted from large data sets (Forni and Gambetti, 2014). However, a wide range of macroeconomic indicators does not necessarily include information on fiscal anticipation, either. This chapter proceeds directly with a non-fundamental representation, which relaxes the need for sufficient information but requires either theoretical structure (Lippi and Reichlin, 1994; Mertens and Ravn, 2010) or non-Gaussianity (Rosenblatt, 2000; Lanne and Saikkonen, 2013) for identification. Similar to the framework of the essay, Chen, Choi, and Escanciano (2017) assume non-Gaussianity and propose a general test for nonfundamentalness. By the strategy of this essay, testing of nonfundamentalness is unnecessary, as the dynamic effects of government spending can be estimated regardless of the nonfundamentalness issue. In other words, the impulse responses can be produced even if a test rejects invertibility.

Understanding how government spending influences the economy is im-

---

3For recent review, see Kilian and Lütkepohl (2017), Ch. 17.
The effects of government spending under anticipation: the noncausal VAR approach

Important for validating the consistency of macroeconomic models as well as for designing fiscal policy. In the neoclassical theory, government spending may either crowd out private consumption and investment or stimulate the economy. The latter occurs when the economy involves nominal rigidities and households are non-Ricardian, eventually leading to a fiscal multiplier larger than one. Under flexible prices, instead, spending causes negative consumption and real wage responses as crowding out dominates. In the empirical literature, the exclusion restrictions based on the predetermined fiscal policy tend to produce positive consumption and real wage responses, whereas the studies employing a proxy for expectations document a decline of these variables (Ramey, 2016). In a straightforward manner, the noncausal model gives insights into the extent to which these differences stem from the anticipation of fiscal shocks.

Using the U.S. postwar data, I document non-negligible anticipation of macroeconomic variables in the face of a spending shock. Investment mildly rises during the anticipation phase before returning to its long-run trend, and the shock increases consumption, employment and the real wage. These reactions imply a fiscal multiplier above but close to one, although it is estimated with high uncertainty. The identified shock is also closely related to defence spending, commonly held as a source of exogeneity in the fiscal policy literature. Finally, I extensively compare my results on different, previously used identification strategies. First, the shock I identify coincides over time with the one obtained by the short-run restrictions from the causal VAR model. However, the impulse responses of the causal VAR incorrectly ignore the anticipation phase and underestimate the size of fiscal multipliers. Second, the results are insensitive to the inclusion of a variable that is informative about agents’ expectations, in line with the theory. In particular, identification schemes relying on the narrative measure of Ramey (2011b) or the excess returns used by Fisher and Peters (2010) do not alter the conclusions in the noncausal framework.

This chapter proceeds as follows. The next section presents the methodology to identify government spending shocks based on the noncausal VAR and illustrates the approach with a model of fiscal foresight. Section 3.3 explores the effects of government spending shocks in the US economy. The last section concludes.
3.2 Theory

By the institutional structure of the government, introducing a new policy involves a lag between legislation and implementation. When economic agents see the forthcoming policies, they are likely to hold richer information for decision-making than an econometrician observes. As a result, a structural VAR model is unable to extract exogenous policy changes from fiscal variables only. In this section, I propose an approach to recover a government spending shock when allowing for misalignment between the information sets of the economic agents and the econometrician. First, I show how impulse responses to the anticipated spending shock can be reproduced by means of noncausality. Second, I illustrate the proposed approach analytically in a model of fiscal foresight. Finally, I review the estimation of the model.

3.2.1 Identification of government spending shocks under anticipation

Let \( y_t = (g_t, y_{2,t}')' \) be an \( n \)-dimensional vector of observables with the detrended quarterly real government spending \( g_t \) and the \( n - 1 \) variables of interest collected in vector \( y_{2,t} \). Assume the mutually uncorrelated structural shocks in \( u_t \) propagate to \( y_t \) through the MA representation

\[
y_t = \sum_{k=0}^{\infty} B_k u_{t-k} = B(L)u_t,
\]

where \( u_t = (u_{g,t}, u_{2,t}')' \) consists of the government spending shock \( u_{g,t} \) and \( (n - 1) \) other structural shocks \( u_{2,t} \), \( L \) is the usual lag operator and \( B(L) = \sum_{k=0}^{\infty} B_k L^k \) an \( (n \times n) \) matrix polynomial convergent in the powers of \( L \).

Conventionally, the identification of the government spending shock and the derivation of the impulse responses of \( y_t \) are based on the causal VAR(\( p \)) model

\[
A(L)y_t = \varepsilon_t, \quad A(L) = I_n - A_1 L - \ldots - A_p L^p, \quad \varepsilon_t \sim (0, \Gamma),
\]

61
with an MA representation

$$y_t = A(L)^{-1} \varepsilon_t = \sum_{k=0}^{\infty} C_k \varepsilon_{t-k} = C(L) \varepsilon_t,$$

(3.3)

which coincides with (3.1) as long as its one-step ahead forecast error is a linear combination of structural shocks, i.e. $$y_t - E[y_t|y_{t-1}, y_{t-2}, \ldots] = A(L) y_t = \varepsilon_t = B_0 u_t.$$ Let the first row of the structural VAR model (3.2) correspond to a fiscal or spending rule of the government,

$$g_t = \sum_{i=1}^{p} a_{1,i} y_{t-i} + b_{11} u_{g,t} + b_{12} u_{2,t},$$

(3.4)

where $$a_{1,i}$$ ($$n \times 1$$) collects the autoregressive coefficients and $$(b_{11}, b_{12})$$ is the first row of matrix $$B_0$$. Government spending is therefore determined by changes in the lags of observables and by the current structural shocks $$u_{g,t}$$ and $$u_{2,t}$$. Tracing back to Blanchard and Perotti (2002) (BP, henceforth), the spending shock is identified from the system of equations (3.2) with $$\varepsilon_t = B_0 u_t$$ by imposing exclusion restrictions $$b_{12} = 0_{1 \times (n-1)}$$. Accordingly, based on the predetermined nature of economic policy, it takes at least a quarter for the government to learn about the state of the economy and to implement any measures in response. The spending shock $$u_{g,t}$$ is thus the only exogenous change that both drives current spending $$g_t$$ and is unrelated to the past states of the economy.4

However, the strategy of BP potentially fails to recover an unexpected exogenous shock to government spending. In particular, the identified shock in the U.S. data is related to defence spending and shown to be predictable by war dates as well as by professional forecasts (Ramey, 2011b), i.e. by information available to the public. Consequently, such fiscal foresight prevents the VAR model (3.2) from producing a forecast error unexpected both to the economic agents and the econometrician. The error is instead a linear combination of past errors from which a static impact matrix $$B$$ alone cannot recover the structural shocks (Lippi and Reichlin, 1994). Consequently, the measured effects of government spending may be starkly distorted as the MA represen-

---

4Originally, Blanchard and Perotti (2002) identify both spending and tax shocks using a non-recursive system that combines the exclusion restrictions $$b_{12} = 0_{1 \times (n-1)}$$ with information about elasticities. Nevertheless, ignoring the tax shock and identifying the spending shock through recursive restrictions produces similar results (Ramey, 2011b, 2016). See also Caldara and Kamps (2017) for non-zero estimates of the spending rule elasticity with respect to output.
3.2 Theory

tation (3.3) cannot reveal the underlying impulse responses to the shocks $u_t$ (Ramey, 2011b; Leeper et al., 2013).

The above nonfundamentalness problem boils down to the noninvertibility of the MA representation as economic agents react based on broader information than the history of $y_t$ contains. The invertibility of the MA representation (3.1) could be attained by enriching $y_t$ with variables reflecting the information set of economic agents (Ramey, 2011b; Fisher and Peters, 2010; Auerbach and Gorodnichenko, 2012; Caggiano et al., 2015) or by using a large-scale VAR model (Ellahie and Ricco, 2017). Alternatively, by imposing dynamic structure on the nonfundamental error term, a Blaschke matrix would recover the spending shock (Mertens and Ravn, 2010). However, while the former approach is subject to the ability of the additional variables to establish invertibility and to identify relevant sources of exogeneity, the latter approach may implicitly impose restrictive structure on the underlying economy.

As an alternative to the above approaches, consider a representation

$$
[g_t \ y_{2,t}] = \sum_{k=0}^{\infty} M_k \epsilon_{t-k} + \begin{bmatrix} 0 \\ f_{2,t} \end{bmatrix},
$$

(3.5)

where $M(L) = \sum_{k=0}^{\infty} M_k L^k$, $M_0 = I_n$, is a convergent $(n \times n)$ MA polynomial invertible in $L$ and $\epsilon_t \sim (0, \overline{\Gamma})$ is an independent and identically distributed (iid) error term. In turn, the $(n-1)$-dimensional vector $f_{2,t}$ depends directly on the future values of $y_t$:

$$
f_{2,t} = \Phi_{21,1} g_{t+1} + \Phi_{22,1} y_{2,t+1} + \ldots + \Phi_{21,s} g_{t+s} + \Phi_{22,s} y_{2,t+s},
$$

(3.6)

where $\Phi_{21,k}$ and $\Phi_{22,k}$ for $k = 1, \ldots, s$ are $((n-1) \times 1)$ and $((n-1) \times (n-1))$ matrices, respectively. In particular, the inclusion of $f_{2,t}$ tackles the noninvertibility of the MA representation (3.1) by allowing $y_{2,t}$ to depend on the future structural shocks. That is, when the observables induce nonfundamentalness in (3.1), $f_{2,t}$ ensures that the dynamics can be correctly captured with respect

---

5. In other words, $y_t$ is noninvertible in the past as there exist roots inside the unit circle for $|B(z)|$.

6. In detail, Mertens and Ravn (2010) derive from a rational expectations model a Blaschke matrix that maps the nonfundamental error term of the VAR model to the anticipated spending shock.

7. The information deficiency could also be tested based on factors (Forni and Gambetti, 2014) or non-normality (Chen et al., 2017). However, concluding nonfundamentalness from these tests may not necessarily be due to the fiscal foresight but equally well be a result of other misspecification issues or omitted factors.

---

63
to an anticipated error term $\epsilon_t$. Hence, the future terms control for the effects of omitted factors and expectations dismissed by the invertible MA polynomial $M(L)$. On the contrary, when the underlying MA representation (3.1) is invertible to the past and the causal VAR model is valid, $f_{2,t}$ is approximately zero as the lead terms become superfluous, and the equation (3.5) reduces to the implicitly invertible MA representation (3.3) of the causal VAR model with an unpredictable error term $\epsilon_t = \epsilon_t$.  

Now, from the representation (3.5), a government spending shock $\tilde{u}_g,t$ that is allowed to be anticipated by the variables in $y_{2,t}$ can be identified. Assume the error term $\epsilon_t$ is a static rotation of the shocks $\tilde{u}_t$, containing current or lagged values of the underlying, unanticipated structural shocks $u_t$. The uncorrelated structural shocks $\tilde{u}_t$ with unit variance are mapped into the error term as 

$$
\epsilon_t = \tilde{B} \tilde{u}_t, \quad (3.7)
$$

and $\tilde{B}$ satisfies $E[\epsilon_t \epsilon_t'] = \tilde{\Gamma} = \tilde{B} \tilde{B}'$. Let the first row of $\tilde{B}$ be $(\tilde{b}_{11}, \tilde{b}_{12})$ with scalar $\tilde{b}_{11}$ and a row vector $\tilde{b}_{12}$ of dimension $n - 1$. Noting that $M_0 = I_n$, by (3.5), the impact effect of the current structural shocks on government spending $g_t$ is equal to 

$$
\epsilon_{1,t} = \tilde{b}_{11} \tilde{u}_g,t + \tilde{b}_{12} \tilde{u}_{2,t}. \quad (3.8)
$$

Imposing $\tilde{b}_{12} = \mathbf{0}_{1 \times (n-1)}$, spending is predetermined within one quarter except for exogenous changes due to $\tilde{u}_{g,t}$. In other words, the fiscal policy responds contemporaneously only to its own shock in addition to the past variation.

By the above scheme, the spending shock is identified by the strategy of BP but relaxed to be anticipated through the term $f_{2,t}$, leaving $g_t$ unchanged prior to $t$. Accordingly, the government follows the spending rule (3.4), but the non-systematic deviation, $u_{g,t}$, may now be anticipated. For $f_{2,t} = \mathbf{0}_{(n-1) \times 1}$, the identification reduces to the original scheme of BP and recovers an unanticipated spending shock. Under noninvertibility, instead, the past observables are incapable of recovering the fundamental shock, which can then be obtained with the help of $f_{2,t}$. It should, however, be noted that the representations (3.1) and (3.5) are not necessarily equivalent in a way that the former could directly be rewritten in terms of $\tilde{u}_t$ and $f_t$ to produce the latter. An

---

8Specifically, the future terms account for noninvertibility of (3.1), when the unstable roots of the underlying MA polynomial $B(L)$ are inverted to the future.

9Under noncausality, the conditional expectation $E[\epsilon_{t+j}|y_t, y_{t-1}, \ldots], j > 0$ is nonzero such that errors can be predicted by the past observed variables, in contrast to the forecast error $\epsilon_t$ of the causal VAR.
3.2 Theory

exact, direct mapping may instead only exist for a particular set of economic models. The approach is rather a parsimonious departure from invertibility to approximate the true process and to flexibly identify an anticipated government spending shock. The representation (3.5) additionally covers a wider range of underlying economic dynamics and forms of anticipation than a causal VAR model (3.2) alone.

The model (3.5) can be written as the noncausal VAR\((r,s)\) model of Lanne and Saikkonen (2013)

\[
\Pi(L)\Phi(L^{-1})y_t = \epsilon_t, \tag{3.9}
\]

where

\[
\Phi(L^{-1}) = I - \Phi_1L^{-1} - \ldots - \Phi_s^{-s},
\]

\[
\Phi_i = \begin{bmatrix} 0 & 0_{1 \times (n-1)} \\ \Phi_{21,i} & \Phi_{22,i} \end{bmatrix}, i = 1, \ldots, s.
\]

and \(\Pi(L) = I - \Pi_1L - \ldots - \Pi_rL^r\). To see this, write the representation (3.5) equivalently as

\[
\Phi(L^{-1})y_t = M(L)\epsilon_t,
\]

where \(M(L)\) can be inverted to the left-hand side and its inverse be approximated up to a truncation error with the causal polynomial \(\Pi(L)\). As a result, the fiscal rule of the noncausal model (3.9) coincides with (3.4), as

\[
g_t = \sum_{i=1}^r \pi_{1,i}^\prime y_{t-i} + \bar{b}_{11} \bar{u}_{g,t} + \bar{b}_{12} \bar{u}_{2,t}, \tag{3.10}
\]

where \(\pi_{1,i}^\prime\) is the first row of matrix \(\Pi_i, i = 1, \ldots, r\). Simultaneously, anticipation of \(g_t\) by \(y_{2,t}\) is allowed through the future terms of \(y_t\).

The impulse responses to the identified shock \(\bar{u}_{g,t}\) are derived from the two-sided MA representation of the model,

\[
y_t = \Phi(L^{-1})^{-1}\Pi(L)^{-1}\epsilon_t = \sum_{k=-\infty}^{\infty} \Psi_k \bar{B}\bar{u}_{t-k} \tag{3.11}
\]

through which \(y_t\) generally depends both on the past and future shocks. Hence, the impulse responses to a government spending shock read as

\[
\frac{\partial y_{t+k}}{\partial \bar{u}_{g,t}} = \Psi_k \bar{b}_1, \ k = \ldots, -1, 0, 1, \ldots \tag{3.12}
\]
The effects of government spending under anticipation: the noncausal VAR approach

where $\tilde{b}_1$ is the first column of matrix $\tilde{B}$, obtained from $\tilde{\Gamma} = \tilde{B}\tilde{B}'$ after imposing the exclusion restrictions.\textsuperscript{10} By the stability of the matrix polynomials $\Pi(L)$ and $\Phi(L^{-1})$, the coefficients $\Psi_k$ decay to zero as $k \to \pm\infty$.\textsuperscript{11} Despite the infinite number of lead terms, the two-sided representation approximates the true model (3.1) but shows the responses of the most recent shocks at the negative lags. Beyond the anticipation horizon of economic agents, the corresponding MA coefficients in (3.12) are close to zero.

3.2.2 Analytical example: a model of fiscal foresight

Next, I illustrate how the above noncausal approach resolves the noninvertibility issue in an example of fiscal foresight from Leeper et al. (2013). In this particular setting, a mapping from the theoretical noninvertible model to the noncausal VAR exists.

In the model, a representative household maximises the expected welfare by deciding upon consumption under perfect depreciation of capital $K_t$, exogenous productivity $A_t$ and a proportional tax $\tau_t$ on production, $\tau_tY_t = \tau_tA_tK_{t-1}^\alpha$. The tax revenue is redistributed by the government through lump-sum transfers $T_t$. Maximising the expected welfare $E_0\sum_{t=0}^\infty \beta_t \log C_t$ subject to the budget constraint $C_t + K_t + T_t \leq (1 - \tau_t)A_tK_{t-1}^\alpha$, log-linearising and assuming that $a_t = \log A_t - \log \tilde{A}$ is uncorrelated, the solution for capital becomes

$$k_t = ak_{t-1} + a_t + (1 - \theta)\frac{\tau}{1 - \tau} \sum_{i=0}^\infty \theta^i \hat{E}_t \hat{\tau}_{t+i+1},$$

(3.13)

where $\theta = a\beta(1 - \tau) < 1$ and $k_t$, $a_t$ and $\hat{\tau}_t$ are log-deviations from the steady state.

Consider now agents observing a perfect signal on the tax rate $q$ periods forward, i.e. $\hat{\tau}_t = u_{t,t-q}$. Furthermore, let $a_t = u_{A,t}$ and assume $u_{A,t}$ and $u_{t,t}$ are uncorrelated. Substituting these to the solution yields

$$k_t = ak_{t-1} + u_{A,t} - \kappa(u_{t,t-q+2} + \theta u_{t,t-q+1} + \ldots + \theta^{q-1}u_{t,t}),$$

(3.14)

\textsuperscript{10}In practice, $B$ is derived as a lower triangular matrix from the Cholesky decomposition of $\Gamma$. The essay considers only partial identification of $B$, i.e. its first column. The remaining shocks of the economy are contained in the reduced-form term $\bar{B}_{22}\bar{u}_{2,t}$, where $\bar{B}_{22}$ is the lower-triangular lower-right $((n-1) \times (n-1))$ block of matrix $B$. By the structure of the model, the remaining shock are restricted to be neither anticipated nor unanticipated.

\textsuperscript{11}The stability is ensured by $\det \Pi(z) \neq 0$ and $\det \Phi(z) \neq 0$ for $|z| \leq 1$.
where $\kappa = (1 - \theta) \tau / (1 - \tau)$. Under foresight, i.e. $q > 0$, capital is determined by a pattern where the most recent news, $u_{\tau,t}$, informative about the most distant tax rates is discounted the heaviest by an anticipation rate $\theta$. By this inverse discounting, the history of observables is likely insufficient to recover the most recent shocks as they have the least weight on the current dynamics.\textsuperscript{12} In particular, the MA representation of the observables $y_t = (\hat{\tau}_t, k_t)'$

$$
\begin{bmatrix}
\hat{\tau}_t \\
k_t
\end{bmatrix}
= \begin{bmatrix}
\frac{L^q}{1 - \kappa L^{-1} + \theta L^{-2} + \ldots + \theta^{-1}} & 0 \\
\frac{1}{1 - \alpha L}
\end{bmatrix}
\begin{bmatrix}
u_{\tau,t} \\
u_{A,t}
\end{bmatrix}
= B(L)u_t,
$$

is noninvertible in the past since $|B(z)| = z^q = 0$ for $z = 0$. Hence, no causal VAR representation exists for the observables.\textsuperscript{13}

Nonetheless, $y_t$ can be written as in (3.5). For $q = 2$, rewrite $k_t$ as

$$
k_t = -\kappa \frac{L + \theta}{1 - \alpha L} u_{\tau,t} + \frac{1}{1 - \alpha L} u_{A,t}
= -\kappa \theta \tau_{t+2} - \kappa (1 + \theta \alpha) \tau_{t+1} - \kappa \alpha \frac{1 + \theta \alpha}{1 - \alpha L} u_{\tau,t-2} + \frac{1}{1 - \alpha L} u_{A,t},
$$

where $u_{\tau,t}$ and $u_{\tau,t-1}$ as signals about future tax rates are substituted out using $u_{\tau,t} = \hat{\tau}_{t+2}$. This leads to a representation of the form (3.5) for $y_t$,

$$
\begin{bmatrix}
\hat{\tau}_t \\
k_t
\end{bmatrix}
= \begin{bmatrix}
\kappa ((1 + \theta \alpha) L^{-1} + \theta L^{-2}) & 0 \\
0 & 1 - \alpha L
\end{bmatrix}
\begin{bmatrix}
1 \\
0
\end{bmatrix}
\begin{bmatrix}
\hat{\tau}_t \\
k_t
\end{bmatrix}
= \begin{bmatrix}
1 & 0 \\
0 & 1 - \alpha L
\end{bmatrix}
\begin{bmatrix}
\kappa ((1 + \theta \alpha) L^{-1} + \theta L^{-2}) & 0 \\
0 & 1
\end{bmatrix}
\begin{bmatrix}
u_{\tau,t-2} \\
u_{A,t}
\end{bmatrix}.
$$

Multiplying from the left by the inverse of the MA polynomial on the right-hand side yields the noncausal VAR(1,2) model (3.9)

$$
\begin{bmatrix}
1 & 0 \\
0 & 1 - \alpha L
\end{bmatrix}
\kappa ((1 + \theta \alpha) L^{-1} + \theta L^{-2})
\begin{bmatrix}
\hat{\tau}_t \\
k_t
\end{bmatrix}
= \begin{bmatrix}
0 & 1 \\
-\kappa \alpha (1 + \theta \alpha) & 1
\end{bmatrix}
\begin{bmatrix}
u_{\tau,t-2} \\
u_{A,t}
\end{bmatrix} = \epsilon_t,
$$

with

$$
\epsilon_t = \tilde{B} \bar{u}_t, \quad \tilde{B} = \begin{bmatrix}
1 & 0 \\
-\kappa \alpha (1 + \theta \alpha) & 1
\end{bmatrix}, \quad \bar{u}_t = \begin{bmatrix}
u_{\tau,t-2} \\
u_{A,t}
\end{bmatrix}.
$$

\textsuperscript{12}See Leeper et al. (2013) for a more thorough discussion.

\textsuperscript{13}Leeper et al. (2013) additionally show that the nonfundamental representation produced by a causal VAR can severely misspecify the tax shocks. Ramey (2009) demonstrates with Monte Carlo evidence that noninvertibility has equally adverse consequences on the inference about government spending shocks.
The effects of government spending under anticipation: the noncausal VAR approach

Figure 3.1: Impulse response functions of the fiscal foresight model to a technology and an anticipated tax shock
The upper panel corresponds to the theoretical impulse responses obtained from the solution. The lower graphs plot the impulse responses obtained from the noncausal VAR(1,2) model.

The error term $\epsilon_t$ in (3.16) now contains shocks $u_{\tau,t-2}$ and $u_{A,t}$, the former being anticipated by the economic agents. Moreover, $(\hat{\tau}_t, k_t)'$ has a two-sided MA representation (3.11) and, as a result, the impulse responses with respect to a tax shock $u_{\tau,t-2}$ read as

$$\frac{\partial y_{t+k}}{\partial u_{\tau,t-2}} = \Psi_{k} b_{1}, \ k = \ldots, -1, 0, 1, \ldots .$$

The impulse responses from the noncausal VAR model can thus be interpreted in a conventional manner but they are located additionally at the leads, i.e. at $k < 0$, due to the different time-indexing of the shock. In particular, this time-shifting occurs as noninvertibility prevents obtaining the shock as unanticipated using the current and past values of $y_t$ only. Noncausality facilitates then the recovery of an anticipated shock corresponding to the lagged underlying shock. Figure 3.1 depicts the impulse responses from the theoretical and noncausal models in the upper and lower plots, respectively, and confirms the equivalence of the MA representations (3.1) and (3.11) in this set-up.
3.2 Theory

Evidently, the impulse responses coincide, but through the two-sided MA representation of the noncausal VAR, the timing of the tax shock differs. The anticipation effects in capital can now be seen before \( k = 0 \). The policy shock thus influences capital already at the coefficients corresponding to the lead terms of errors, but those responses are zero at leads beyond \( k = -2 \).

3.2.3 Estimation

This subsection outlines the estimation of the noncausal VAR\((r,s)\) model (3.9). To identify a noncausal VAR\((r,s)\) from a causal VAR\((r+s)\) model in terms of likelihood, it is necessary deviate from Gaussianity as the models are observationally equivalent by their first and second moments and cannot be distinguished under normality. In what follows, the error term \( \epsilon_t \) is assumed to be multivariate t-distributed, implying unique identification of the model parameters through its likelihood function.\(^{14}\) The noncausal VAR is then equivalently written as

\[
\omega_t^{1/2} \Pi(L) \Phi(L^{-1}) y_t = \eta_t, \tag{3.18}
\]

where \( \lambda \omega_t \) is \( \chi^2_\lambda \)-distributed and \( \eta_t \sim \text{N}(0, \Sigma) \) such that \( \tilde{\Gamma} = \mathbb{E}[\epsilon_t \epsilon_t'] = \frac{\lambda}{\lambda-2} \Sigma \).

Hence, the error distribution is Gaussian conditional on the scalar volatility term \( \omega_t^{-1/2} \) that controls for leptokurtosis of the time series, i.e. \( \omega_t \) catches exogenous, common volatility in observables. For small \( \lambda \), the distribution has fatter tails than under normality. On the other hand, the distribution is closer to Gaussianity for large values of \( \lambda \). Moreover, \( \eta_t \) is a linear combination of the Gaussian structural shocks, which are recovered through the rotation matrix \( \bar{B} \).\(^{15}\)

I rely on Bayesian methods to tackle the large parameter space arising due to the additional lead terms. As shown in Appendix 3.A, the model has a conditional likelihood function and a computationally feasible posterior distribution under standard prior distributions. In particular, exploiting the multiplicative structure of the model and the conditional normality in (3.18), the model can be estimated using a Metropolis-within-Gibbs sampler of Lanne and Luoto (2016) with which the parameters are drawn from the posterior distribution as follows. First, the lag and lead coefficients in \( \Pi(L) \) and \( \Phi(L^{-1}) \)

\(^{14}\)For details on identifiability, see Lanne and Saikkonen (2013).

\(^{15}\)The distributional assumption implies that \( \eta_t \) may contain both anticipated and unanticipated structural shocks that share the same volatility term. Avoiding this potential caveat would require a less parsimonious empirical strategy such as considering an alternative non-Gaussian distribution.
are conditionally normally distributed. Second, the scale matrix $\Sigma$ follows an inverse Wishart distribution conditional on the remaining parameters of the model. Last, $\omega = (\omega_{t+1}, \ldots, \omega_{T-s})$, and $\lambda$ can be conditionally drawn using a $\chi^2_{\lambda+n}$-distribution and a kernel for $\lambda$, with the latter solely depending on the volatility terms $\omega$.

3.3 The impact of government spending in the U.S. economy

This section examines the dynamic effects of an exogenous change in government spending in the U.S. economy. According to economic theory, the effects of a government spending shock hinge on the interaction of wealth, intertemporal and distortionary effects (Ramey, 2011a). If the wealth effect of labour supply dominates, Ricardian households decrease both consumption and investment in response to increased spending, and crowding-out effects imply a fiscal multiplier smaller than one. In contrast, when the economy involves non-Ricardian and Keynesian elements, a spending shock is followed by increasing marginal product of labour and, consequently, rising wages lead to a positive consumption response and stimulative effects of fiscal policy. Finally, distortionary taxation to finance spending dampens the positive effects on consumption, employment and output.

However, validating the effects of government spending poses an econometric challenge due to fiscal foresight. According Ramey (2011b), inference on the effects of government spending essentially depends on the timing of the shocks and their anticipation. Ellahie and Ricco (2017) show that the use of large-scale VARs mitigate the distortionary effects of insufficient information. Chen et al. (2017) analyse the presence of nonfundamentalness in fiscal VAR models with their test for noninvertibility. According to their results, both baseline specifications of Blanchard and Perotti (2002) and Ramey (2011b) survive the hypothesis of invertibility. However, this observation is in contrast with the results of Ramey (2011b) who document forecasting power of the constructed narrative measure with respect to the government spending of BP. The noncausal VAR overcomes the noninvertibility issue as the government spending shock may but need not be anticipated. In particular, the approach retains the VAR methodology with the conventional exclusion re-

---

16See Gali et al. (2007) and Ramey (2011a) for more in-depth discussion on propagation mechanisms of government spending.
strictions on the fiscal rule, without assuming additional variables to account for the foresight.

### 3.3.1 Data and estimation

I use the following U.S. quarterly macroeconomic data. My measure of public spending is government consumption expenditures and gross investment. Output is the Gross Domestic Product (GDP), private consumption is the sum of consumption of nondurables and services, and investment consists of fixed private investment and consumption of durables. These national accounts variables, taken from the National Income and Product Accounts (NIPA) Tables of the Bureau of Economic Analysis, are transformed into real values by the GDP deflator, into per-capita terms by the civilian noninstitutional population and expressed in logs. Employment and wages are the log per-capita hours and the log real hourly compensation, respectively, from the nonfarm business sector. I derive the average tax rate as all federal receipts divided by the nominal GDP. These seven variables, from which I subtract their quadratic trend, form the baseline specification and span the quarters from 1945Q1 until 2013Q4. Additionally, I consider annualised inflation, computed as a log difference of the GDP deflator, and two interest rates, the 3-month T-bill rate and the Moody’s Seasoned Baa Corporate Bond Yield.17

The noncausal VAR\(r,s\) models I estimate include the above variables with the number of lags and leads set to four. Four lags, on the one hand, allows observables to have an invertible MA polynomial \(M(L)\) in (3.5) general enough to fully catch the variation of structural shocks in the absence of nonfundamentalness. On the other hand, \(s = 4\) leads imply a rich structure for the noncausal part \(f_2\), if nonfundamentalness arises as a result of anticipation. In particular, the structural shocks are recovered by the lead terms as anticipated (3.1) in case the underlying MA representation is noninvertible to the past but invertible to the past and future. As a whole, four lags and leads are then expected to be sufficient to account for the full dynamics of observables.

I estimate the model with Bayesian methods and set a Minnesota-Litterman-type prior distribution as also used by Lanne and Luoto (2016), explained in Appendix 3.A. Specifically, I control for the tightness of the prior distribution separately for the lag and lead coefficients. By adjusting the overall tightness parameters, the prior about the lag coefficients is less informative, whereas the lead coefficients are shrunk more strongly towards zero. Hence, a priori,

---

17 Apart from the NIPA variables, data are retrieved from the FRED database.
The effects of government spending under anticipation: the noncausal VAR approach

Figure 3.2: Posterior distribution of degrees-of-freedom parameter $\lambda$
Posterior draws of $\lambda$ from the baseline noncausal VAR. The dashed vertical line is the posterior mean.

the lag terms are more important to determine the dynamics of variables. I proceed by drawing 50,000 times from the posterior distribution. For each draw, I derive the MA representation (3.11) and impose the exclusion restrictions using the Cholesky decomposition $\bar{\Gamma} = \bar{B}\bar{B}'$. As for any Gibbs sampler, the algorithm to obtain posterior draws performs well when the distribution is unimodal. Multimodality, however, easily arises in the estimation of the noncausal VAR, as observed by Lanne and Luoto (2016). Nonetheless, the less loose prior distribution for the lead coefficients by the greater overall tightness and the restrictions imposed in (3.9) are powerful in attaining a unimodal posterior distribution.\(^\text{18}\)

For identifying a unique VAR($r,s$) specification, it is necessary to assume non-Gaussianity of the error term. Conveniently, the assumed multivariate $t$-distribution nests Gaussianity for a large degrees-of-freedom parameter $\lambda$. Low estimates of $\lambda$ thus immediately suggests the validity of the distributional assumption compared to Gaussianity, implied by excess kurtosis in the error distribution.\(^\text{19}\) In Figure 3.2, I plot the histogram of the posterior draws of $\lambda$ from the baseline VAR model. The histogram clearly indicates that a large probability mass is located at low degrees of freedom. Moreover, the data strongly dominate the assumed prior mean 10 with a posterior mean of

\(^{18}\)Details on estimation and convergence of the algorithm are found in Appendices 3.A and 3.B.

\(^{19}\)Similarly, Chen et al. (2017) document significant non-normality in fiscal VAR models.
4.2: the probability of $\lambda$ being greater than 6 is extremely low. Therefore, the data lend support for the multivariate t-distribution, which facilitates the identification of the noncausal model.

### 3.3.2 Impulse responses to a government spending shock

In Figure 3.3, the solid lines depict the estimated impulse responses of the seven variables included in the baseline noncausal VAR(4,4) model to a one standard deviation government spending shock. Therein, I also report the posterior medians of the estimates together with the 68 and 90 percent credible sets. Because of noncausality, the responses are additionally located at the negative lags, corresponding to the lead terms of the MA representation. Beyond lead 10, these reactions are close to zero. In Figure 3.12 of Appendix 3.C, I additionally plot the impulse responses from the models for various lead orders, showing that the results remain similar irrespective of the number of included leads.

According to the noncausal model, government spending increases by one percent relative to its trend and output peaks at 0.4 percent in response to a spending shock $u_{g,t}$. The shock materialises in spending from time 0 onwards, implied by the zero restrictions imposed on the lead terms of $y_t$ in the first equation of the model. The other variables anticipate this increase from quarter -6 onwards. While all variables increase at these anticipatory lags, the most reactions are statistically insignificant. In contrast, GDP significantly reacts approximately one and a half years ahead of the future spending increase. Simultaneously, investment starts to increase and peaks at the realisation of the shock, converging afterwards towards its trend level. On the other hand, consumption remains close to its trend level before the shock arrives after which it rises in a hump-shaped pattern. Hours worked rise fast at the anticipation lags and stay positive for the following 10 quarters after starting to decrease at the materialisation of the shock. The real wage exhibits a hump-shaped increase, which occurs simultaneously with the decline of hours. Last, the average tax rate rises both at the anticipation phase and at the materialisation of the shock, potentially induced by automatic stabilisers.

I continue by augmenting the baseline VAR model with inflation, the short-term rate and the corporate bond yield. In panel (a) of Figure 3.4, I report for the sake of space only the responses of the three additional variables from

---

20 Both the posterior medians and the credible sets are computed as periodwise quantiles from the impulse responses derived from the posterior draws.
The effects of government spending under anticipation: the noncausal VAR approach

Figure 3.3: Impulse responses to the government spending shock from the baseline model
Solid lines and dashed lines are the median impulse responses from the noncausal VAR(4,4) and causal VAR(4) models, respectively, to a one standard deviation government spending shock. The dark and light grey shaded regions are the 68 and 90 percent, respectively, credible sets of the estimated impulse responses from the noncausal model.

this ten-variable VAR(4,4) model. A one percent, exogenous increase in government spending has a small, negative impact on inflation, which decreases by 0.1 annualised percentage points.\textsuperscript{21} The 3-month rate shows no significant movements whereas the corporate bond rate mildly declines after the shock materialises, although the effect is statistically insignificant.

Above, the spending shock caused an initially increasing investment and a positive reaction of consumption at the realisation of the shock. For a more in-depth analysis, I replace consumption and investment with their subcomponents in the baseline VAR. Panel (b) of Figure 3.4 plots the responses of disaggregated consumption and investment components from the re-estimated noncausal VAR(4,4). Both services and nondurable consumption respond significantly and positively to a spending shock, which translates to the previously documented increase of the aggregate consumption. Suggesting that consumption at least partly anticipates the spending increase, consumption of services rises to some extent at its leads. On the other hand, the observed

\textsuperscript{21}This slight but somewhat counter-intuitive decline is potentially resulted by other factors such as oil price changes and has been found in earlier studies as well. See also Mountford and Uhlig (2009) for discussion.
3.3 The impact of government spending in the U.S. economy

Figure 3.4: Impulse responses to the government spending shock from the VAR models with additional variables

Solid lines and dashed lines are the median impulse responses from the noncausal VAR(4,4) and causal VAR(4) models, respectively, to a one standard deviation government spending shock. The dark and light grey shaded regions are the 68 and 90 percent, respectively, credible sets of the estimated impulse responses from the noncausal model. The impulse responses in panel (a) are from the 10-variable VAR including the baseline and the plotted variables. The impulse responses in panel (b) are computed from the 10-variable VAR including the baseline variables but consumption and investment replaced by the variables shown.
The effects of government spending under anticipation: the noncausal VAR approach

Figure 3.5: Spending shock, defence and non-defence expenditures
The solid, light grey line depicts the posterior median of the 4-quarter moving average of the spending shock identified from the baseline noncausal VAR(4,4). Dot-dashed and dashed lines are the log real national defense consumption expenditures and gross investment, respectively. Dotted line refers to the log of real federal non-defence consumption expenditures and gross investment. All variables are demeaned and standardised.

increase in investment is due to durable consumption and nonresidential investment.\textsuperscript{22}

To illustrate movements of the identified shock over time, Figure 3.5 plots its four-year moving average from the baseline noncausal model. I additionally plot federal spending divided into three components, consumption expenditures and gross investment on national defence, and non-defence federal spending. Accordingly, the spending shock is closely related to the U.S. defence expenditures in the medium run. The spending shock series leads consumption and investment components of defence expenditures and is unrelated to non-defence spending. However, the linkage between the shock and defence expenditures has faded since the 1990s, after which the identified exogenous changes have evolved in a less systematic pattern. By these observations, the shock can be characterised by two insights. First, the great

\textsuperscript{22}In both of the augmented specifications, the estimated responses of the seven remaining variables not shown in Figure 3.4 are indistinguishable from those obtained from the baseline model and reported in in Appendix 3.C.
3.3 The impact of government spending in the U.S. economy

The persistence of the shock observed in the impulse responses is likely to stem from the nature of defence spending, from decade-lasting military build-ups and wars that the United States was engaged in. Second, the shock is unrelated to the non-defence component of federal spending. It induces instead variation that mainly belongs to a particular class of events, the U.S. military expenditures, which likely are orthogonal to the present state of the economy and which have been used to identify exogenous events in spending in the existing literature.\(^\text{23}\)

In the noncausal model, the identified spending shock may be predictable to the economic agents, without ruling out causality a priori. To assess the importance of the lead terms of \(y_t\), Figures 3.3 and 3.4 include in dashed lines the impulse responses from the causal VAR(4) models to an unanticipated spending shock following the identification of Blanchard and Perotti (2002). Setting \(s = 0\) and disregarding noncausality, conclusions about the spending shock are slightly altered, despite the general pattern of the estimates remaining close to their noncausal counterparts at the positive lags. Most importantly, the causal VAR is unable to reveal the positive reaction of investment located at the negative lags, causing a negative response estimate. Similarly, the tax rate is estimated to decline from quarter 0 onwards, and the negative responses of inflation and interest rates are estimated stronger.

In general, the causal VAR ignores – by construction – the responses at the leads, despite the fact that the shock may well be anticipated. Given the statistical significance of these reactions, the causal model suffers from deficient information. Moreover, as shown in Figure 3.14 in Appendix 3.C, the shock identified from the causal VAR virtually coincides over time with the one from the noncausal model. In other words, the causal VAR recovers a shock that aligns with the shock from the noncausal model, but only the latter is able to uncover the reactions that occur at the anticipatory, negative lags. A causal VAR model under the exclusion restrictions is thus at high risk to catch defence-spending-related events that the economic agents are able to forecast.\(^\text{24}\)

\(^{23}\)U.S. defence spending and military events are regarded as a source of exogenous variation, amongst others, by Ramey and Shapiro (1998), Fisher and Peters (2010), Ramey (2011b) and Ben Zeev and Pappa (2017).

\(^{24}\)The results obtained in this section are robust to various perturbations in the research setting. First, the conclusions remain the same when less informative prior distributions for the lag coefficients are used. The results are also invariant to a somewhat less informative prior distribution of the lead coefficients, although this loosening is subject to the emergence of multimodality. Second, in the causal VAR, the corresponding least squares estimates are close to the reported posterior medians, and the results are qualitatively similar to those obtained in a model where
In view of the economic theory, I interpret the results as follows. The increase of both output and employment in response to the anticipated spending shock suggests the dominance of the wealth over substitution effects of the households. Moreover, as the shock induces a profound increase in real wage and no significant decline of consumption or investment, there is evidence, to some extent, on the existence of non-Ricardian and Keynesian mechanisms. On the other hand, the absence of strong positive consumption and investment responses suggests certain degree of crowding out of private business. The path of investment is mostly due to the non-residential investment, whereas residential investment show no significant reactions. Somewhat surprisingly, the shock has deflationary effects on the price level in contrast with the neoclassical theory and positively reacting consumption.

3.3.3 The size of the fiscal multiplier

Does government spending stimulate the economy? A fiscal multiplier greater than unity indicates that an increase in government spending boosts private economy in a way that the benefits dominate the crowding-out and distortionary effects of public consumption and taxation. The spending literature calculates the multiplier in two alternative ways, either as a peak output response relative to the initial government spending impact effect or as a ratio between the present value integrals of output and government spending responses. I follow the latter technique, also suggested by Mountford and Uhlig (2009) and Ramey (2016), as the former method tends to overestimate the size of the fiscal multiplier. In addition, the latter takes into account more flexibly timing, persistence and anticipation of the shock. The fiscal multiplier is defined as

\[
\sum_{k=H_1}^{H_2} (1 + r)^{-k} \frac{\partial \log GDP_t + k}{\partial \hat{u}_{G,t}} \frac{GDP}{G},
\]

where \( H_1 \) is set to -10, to the time point of initial reactions according to Figure 3.4, \( H_2 \) is the length of horizon after the shock has realised and \( r \) is the long-run real interest rate computed as a sample mean of the difference between the T-bill rate and inflation. Last, \( GDP/G \) is the sample mean of the ratio of GDP to government spending and converts the percentage deviations to variables are included in log levels with no quadratic trend. Third, the estimated noncausal impulse responses remain the same when produced from a larger set of smaller-dimensional models which included the above variables.
### 3.3 The impact of government spending in the U.S. economy

The impact of government spending in the U.S. economy

![Histograms for fiscal multipliers on different horizons](image)

Figure 3.6: Histogram of posterior draws for fiscal multipliers on different horizons $H_2$

The multipliers are computed using (3.19) with $H_1 = -10$ and based on the baseline noncausal VAR(4,4) model. Quantities on the y-axis are normalised such that histograms integrate to 1. Dotted vertical lines are the medians of the multipliers. The red dashed lines are the median fiscal multipliers from the corresponding causal VAR(4) model.

Monetary units in real terms.

Figure 3.6 reports the posterior distribution of the fiscal multiplier (3.19) and their medians in dotted lines for various horizons. Under shorter horizons, $H_2 \in \{0, 5, 10\}$, in the upper plots, the median multiplier is significantly above one, being the largest when only the impact effect of government spending is included, $H_2 = 0$. Eventually, when the horizon lengthens, the posterior distribution of multiplier converges to a distribution with a median of 1.4, as seen in the lower graphs. The impulse responses of Figure 3.3 generate the mechanism behind this pattern. Since the government spending response is prolonged relative to the reaction of GDP, the size of multiplier decreases as more inputs are added to the denominator of (3.19).

Overall, there is great uncertainty on whether government spending can be stimulative beyond a short horizon, seen as large dispersions in the posterior distributions of Figure 3.6. As soon as the horizon is longer, a significant portion of the probability mass is concentrated on the region below one, and the long-run multiplier with $H_2 = 40$ reaches with high probability values both
above and below one.\textsuperscript{25} It is also noteworthy that the multiplier computed here is not purely deficit-based as government spending may be followed by a distortionary increase in tax rate. In light of this evidence, the overall impact of the identified exogenous government spending on the private economy remains imprecise.\textsuperscript{26}

Finally, I draw in dashed lines the posterior medians of fiscal multipliers computed from the causal VAR(4) model with the baseline variables. At all horizons, the multipliers are quantified to be smaller than their corresponding estimates from the noncausal model. Strikingly, despite the positive response of consumption, the causal VAR model with the BP identification has a tendency to produce small multipliers (See also, Ramey, 2016). This well-known controversy can be explained from the causal responses shown in Figure 3.3. As the causal structural VAR model disregards the anticipation effect in GDP, the nominator of (3.19) is necessarily smaller relative to the denominator, which results in a smaller multiplier.

### 3.3.4 Relation to government spending shock measures

In the government spending literature, fiscal foresight and nonfundamentalness have been tackled by using a measure of news either to enrich the information set of a VAR model or to derive the responses to a shock using local projections (Jordà, 2005). In the noncausal model (3.9) instead, the anticipated spending shock can be recovered independent of the nonfundamentalness issue exploiting the predetermined nature of government policy, i.e. through the exclusion restrictions imposed on the error term. Including a proxy to the noncausal VAR can then shed light on how informative the variable is about the identified shock.

As concluded in Figure 3.5, the shock identified in the noncausal VAR reflects defence expenditures, and I thus consider two prominent measures of shocks applied in the literature, the narrative defence news and the excess returns of military contractors. First, Ramey’s narrative news (Ramey, 2011b)

\textsuperscript{25}Owyang, Ramey, and Zubairy (2013) and Ramey (2016) argue that a trend in the GDP-to-spending ratio leads to a bias in the multiplier estimates computed using (3.19). In the sample, the mean value of $Y/G$ is 4.8, while the ratio varies over time between 4 and 6.5. I reproduced, for robustness, impulse responses and fiscal multipliers by transforming the national accounts variables with the Gordon-Krenn transformation as the authors suggest. The results are of the same magnitude as those reported in this section.

\textsuperscript{26}One possible explanation for this uncertainty may be the time-dependence in the effectiveness of fiscal policy, as examined by Auerbach and Gorodnichenko (2012), Owyang et al. (2013) and Caggiano et al. (2015).
3.3 The impact of government spending in the U.S. economy

Figure 3.7: The identified spending shock, Fisher-Peters excess returns and Ramey’s narrative news

Grey solid line depicts the median of identified anticipated spending shocks from the baseline VAR(4,4) model, the dark solid line the Ramey (2011b) narrative news of anticipated government spending, measured as a share of future government spending of GDP and the black dashed line the excess returns of military contractors constructed by Fisher and Peters (2010).

Captures information held by the public about the expected discounted value of government spending changes due to foreign policy events relative to GDP. On the other hand, the Fisher-Peters excess returns (Fisher and Peters, 2010) aims to gauge the market expectations about future spending by the asset prices of top three U.S. military contractors.27 According to Fisher and Peters (2010), the difference between these series stems from the fact that market expectations about military spending evolve in a more nuanced way than the immediate changes seen in the Ramey’s news series. However, Ramey (2016) argues the excess returns series has low instrumental relevance for government spending.

In Figure 3.7, I plot these two variables along with the identified govern-

27The defence news variable and Fisher-Peters excess returns are available as supplementary data of Ramey (2016) in Valerie Ramey’s webpage. The narrative news series, extended by Owyang et al. (2013), spans the whole post-war period until 2013Q4 whereas the last observation for Fisher-Peters data is 2007Q4.
ment spending shock estimated from the baseline noncausal VAR(4,4) model. The shock spikes during military events, similar to the narrative defence news, most notably during the Korean and Vietnam wars at the beginning of 1950s and 1970s, respectively. The Fisher-Peters excess returns comoves with the shock during the 1960s and 1970s as well as during Ronald Reagan’s presidency. However, neither of the series is a direct empirical counterpart of the identified shock.

I continue by adding the variables to the baseline noncausal VAR and allow them to anticipate the shock identified by the standard exclusion restrictions. Given its relevance, a shock-related measure would respond positively to the future spending increase. Figure 3.8 graphs the median impulse responses of government spending, GDP and the two respective measures of defence news to a one standard deviation shock identified by the exclusion restrictions in both models where the eighth variable is either Ramey’s narrative news or the Fisher-Peters excess returns. Accordingly, the exclusion or inclusion of either of the proxies does not alter the estimates about the responses of government spending and GDP – in line with the prediction that the noncausal model remains valid regardless of information contained in the observables. Interestingly, before the realisation moment at 0, Ramey’s news variable reacts slightly positively whereas the excess returns move statistically significantly upwards, implying that the latter series is able to predict the future spending increase.

The relevance of the spending shock can also be analysed by means of its relative contribution to the overall movements in a variable. Formally, the $i$th variable in the noncausal VAR has an MA representation

$$y_{i,t} = e_i' \sum_{k=-\infty}^{\infty} \Psi_k \left( \bar{b}_1 \bar{u}_{1,t-k} + \bar{b}_2 \bar{u}_{2,t-k} \right), \quad (3.20)$$

where $e_i = (0, ..., 1, ..., 0)'$ with 1 in its $i$th element and an $(n \times (n - 1))$ matrix $\bar{b}_2$ consists of columns of $\bar{B}$ corresponding to the $n - 1$ remaining shocks contained in $\bar{u}_{2,t-k}$. Now, define the fraction of variance of $y_{i,t}$ due to the

---

28 The excess returns being available only until 2007Q4, the second model spans a shorter time period. However, the results from the baseline model or from the specification with Ramey’s narrative news do not alter when using data up to 2007Q4. The model with the excess returns includes four lags and leads. The inclusion of Ramey’s news to the model, however, induces convergence problems to the algorithm, caused by the fact that the series is dominated by zero values. A smooth running of the algorithm can be achieved by diminishing the lag and lead orders to three. The results would remain qualitatively the same if four lag and leads were used.
3.3 The impact of government spending in the U.S. economy

Figure 3.8: Impulse responses from the VAR models with a measure of news
Marked and solid lines are the posterior median impulse responses to the anticipated spending shock from the noncausal VAR models augmented with Ramey’s narrative news and the Fisher-Peters excess returns, respectively. The dark and light grey shaded regions are the 68 and 90 percent posterior credible sets, respectively, shown in the responses of spending, GDP and Ramey’s news from the former and in the response of excess returns from the latter VAR model.

<table>
<thead>
<tr>
<th>Panel (a) Identification by the exclusion restrictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizon ((H_2))</td>
</tr>
<tr>
<td>(-10)</td>
</tr>
<tr>
<td>(-5)</td>
</tr>
<tr>
<td>(0)</td>
</tr>
<tr>
<td>(5)</td>
</tr>
<tr>
<td>(10)</td>
</tr>
<tr>
<td>(20)</td>
</tr>
<tr>
<td>(30)</td>
</tr>
<tr>
<td>(40)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel (b) Identification by the proxy variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizon ((H_2))</td>
</tr>
<tr>
<td>(-10)</td>
</tr>
<tr>
<td>(-5)</td>
</tr>
<tr>
<td>(0)</td>
</tr>
<tr>
<td>(5)</td>
</tr>
<tr>
<td>(10)</td>
</tr>
<tr>
<td>(20)</td>
</tr>
<tr>
<td>(30)</td>
</tr>
<tr>
<td>(40)</td>
</tr>
</tbody>
</table>

Table 3.1: Fractions of variance contributed by the spending shock
The shares shown in percentages are computed using (3.21) with \(H_1 = -10\) as posterior medians. Baseline + Ramey refers to the VAR(3,3) model with the seven baseline variables and the Ramey news, Baseline + F-P to the VAR(4,4) model with the baseline variables and the Fisher-Peters excess returns. In panel (b), the identification is proceeded by the approach maximising the contribution of the shock to the eighth variable, explained in text.
government spending shock $\bar{u}_{g,t}$ over horizon $[H_1, H_2]$ as

$$
\rho(y_{i,t}; H_1, H_2) = \frac{\sum_{k=H_1}^{H_2} e_i' \Psi_k \bar{b}_1 \bar{e}_i' \Psi_k' e_i}{\sum_{k=H_1}^{H_2} e_i' \Sigma \Psi_k' e_i}.
$$

(3.21)

Under causality, the fraction (3.21) reduces to the forecast error variance decomposition of a VAR model over horizon $H_2$. Once $s > 0$, $\rho(y_{i,t}; H_1, H_2)$ generally gives the fraction of the unconditional variance of variable $y_{i,t}$ explained by the spending shock.

In panel (a) of Table 3.1, I report the fractions (3.21) in the baseline model and in the models with a proxy variable. The identified shock explains over 50 percent of the detrended variation in government spending and approximately 15 percent in output. Anticipation of the government spending shocks, observed at negative $H_2$, accounts for five percent of the overall variation in output. Strikingly, the shock is able to explain a part of movements in Ramey’s news before time 0 after which it contributes minimally to the movements of the variable. The shock accounts, in turn, for a fifth of the variation in the excess returns both before and after the time point 0. Overall, the both measures are to a great extent explained by factors other than those caught by the shock from the standard identification. However, given their positive contributions for $H_2 < 0$, both series are able to predict the government spending shock to some extent.

Does the use of identification based on the proxies of news lead to different conclusions about the effects of government spending? The mild responses of these variables to the spending shock and their small contributions suggest that the changes in the measures consist of events different from those captured by the identified shock. Nonetheless, these events, provided their exogeneity, may induce effects similar to the structurally identified shock. I therefore derive impulse responses to a shock identified by the variation of a news measure. In both of the augmented eight-variable models above, I proceed by finding a shock that explains the most of the overall movements of the variable informative about the defence news, i.e. it maximises the fraction of variance (3.21) among all possible linear mappings from structural shocks to the reduced-form error term. Given that the variable is mainly driven by the

29 Appendix 3.D shows in detail that this identification can be achieved through an eigenvalue problem, similar to the Max Share approach (Uhlig, 2004; Francis et al., 2014) which rotates the error of a VAR model to find the shock that maximises its amount to the forecast error variance. In the identification, I use a horizon of $[-20, 20]$. The use of a shorter horizon has no effect on the
exogenous variation that translates to changes in spending, the identification strategy is valid recovering the causal effects.

Panel (b) of Table 3.1 reports the shares contributed by these two alternatively identified shocks to the unconditional variance of government spending, GDP and the respective news measure. A single shock is able to explain the major part of the variance of the variable. While the Fisher-Peters shock also moderately contributes to the overall detrended variation in spending and GDP, the Ramey shock barely influences these variables. In Figure 3.9, I report the results from all identification strategies, in Panel (a) for Ramey’s news and in Panel (b) for the Fisher-Peters excess returns. First, the solid and marked solid lines draw the impulse responses to the shock identified by the standard exclusion restriction from the augmented models. They are identical to those in Figure 3.8 but are rescaled by normalising the maximum impact on spending to one percent. Second, the dashed and dashed-dotted lines depict the responses to the Ramey and Fisher-Peters shocks, derived from the alternative identification strategy with respect to the these two measures.

In response to the Ramey shock, shown in Panel (a), the narrative news variable jumps by three percentage points, as expected from the identification strategy that maximises the contribution of the shock to the variable. The jump is followed by a gradual increase of spending which peaks after a year. Broadly, the responses to the Ramey shock are also similar to those from the standard identification and within the credible bands, despite the shocks being empirically unrelated. Notably, the Ramey shock induces a delayed increase of government spending, and the timing of the shock differs from the baseline shock. The increase of the news variable is, however, anticipated in real wages and consumption already at the negative lags of the horizon, which can now be observed as a result of noncausality. Correspondingly, the reactions to the Fisher-Peters shock are remarkably similar to the other identification strategies. The Fisher-Peters shock induces a substantial and hump-shaped increase of the excess returns, occurring both at the positive and negative lags. The spending increase is also significantly anticipated by GDP, investment, real wages and tax rate. These reactions are stronger but broadly consistent with the standard identification strategy – despite being driven by separate factors as suggested by Table 3.1.

Unlike the previously employed empirical strategies, the noncausal VAR approach can assess the connection between the structurally identified shock and the empirical measures of the latent shock, as the error term need not
The effects of government spending under anticipation: the noncausal VAR approach

Figure 3.9: Impulse responses to a spending shock identified by two alternative strategies

Black marked lines in panel (a) and solid lines in panel (b) are the posterior median impulse responses to the shock identified by the standard exclusion restrictions from the noncausal models augmented with Ramey’s narrative news and the Fisher-Peters excess returns, respectively. The dashed and dot-dashed lines are the responses to the proxy-based identified Ramey and Fisher-Peters shocks, respectively. The dark and light grey shaded regions are the 68 and 90 percent posterior credible sets, respectively, from the standard identification.
be unanticipated. According to the evidence presented here, the variation measured by the narrative news and excess returns unlikely originate from the source that drives the structurally identified shock. Therefore, the use of these variables does not directly alleviate the predictability problem of the shock of Blanchard and Perotti (2002). The different results rather stem from the sources of exogenous variation used for identification and their validity.

3.4 Conclusions

This essay addressed the question of the implications of a government spending shock in the face of anticipation. Any empirical strategy attempting to quantify these effects confronts an econometric issue with the timing of the shock as the economic agents are likely to have a larger information set than the model assumes. I resolved the issue of deficient information with the noncausal VAR that is able to incorporate fiscal foresight, while it simultaneously retains the advantages of the VAR methodology by imposing few assumptions on the underlying economy. The analysis of fiscal policy could then be employed with a standard identification strategy based on exclusion restrictions, and an anticipated government spending shock could be recovered.

The noncausal VAR methodology deviates from the forecast error interpretation of the residual but – despite anticipation – facilitates the conduct of conventional structural analysis. In a simple model of fiscal foresight, I analytically showed that the noncausal model is able to solve the noninvertibility problem. In a more general setting, the lead terms of the model are expected to capture flexibly anticipation which would be misinterpreted when relying on the lagged observables only. Essentially, the approach does not rule out the causal case a priori as invertibility of the underlying MA representation is nested in the framework.

In the U.S. postwar economy, the estimated spending shock induced an increase in the forward-looking variables during the anticipatory phase. Spending also turned out to be followed by rising consumption, nonresidential investment, worked hours and wages. Together, these movements implied a fiscal multiplier above but close to unity. In addition, anticipation is important to take anticipation into consideration as it affects the measured fiscal multiplier and the overall impact of forward-looking variables. I also revisited two prominent alternative strategies based on measures of spending news that attempt to circumvent the nonfundamentalness problem. Notably, a variable to catch the expectations of economic agents is unlikely to measure the
same variation identified by the exclusion restrictions, although they produce broadly similar results.

Finally, I consider the following areas useful for further research. First, the noncausal approach is readily available for the study of government spending shocks in other economies, as research can be done using conventional macroeconomic data only, without engaging in costly and demanding data collection of proxy variables. Second, the examination of tax policy with the noncausal model, after imposing adequate structure, can be viewed as a useful extension. Finally, the estimation of the model was based on a simple deviation from Gaussianity which, however, assumed cross-dependent volatility for the structural shocks. Furthermore, detrended variables were used because the implications of stochastic trends in the model are yet unknown. Using alternative distributions and allowing for nonstationarity could strengthen the robustness of the approach.

References


89
The effects of government spending under anticipation: the noncausal VAR approach


3.4 Estimation of the noncausal VAR(r, s)

I outline here a version of the Gibbs sampler constructed by Lanne and Luoto (2016). Let \( \Pi \) and \( \Phi \) be matrices stacking \( \Pi_i' \) for \( i = 1, \ldots, r \) and \( \Phi_i' \) for \( i = 1, \ldots, s \), respectively. Furthermore, define the following matrices. First, stack \( iW \) denotes the inverse Wishart distribution. Furthermore, define the following matrices. First, stack \( \Phi \) to satisfy (3.9), introduce an \( ((n^2s - s^*) \times 1) \) vector \( \phi_r \) containing the unrestricted parameters of \( \Phi \) and an \( (n^2s \times (n^2s - s^*)) \) deterministic matrix \( R_\Phi \) which maps the unrestricted parameters to the matrix \( \Phi \) as \( \phi = R_\Phi \phi_r \).

The approximate conditional joint density of \( y = (y_1, \ldots, y_T) \) on \( \omega = (\omega_{r+1}, \ldots, \omega_{T-s}) \) is

\[
p(y|\omega, \theta) \approx \prod_{r+1}^{T-s} p(e_t(\theta)|\omega_t, \Sigma)
\]

with

\[
p(e_t|\omega_t, \Sigma) = \frac{\omega_t^n}{(2\pi)^{n/2}\Sigma^{1/2}} \exp \left( -\frac{1}{2} \omega_t e_t(\theta)' \Sigma^{-1} e_t(\theta) \right),
\]

\[
e_t(\theta) = \nu_t(\phi) - \sum_{j=1}^{r} \Pi_j(\pi) \nu_{t-j}(\phi),
\]

and

\[
\nu_t(\phi) = y_t - \Phi_1(\phi)y_{t+1} - \ldots - \Phi_s(\phi)y_{t+s}.
\]

The prior distributions are set as follows: \( \pi \sim N(\pi, \Sigma_\pi)I(\pi), \phi_r \sim N(\phi_r, V_{\phi_r})I(\phi), \Sigma \sim iW(S, \nu) \) and \( \lambda \sim \text{Exp}(\Lambda) \), where \( \text{Exp}(\cdot) \) is indicator function equal to 1 when the polynomial to which \( \pi \) or \( \phi \) is mapped is stable and \( iW \) denotes the inverse Wishart distribution. Furthermore, define the following matrices. First, stack \( y_t^* = \omega_t^{1/2} \Pi(L)y_t \) to a \((T - r - s)n \times 1 \) vector \( y^* \), and \( X_t^* = \omega_t^{1/2} \Pi(L)X_t \) to a \((T - r - s)n \times sn^2 \) matrix \( X^* \), where \( X_t = I_n \otimes [y_{t+1}' \cdots y_{t+s}'] \). Define similarly matrices \( \mathbf{Y} \) and \( \mathbf{U} \) by stacking \( \nu_t^* = \omega_t^{1/2} \nu_t(\phi) \) and \( U_t^* = \omega_t^{1/2} [\nu_{t-1}(\phi) \cdots \nu_{t-r}(\phi)]' \), respectively, for \( t = r + 1, \ldots, T - s. \)


### Appendices

#### 3.A Estimation of the noncausal VAR(r, s)

I outline here a version of the Gibbs sampler constructed by Lanne and Luoto (2016). Let \( \Pi \) and \( \Phi \) be matrices stacking \( \Pi_i' \) for \( i = 1, \ldots, r \) and \( \Phi_i' \) for \( i = 1, \ldots, s \), respectively. Furthermore, define the following matrices. First, stack \( iW \) denotes the inverse Wishart distribution. Furthermore, define the following matrices. First, stack \( \Phi \) to satisfy (3.9), introduce an \( ((n^2s - s^*) \times 1) \) vector \( \phi_r \) containing the unrestricted parameters of \( \Phi \) and an \( (n^2s \times (n^2s - s^*)) \) deterministic matrix \( R_\Phi \) which maps the unrestricted parameters to the matrix \( \Phi \) as \( \phi = R_\Phi \phi_r \).

The approximate conditional joint density of \( y = (y_1, \ldots, y_T) \) on \( \omega = (\omega_{r+1}, \ldots, \omega_{T-s}) \) is

\[
p(y|\omega, \theta) \approx \prod_{r+1}^{T-s} p(e_t(\theta)|\omega_t, \Sigma)
\]

with

\[
p(e_t|\omega_t, \Sigma) = \frac{\omega_t^n}{(2\pi)^{n/2}\Sigma^{1/2}} \exp \left( -\frac{1}{2} \omega_t e_t(\theta)' \Sigma^{-1} e_t(\theta) \right),
\]

\[
e_t(\theta) = \nu_t(\phi) - \sum_{j=1}^{r} \Pi_j(\pi) \nu_{t-j}(\phi),
\]

and

\[
\nu_t(\phi) = y_t - \Phi_1(\phi)y_{t+1} - \ldots - \Phi_s(\phi)y_{t+s}.
\]

The prior distributions are set as follows: \( \pi \sim N(\pi, \Sigma_\pi)I(\pi), \phi_r \sim N(\phi_r, V_{\phi_r})I(\phi), \Sigma \sim iW(S, \nu) \) and \( \lambda \sim \text{Exp}(\Lambda) \), where \( \text{Exp}(\cdot) \) is indicator function equal to 1 when the polynomial to which \( \pi \) or \( \phi \) is mapped is stable and \( iW \) denotes the inverse Wishart distribution. Furthermore, define the following matrices. First, stack \( y_t^* = \omega_t^{1/2} \Pi(L)y_t \) to a \((T - r - s)n \times 1 \) vector \( y^* \), and \( X_t^* = \omega_t^{1/2} \Pi(L)X_t \) to a \((T - r - s)n \times sn^2 \) matrix \( X^* \), where \( X_t = I_n \otimes [y_{t+1}' \cdots y_{t+s}'] \). Define similarly matrices \( \mathbf{Y} \) and \( \mathbf{U} \) by stacking \( \nu_t^* = \omega_t^{1/2} \nu_t(\phi) \) and \( U_t^* = \omega_t^{1/2} [\nu_{t-1}(\phi) \cdots \nu_{t-r}(\phi)]' \), respectively, for \( t = r + 1, \ldots, T - s. \)
Following Lanne and Luoto (2016), the full conditional posterior distribution of $\phi_r$ can be derived as

$$
\phi_r | y, \pi, \Sigma, \omega \sim N(\bar{\phi}_r, \bar{V}_{\phi_r} I(\phi_r), \phi = R_{\phi} \phi_r)
$$

$$
\bar{V}_{\phi_r}^{-1} = \bar{V}_{\phi_r}^{-1} + R_{\phi}^t X^* \Omega X^* R_{\phi}, \bar{\phi} = \bar{V}_{\phi_r}^{-1} \phi_r + R_{\phi}^t X^* \Omega Y^*
$$

and $\Omega = I_{T-r-s} \otimes \Sigma^{-1}$. The conditional distribution of $\pi$ reads as

$$
\pi | y, \phi, \Sigma, \omega \sim N(\bar{\pi}, \bar{V}_{\pi} I(\pi), \bar{V}_{\pi}^{-1} \pi = \bar{V}_{\pi}^{-1} \phi + \text{vec} \left( U^\prime \Sigma^{-1} \right))
$$

Defining further $\bar{S} = \bar{S} + E^\prime E, E = Y - U\Pi$ and $\bar{v} = v + T - s - r$, the conditional posterior distribution for $\Sigma$ is

$$
\Sigma | y, \pi, \phi, \omega \sim I W(\bar{S}, \bar{v}).
$$

The remaining paremeters $\omega = (\omega_{r+1}, \ldots, \omega_{T-r-s})$ and $\lambda$ are jointly drawn from

$$
\left( \lambda + e_t(\theta)^\prime \Sigma^{-1} e_t(\theta) \right) \omega_t | y, \pi, \phi, \Sigma, \lambda \sim \chi^2(\lambda + n), \ t = r + 1, \ldots, T - s
$$

and with Metropolis-within-Gibbs step from kernel

$$
p(\lambda | y, \omega) \propto \left( 2^{\lambda/2} \Gamma(\lambda/2) \right)^{-T-r-s} \lambda^{(T-r-s)/2} \left( \prod_{t=r+1}^{T-s} \omega_t^{(\lambda-2)/2} \right) \exp \left[ -\left( \frac{1}{\lambda} + \frac{1}{2} T-r+1 \omega_t \right) \lambda \right].
$$

(3.22)

The last step uses the univariate normal distribution with mean equal to the mode and variance equal to the inverse of the second hessian of the above kernel as a candidate distribution. The standard Metropolis-Hastings acceptance probability is computed using (3.22).

In the empirical analysis, I use the following Minnesota type prior distribution. I set the means of $\pi$ and $\phi_r$, $\bar{\pi}$ and $\bar{\phi}_r$, to 0, and the coefficients are
3.B Convergence of the posterior sampler

assumed, a priori, independent by having zeros on the off-diagonals of covariance matrices $\nabla_\pi$ and $\nabla_\phi$. On the other hand, $\sigma_{\pi,ijl}^2$ and $\sigma_{\phi,ijl}^2$, the diagonal elements of $\nabla_\pi$ and $\nabla_\phi$ corresponding to the $l$th lag or lead of variable $j$ in equation $i$ are given by

$$
\sigma_{\pi,ijl} = \frac{\gamma_1,\pi}{\gamma_3} \frac{\gamma_1,\pi}{\gamma_3} \sigma_i \frac{\gamma_1,\pi}{\gamma_3} \sigma_j, \quad i = 1, \ldots, n, \quad j = 1, \ldots, n, \quad l = 1, \ldots, r,
$$

$$
\sigma_{\phi,ijl} = \frac{\gamma_1,\phi}{\gamma_3} \frac{\gamma_1,\phi}{\gamma_3} \sigma_i \frac{\gamma_1,\phi}{\gamma_3} \sigma_j, \quad i = 2, \ldots, n, \quad j = 1, \ldots, n, \quad l = 1, \ldots, s,
$$

where $\sigma_i$ is estimated as the residual standard error from a univariate autoregression with $r$ lags on the $i$th variable, $\gamma_{1,\pi}$ and $\gamma_{1,\phi}$ control for overall tightness, $\gamma_2$ for relative tightness and $\gamma_3$ is a decay parameter for more distant lags and leads. For these hyperparameters, I use values $\gamma_{1,\pi} = 0.2$, $\gamma_2 = 0.5$ and $\gamma_3 = 1$, standard in the Bayesian VAR literature. Additionally, I set $\gamma_{1,\phi} = 0.15$, which shrinks the lead coefficients moderately but somewhat more towards zero. Last, I use the following values for the remaining hyperparameters: $S = (\nu - n - 1) \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ with degrees-of-freedom parameter $\nu = n + 2$ and $\lambda = 10$.

3.B Convergence of the posterior sampler

Figures 3.10 and 3.11 plot the paths of the Markov chains obtained from the Gibbs sampler in the estimation of the baseline model. The model is estimated with 50,000 draws, and every 10,000th draw is started from new initial values with 1,000 burn-in draws. As can be seen from the plots, the sampler converges fast to the ergodic distribution and is invariant to the starting values.

3.C Further empirical results

Figure 3.12 plots results from the baseline specification when the number of leads is changed. In Figure 3.13, I plot the impulse responses of the remaining variables in the additional specification not reported in text. Figure 3.14 plots the shock from the noncausal and causal VAR of the baseline specification over time.
Figure 3.10: Paths of the Markov chains for the draws of elements in $\pi$ and $\phi$ of the baseline noncausal VAR(4,4) model

The x-axes correspond to the draws, the y-axes to the parameter values.
3.C Further empirical results

Figure 3.11: Paths of the Markov chains for the draws of scale matrix and the degrees-of-freedom parameter of the baseline noncausal VAR(4,4) model
The x-axes correspond to the draws, the y-axes to the parameter values

Figure 3.12: Impulse responses of the baseline variables from the noncausal VAR(4,s) models for s = 0, 1, 2, 3, 4.
Impulse responses produced from the noncausal VAR models with different lead lengths. The dark and light grey shaded regions are the 68 and 90 percent, respectively, credible sets of the estimated impulse responses from the VAR(4,4) model.
The effects of government spending under anticipation: the noncausal VAR approach

The graphs show the impulse responses of the remaining variables of the model not shown in Figure 3.4. Solid lines and dashed lines are the median impulse responses from the noncausal VAR(4,4) and causal VAR(4) models, respectively, to a one standard deviation government spending shock. The dark and light grey shaded regions are the 68 and 90 percent, respectively, credible sets of the estimated impulse responses from the noncausal model. The impulse responses in Panel (a) of Figure are from the 10-variable VAR including the baseline variables, inflation and interest rates. The impulse responses in Panel (b) are computed from the 10-variables VAR including the baseline variables but consumption and investment replaced by the variables in Panel (b) of Figure 3.4.

Figure 3.13: Impulse responses of the remaining variables to the government spending shock in the additional specifications
3.D Alternative identification scheme

To identify a shock contributing the most to a news variable, the starting point is the two-sided MA representation of \( y_t \)

\[
y_t = \sum_{k=-\infty}^{\infty} \Psi_k \bar{B} \bar{u}_{t-k},
\]

where matrix \( \bar{B} \) rotates the structural shocks \( \bar{u}_t \) to the reduced-form errors \( \epsilon_t \) as

\[
\epsilon_t = \bar{B} \bar{u}_t.
\]

\( \bar{B} \) can now be found from

\[
\bar{B} \bar{B}' = \bar{\Gamma}
\]

as \( E[\epsilon_t \epsilon_t'] = \bar{\Gamma} = \frac{\lambda}{\lambda^2} \Sigma = E[\bar{B} \bar{u}_t \bar{u}'_t \bar{B}'] = \bar{B} \bar{B}' \). On the other hand, by Cholesky decomposition, \( \bar{\Gamma} = \bar{A} \bar{A}' \), or, by introducing an orthogonal matrix \( W \), \( \bar{\Gamma} = \bar{A} W W' \bar{A}' \). Consequently, rotation of \( W \) yields \( \bar{B} = \bar{A} W \). As the interest is in one shock only, it suffices to find the first column of \( W \), \( w_1 \) such that \( \gamma_1 = \bar{A} w_1 \).
The effects of government spending under anticipation: the noncausal VAR approach

is the first column of $\tilde{B}$.

The MA representation of the $i$th variable in $y_t$ is then

$$y_{i,t} = \sum_{k=-\infty}^{\infty} e_i' \Psi_k \tilde{A} W \tilde{u}_{t-k}$$

with variance

$$\text{Var}(y_{i,t}) = \frac{\lambda}{\lambda - 2} \sum_{k=-\infty}^{\infty} e_i' \Psi_k \tilde{A} \tilde{A}' \Psi_k' e_i$$

$y_{i,t}$ can further be decomposed to the contributions by the $n$ structural shocks

$$y_{i,t} = \sum_{k=-\infty}^{\infty} e_i' \Psi_k \tilde{A} [w_1 \tilde{u}_{1,t-k} + \ldots + w_n \tilde{u}_{n,t-k}]$$

such that the contribution of the first shock to the variable reads as

$$y_{i,t}^1 = \sum_{k=-\infty}^{\infty} e_i' \Psi_k \tilde{A} w_1 \tilde{u}_{1,t-k} = \sum_{k=-\infty}^{\infty} e_i' \Psi_k \tilde{A} w_1 \tilde{u}_{1,t-k}.$$

As the aim is to find a shock with the greatest contribution to the $i$th variable, $w_1$ is found, similar to Uhlig (2004), by maximising

$$\frac{\text{Var}(y_{i,t}^1)}{\text{Var}(y_{i,t})} = \frac{\sum_{k=-\infty}^{\infty} e_i' \Psi_k \tilde{A} w_1 w_1' \tilde{A}' \Psi_k' e_i}{\sum_{k=-\infty}^{\infty} e_i' \Psi_k \tilde{A} \tilde{A}' \Psi_k' e_i}$$

subject to the orthogonality of $W$, $w_1' w_1 = 1$. By rewriting

$$e_i' \Psi_k \tilde{A} w_1 w_1' \tilde{A}' \Psi_k' e_i = \text{tr} (e_i' \Psi_k \tilde{A} w_1 w_1' \tilde{A}' \Psi_k' e_i)$$

$$= \text{tr} (w_1' \tilde{A}' \Psi_k' e_i e_i' \Psi_k \tilde{A} w_1)$$

$$= \text{tr} (w_1' \tilde{A}' \Psi_k' E_{ii} \Psi_k \tilde{A} w_1)$$

$$= \text{tr} (w_1' S_k w_1),$$

the nominator of the objective function is

$$\sum_{k=-H}^{H} w_1' S_k w_1 = w_1' \tilde{S} w_1$$

98
for large $H$. As the denominator is independent of $w_1$, the problem can be solved by setting up the Lagrangian

$$L = w_1' \bar{S} w_1 - \mu (w_1' w_1 - 1).$$

The first-order condition is

$$\bar{S} w_1 = \mu w_1,$$

and since $w_1' \mu w_1 = \mu$, the eigenvector corresponding to the maximal eigenvalue of the positive definite matrix $\bar{S}$ is the optimum.
THE EFFECTS OF GOVERNMENT SPENDING UNDER ANTICIPATION: THE NONCAUSAL VAR APPROACH
4 Evidence on news shocks under information deficiency

4.1 Introduction

In news-driven business cycles, economic agents receive signals about future productivity, creating fluctuations in forward-looking variables before the news materialises. However, dynamics driven by these news shocks likely generates a situation where the information set of economic agents relevant for decision making is broader than what an econometrician observes. Consequently, a conventional vector autoregressive model (VAR) used for empirical validation may produce misleading results about the implications and significance of the news shocks.

Starting from the seminal paper of Beaudry and Portier (2006), the structural VAR (SVAR) methodology identifies the news shock as a shock that changes stock prices on impact but has delayed but persistent effects on total factor productivity (TFP). However, by the expectations of forward-looking agents, the news shocks imply noninvertibility of the theoretical moving average (MA) representation for a typical small number of variables included to the VAR model (Leeper, Walker, and Yang, 2013; Forni, Gambetti, and Sala, 2014). The resulting nonfundamentalness problem prevents obtaining the structural shocks and their impulse responses from a causal autoregressive representation of the observed variables.

---

1 An early version of the essay is released as HECER Discussion Paper No. 415.
2 Beaudry and Portier (2014) review the recent literature on news-driven business cycles.
3 Hansen and Sargent (1991) and Lippi and Reichlin (1994b) provide earlier discussion on the topic. For a more recent review, see Lütkepohl (2014) and Beaudry and Portier (2014). In addition to technology-related news shocks, nonfundamentalness is in the fiscal policy under anticipation.
Evidence on news shocks under information deficiency

Nonfundamentalness or noninvertibility eventually boils down to the fact that the observables do not contain all relevant state variables of the economy (Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson, 2007). The problem can thus be avoided by including enough relevant variables to the VAR which then approximates the underlying MA representation. In the studies of Beaudry and Portier (2006) and Barsky and Sims (2011), nonfundamentalness is absent if the variables included are sufficiently forward-looking to achieve invertibility. Alternatively, augmenting the VAR with factors extracted from large-dimensional macroeconomic data can resolve the information deficiency (Forni and Gambetti, 2014; Forni et al., 2014).

This chapter introduces a new approach to structural analysis of news shocks under insufficient information while still imposing few restrictions on the underlying economic process. In place of leaning on information in the observables only, I consider a noncausal representation that includes lead terms, arising as a result of nonfundamentalness. I make the representation operational by estimating the noncausal VAR model of Lanne and Saikkonen (2013). Under nonfundamentalness, a causal model produces errors that are linear combinations of the past and present shocks (Lippi and Reichlin, 1994b), while the noncausal VAR model filters out, through its distinct lag and lead polynomials, an error term consisting of fundamental shocks anticipated by the economic agents.

As the noncausal VAR nests a causal VAR model by its lag terms, the approach conveniently complements the VAR analysis and facilitates checking fundamentalness of observables for the underlying economy. If data lend support for nonzero lead terms, nonfundamentalness is present and the model involves anticipated shocks. However, to distinguish noncausal and causal representations, it is necessary to deviate from Gaussianity of the error term. As one particular deviation, I use a multivariate t-distribution, which adds a

---

4 The issue of nonfundamentalness can also be avoided by imposing enough theoretical structure such as in Schmitt-Grohé and Uribe (2012) and Görtz and Tsoukalas (2017) who directly estimate theoretical models with news shocks. Plagborg-Møller (2018) and Barnichon and Matthes (2018) have proposed alternative methodologies to estimate possibly noninvertible representations directly. However, these strategies require strong prior information on the propagation mechanism of news shocks, making the results heavily dependent on the structural assumptions. A study of Arezki, Ramey, and Sheng (2017) derives the impulse responses using a narrative approach by collecting proxies for news shocks from oil discoveries.

common volatility term to the Gaussian structural shocks. The noncausal VAR model can then be identified and estimated by maximum likelihood (Lanne and Saikkonen, 2013) or by Bayesian methods (Lanne and Luoto, 2016).

I make two methodological contributions in this essay. First, I show how to conduct structural analysis under nonfundamentalness with the noncausal VAR model. In particular, the error term can be mapped into anticipated structural shocks, and the impulse response analysis is based on the two-sided MA representation of the model. Second, I propose the use of a medium-run identification following Francis, Owyang, Roush, and DiCecio (2014) and Uhlig (2004) to identify a news shock in the noncausal framework. I find the news shock as a shock driving total factor productivity the most. Once the observables induce fundamentalness and causality, the identification scheme reduces to the Max Share approach of Francis et al. (2014) and nests the strategy of Kurmann and Sims (2017) to identify news shocks. Importantly, the identification strategy is consistent with the observation that news shocks have long-lasting effects on TFP. However, unlike the recent literature, it imposes no orthogonality of the shock with productivity but finds the news shock indirectly if the identified shock induces early reactions.

I examine the performance of the approach by means of Monte Carlo simulations of a New Keynesian model augmented with news shocks. When the model implies fundamentalness for observables, the noncausal model is outperformed by the causal model that is capable of replicating the true impulse responses. On the other hand, noncausality arises through choosing lead terms to the VAR model under nonfundamentalness. Moreover, the noncausal model reproduces the theoretical impulse responses while a causal VAR fails to reveal the initial reactions as the identification is based on a misspecified error term strongly weighted by the lagged shocks. Hence, although the nonfundamentalness issue may only slightly distort the results on news shocks in situations with a persistent technological process and the discount factor close to unity as argued by Sims (2012), Beaudry and Portier (2014) and Beaudry, Fève, Guay, and Portier (2015), the performance of a causal model may well be considerably deteriorated when not all relevant variables are included to the VAR.

In the postwar U.S. data, I find support for non-Gaussianity of the error term.
Evidence on news shocks under information deficiency

term of the estimated VAR, which allows me to compare causal and noncausal models. Moreover, noncausality is found to be a strong feature of data. In response to the news shock, investment, hours, consumption and output increase on impact and inflation turns negative. The news shock prompts a continuous, steady improvement of TFP and smooth responses of macroeconomic variables, in contrast to the stronger reactions found by Beaudry and Portier (2006). In this respect, my results are in line with Barsky and Sims (2011) and Forni et al. (2014) who measure limited role for the news shock in the short run.\(^7\) Hence, under information deficiency, a causal VAR model may overemphasise the relevance of news shocks, as also found by Forni et al. (2014), when nonfundamental errors capture the timing of the shock incorrectly.

The chapter is organised as follows. Next, in Section 4.2, I review general results on noncausality and nonfundamentalness, present the methodology based on the noncausal VAR, and propose the identification scheme for finding news shocks. Section 4.3 presents Monte Carlo simulation results. In Section 4.4, I present empirical evidence on news shocks. The last section concludes.

4.2 Theory

When news shocks drive the economy, the past observables may not contain sufficient information to recover the structural shocks of interest, which eventually leads to the existence of a noncausal representation. This section presents an approach to study the effects of news shocks based on a noncausal VAR model. The implied two-sided MA representation of the model facilitates the derivation of impulse responses to the shocks that affect the current state of the economy before being observed by the econometrician but already anticipated by the forward-looking economic agents.

4.2.1 Nonfundamentalness and noncausality

I start by reviewing general results on noninvertibility and nonfundamentalness, and demonstrate how they give rise to noncausality.\(^8\) Consider the

\(^7\)For further empirical results, see Schmitt-Grohé and Uribe (2012), Beaudry et al. (2015), Barsky et al. (2015) and Kurmann and Otrok (2013).

\(^8\)Throughout, I use terms noninvertibility and nonfundamentalness interchangeably.
4.2 Theory

equilibrium of a linearised macroeconomic model for \( k \) observed variables in \( y_t \) with a vector autoregressive moving average (VARMA) representation

\[
A(L)y_t = B(L)u_t,
\]

where \( u_t \) is a vector containing the \( k \) uncorrelated structural shocks driving the economy, and \( E_t[u_{t+j}] = 0 \) when \( j > 0 \) and \( E_t[u_{t+j}] = u_{t+j} \) for \( j \leq 0 \). \( E_t[\cdot] \) denotes the expectation conditional on the information set of the agents. \( A(L) \) and \( B(L) \) are \((k \times k)\) matrix polynomials, with \( L \) the usual lag operator, that determine the unique equilibrium of the model in terms of finite lags up to a truncation. \( A(L) \) is assumed to be stable, implying an MA representation

\[
y_t = A(L)^{-1}B(L)u_t.
\]

When the MA polynomial \( B(L) \) in (4.1) is invertible in the past, i.e. \(|B(z)|\) has no roots inside the unit circle, the structural shocks and the impulse responses can be obtained with a conventional causal VAR(\( p \)) model,

\[
C(L)y_t = \epsilon_t, \quad C(L) = I - C_1 L - \ldots - C_p L^p,
\]

from the reduced-form error term \( \epsilon_t = B_0 u_t \) after imposing identifying restrictions on matrix \( B_0 \).\(^9\) However, under nonfundamentalness, the polynomial \( B(L) \) is noninvertible in the past, implying that there exists no VAR(\( \infty \)) representation to recover the shocks \( u_t \) from the history of \( y_t \) only. In that case, fitting a conventional VAR model to \( y_t \) produces a nonfundamental error term which is a linear combination of the past shocks (Lippi and Reichlin, 1994b; Fernández-Villaverde et al., 2007), distorting conclusions drawn from the estimated impulse responses.

Noninvertibility of the MA polynomial is potentially caused by the existence of news shocks, when the forward-looking agents see exogenous changes not contained in the empirical model. For the sake of illustration, let the observables \( y_t \) contain all state variables except \( k \) uncorrelated exogenous vari-

\(^9\)There may be a truncation error when the inverse of \( B(L) \) is of infinite order, which can be diminished by increasing lag order \( p \).
Evidence on news shocks under information deficiency

By Sims (2002), \( y_t \) has a forward-looking solution

\[
y_t = \Theta_1 y_{t-1} + \Theta_c + \Theta_0 z_t + \Theta_y \sum_{s=1}^{\infty} \Theta_f^{s-1} \Theta_z E_t z_{t+s}.
\]

(4.3)

The exogenous variables are driven by unanticipated shocks when \( z_t = u_t \), and the last term vanishes, leading directly to a VAR(1) representation. In contrast, when agents have foresight on the exogenous variables \( q \) periods ahead, \( z_t = u_t - q \), and the equilibrium is determined by

\[
y_t = \Theta_1 y_{t-1} + \Theta_c + \Theta_0 u_{t-q} + \Theta_y \Theta_z u_{t-q+1} + \Theta_y \Theta_f \Theta_z u_{t-q+2} + \ldots + \Theta_y \Theta_f^{q-1} \Theta_z u_t,
\]

(4.4)

which corresponds to the VARMA representation (4.1) of the model. Strikingly, even though the more distant expected events of \( z_t \) obtain a weaker weight in the forward-looking solution (4.3), the most recent innovation \( u_t \) informative about the future event \( z_{t+q} \) is discounted the heaviest by factor \( \Theta_y \Theta_f^{q-1} \Theta_z \) in (4.4). This reverse discounting easily causes noninvertibility of the MA polynomial \( B(L) = \Theta_0 L^q + \Theta_y \Theta_z L + \ldots + \Theta_y \Theta_f^{q-1} \Theta_z \), the most recent shocks having the least influence on the overall dynamics of \( y_t \).

The noninvertible solution prevents the recovery of the news shock contained in \( u_t \) based on the past and current values of \( y_t \) only. To gain fundamentality for (4.4), an obvious strategy is to include in \( y_t \) variables such as measured proxies for news or other forward-looking variables that do not suffer from the inverse discounting of the shock term. By the strategy, \( B(z) \) would eventually become invertible. However, it may be difficult to come up with suitable forward-looking variables and ascertain their validity. The error term \( \epsilon_t \) and the structural shocks \( u_t \) could alternatively be restored from the nonfundamental error term of a causal VAR by a known Blaschke matrix (Lippi and Reichlin, 1994b). Unfortunately, this dynamic rotation is not unique, and the set of Blaschke matrices can be shrunk only by means of economic theory (See, e.g., Mertens and Ravn, 2010; Forni, Gambetti, Lippi, 2014).

10This illustration follows Leeper et al. (2013). See also, Walker and Leeper (2011), Forni and Gambetti (2014), Forni et al. (2014), Beaudry and Portier (2014) and Sims (2012).

11Matrices \( \Theta_1, \Theta_c, \Theta_0, \Theta_y, \Theta_f, \Theta_z \) are functions of parameters of the model of dimensions \((k \times k), (k \times 1), (k \times k), (k \times m), (m \times m) \) and \((m \times k)\), respectively. \( m \) is a dimension of the unstable block of the system, defined in Sims (2002).
and Sala, 2017), which may be infeasible or set restrictive assumptions on the underlying structure.

As an alternative to the above approaches, it is possible to rewrite the VARMA model (4.1) under noninvertibility as a noncausal autoregressive representation for $y_t$, representing the structural shocks in terms of past and future terms. Let $l$ roots of $|B(z)|$ lie within the unit circle. Then $y_t$ is noncausal as

$$\check{c}_l \beta(L)^{-1} B^\text{adj}(L) \alpha(L^{-1})^{-1} A(L) y_t = u_{t-l},$$

(4.5)

where $\check{c}_l$ is constant, $B^\text{adj}(z)$ is the adjoint matrix of $B(z)$, and $\alpha(z^{-1})^{-1}$ and $\beta(z)^{-1}$ are scalar convergent power series expansions in $z^{-1}$ and $z$, respectively (see Appendix 4.A for details). Through the lead polynomial $\alpha(z^{-1})^{-1}$, the time-shifted structural shocks $u_{t-l}$ are functions of the past, current and future terms of $y_t$. While the history of $y_t$ lacks information to catch the variation of $u_t$, movements of the lagged shocks are captured by a linear weighted sum of the past and future values of $y_t$.

Hence, both lags and leads of observables are sufficient to recover the structural shocks that are now anticipated due to the time-shifting.

### 4.2.2 Noncausal VAR

Noncausality implied by nonfundamentalness facilitates the recovery of an anticipated but exogenous error term and the derivation of impulse responses to structural shocks. However, direct inference on the noncausal representation (4.5) is infeasible. In this section, I present the noncausal VAR model proposed by Lanne and Saikkonen (2013) which I use to make inference on the propagation of the news shock.

The noncausal VAR model of Lanne and Saikkonen (2013),

$$\Pi(L)\Phi(L^{-1}) y_t = \epsilon_t,$$

(4.6)

includes distinct lag and lead polynomials $\Pi(z) = I_k - \Pi_1 z - \ldots - \Pi_r z^r$ and $\Phi(z^{-1}) = I_k - \Phi_1 z^{-1} - \ldots - \Phi_s z^{-s}$. The error term $\epsilon_t$ is independent and identically distributed (iid) with zero mean and positive definite covariance

---

12To establish an exact mapping between (4.1) and (4.5), an infinite number of terms has to be included. However, as the more distant terms of the both scalar polynomials converge to zero, a finite number of terms are sufficient to obtain the structural shocks up to a truncation error. If all roots of $|B(z)|$ within the unit circle are equal to zero, the noncausal representation is finite in its leads.
Evidence on news shocks under information deficiency

Hence, the observed variables are written in a form with separate past and future-relevant parts. In the rest of the essay, model (4.6) is referred to as VAR(r, s). If \( s = 0 \) or \( \Phi_i = 0, i = 1, \ldots, s \), the model reduces to a standard causal VAR(r).

To guarantee stationarity and the existence of an MA representation, the following stability conditions hold:

\[
\text{det} \Pi(z) \neq 0, |z| \leq 1 \text{ and } \text{det} \Phi(z) \neq 0, |z| \leq 1,
\]

i.e. the polynomials \( \Pi(z) \) and \( \Phi(z) \) have well-defined inverses convergent in the powers of \( z \). The process \( \Phi(L^{-1})y_t = y_t - \Phi_1 y_{t+1} - \ldots - \Phi_s y_{t+s} \) is by the former condition stationary and has an MA representation

\[
\Phi(L^{-1})y_t = \Pi(L)^{-1} \epsilon_t = M(L) \epsilon_t = \sum_{j=0}^{\infty} M_j \epsilon_{t-j}.
\] (4.7)

The process \( y_t \) can also be decomposed as

\[
y_t = \Phi_1 y_{t+1} + \ldots + \Phi_s y_{t+s} + \sum_{j=0}^{\infty} M_j \epsilon_{t-j} = f_t + \sum_{j=0}^{\infty} M_j \epsilon_{t-j},
\] (4.8)

which highlights the dependence of \( y_t \) on the future through the lead terms \( \Phi_i, i = 1, \ldots, s \). Conveniently, when \( y_t \) is fundamental, the lead coefficients are zeros and the decomposition (4.8) reduces to an MA representation of the causal VAR model (4.2). A nonzero future-dependence \( f_t \) in (4.8) indicates instead the insufficiency of the lags to recover the structural shocks. Consequently, \( \epsilon_t \) consists of anticipated shocks – lagged and current components of \( u_t \) – and is generally different from the fundamental error term \( \epsilon_t \) recovered by a valid VAR model.

It should, however, be emphasised that no direct mapping between a non-invertible and noncausal model exists in general, but the latter may include representations which lack economic interpretation. To prevent the emergence of such representations, it is possible to set parameter restrictions on the lead coefficients of the model. Specifically, I refine the set of noncausal representation by assuming that total factor productivity \( y_{1,t} = a_t \), ordered as the first variable in \( y_t \), has a future-dependent term \( f_{1,t} \) of the form

\[
f_{1,t} = \phi_{11,1} a_{t+1} + \ldots + \phi_{11,s} a_{t+s},
\] (4.9)
which is the first element of \( f_t \) in (4.8). Hence, total factor productivity is backward-looking with respect to the other observables but may be noninvertible with respect to its own innovation.\(^{13}\)

Finally, inverting the stable polynomial \( \Phi(L^{-1}) \) in (4.7) produces a two-sided MA representation for \( y_t \):

\[
y_t = \sum_{j=-\infty}^{\infty} \Psi_j \epsilon_{t-j}.
\] (4.10)

Thus \( y_t \) depends, in general, both on the past and future error terms. Given nonzero \( \Psi_j \) resulting from lead terms in \( \Phi(L^{-1}) \) different from zero, \( \epsilon_t \) has an effect on \( y_t \) both before and after period \( t \). The effect of more distant error terms disappears since \( \Psi_j \) converges to zero when \( j \to \infty \) or \( j \to -\infty \).

If the error term of the model is Gaussian, a noncausal \( \text{VAR}(r,s) \) is observationally equivalent to a causal \( \text{VAR}(r+s) \) model as they cannot be statistically distinguished by the properties of the first and second moments alone.\(^{14}\) Therefore, estimation of the noncausal model necessarily requires departure from Gaussianity. In what follows, the error term is assumed to be multivariate t-distributed, i.e. it can be characterised by \( \epsilon_t = \omega_t^{-1/2} \eta_t, \eta_t \sim N(0, \Sigma) \) and \( \lambda \omega_t \) is \( \chi^2_\lambda \)-distributed. As a consequence,

\[
\omega_t^{1/2} \Pi(L) \Phi(L^{-1}) y_t = \eta_t.
\]

Accordingly, the non-Gaussian assumption adds a stochastic volatility factor \( \omega_t^{1/2} \) that affects \( y_t \) in addition to the normally distributed error term \( \eta_t \). Conditional on \( \omega_t \), the error term is Gaussian, but unconditionally, higher moments of data are relaxed to be determined by the degrees-of-freedom parameter \( \lambda \). For small \( \lambda \), the variables exhibit leptokurtic pattern, as also observed in economic time series.\(^{15}\) On the other hand, when \( \lambda \) increases, the distribution approaches the Gaussian distribution. Estimating \( \lambda \) thus allows

---

\(^{13}\)This type of restriction is not fully exhaustive in ruling out representations that are unlikely to be generated by the underlying noninvertible model. Nevertheless, experience from Monte Carlo simulations suggests that imposing such restrictions performs well in practice.

\(^{14}\)This nonidentifiability holds for noninvertible models as well. See, e.g., Rosenblatt (2000).

\(^{15}\)Various studies consider non-Gaussian distributions using macroeconomic data. Distribution of growth rates of output in OECD have been observed to be fat-tailed by Fagiolo et al. (2008). The estimation results of Cúrdia et al. (2014) and Chib and Ramamurthy (2014) suggest a t-distribution for innovations in DSGE models in the low-frequency data. See also Ascari et al. (2015) for fat-tailed distributions in macroeconomic time series.
for checking the validity of the departure from normality.

The model can be estimated by maximising the log-likelihood function using standard inference, as shown by Lanne and Saikkonen (2013), or by Bayesian methods (See Appendices 4.B and 4.F.1 for details). The latter approach is particularly useful as the prior distribution of VAR coefficients can then be shrunk towards zero when the number of parameters is large. Finally, selecting orders $r$ and $s$ can be based on conventional information criteria by comparing all nested $\text{VAR}(r,s)$ models satisfying $r + s \leq p_{\text{max}}$ with $s \geq 0$ and $r > 0$, i.e. causality is not ruled out in advance. Alternatively, a VAR model with a sufficient number of lag and lead terms directly covers both causal and noncausal dynamics and, under fundamentalness, the lead terms become innocuous.

Consequently, selecting $s > 0$ or estimating significant values for lead coefficients directly suggest, by (4.8), the inadequacy of the causal VAR and its invertible MA representation to capture the underlying shocks. In this respect, the estimation of lead terms can be seen as a test for nonfundamentalness, similar to the more general framework of Chen et al. (2017). Independent of the conclusion of a test, however, the noncausal VAR is able to reproduce the underlying impulse response function via its MA representation (4.10).

### 4.2.3 Noncausality in a stylised model with news shocks

Next, I illustrate in a stylised, two-variable rational expectations model similar to that used by Beaudry and Portier (2014), how noninvertibility and noncausality arise when news shocks drive exogenous technology. This leads to an exact mapping between the noncausal representation (4.5) and the noncausal VAR model (4.6) that reproduces the true impulse responses. Let the equilibrium conditions for an endogenous variable $x_t$ and a technology process $a_t$ be

$$
a_t = \rho a_{t-1} + \epsilon^a_{t-2}, \quad (4.11)
$$

$$
x_t = \beta E_t[x_{t+1}] + a_t + v_t, \quad (4.12)
$$

where $E_t[\cdot]$ denotes conditional expectation with respect to the information set containing history of $\{a_t, x_t, \epsilon^a_t, v_t\}$. $\epsilon^a_t$ and $v_t$ are mutually uncorrelated structural shocks with $\epsilon^a_t$ being a news shock that affects productivity two periods later and $v_t$ an unexpected nominal shock on $x_t$. 

110
4.2 Theory

Assuming $\beta < 1$, $x_t$ has a forward-looking solution

$$x_t = \sum_{j=0}^{\infty} \beta^j E_t a_{t+j} + v_t$$

(4.13)

and together with (4.11),

$$x_t = \theta a_t + \theta \beta^2 \varepsilon_t^a + \theta \beta \varepsilon_{t-1}^a,$$

(4.14)

where $\theta = (1 - \beta \rho)^{-1}$. Thus, $a_t$ and $x_t$ have a VARMA representation

$$\begin{bmatrix} 1 - \rho L & 0 \\ -\theta \rho L & 1 \end{bmatrix} \begin{bmatrix} a_t \\ x_t \end{bmatrix} = \begin{bmatrix} \theta \beta^2 + \theta \beta L + \theta L^2 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_t^a \\ v_t \end{bmatrix} = B(L)u_t,$$

(4.15)

which is noninvertible in the past since $|B(z)| = z^2 = 0$ for $z = 0$. Hence, the history of $y_t = (a_t, x_t)'$ is insufficient to recover the structural shocks $u_t = (\varepsilon_t^a, v_t)'$, and impulse responses cannot be derived from a causal VAR representation for $y_t$.

However, as in (4.5), $y_t$ can be written as noncausal in terms of an anticipated error term. By (4.11), $\varepsilon_t^a$ and $\varepsilon_{t-1}^a$ reveal productivity perfectly two periods forward, and future terms of productivity can be directly substituted to $x_t$ in (4.14) using $\varepsilon_t^a = -a_{t+2} + \rho a_{t+1}$ such that

$$x_t = \rho a_{t-1} + \beta a_{t+1} + \theta \beta^2 a_{t+2} + \varepsilon_{t-2}^a + v_t.$$

Hence, $x_t$ is noncausal with finite leads. Together with (4.11), $y_t$ is equivalently expressed as

$$\begin{bmatrix} I_2 \left[ \rho \ 0 \right] L \end{bmatrix} \begin{bmatrix} I_2 \left[ \beta \ 0 \right] L^{-1} + \left[ \theta \beta^2 \ 0 \right] L^{-2} \end{bmatrix} \begin{bmatrix} a_t \\ x_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{t-2}^a \\ v_t \end{bmatrix},$$

(4.16)

which is the noncausal VAR(1,2) model (4.6) with the right-hand side error term containing the anticipated shock $\varepsilon_{t-2}^a$.

Since both matrix polynomials on the left-hand side of (4.16) are stable, $y_t$ has the two-sided MA representation (4.10). I show in Appendix 4.C.1 that this representation collects the impulse response coefficients to the structural shocks $\varepsilon_{t-2}^a$ and $v_t$, which analytically coincide with their theoretical counterparts. In Figure 4.1, I plot the theoretical and empirical impulse responses.
Evidence on news shocks under information deficiency

of \( a_t \) and \( x_t \) in the upper and lower panels, respectively, when \( \beta = 0.9 \) and \( \rho = 0.9 \). Comparing the panels reveals that the noncausal VAR reproduces the theoretical impulse responses with respect to the anticipated shock \( \epsilon_{t-2}^a \). This shock realises in \( a_t \) at lag 0 but affects \( x_t \) at the negative lags due to terms \( \epsilon_{t-1}^a \) and \( \epsilon_t^a \) in (4.14). Importantly, the theoretical and empirical responses differ only in the timing of the shock \( \epsilon_t^a \), and the inclusion of negative lags allows to track the full propagation of the shock. Despite this time-shifting which occurs due to the absence of a causal representation, the impulse responses can be interpreted from the two-sided MA representation in a conventional manner, and they provide conclusions consistent with the underlying economic process.

This example established an exact mapping from the theoretical model under noninvertibility to the noncausal VAR (4.6). In Appendix 4.C.2, I show that this mapping holds for a more general technology process as well. However, due to the multiplicative structure of the noncausal VAR, the equivalence does not hold in general for all noninvertible models as the former does not cover all possible noncausal representations. Nevertheless, as confirmed in the Monte Carlo simulations of the next section, the empirical model approximates the theoretical representation accurately enough.

4.2.4 Identification of news shocks

I turn now to the structural analysis and identification of a news shock with the noncausal VAR, which is the main contribution of the essay.\(^{16}\) The causal SVAR methodology identifies the news shock as a shock with an immediate effect on forward-looking variables, most notably investment and stock prices, followed by a delayed, long-term impact on total factor productivity (TFP). However, the strategy fails to recover the true shock under nonfundamentalness. The identification scheme I propose extracts from the reduced-form error of a noncausal VAR model a shock that explains the most of the movements in productivity at a long but finite horizon.

Decompose the anticipated, multivariate t-distributed error term of the noncausal VAR to structural shocks as

\[
\epsilon_t = \omega_t^{-1/2} \eta_t = \bar{B} \bar{u}_t,
\]

\(^{16}\)Prior to this study, the noncausal VAR model has not been used for identifying structural shocks. Davis and Song (2012) also conduct impulse response analysis based on the two-sided MA representation. However, with the recursive identification through Cholesky decomposition they use, it is difficult to draw a structural interpretation.
Figure 4.1: Theoretical and empirical impulse response functions of the example \( \beta = 0.9 \) and \( \rho = 0.9 \). The upper row shows the theoretical impulse responses of \( a_t \) and \( x_t \) to shocks \( \epsilon^*_t \) and \( \nu_t \). The lower row plots the impulse responses obtained from the MA representation of the noncausal VAR to the shocks of the error term, \( \epsilon^*_{t-2} \) and \( \nu_t \).

where the t-distributed structural shock vector, \( \bar{u}_t = \omega_t^{-1/2} u^*_t = [\bar{u}_{1,t} \cdots \bar{u}_{k,t}]' \sim t_{\lambda}(I_k) \), is a product of two latent factors, a \( k \)-dimensional vector of Gaussian shocks \( u^*_t \sim N(0,I_k) \) and the volatility term \( \omega_t^{-1/2} \). Hence, compared to the standard Gaussian setting, the term \( \omega_t^{-1/2} \) is added to the normally distributed structural shocks in \( u^*_t \) to control for overall volatility. As a whole, the structural shocks are mapped by a rotation matrix \( \bar{B} \) to the reduced-form error term \( \epsilon_t \). From the two-sided MA representation (4.10), a reaction to a unit shock in \( \bar{u}_{i,t} \) is

\[
\frac{\partial y_{t+j}}{\partial \bar{u}_{i,t}} = \Psi_j \gamma_{i}, \; j = \ldots, -1,0,1,\ldots, \tag{4.18}
\]

where \( \gamma_{i} \) denotes the \( i \)th column of matrix \( \bar{B} \).

The identification of structural shocks follows the standard strategy by finding a nonsingular matrix \( \bar{B} \) such that \( E[\epsilon_t \epsilon'_t] = E[\bar{B} \bar{u}_t \bar{u}'_t \bar{B}'] \), or

\[
\Sigma = \bar{B} \bar{B}'. \tag{4.19}
\]
Furthermore, let $W$ be an orthogonal $(k \times k)$ matrix with $w_i$ its $i$th column, and $\tilde{A}$ a Cholesky factor satisfying $\Sigma = \tilde{A} \tilde{A}'$. Assume now the news shock $\bar{u}_{1,t}$ is ordered first in $\bar{u}_t$. By rewriting the rotation matrix as $B = \tilde{A}W$, the impulse responses to the news shock are obtained by solving for $w_1$ such that $\gamma_1 = \tilde{A}w_1$.

Finding $w_1$ builds in turn on the view of Beaudry and Portier (2006) that the news shock induces a persistent, long-term change in productivity. For this purpose, consider the MA representation (4.10) for TFP, the first variable in $y_t$,

$$y_{1,t} = \sum_{j=-\infty}^{\infty} e_1' \Psi_j \tilde{A} \bar{u}_{t-j}$$

with the vector $e_i = [0 \ldots 1 \ldots 0]'$ having one in its $i$th element. As $W\bar{u}_{t-j} = w_1\bar{u}_{1,t} + \ldots w_k\bar{u}_{k,t}$, TFP is a sum of the contributions of the $k$ structural shocks

$$y_{1,t} = y^1_{1,t} + \ldots + y^k_{1,t}, \quad y^i_{1,t} = \sum_{j=-\infty}^{\infty} e_1' \Psi_j \tilde{A} w_i \bar{u}_{i,t-j}.$$ (4.21)

Hence, the share of variance of $y_t$ due to a news shock over a horizon $[H_1, H_2]$ reads as

$$\Omega^{[H_1, H_2]}_{1,y_{1,t}} = \frac{E[\sum_{j=H_1}^{H_2} e_1' \Psi_j \tilde{A} w_i \bar{u}_{i,t-j} \bar{u}_{i,t-j}' w_i' \tilde{A}' \Psi_j' e_1]}{E[\sum_{j=H_1}^{H_2} e_1' \Psi_j \Sigma \Psi_j' e_1]} = \frac{\sum_{j=H_1}^{H_2} e_1' \Psi_j \tilde{A} w_i w_i' \tilde{A}' \Psi_j' e_1}{\sum_{j=-\infty}^{\infty} e_1' \Psi_j \Sigma \Psi_j' e_1},$$ (4.22)

which reduces to the share of the forecast error variance if $y_t$ is causal. Now, I find the single component among the possible orthogonal structural shocks weighting the most on the variance of TFP over the finite horizon $[H_1, H_2]$. In other words, $\bar{u}_{1,t}$ emerges as a shock that accounts for the largest share of the variance of $y_{1,t}$. Formally, $w_1$ is found by maximising (4.22) subject to the orthogonality of $W$, $w_1'w_1 = 1$. Appendix 4.D shows the maximisation problem has a unique solution as an eigenvalue problem.

The above strategy is a close variant of the medium-run identification approach of Francis et al. (2014), which builds on Uhlig (2004). For studying news shocks, similar identification schemes have been employed by Barsky and Sims (2011), Kurmann and Otrok (2013), Forni et al. (2014) and Kurmann and Sims (2017). However, instead of the forecast error variance decomposition, I use the share of unconditional variance (4.22) explained by one shock.
As soon as the VAR model is causal, the identification scheme coincides with the Max Share approach of Francis et al. (2014).

In contrast with the existing literature, I do not impose orthogonality of the shock to current productivity, motivated by two insights. First, due to the two-sided MA representation (4.10), identifying restrictions on the impact effects are generally difficult to implement in the noncausal VAR as the timing of the shock is unknown a priori. Second, as extensively discussed in Kurmann and Sims (2017), relaxing the impact effects is potentially more robust to revisions and noise in the TFP measure as well as to a small sample bias involved in long-run identification schemes.

Essentially, the identification is consistent with the view that permanent changes in productivity stem mainly, but not solely, from a single shock that is interpreted as the news shock if it has anticipatory effects on the forward-looking variables. When the impact effect on TFP is negligible and the observables induce fundamentalness, the shock is equivalently recovered by the approach of Barsky and Sims (2011), where the news shock maximises the share of forecast error variance and is orthogonal to current productivity.\footnote{The news shocks identified by the strategy may also have impact effects on productivity and in fact nest both smoothly diffusing technology shocks as in Lippi and Reichlin (1994a) or Walker and Leeper (2011) as well as shocks with delayed effects on TFP as mostly considered in the VAR literature starting from Beaudry and Portier (2006).} Finally, it is worth noting that the proposed identification of news shocks does not exclude the existence of surprise technology changes. The latter shocks need, however, be relatively less important for the variation in total factor productivity in the medium run.

Overall, the key advantage of the approach outlined here is its invariance to the information in the observables. Estimating a causal VAR model under nonfundamentalness instead produces reduced-form errors from which any static identification scheme extracts misspecified structural shocks. In contrast, the identification in the noncausal VAR remains valid independent of nonfundamentalness as the model nests a causal VAR model by its lag terms.

### 4.3 Monte Carlo simulation

In this section, I analyse with Monte Carlo simulations how causal and noncausal VAR models are able to identify news shocks and match the true impulse responses of a small-scale macroeconomic model.
4.3.1 New Keynesian model with news shocks

Consider a textbook New Keynesian (NK) model (See Galí, 2009). The equilibrium is characterised by the dynamic IS equation

$$\tilde{y}_t = -\frac{1}{\sigma} (i_t - E_t \pi_t - r^n_t) + E_t \tilde{y}_{t+1},$$

where $\tilde{y}_t$ is the output gap relative to the flexible price outcome, $\pi_t$ is inflation, $i_t$ is the interest rate and $r^n_t = \rho + \sigma \psi_n E_t \Delta a_{t+1}$ is the natural interest rate. The NK Phillips curve determines inflation:

$$\pi_t = \beta E_t [\pi_{t+1}] + \kappa \tilde{y}_t,$$

and the central bank sets the nominal interest rate according to the Taylor rule

$$i_t = \rho (1 - \rho_m) + \rho_m i_{t-1} + \phi_\pi \pi_t + \phi_y \tilde{y}_t + \epsilon_t^m.$$

The economy is driven by three exogenous shocks, the anticipated and unanticipated shocks on technology, $\epsilon_t^a$ and $\epsilon_t^u$, respectively, and the monetary policy shock $\epsilon_t^m$. Technology follows

$$a_t = \rho_a a_{t-1} + \epsilon_t^a - q + \sigma_u \epsilon_t^u$$

so through the news shock $\epsilon_t^a$, agents are able to predict technology $q$ periods forward. The surprise technology, or noise shock $\epsilon_t^u$ prevents the agents from perfectly observing future technology and $\sigma_u$ determines its relative importance.

The inclusion of the news shock alters the dynamics of the sticky-price model as follows. In response to a positive news shock, firms anticipate the future improvement in productivity and have an incentive to decrease prices beforehand. As a result, the economy initially experiences a boom with positive output gap, which accelerates inflation. To stabilise inflation and output gap, the Taylor rule responds by increasing the interest rate. Consistent with sticky-price models, the news shock at the time of its materialisation eventually turns the output gap and inflation negative, similar to the effects of a surprise technology shock in a basic NK model.
4.3.2 Simulation design

I solve the model numerically using standard calibration (see Appendix 4.E.1). In addition, I vary the significance of news shocks and the anticipation horizon in the model by considering values $\sigma_u \in \{0.5, 1\}$ and $q \in \{3, 16\}$. These values are motivated as follows. First, increasing the standard deviation of the noise shock from $\sigma_u = 0.5$ to $\sigma_u = 1$ decreases the significance of news shocks in driving technology. The nonfundamentalness-inducing shock comprises then a smaller share of exogenous variation, which may alleviate the noninvertibility problem. Second, compared to the benchmark anticipation horizon $q = 3$ also used by Sims (2012), the longer anticipation horizon $q = 16$ is more consistent with the empirical estimates starting from Beaudry and Portier (2006) who measure news shocks moving the economy several years ahead. However, lengthening the horizon relates news to more distant changes in technology, which worsens the ability of the current observables to recover the underlying shocks and to replicate the true impulse responses from the causal model.

I simulate from the model 1,000 samples of time series of length $T = 250$ with structural shocks $u_t = (\varepsilon_a, \varepsilon_m, \varepsilon_u)'$ drawn from multivariate t-distribution $t_{\lambda, (\mu^{\omega^{-1/2}} I_3)}$ with $\lambda = 4$. The structural shocks are therefore uncorrelated and Gaussian conditional on the latent volatility factor $\omega_t^{-1/2}$. Due to the low degrees-of-freedom parameter $\lambda$, the distribution of shocks has more weight on tails than in the purely Gaussian case, extreme shocks being more frequent. For each simulated sample, I estimate both causal and noncausal VAR models with two sets of observables, $y_1^t = (a_t, y_{1t}, \pi_t)$ and $y_2^t = (a_t, i_t, \pi_t)$ as the number of included state variables influences the nonfundamentalness issue. The first set includes two forward-looking variables, inflation and output gap, whereas the latter includes two state variables, technology and the interest rate. The models are estimated by maximum likelihood with the restrictions (4.9) imposed on productivity $a_t$.

---

18 However, see Sims (2012) for a contrasting view.
19 The term $\mu \omega^{-1/2} = \sqrt{\frac{\lambda-2}{\lambda}}$ induces a unit variance for $(\varepsilon_a, \varepsilon_m, \varepsilon_u)'$.
20 Since structural shocks are uncorrelated, the additional assumption compared to the Gaussian case has no consequences to the equilibrium conditions of the DSGE model up to the first order approximation.
Evidence on news shocks under information deficiency

4.3.3 Nonfundamentalness and model selection

In Table 4.1, I summarise statistics from the simulated models. I report the maximum absolute eigenvalue of the invertibility condition of Fernández-Villaverde et al. (2007), computed from the theoretical state space representation for the observables. When this eigenvalue is greater than one in modulus, the observables suffer from nonfundamentalness.\(^{21}\) Examining the nonfundamentalness issue further, I estimate all VAR\((r, s)\) models satisfying \(r + s \leq p_{\text{max}} = 6\) with \(s \geq 0\) and \(r > 0\) and compare them by the Akaike Information Criterion (AIC) in each simulated sample.\(^{22}\) The shares of selected specifications are collected in Table 4.1.

For the anticipation horizon \(q = 3\), Table 4.1 shows that including two forward-looking variables, output gap and inflation, in \(y_t\) is insufficient to attain invertibility, the maximum absolute eigenvalue remaining greater than one. The absence of an invertible MA representation is consequently reflected by primarily selecting models with lead terms. In turn, the invertibility condition is satisfied for observables \(y_t^2\). By replacing one forward-looking variable from the observable vector by a state variable, the interest rate, the causal VAR remains valid with respect to the structural shocks. In line with this fact, fundamental observables \(y_t^2\) induce the dominance of causal specifications over noncausal variants in the model selection. When lengthening the horizon to \(q = 16\), the interest rate is, however, unable to recover the state space, which now includes 16 lags of \(\varepsilon^a_t\), and the noninvertibility problem emerges irrespective of the used variables. In fact, avoiding nonfundamentalness can only be achieved, if feasible, by including a sufficiently precise proxy for the news shock \(\varepsilon^a_{t'}\), which would revert an invertible state space representation. As a consequence of nonfundamentalness, over 90 percent of the selected models are noncausal.

Although the above model selection performs well in checking the sufficiency of information, it has a slight bias towards including additional lead terms under fundamentalness. Hence, the procedure cannot be regarded as a formal test for nonfundamentalness but it rather provides first-hand infor-

---

\(^{21}\)That is, under a state space representation with \(x_t\), an \((n \times 1)\) vector of state variables,

\[
\begin{align*}
x_t &= Ax_{t-1} + Bu_t, \\
y_t &= Cx_{t-1} + Du_t,
\end{align*}
\]

where \(A, B, C\) and \(D\) are \((n \times k)\), \((n \times k)\), \((k \times n)\) and \((k \times k)\) matrices, respectively, \(y_t\) is fundamental, if matrix \(F = A - BD^{-1}C\) has all eigenvalues inside the unit circle.

\(^{22}\)Increasing the maximum number of lags and leads does not change the results.
4.3 Monte Carlo simulation

<table>
<thead>
<tr>
<th>q</th>
<th>3</th>
<th>16</th>
<th>3</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>σ_a</td>
<td>0.5</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Observables</th>
<th>y_1^1</th>
<th>y_1^2</th>
<th>y_1^3</th>
<th>y_1^4</th>
<th>y_2^1</th>
<th>y_2^2</th>
<th>y_2^3</th>
<th>y_2^4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum eigenvalue</td>
<td>14.13</td>
<td>0.71</td>
<td>1.59</td>
<td>1.37</td>
<td>14.13</td>
<td>0.71</td>
<td>1.59</td>
<td>1.37</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Orders</th>
<th>Specification selected by AIC (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>r = 1</td>
<td>s = 0</td>
</tr>
<tr>
<td>r = 2</td>
<td>s = 0</td>
</tr>
<tr>
<td>r = 3</td>
<td>s = 0</td>
</tr>
<tr>
<td>r = 4</td>
<td>s = 0</td>
</tr>
<tr>
<td>r = 5</td>
<td>s = 0</td>
</tr>
<tr>
<td>r = 6</td>
<td>s = 0</td>
</tr>
<tr>
<td>r = 1</td>
<td>s &gt; 0</td>
</tr>
<tr>
<td>r = 2</td>
<td>s &gt; 0</td>
</tr>
<tr>
<td>r = 3</td>
<td>s &gt; 0</td>
</tr>
<tr>
<td>r = 4</td>
<td>s &gt; 0</td>
</tr>
<tr>
<td>r = 5</td>
<td>s &gt; 0</td>
</tr>
</tbody>
</table>

| Causal (%) | 1.2 | 79.5 | 3.1 | 0.0 | 3.1 | 78.0 | 3.3 | 0.0 |
| Noncausal (%) | 98.8 | 20.5 | 96.9 | 100.0 | 96.9 | 22.0 | 96.7 | 100.0 |

(r_{AIC}^{r,s}, s_{AIC}^{r,s}) (1,2) (5,0) (5,1) (1,2) (1,2) (5,0) (5,1) (1,2)

p_{AIC}^{r,s} 3 5 6 3 3 5 6 3

Table 4.1: Estimation results for the simulated models: causal and noncausal specifications selected by AIC

Percentages refer to the frequency of a VAR(r,s) model being selected from minimising the Akaike information criterion. Models are estimated by maximum likelihood with multivariate t-distributed residuals and restrictions (4.9) on a_t for 1,000 simulated Monte Carlo samples with T = 250. Noncausal specification with r = r,* s,* > 0 nests all leads satisfying r* + s* ≤ 6, (r_{AIC}^{r,s}, s_{AIC}^{r,s}) is the most selected model with s ≥ 0, p_{AIC}^{r,s} the most selected causal model (s = 0).

4.3.4 Impulse responses

In the simulated samples, I estimate the impulse responses to the news shock from the causal VAR(p) and noncausal VAR(r,s) models using the orders of
Evidence on news shocks under information deficiency

Figure 4.2: Impulse responses to a news shock from a causal and noncausal VAR in the NK model with \( q = 3 \) and \( \sigma_u = 0.5 \)

The estimates are the periodwise medians of impulse responses from the Monte Carlo samples. In the causal VAR, the dashed lines and solid marked lines refer to the Barsky-Sims (B-S) and Max Share (MS) identification, respectively. The light and dark grey shaded areas border the middle 90 and 68 percent, respectively, of the distribution for the estimated impulse responses. The solid lines are the theoretical impulse responses, aligned in panel (b) with the estimated noncausal impulse responses according to the maximum impact on technology.

In the causal models, the news shock is identified by two alternative strategies, the Barsky-Sims strategy (Barsky and Sims, 2011), where the shock has no impact effect on technology, and by the Max Share approach of Subsection 4.2.4. These schemes are consistent with the theoretical model as long as observables are fundamental for the underlying process and \( \sigma_u < 1 \). Specifically, the latter strategy does not impose a zero impact effect on technology. If a noncausal VAR is selected, I identify the news shock as described in Subsection 4.2.4: the news shock is a shock explaining the most of the movements in technology \( y_{1,t} = a_t \) over a finite horizon \([H_1, H_2]\) with \( H_1 = -20 \) and \( H_2 = 40 \).

Panel 8a) of Figure 4.2 plots the impulse responses of \( y_{1,t} \) to a news shock from the estimated VAR(3) models for \( q = 3 \) and \( \sigma_u = 0.5 \). As an immediate

---

23In the former, the shock maximises the sum of the forecast error variance of the causal VAR (4.2). The first column of matrix \( b_0 \) is then found by maximising \( \sum_{h=0}^{H_2} \left( \sum_{j=0}^{H_1} c^j_1 \Psi_j A_0 \gamma^j_1 A'_0 \gamma'_1 e^1_1 \right) / \left( \sum_{j=0}^{H_1} c^j_1 \Psi_j A_0 A'_0 \Psi'_1 e^1_1 \right) \) subject to \( \gamma'_1 \gamma_1 = 1 \) and \( \gamma_{1,1} = 0 \), \( A_0 \) being a lower-triangular satisfying \( \mathbb{E}[e_t e'_t] = A_0 A'_0 \).
consequence of nonfundamentalness, the dashed (B-S, Barsky-Sims) and solid marked (MS, Max Share) median responses from the causal VAR in panel (a) fail to coincide with the theoretical counterparts. The distortion follows a pattern where the theoretical responses are lagged by two periods, originating from the nonfundamental error term that strongly weighs the past shocks. As the causal VAR cannot then retrieve the most recent shock, the model is incapable to reproduce the initial reactions. This drawback disappears only if the observables imply fundamentalness, which can be achieved by including the interest rate. The causal VAR with $y^2_t$ and both the Barsky-Sims and Max Share identification schemes are then able to reproduce the true impulse responses, shown in Appendix 4.E.2. Hence, although both observables could well be used for an empirical analysis on news shocks, their capability to derive the true responses heavily depends on invertibility.

Instead, the impulse responses of $y^1_t$ can be recovered by the noncausal VAR selected by AIC, shown in panel (b) of Figure 4.2, where the timing of the theoretical responses are changed such that the peak response of technology is aligned with those from the empirical model. Notice that no information about this shifting of the error term was needed in the estimation. Evidently, the noncausal VAR is able to match the theoretical impulse responses and measures the early reactions to a news shock at negative lags, as a result of the two-sided MA representation including the leads of time-shifted shocks. In Appendix 4.E.2, I additionally show results when the relative weight of the news shock is greater, $\sigma_u = 1$, referring to a situation when the identification strategy is not consistent with the underlying model as the noise and news shocks are equally important. In that case, the impulse responses can still be accurately enough recovered from the noncausal VAR while the distortion in the causal VAR prevails.

Once the anticipation horizon is $q = 16$, fundamentalness becomes unattainable even with the inclusion of the interest rate. This information deficiency has adverse effects on the performance of causal VAR models as seen in panels (a) and (b) of Figure 4.3. For both $y^1_t$ and $y^2_t$, the estimated impulse responses lead the theoretical counterparts by several periods and are therefore unable to reveal the initial, smooth responses to a news shock. Concluding from the causal VAR with the Barsky-Sims identification, the news shock would induce an immediate, strong response of output gap, interest rate and inflation. Similarly, with the Max Share strategy, the effects of the shock are falsely measured. The risk of nonfundamentalness thus concerns missing the responses

---

24See Lippi and Reichlin (1994b); Leeper et al. (2013) for deeper evaluation.
Evidence on news shocks under information deficiency

Figure 4.3: Impulse responses to a news shock from a causal and noncausal VAR in the NK model with $q = 16$ and $\sigma_u = 0.5$

The estimates are the periodwise medians of impulse responses from the Monte Carlo samples. In the causal VAR, the dashed lines and solid marked lines refer to the Barsky-Sims (B-S) and Max Share (MS) identification, respectively. The light and dark grey shaded areas border the middle 90 and 68 percent, respectively, of the distribution for the estimated impulse responses. The solid lines are the theoretical impulse responses, aligned in panels (c) and (d) with the estimated noncausal impulse responses according to the maximum impact on technology.
4.4 News shocks and U.S. business cycles

News shocks generate fluctuations stemming from reactions to productivity changes that materialise in the future. Beaudry and Portier (2006) concluded that they account for a half of the output variation and increase stock prices, investment, consumption and hours on impact. Nonetheless, the estimated impulse responses and the significance of the news shock starkly depend on the choice of variables and identification strategy. Using the medium-run identification, Barsky and Sims (2011) find that news shocks increase consumption, while it initially decrease investment, hours and prices. Moreover, the news shocks play a less important role for the short-run fluctuations. Using the factor-based test of Forni and Gambetti (2014), Forni et al. (2014) reject

Distortion in causal VAR models is also independent of the chosen lag length. Moreover, least squares estimation of the causal VAR model produces similar results but with wider confidence bands.

On the other hand, when I simulate the theoretical model with Gaussian shocks, I estimate the degrees-of-freedom parameter to be considerably larger, above 30. Higher estimates of $\lambda$ thus lead to a failure of the distributional assumption and weak identification of the noncausal VAR.
Evidence on news shocks under information deficiency

the hypothesis of fundamentalness in small-scale VAR models. In a factor-augmented VAR they use, the news shock signifies less at the business cycle frequencies, and positive news shocks do not produce such economic expansions as suggested by Beaudry and Portier (2006).

On the other hand, according to Beaudry et al. (2015), nonfundamentalness does not alter the conclusions about the news but the identification scheme is a vital issue in reproducing the results of Beaudry and Portier (2006). Similarly, Sims (2012) and Forni et al. (2018) argue that an informationally deficient VAR may well recover a news shock sufficiently accurately despite the underlying model implies noninvertibility. However, the simulation results from the previous section suggest that the nonfundamentalness issue may be a consequential factor for the results drawn from the causal VAR.

Here, I revisit the evidence on news shocks in the U.S. economy with the noncausal VAR. The approach tackles the above issues with inference robust against the potentially insufficient information set.

4.4.1 Data

I use quarterly U.S. macroeconomic time series. Total factor productivity ($TFP_t$) is the capacity-utilisation adjusted measure constructed by Fernald (2012). Consumption is defined as a sum of consumption of nondurables and services, output as the real gross domestic product (GDP) and investment as the sum of fixed private investment and consumption of durables. Hours worked and the real compensation per hour, measuring the real wage, are from the nonfarm business sector. The variables are seasonally adjusted, expressed in logs and per-capita terms by the civilian noninstitutional population. Annualised inflation is computed from the consumer price index. Following the literature, stock prices, measured as the log of the real S&P 500 index, is transformed to per-capita terms. The above data span quarters from 1948:1 to 2015:4, where the first quarter is lost because of differencing.

In the further analysis, I additionally use the following variables. The effective Federal Funds rate ($FFR_t$), available from quarter 1954:3 onwards, is aggregated to the quarterly frequency. As the long-term rate ($r_{10}^t$), I use the 10-year Constant Treasury Maturity Rate, available from quarter 1953:2 onwards. Last, consumer confidence is measured by the Index of Consumer Sentiments, released by the Michigan Surveys of Consumers and starting from quarter 1961:1.27

---

27The national accounts variables are taken from the National Income and Product Accounts
4.4 News shocks and U.S. business cycles

4.4.2 Estimation and identification

To avoid the estimation of large models, I include the above variables in different specifications which all include TFP, either investment or stock prices, consumption and hours in addition to the variables of interest. I difference all except hours and interest rates, as the lag and lead polynomials of the noncausal model are required to be stable.

I use Bayesian inference to tackle the large parameter space and also to circumvent the model selection problem found to be somewhat less robust in Section 4.3. Specifically, instead of relying on information criteria, I fix the number of lags and leads to four. The resulting specification is then able to cover a rich class of both fundamental and nonfundamental representations. Moreover, Bayesian prior shrinkage implicitly facilitates the model selection by weighing those lag and lead coefficients supported by data. Importantly, setting a sufficiently high number of lags, $r = 4$, does not rule causality out a priori, as the invertible MA part in (4.7) is then able to capture the full dynamics. If invertibility is instead not supported by data, $s = 4$ leads with restrictions (4.9) imposed on TFP accommodate noncausality induced by nonfundamentalness. As a consequence of noncausality, the impulse responses will generally be located both on the positive and negative lags but produce results consistent with the underlying process.

I set a Minnesota-type prior distribution, discussed in Appendix 4.F.1 in detail. In particular, I use tightness parameters to adjust the shrinkage separately for the lag and lead coefficients. I set their values in a way that less prior information is imposed on the lag terms whereas the lead terms are shrunk more towards zero. Hence, a priori, the lag terms are more important to determine the dynamics of variables. The estimation is eventually based on a Gibbs sampler derived by Lanne and Luoto (2016), where conditional normality of the likelihood is exploited to draw efficiently from the posterior distribution (See Appendix 4.F.1). The noncausal VAR, however, easily involves a multimodal posterior distribution, as observed by Lanne and Luoto (2016). By a stronger degree of tightness for the lead terms and the restrictions (4.9), the multimodality can be avoided in practice.

As was discussed in Subsection 4.2.2, the identification of a unique non-
Evidence on news shocks under information deficiency

Figure 4.4: Posterior distribution of degrees-of-freedom parameter $\lambda$
Posterior draws of $\lambda$ from the baseline noncausal VAR. The dashed vertical line is the posterior mean.

causal VAR model requires deviation from Gaussianity. Fortunately, the assumed multivariate $t$-distribution nests Gaussianity for large values of degrees-of-freedom parameter $\lambda$ and provides thereby a measure of non-normality and excess kurtosis. In other words, a low estimate of $\lambda$ will support the validity of the distributional assumption compared to Gaussianity. In Figure 4.4, I plot the histogram of the posterior draws of $\lambda$ for the baseline specification which includes TFP, investment, consumption, hours, output and inflation. The histogram evidently reveals that the posterior distribution heavily weights low degrees of freedom. Data also strongly dominate the prior mean of $\lambda$ set to 10 with a posterior mean lower than 5. Moreover, although not reported, degrees of freedom is estimated low in the causal VAR(4) model as well. Hence, allowing for fatter tails improves the fit of the model and facilitates the identification of the noncausal model.

In the VAR models I estimate, the identification of news shocks follows the strategy outlined in Section 4.2.4. The news shock is then a shock explaining
the most of the variation of total factor productivity, $y_{1,t}$

$$y_{1,t}^1 = \log TFP_t = \sum_{j=-H_1}^{H_2} e_1' \Psi_j \tilde{A} w_{1,t-j}$$

where $\Psi_j = \sum_{k=-\infty}^{j} \Psi_k$ is a cumulative sum that transforms the differenced variables to levels. In other words, the news shock maximises the share of variance of TFP in levels, $\Omega_{1,y_{1,t}}^{[H_1,H_2]}$ in (4.22), over the horizon $[H_1,H_2] = [-20,40]$. The results are robust at a broad range of values of $H_1$ and $H_2$.

### 4.4.3 Results from the baseline model

Figure 4.5 plots the impulse responses to a one standard deviation news shock from the baseline VAR(4,4) model. The posterior median responses are shown in solid lines together with the 68 and 90 percent periodwise credible sets. The news shock triggers an increase of total factor productivity, and investment, consumption, hours and output respond positively and persistently. The increase of TFP is anticipated by the forward-looking variables which move at the negative lags of the horizon.

The identified shock diffuses to the economy as follows. First, at the negative lags, TFP slightly starts to absorb the shock. At these anticipatory lags, investment and hours move from their initial levels modestly but statistically significantly, and inflation starts to decline two years ahead of the jump of TFP. At quarter 0, investment, consumption and output permanently increase, and hours continue to rise. Simultaneously, inflation starts to recover and converges to zero after a year. During the final phase, at the positive lags, the shock continues to diffuse to TFP, investment and hours experience hump-shaped increases, while consumption and GDP remain at their new levels.

Overall, the estimated shock identifies persistent changes in TFP – consistent with the view that a news shock signals long-run productivity improvement. The shock also influences the forward-looking variables, most notably inflation, before it materialises in productivity. The responses at the negative lags before productivity attains its new level are, however, modest. Instead of strong reactions, the long-run, not-yet-materialised productivity movements induce smooth anticipation effects. Hence, it is difficult to view the news shock as a strong signal from the future productivity and as a trigger of short-run fluctuations. Rather, the comovement of TFP and the forward-looking variables suggests that the permanent technology shock propagates not only
Evidence on news shocks under information deficiency

Figure 4.5: Impulse responses to a news shock from the baseline model

The black solid lines are the posterior median responses to the news shock from the noncausal VAR(4,4). The dashed and marked lines depict the impulse responses from the causal VAR(4) model to the shock identified with the Barsky-Sims (B-S) and Max Share (MS) strategy, respectively. All responses are shown in levels. Light and dark grey shaded areas are the 90 and 68 percent periodwise credible sets of the noncausal model.

through expectations but also through the simultaneous productivity growth.

The plots of Figure 4.5 included additionally in dashed and solid marked lines the impulse responses to the news shock identified from the causal VAR(4) model by the Barsky-Sims (B-S) and the Max Share (MS) schemes, respectively. Considering the both identification schemes allows to examine, how crucial is the zero impact restriction of the former strategy for the results. Regardless of the identification scheme, the responses are remarkably similar for all variables except TFP. The Barsky-Sims strategy produces a constant growth path of TFP, while the Max Share scheme increases the variable on impact. However, the implied growth rates of TFP coincide from the both identification schemes after quarter 0. The latter observation casts doubt on the validity of the zero impact restriction as also argued by Kurmann and Sims (2017). In particular, the Barsky-Sims strategy involves a risk that the potential initial effect of the shock on TFP will be ignored.

\[28\]

Results remain similar when estimating a causal VAR in levels with a constant term as well as estimating a causal VAR with least squares. In particular, the risk of distortion caused by differencing variables, pointed out by Beaudry and Portier (2014), can be regarded as low.

\[29\]
4.4 News shocks and U.S. business cycles

Comparing the noncausal and causal responses from the Max Share identification reveals, in turn, that the reactions from both models align at the positive lags of the horizon. However, the causal VAR is unable to recover the initial effects of the forward-looking variables observed at periods before time 0. In particular, the estimates from the noncausal VAR document smoother responses of productivity and forward-looking variables starting at the negative lags, whereas the causal VAR produces somewhat stronger immediate reactions. In fact, the causal responses potentially misinterpret the impact effects under nonfundamentalness. When noninvertibility matters, a causal VAR depicts reactions to a shock from the incorrectly identified, nonfundamental error. The model may then overestimate the initial responses, as the nonfundamental shock is a compound of the past shocks already anticipated by the economic agents. In particular, this observation is also consistent with the findings of Forni et al. (2014) who measure the reactions to a news shock to become smoother after conditioning on more information.

How much do news shocks contribute to the economic fluctuations? In place of the forecast error variance decomposition of causal VAR models, the noncausal VAR measures the significance of the news shock by the relative conditional variance of \( y_{i,t} \) due to news shocks, \( \Omega_{\tilde{H}_1,\tilde{H}_2}^{[H_1,H_2]} \) in (4.22), over the horizon from \( H_1 = -20 \) until \( H_2 \). In Figure 4.6, these shares contributed by the news shock along with the credible sets are plotted for the baseline specification. The identified shock determines the major part of TFP at longer horizon, consistent with Beaudry and Portier (2006) who identify the news shock as a long-run component of productivity. The identified shock is also central in explaining fluctuations in investment, consumption, employment, output and inflation. However, the evidence contrasts the news shock view and the results from the Barsky-Sims identification in one important aspect: the shock explains a considerable part of movements in TFP also in the short run.

Figure 4.7 plots the four-quarter moving average of the identified news shock \( \tilde{u}_{1,t} \) along with a cycle component of capital-to-TFP measure. The series is detrended using the Hodrick-Prescott (HP) filter with smoothing parameter 1,600. Capital-to-TFP ratio is computed from the investment and TFP series of Fernald (2012).

---

30 In particular, this observation is also consistent with the findings of Forni et al. (2014) who measure the reactions to a news shock to become smoother after conditioning on more information.

31 The series is detrended using the Hodrick-Prescott (HP) filter with smoothing parameter 1,600. Capital-to-TFP ratio is computed from the investment and TFP series of Fernald (2012).
Evidence on news shocks under information deficiency

Figure 4.6: Share of variation due to the identified news shock
The shares are computed from the baseline VAR model using equation (4.22) with all variables in levels. \( H_1 = -20 \) and \( H_2 \) is varied on the x-axes. The dashed and marked lines depict the shares due to the news shock identified by the Barsky-Sims (B-S) and Max Share (MS) strategy, respectively, from the causal VAR. The solid lines are the posterior medians. Light and dark grey shaded areas are the 90 and 68 percent credible sets of the noncausal model.

In conclusion, the news shock generates a growth of TFP together with modest positive reactions of forward-looking variables. The results confirm the positive comovement of consumption and investment conditional on the news shock, in line with Beaudry and Portier (2006) and the news view of business cycle, but difficult to be theoretically generated in dynamic general equilibrium models. Given the less significant anticipation of future technology, my results also align with Barsky and Sims (2011) and Forni and Gambetti (2014). Inferring from the above results, the comovement can be explained by the contemporaneous improvement of TFP, which gives the news shock a character similar to a shock that slowly diffuses to technology as in Lippi and Reichlin (1994a) or Walker and Leeper (2011). As also interpreted by Kurmann and Sims (2017), the news shock is related to sustained productivity growth.
4.4 News shocks and U.S. business cycles

Figure 4.7: News shock and investment over time
The solid line is the posterior median of the four-quarter moving average of the identified shock from the baseline noncausal VAR(4,4) model. The cycle component of the capital-to-TFP ratio in dashed lines is a percentage deviation from the HP-trend with smoothing parameter 1,600. The shaded grey areas are the NBER recessions.

and affects TFP both on impact and with a lag.

4.4.4 Reactions of further forward-looking variables
The above results documented that persistent technology shocks are smoothly predicted by the main macroeconomic indicators. However, since the identification imposes no impact effects, the news shock is revealed indirectly under the assumption that surprise technology shocks are less important for the changes in TFP. Next, I confirm with a set of forward-looking variables that the permanent TFP growth is indeed anticipated.

The plots of Figure 4.8 show the impulse responses of four additional variables, the stock price index, the real wage, consumer confidence and research and development (R&D) expenditures, the solid lines depicting the results from the noncausal model.\textsuperscript{32} The responses are separately estimated in the

\textsuperscript{32}The R&D expenditures are measured as per-capita investment to intellectual property products divided by the GDP deflator, taken from the NIPA tables.
Evidence on news shocks under information deficiency

Figure 4.8: Impulse responses of the forward-looking variables to a news shock

The impulse response of each variable is derived from a specification that additionally includes five baseline variables shown in Figures 4.16 and 4.17 in Appendix 4.F.3. The black solid lines are the posterior median responses to the news shock from the noncausal VAR(4,4) model. The dashed and marked lines depict the impulse responses from the causal VAR(4) model to the shock identified with the Barsky-Sims (B-S) and Max Share (MS) strategy, respectively. All responses are shown in levels. Light and dark grey shaded areas are the middle 90 and 68 percent credible sets from the noncausal model.

VAR models that include five variables from the baseline specification—excluding inflation—in addition to the variable of interest. The estimates regarding the remaining variables coincide with those shown in Figure 4.5 and are only reported in Appendix 4.F.3. In response to the technology shock, stock prices start to increase two years before TFP jumps, eventually peaking at quarter 0. Similarly, consumer confidence significantly rises five quarters before the main realisation moment of the shock, supporting the view that the public is able to foresee the improvement of technology. Simultaneously, the real wage is prompted to a steady growth path. R&D expenditures gradually increase prior to TFP: the anticipation of the shock fosters development of complementary technologies. On the contrary, the causal responses plotted in dashed and marked lines are generally unable to show the initial anticipatory path of dynamics tracked in the noncausal model on the negative lags.

Finally, to analyse the behaviour of interest rates, Figure 4.9 depicts the impulse responses from a specification that includes the term spread $r_{10} - FFR_t$ and the 10-year rate as well as variables from the baseline specification. I additionally back out the response of the Federal Funds rate as a difference
4.5 Conclusions

The presence of news shocks permits economic agents to respond not only to the current but also to the future changes in productivity. However, causal VAR models fail to provide reliable evidence in validating these dynamics between the long-term rate and the spread.\textsuperscript{33} According to the solid lines depicting the noncausal responses, the Federal Funds rate initially declines in response to the deflationary effects, consistently with inflation-targeting monetary policy. As the 10-year rate changes only mildly, the constant increase of the term spread is caused by the decreasing policy rate. These observations are also in line with Kurmann and Otrok (2013) who use a similar set of variables to examine the effects of news shocks on the term structure of interest rate. At date 0, once TFP experiences the fast growth, the term spread peaks after which it returns to its long-run mean within the five subsequent quarters. The news shock thus has deflationary effects, leading to a decrease of the Federal Funds rate and an increase in the slope of the term structure. However, compared to the causal responses, the jump of the spread is explained by the anticipatory movements of the interest rates at the negative lags, which cannot be measured in the causal framework.

\textsuperscript{33}The conclusions about the baseline variables are consistent with Figure 4.5 and shown in Appendix 4.F.3
Evidence on news shocks under information deficiency

under nonfundamentalness, when economic agents possess more information than an econometrician. Consequently, the inference drawn from a VAR model may be based on misinterpreted structural shocks.

By allowing for noncausality, the nonfundamentalness issue is resolved by including future terms that capture variation omitted in an informationally deficient causal VAR. Structural analysis on the anticipated technology shocks can then be conducted using the two-sided MA representation with the identification scheme I introduced. I identified a technology shock as the most significant factor moving productivity, which indirectly reveals the news shock. According to the Monte Carlo simulations, the approach is able to, first, detect nonfundamentalness in the form of noncausality and, second, recover the responses to a news shock under nonfundamentalness.

The evidence from the U.S. economy suggested that the news shock induced contemporaneous increases in investment, hours, output and consumption as well as in total factor productivity, which may be the determining factor behind the strong anticipating responses usually observed in the VAR literature. Rather than being a pure signal about productivity in the distant future, the news shock diffuses into TFP both in the short and long run. Mirroring this evidence with the news shock literature, the results confirm that the news shock explains the major part of long-run movements in TFP and is as such central in contributing to business cycle fluctuations. Nonetheless, it is difficult to view the news shock as generating strong short-run reactions with materialisation in productivity only in the long run.

Finally, the analysis proceeded with the following limitations in mind. First, the noncausal VAR model is not a one-to-one empirical alternative to a noninvertible model but provides a sufficiently accurate approximation as its multiplicative form may rule out certain representations. The link between noncausality and noninvertibility could be shown to exist in the Monte Carlo simulations, and the risk of misspecification has been minimised by the selection of lag and lead orders. Second, the multivariate t-distribution provides a simple departure from Gaussianity by adding a single volatility term to the error term that may contain both anticipated and unanticipated shocks. Hence, establishing identifiability and estimation theory for more general non-Gaussian distributions or conditional heteroskedasticity can be viewed as useful extensions. I leave these considerations for subsequent research.
References


Evidence on news shocks under information deficiency


Appendices

4.A Noncausal representation of a noninvertible VARMA model

To derive (4.5), let \( l \) roots of \( |B(z)| \) lie inside the unit circle with \( l_0 \) roots equal to zero. By standard algebra, \( |B(z)| \) as scalar polynomial of order \( kd \) can be factorised to \( |B(z)| = z^{l_0} (1 - z_0^{-1} z) \ldots (1 - z_{kd}^{-1} z) \), and

\[
|B(z)|^{-1} = z^{-l_0} \left( \prod_{i=l_0+1}^{l} (1 - z_i^{-1} z) \right)^{-1} \left( \prod_{i=l_0+1}^{kd} (1 - z_i^{-1} z) \right)^{-1} \\
= z^{-l} \bar{c}_l \alpha(z^{-1})^{-1} \beta(z)^{-1}
\]

with

\[
\alpha(z^{-1}) = \prod_{i=l_0+1}^{l} (1 - z_i z^{-1}), \quad \beta(z) = \prod_{i=l_0+1}^{kd} (1 - z_i z^{-1}),
\]

and \( \bar{c}_l = (-1)^{l-l_0} \prod_{i=l_0+1}^{l} z_i \). Now, scalars \( \alpha(z^{-1})^{-1} \) and \( \beta(z)^{-1} \) are well-defined power series expansions in \( z^{-1} \) and \( z \), respectively, decaying to zero at geometric rate. Consequently, the inverse of \( B(z) \) is

\[
B(z)^{-1} = z^{-l} \bar{c}_l \alpha(z^{-1})^{-1} \beta(z)^{-1} B_{adj}(L),
\]

where \( B_{adj}(z) \) is the adjoint matrix of \( B(z) \) of degree at most \( (k - 1)d \). Hence,

\[
L^{-l} \bar{c}_l \beta(L)^{-1} B_{adj}(L) \alpha(L^{-1})^{-1} A(L)y_t = u_t
\]

or

\[
\bar{c}_l \beta(L)^{-1} B_{adj}(L) \alpha(L^{-1})^{-1} A(L)y_t = u_{t-l}
\]
4.B Maximum likelihood estimation of the noncausal VAR model

The log-likelihood function of the noncausal VAR is

$$ l(\theta) = \sum_{t=r+1}^{T-s} \log f(\epsilon_t(v)'; \Sigma^{-1}\epsilon_t(v); \lambda) $$

with density $f(\cdot; \lambda)$ of the multivariate $t$, $v = (\pi, \phi)$, $\pi = \text{vec}(\Pi)$, $\phi = \text{vec}(\Phi)$, scale matrix $\Sigma$, degrees-of-freedom parameter $\lambda$, and residuals

$$ \epsilon_t(v) = v_t(\phi) - \sum_{j=1}^{r} \Pi_j(\pi)v_{t-j}(\phi), $$

where

$$ v_t(\phi) = y_t - \Phi_1(\phi)y_{t+1} - \ldots - \Phi_s(\phi)y_{t+s}. $$

Derived by Lanne and Saikkonen (2013), the maximum likelihood estimator is asymptotically normal as

$$ \sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \mathcal{I}_{\theta\theta}(\theta_0)^{-1}) $$

with $\mathcal{I}_{\theta\theta}(\theta_0) = -(T-r-s)^{-1}E\left[\frac{\partial^2 l(\theta_0)}{\partial \theta \partial \theta}\right]$. $\mathcal{I}_{\theta\theta}(\theta_0)$ can be consistently estimated by $-(T-r-s)^{-1}\frac{\partial^2 l(\hat{\theta})}{\partial \theta \partial \theta}$.  

4.C Details on the stylised example

4.C.1 Derivation of the MA coefficients

The MA coefficients $\Psi_j$ of the noncausal VAR (4.16) solve recursion

$$ \Psi_j = \Pi^j_1 + \Phi_1 \Psi_{j+1} + \Phi_2 \Psi_{j+2}, \quad j \geq 0 $$

$$ \Psi_j = \Phi_1 \Psi_{j+1} + \Phi_2 \Psi_{j+2}, \quad j < 0. $$
Evidence on news shocks under information deficiency

By solving for each $j$,

$$\Psi_j = \begin{bmatrix} \rho_j & 0 \\ \theta \rho_j & 0 \end{bmatrix}, j > 0,$$

$$\Psi_0 = \begin{bmatrix} 1 & 0 \\ \beta \rho \theta & 1 \end{bmatrix}$$

$$\Psi_j = \begin{bmatrix} 0 & 0 \\ \theta \beta^{-j} & 0 \end{bmatrix}, j = -1, -2$$

$$\Psi_j = 0, j < -2,$$

the impulse responses are computed from

$$y_t = \sum_{j=-\infty}^{\infty} \Psi_j \epsilon_{t-j} = \sum_{j=-\infty}^{\infty} \Psi_j \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{t-2-j} \\ v_{t-j} \end{bmatrix}$$

$$= \sum_{j=-\infty}^{\infty} \left[ \psi_{j,11} \epsilon_{t-2-j} + \psi_{j,12} (\epsilon_{t-2-j} + v_{t-j}) \right] + \left[ \psi_{j,21} \epsilon_{t-2-j} + \psi_{j,22} (\epsilon_{t-2-j} + v_{t-j}) \right]$$

as

$$\frac{\partial a_{t+j}}{\partial \epsilon_{t-2}} = \psi_{11,j} + \psi_{12,j} = \psi_{11,j}, \quad \frac{\partial a_{t+j}}{\partial v_t} = \psi_{12,j} = 0,$$

$$\frac{\partial x_{t+j}}{\partial \epsilon_{t-2}} = \psi_{21,j} + \psi_{22,j}, \quad \frac{\partial x_{t+j}}{\partial v_t} = \psi_{22,j},$$

where $\psi_{k,mn}$ is the $(m,n)$ element of matrix $\Psi_j$. The impulse responses from the noncausal model to the anticipated shock $\epsilon_{t-2}^a$ are thus

$$\frac{\partial a_{t+j}}{\partial \epsilon_{t-2}^a} = \begin{cases} \rho, & j \geq 0 \\ 0, & j < 0 \end{cases}, \quad \frac{\partial x_{t+j}}{\partial \epsilon_{t-2}^a} = \begin{cases} \theta \rho, & j \geq 0 \\ \theta \beta^{-j}, & -2 \leq j < 0 \\ 0, & j < -2 \end{cases},$$

which coincide with the theoretical impulse responses.
Noncausal representation with a more general technology process

Instead of (4.11), consider a more general technology process

\[ a_t = \rho a_{t-1} + \varepsilon_t^{a} + \chi \varepsilon_{t-1} \tag{4.24} \]

and assume it determines the equilibrium together with (4.12). Accordingly, a shock \( \varepsilon_t^{a} \) affects both \( a_t \) and \( a_{t+q} \) such that agents gradually learn about the future technology \( q \) periods forward.\(^{34}\) If \( \chi < 1 \), \( \varepsilon_t^{a} \) has a greater contribution to \( a_{t+q} \) than to \( a_t \), in which case the ARMA process for \( a_t \) is primarily driven by shocks observed by the economic agents before \( t \).

The solution can now be derived from (4.13) such that \( x_t = \theta z_t + \nu_t \) for \( q = 0 \) and

\[ x_t = \theta a_t + \theta \sum_{j=0}^{q-1} \beta^q \varepsilon_{t-j} \tag{4.25} \]

for \( q > 0 \). Let now the anticipation horizon be \( q = 2 \). By inserting (4.11) to the latter equation and collecting terms, \( y_t = (a_t, x_t)' \) follows

\[ y_t = \begin{bmatrix} \rho & 0 \\ \theta \rho & 0 \end{bmatrix} y_{t-1} + \begin{bmatrix} \chi + L^2 \\ \theta (\beta^2 + \chi) + \theta \beta L + \theta L^2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} \varepsilon_t^{a} \\ \nu_t \end{bmatrix} \tag{4.26} \]

If movements in the exogenous process for \( a_t \) are dominated by the anticipated lag term, i.e. \( \chi < 1 \) and \( q > 0 \), \( y_t = (a_t, x_t)' \) is noninvertible in the past since \( |B(z)| = 0 \) for \( z = \pm \sqrt{\chi} i \). In other words, as soon as the news shock contributes to \( a_t \) relatively more, \( \chi < 1 \), the observables suffer from nonfundamentalness and no causal VAR representation in terms of \( y_t \) exists for structural shocks \( u_t = (\varepsilon_t^{a}, \nu_t) \).

As before, the solution can be written in terms of future observables. In particular, rewriting the noninvertible process (4.24) as

\[ (1 - \rho L)a_t = (L^2 + \chi) \varepsilon_t^{a} \tag{4.27} \]

\(^{34}\)Walker and Leeper (2011) describe \( \varepsilon_t^{a} \) as a correlated news shock. An example with a similar process is also considered in Beaudry and Portier (2014).
and its right-hand side as \((1 + \chi L^{-2})\varepsilon_{t-2}^a\), the modified lag polynomial, \(1 + \chi z^2\), has no roots smaller than one in modulus such that (4.27) implies

\[
(1 - \rho L)(1 + \chi L^{-2})^{-1}a_t = \varepsilon_{t-2}^a. \tag{4.28}
\]

Therefore, \(a_t\) is noncausal with a time-shifted error, and the shock \(\varepsilon_{t-2}^a\) is a function of the past and future values of \(a_t\). Using the representation (4.28) to substitute the shocks \(\varepsilon_{t-1}\) and \(\varepsilon_t\) in (4.26), \(x_t\) has a noncausal form,

\[
x_t = \rho a_{t-1} + \left(\beta - \chi \rho + \theta(\beta^2 + \chi)L^{-1}\right) \sum_{j=0}^{\infty}(-\chi L^{-2})^j a_{t+1} + \varepsilon_{t-2}^a + \nu_t \tag{4.29}
\]

consisting of an infinite number of leads of \(a_t\). Furthermore, \(a_t\) and \(x_t\) in (4.28) and (4.29) satisfy

\[
(I_2 - \Pi_1 L)(I_2 - \Phi_1 L^{-1} - \Phi_2 L^{-2} - \ldots)y_t = \epsilon_t, \tag{4.30}
\]

with

\[
\Phi_j = \begin{bmatrix} \phi_{j,11} & 0 \\ \phi_{j,21} & 0 \end{bmatrix},
\]

\[
\phi_{j,11} = \begin{cases} 0, & j = 1,3,\ldots \\ -(-\chi)^{j-2}, & j = 2,4,\ldots \end{cases}
\]

\[
\phi_{j,21} = \begin{cases} \beta(-\chi)^{j-1}, & j = 1,3,\ldots \\ \theta(\beta^2 + \chi)(-\chi)^{j-1}, & j = 2,4,\ldots \end{cases}
\]

and the error term

\[
\epsilon_t = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{t-2}^a \\ \nu_t \end{bmatrix}.
\]

In particular, the noncausal representation (4.30) is directly related to the underlying model when the infinite number of lead terms is truncated by large lead order \(s\), and the noncausal VAR(1,\(s\)) recovers a linear combination of structural shocks \(\varepsilon_{t-2}^a\) and \(\nu_t\).
4.D Solution to the identification problem

Finally, the MA coefficients are numerically obtained from recursions

$$
\Psi_j = \Pi_1^j + \Phi_1 \Psi_{j+1} + \ldots + \Phi_s \Psi_{j+S}, \quad j \geq 0
$$

$$
\Psi_j = \Phi_1 \Psi_{j+1} + \ldots + \Phi_s \Psi_{j+S}, \quad j < 0,
$$

where $S$ is a sufficiently large integer. The impulse responses of the noncausal VAR(1,s) and the theoretical model are now plotted in the lower and upper plots of Figure 4.10, respectively, for $\chi = 0.5$. Similar to Figure 4.1, the noncausal impulse responses replicate the theoretical counterparts shown in the upper panel.

4.D Solution to the identification problem

$w_1$ solves

$$
\max_{w_1} \Omega_{1,y_{1,t},[H_1,H_2]},
$$

(4.31)
Evidence on news shocks under information deficiency

\[ \Omega_{1,H_1,H_2}^{[H_1,H_2]} = \frac{\mathbb{E}_{\Sigma_j=H_1}^\Sigma_{j=H_1} e'_j \Psi_j \tilde{A} w_1 \bar{\alpha}_{t-j} \tilde{A}' \Psi' e_1}{\mathbb{E} \left[ \sum_{j=H_1}^{H_2} e'_j \Psi_j \Sigma \Psi' e_1 \right]} = \frac{\sum_{j=H_1}^{H_2} e'_j \Psi_j \tilde{A} w_1 w'_j \tilde{A}' \Psi' e_1}{\sum_{j=-\infty}^{\infty} e'_j \Psi_j \Sigma \Psi' e_1}, \]

subject to the orthogonality of \( W, w'_1 w_1 = 1 \).

Solving the problem follows Uhlig (2004). By rewriting

\[ e'_j \Psi_j \tilde{A} w_1 w'_j \tilde{A}' \Psi' e_1 = \text{tr} \left( e'_j \Psi_j \tilde{A} w_1 w'_j \tilde{A}' \Psi' e_1 \right) = \text{tr} \left( w'_j \tilde{A}' \Psi' e_1 e'_j \Psi_j \tilde{A} w_1 \right) = \text{tr} \left( w'_j \Psi_j E_{11} \Psi_j \tilde{A} w_1 \right) = \text{tr} \left( w'_j S_{ik} w_1 \right), \]

the nominator of the objective function is

\[ \sum_{j=H_1}^{H_2} w'_j S_{jk} w_1 = w'_j S w_1 \]

As the denominator is independent of \( w_1 \), the problem can be solved by setting up the Lagrangian

\[ \mathcal{L} = w'_j S w_1 - \mu (w'_j w_1 - 1). \]

The first-order condition is

\[ S w_1 = \mu w_1, \]

and since \( w'_j \mu w_1 = \mu \), the eigenvector corresponding to the maximal eigenvalue of the positive definite matrix \( S \) is the optimum.

### 4.5 Monte Carlo simulations

#### 4.5.1 NK model of Section 4.3

Equilibrium is determined by equations

\[
\begin{align*}
\hat{y}_t &= -\frac{1}{\sigma} (i_t - \mathbb{E}_t \pi_{t+1} - \pi_t^m) + \mathbb{E}_t \hat{y}_{t+1} \\
\pi_t &= \beta \mathbb{E}_t [\pi_{t+1}] + \kappa \hat{y}_t, \\
i_t &= \rho (1 - \rho_m) + \rho_m i_{t-1} + \phi \pi_t + \phi_y \hat{y}_t + \epsilon_t^m,
\end{align*}
\]
4.E Monte Carlo simulations

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discount factor β</td>
<td>0.99</td>
</tr>
<tr>
<td>Risk aversion σ</td>
<td>2</td>
</tr>
<tr>
<td>Frisch elasticity φ</td>
<td>1</td>
</tr>
<tr>
<td>Calvo parameter θ</td>
<td>2/3</td>
</tr>
<tr>
<td>Capital share α</td>
<td>1/3</td>
</tr>
<tr>
<td>Elasticity of substitution ε</td>
<td>6</td>
</tr>
<tr>
<td>Taylor rule coefficient for inflation $\Phi_\pi$</td>
<td>1.5</td>
</tr>
<tr>
<td>Taylor rule coefficient for output gap $\Phi_y$</td>
<td>0.5/4</td>
</tr>
<tr>
<td>Persistence of interest rate $\rho_m$</td>
<td>0.8</td>
</tr>
<tr>
<td>Persistence of technology $\rho_o$</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Table 4.2: Calibration of the New Keynesian Model

by $r^n_t = \rho + \sigma \psi_y y a E_t^1 a_{t+1}$ and technology (4.23). Coefficients are functions of deep parameter of the model:

$$\kappa = \lambda (\sigma + \phi + \alpha \frac{1}{1-\alpha}),$$

$$\lambda = \frac{(1-\theta)(1-\beta\theta)}{\theta} \Theta,$$

$$\Theta = \frac{(1-\alpha)}{1-\alpha + \alpha\varepsilon},$$

$$\psi_y = \frac{1 + \phi}{\sigma(1-\alpha) + \phi + \alpha}$$

The calibration is summarised in Table 4.2.

4.E.2 Further simulation results

Figure 4.11 plots the impulse responses in the causal VAR models, where the observables induce fundamentalness. In Figure 4.12, the results are shown for the underlying model, where news and noise shocks are equally important. Figure 4.13 plots the impulse responses when the shocks are generated from an alternative non-normal distribution.
Evidence on news shocks under information deficiency

Figure 4.11: Impulse responses to a news shock from the causal VAR(5) with $y_t^2 = (a_t, i_t, \pi_t)$ in the NK model with $q = 3$ and $\sigma_u \in \{0.5, 1\}$. The VAR estimates are the periodwise medians of impulse responses from the Monte Carlo samples. The dashed lines and solid marked lines refer to the Barsky-Sims (B-S) and Max Share (MS) identification, respectively. The light and dark grey shaded areas border the middle 90 and 68 percent, respectively, of the distribution for the estimated impulse responses. The solid lines are the theoretical impulse responses.

Figure 4.12: Impulse responses to a news shock from a causal and noncausal VAR in the NK model with $q = 3$ and $\sigma_u = 1$. The solid lines are the theoretical impulse responses, aligned in panel (b) with the estimated noncausal impulse responses according to the maximum impact on technology. For additional explanations, see Figure 4.11
4. E Monte Carlo simulations

(a) $q = 3$: Noncausal VAR(1,2) on $y_t^1 = (a_t, \tilde{y}_t, \pi_t)$

(b) $q = 16$: Noncausal VAR(4,2) on $y_t^1 = (a_t, \tilde{y}_t, \pi_t)$

Figure 4.13: Impulse responses to a news shock from noncausal models in the NK Model with $\sigma_u = 0.5$ and fat-tailed shocks.

The model is simulated by structural shocks drawn in the following way. First, draw $u_t$ from $N(0, I)$. Second, multiply $u_t$, with probability 0.1 by 3 to put weight on tails. Last, standardise the new error series and simulate $y_t$. The dashed lines are the median estimated impulse responses from the Monte Carlo samples. Light and dark grey shaded areas border the middle 90 and 68 percent, respectively, of the distribution for estimated impulse responses. The solid lines depict the theoretical impulse responses, aligned with the estimated noncausal impulse responses according to the maximum impact on technology. The orders of the models are those from the most selected specification according to AIC.
4.F Details on the empirical part

4.F.1 Bayesian estimation of the noncausal VAR

I refer to Lanne and Luoto (2016) in the derivation of the following Gibbs sampler algorithm. I additionally consider zero restrictions on the elements of $\Phi_i$, $i = 1, \ldots, s$. Let $\Pi$ and $\Phi$ be matrices stacking $\Pi'_i$ for $i = 1, \ldots, r$ and $\Phi'_i$ for $i = 1, \ldots, s$, respectively. Furthermore, write $\pi = \text{vec}(\Pi)$ and $\phi = \text{vec}(\Phi)$, $\theta = (\pi', \phi')'$ and $\bar{\theta} = (\pi', \phi', \text{vech}(\Sigma)', \lambda)'$. To impose $s^*$ zero restrictions on matrix $\Phi$ to satisfy (4.9), introduce an $((n^2s - s^*) \times 1)$ vector $\phi_r$ containing the unrestricted parameters of $\Phi$ and an $(n^2s \times (n^2s - s^*))$ deterministic matrix $R_\phi$ which maps the unrestricted parameters to the matrix $\Phi$ as $\phi = R_\phi \phi_r$.

The approximate conditional joint density of $y = (y_1, \ldots, y_T)$ on $\omega = (\omega_{r+1}, \ldots, \omega_{T-s})$ is

$$p(y | \omega, \theta) \approx \prod_{r+1}^{T-s} p(\varepsilon_i(\theta) | \omega_i, \Sigma)$$

with

$$p(\varepsilon_i | \omega_i, \Sigma) = \frac{\omega_i^{n/2}}{(2\pi)^{n/2} | \Sigma |^{1/2}} \exp \left( -\frac{1}{2} \omega_i \varepsilon_i(\theta)' \Sigma^{-1} \varepsilon_i(\theta) \right),$$

$$\varepsilon_i(\theta) = v_i(\phi) - \sum_{j=1}^{r} \Pi_j(\pi) v_{i-j}(\phi),$$

and

$$v_i(\phi) = y_i - \Phi_i(\phi) y_{i+1} - \ldots - \Phi_s(\phi) y_{i+s}.$$

The prior distributions are set as follows: $\pi \sim N(\pi, V_\pi) I(\pi)$, $\phi_r \sim N(\phi'_r, V_{\phi_r}) I(\phi)$, $\Sigma \sim \text{iW}(\Sigma, \nu)$ and $\lambda \sim \text{Exp}(\lambda)$, where $I(\cdot)$ is indicator function equal to 1 when the polynomial to which $\pi$ or $\phi$ is mapped is stable and iW denotes the inverse Wishart distribution. Furthermore, define the following matrices. First, stack $y_t^* = \omega_1^{1/2} \Pi(L) y_t$ to a $(T - r - s)n \times 1$ vector $y^*$, and $X_t^* = \omega_1^{1/2} \Pi(L) X_t$ to a $(T - r - s)n \times n^2$ matrix $X^*$, where $X_t = I_n \otimes [y_{t+1} \cdots y_{t+s}]'$. Define similarly matrices $Y$ and $U$ by stacking $v_t^* = \omega_1^{1/2} v'_i(\phi)$ and $U_t^* = \omega_1^{1/2} [v_{t-1}(\phi) \cdots v_{t-r}(\phi)]'$, respectively, for $t = r + 1, \ldots, T - s$.

Following Lanne and Luoto (2016), the full conditional posterior distribu-
4. F Details on the empirical part

tion of $\phi_r$ can be derived as

$$
\phi_r | y, \pi, \Sigma, \omega \sim N(\bar{\phi}_r, \bar{V}_\phi_r) \ I(\phi), \ \phi = R_\phi \phi_r
$$

$$
\bar{V}_\phi_r^{-1} = \bar{V}_\phi_r^{-1} + R_\phi' X_s' \Omega X_s R_\phi, \ \bar{\phi} = \bar{V}_\phi_r \left( \bar{V}_\phi_r^{-1} \phi_r + R_\phi' X_s' \Omega Y_s \right)
$$

and $\Omega = I_{T-r-s} \otimes \Sigma^{-1}$. The conditional distribution of $\pi$ reads as

$$
\pi | y, \phi, \Sigma, \omega \sim N(\bar{\pi}, \bar{V}_\pi) \ I(\pi),
$$

$$
\bar{V}_\pi^{-1} = \bar{V}_\pi^{-1} + \Sigma^{-1} \otimes U' U, \ \bar{\pi} = \bar{V}_\pi \left( \bar{V}_\pi^{-1} \phi + \text{vec} \left( U Y \Sigma^{-1} \right) \right)
$$

Defining further $\bar{S} = S + E'E$, $E = Y - U \Pi$ and $\bar{v} = v + T - s - r$, the conditional posterior distribution for $\Sigma$ is

$$
\Sigma | y, \pi, \phi, \omega \sim iW(\bar{S}, \bar{v}).
$$

The remaining parameters $\omega = (\omega_{r+1}, \ldots, \omega_{T-r-s})$ and $\lambda$ are jointly drawn from

$$
\left( \lambda + \epsilon_t(\theta) \right) \Sigma^{-1} \epsilon_t(\theta) \omega_t | y, \pi, \phi, \Sigma, \lambda \sim \chi^2(\lambda + n), t = r + 1, \ldots, T - s
$$

and with Metropolis-within-Gibbs step from kernel

$$
p(\lambda | y, \omega) \propto \left( 2^{\lambda/2} \Gamma(\lambda/2) \right)^{(T-r-s)} \left( \prod_{t=r+1}^{T-s} (\omega_t^{(\lambda/2)})^{\lambda(T-r-s)/2} \right) \left( \frac{1}{\lambda} \right) \left( \frac{1}{2} \sum_{t=r+1}^{T-s} \omega_t \right)^{\lambda/2}. \quad (4.33)
$$

$$
\exp \left[ - \left( \frac{1}{\lambda} \right) \left( \frac{1}{2} \sum_{t=r+1}^{T-s} \omega_t \right) \lambda \right]. \quad (4.34)
$$

In the last step, I use the univariate normal distribution with mean equal to the mode and variance equal to the inverse of the second hessian of the above kernel as a candidate distribution. The standard Metropolis-Hastings acceptance probability is computed using (4.34).

I use the following Minnesota-Litterman type prior distribution. I set the means of $\pi$ and $\phi_r$, $\bar{\pi}$ and $\phi_r$, to 0, and the coefficients are assumed, a priori, independent by having zeros on the off-diagonals of covariance matrices $V_\pi$.
Evidence on news shocks under information deficiency

and $V_{\phi}$. On the other hand, $\sigma_{\pi,ijkl}^2$ and $\sigma_{\phi,ijkl}^2$, the diagonal elements of $V_{\pi}$ and $V_{\phi}$, corresponding to the $l$th lag or lead of variable $j$ in equation $i$ are given by

$$
\sigma_{\pi,ill} = \frac{\gamma_{1,\pi}}{\gamma_3}, \quad \sigma_{\pi,ijl} = \frac{\gamma_{2,\pi} \sigma_i}{\sigma_j}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, n, \quad l = 1, \ldots, r,
$$

$$
\sigma_{\phi,ill} = \frac{\gamma_{1,\phi}}{\gamma_3}, \quad \sigma_{\phi,ijl} = \frac{\gamma_{2,\phi} \sigma_i}{\sigma_j}, \quad i = 2, \ldots, n, \quad j = 1, \ldots, n, \quad l = 1, \ldots, s.
$$

Last, the prior variance corresponding to the unrestricted lead coefficients of the TFP equation I set

$$
\sigma_{\phi,11l} = \frac{\gamma_{11,\phi}}{\gamma_3}, \quad l = 1, \ldots, s.
$$

Here, $\sigma_i$ is estimated as the residual standard error from a univariate autoregression with $r + s$ lags on the $i$th variable, $\gamma_{1,\pi}$, $\gamma_{1,\phi}$ and $\gamma_{11,\phi}$ control for overall tightness, $\gamma_2$ for relative tightness and $\gamma_3$ is a decay parameter for more distant lags and leads. For the tightness parameters regarding the lag coefficients, I use values $\gamma_{1,\pi} = 0.2$, $\gamma_2 = 0.5$ and $\gamma_3 = 1$, standard in the Bayesian VAR literature. For the lead coefficients, I set $\gamma_{1,\phi} = 0.15$, which shrinks the lead coefficients moderately but somewhat more towards zero. Last, I do not strictly shrink the variance corresponding to the lead coefficients of the TFP equation, with $\gamma_{11,\phi} = 1$, to allow – a priori – TFP to react to its own shock at the negative lags. Last, I use the following values for the remaining hyperparameters: $\Sigma = (\nu - n - 1) \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ with degrees-of-freedom parameter $\nu = n + 2$ and $\lambda = 10$.

4.F.2 On the convergence of the posterior sampling

Using the algorithm, I draw 50,000 draws from the posterior distribution in addition to 1,000 burn-in draws. The algorithm is proceeded in five rounds such that every 10,000th draw is started from a new initial value. The paths of the Markov chains plotted in Figures 4.14 and 4.15 clearly suggest convergence of the algorithm.
4.F Details on the empirical part

Figure 4.14: Paths of the Markov chains for the lag and lead coefficients of the baseline noncausal VAR(4,4) model
The x-axes correspond to the draws, the y-axes to the parameter values
Evidence on news shocks under information deficiency

Figure 4.15: Paths of the Markov chains for the covariance matrix and the degrees-of-freedom parameter of the baseline noncausal VAR(4,4) model

The x-axes correspond to the draws, the y-axes to the parameter values
4.F.3 Impulse responses from the additional specifications

Figure 4.16: Impulse responses of the remaining variables in the specifications with additional variables

Impulse responses to the baseline variables not shown in Figure 4.8. The black solid lines are the posterior median responses to the news shock from the noncausal VAR(4,4) model. The dashed and marked lines depict the impulse responses from the causal VAR(4) model, the news shock identified with the Barsky-Sims (B-S) and Max Share (MS) strategy, respectively. All responses are shown in levels. Light and dark grey shaded areas are the 90 and 68 percent periodwise credible sets.
Evidence on news shocks under information deficiency

Figure 4.17: Impulse responses of the remaining variables in the specifications with additional variables

Impulse responses to the baseline variables not shown in Figures 4.8. The black solid lines are the posterior median responses to the news shock from the noncausal VAR(4,4) model. The dashed and marked lines depict the impulse responses from the causal VAR(4) model, the news shock identified with the Barsky-Sims (B-S) and Max Share (MS) strategy, respectively. All responses are shown in levels. Light and dark grey shaded areas are the 90 and 68 percent periodwise credible sets.
4.F Details on the empirical part

Figure 4.18: Impulse responses of the remaining variables in the specification including the interest rates

Impulse responses to the baseline variables not shown in Figure 4.9. The black solid lines are the posterior median responses to the news shock from the noncausal VAR(4,4) model. The dashed and marked lines depict the impulse responses from the causal VAR(4) model, the news shock identified with the Barsky-Sims (B-S) and Max Share (MS) strategy, respectively. All responses are shown in levels. Light and dark grey shaded areas are the 90 and 68 percent periodwise credible sets.