technische universität
dortmund

IN SEARCH OF Q: RESULTS ON IDENTIFICATION IN STRUCTURAL VECTOR AUTOREGRESSIVE MODELS

GÁBOR B. UHRIN

Dissertation

zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften (rer. nat.)
an der Fakultät Statistik der Technischen Universität Dortmund

Dortmund 2017

fakultät
statistik

Betreut von:	Prof. Dr. Martin Wagner, Professor der Ökonometrie und Statistik an der Fakultät Statistik der TU Dortmund

Erstgutachter:	Prof. Dr. Martin Wagner
Zweitgutachter:	Prof. Dr. Walter Krämer
Drittgutachter:	Prof. Dr. Helmut Herwartz
Tag der mündlichen Prüfung:	01.08.2017

ACKNOWLEDGMENTS

There are a number of people who contributed directly or indirectly to this dissertation. I would like to thank them at this place.

First, and foremost I would like to thank Martin Wagner for providing the opportunity to write a thesis under his guidance. His points and remarks are always sharp and excellent, and I've got to know him as a motivating academic with a lot of energy. His enthusiasm for econometrics and statistics had always had a motivating effect on my wanting to dig deeper, and to look precisely behind the concepts. I believe, that this attitude towards academia and academic research is invaluable. I'm glad to have learned a lot from him.

I would like to thank Walter Krämer for being member of the exam committee and taking on the duties of the second referee. His intuitive approach to statistics, and his wide knowledge of econometrics and statistics have always made me think more thoroughly about how to convey my results in an intuitive manner.

Helmut Herwartz has taken the official role of the third referee, and the unofficial role of a second advisor. I am very grateful for this, and for his guidance in general. The freedom he gave me in pursuing my own ideas, and the occasional "Sie schaffen das schon" have helped me grow a lot as an academic.

Besides being a coathor, I have always enjoyed my conversations with Ludger Linnemann. His excellent macroeconomic intuition has taught me a lot about how to address macro problems. I am also thankful for Jörg Rahnenführer's support as the head of the doctoral council, and chairman of my committee, for the "can-do" and "will-do" attitude, and for his efficiency in untying administrative knots.

During my studies there were many people who were influential in forming my view on econometrics and statistics. In alphabetic order I would like to thank the following people for the excellent insights, explanations, motivations, lecture notes, and late Friday statistics classes (which I was always looking forward to): Dietmar Bauer, Jean-Pierre Florens, Max Kasy, Joachim Kunert, Pascal Lavergne, Wolfgang Scherrer, Derya Uysal, and, at this place again, Martin Wagner. Their contribution to this thesis may be indirect, but is definitely significant.

I thank my colleagues in the "Villa" in Göttingen for the excellent work atmosphere. Particular thanks go to Hannes, who always listened (and hopefully not only because we shared the same office, and he had to).

It has been almost seven years since I moved away from Hungary. I thank Gábor and Tamás for always finding time to meet up in Budapest or elsewhere. My parents have always supported my ideas (even the going abroad part!), and were caring about the family home in Budapest, which does feel like a family home whenever I happen to be there. It would be hard to thank them here for the past decades of support and care. But I do appreciate everything they have done for me. Köszönök mindent!

Thank you, Jana, for everything and beyond. Pizza baking is always fun, and I'm happy to carry your books!

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

"Fed increases interest rates as inflation pressures loom." – *Financial Times*[1]
"Merkel: Germany will raise defense spending, but slowly." – *Daily Mail*[2]
"China rolls out fresh tax cuts in bid to support economic growth." – *Bloomberg*[3]

Headlines like the above often dominate the economic, financial, or even general interest sections of newspapers. Policy makers, industry professionals and the general public are equally interested in how policy actions will affect the economy, and, ultimately them.

The economic crisis of 2007–2008 has shown that there is still a lot to learn about these economic issues, especially from an empirical point of view. Central banks have opted to use "non-standard" monetary policy measures in response to the financial and sovereign debt crisis. Economic theory and empirical economics have started to devote more attention to the study of the interplay between financial markets and monetary and fiscal policy. Oil prices have dropped significantly in the past two years. In light of these developments, it seems ever more important to gain a more precise understanding of the effects of so-called macroeconomic shocks on the economy.

In economic theory, a (macroeconomic) shock is an exogenous innovation to the equation that describes the behavior of the variable under scrutiny. As a simple example, a demand shock is an exogenous event that shifts the demand curve in the basic economic model of supply and demand. More precisely, assume, that the demand and supply curves are characterized by the following equations:

$$q = -\beta p + \varepsilon_D, \tag{Demand}$$
$$p = \gamma q + \varepsilon_S, \tag{Supply}$$

where q is the quantity of an object on the market, and p is the price for one unit of the same object. A demand shock is the random variable ε_D that is uncorrelated with (or independent of) p and ε_S.

The study of economic shocks is also a statistical problem. The above set of equations immediately suggests that we can estimate the supply and demand shocks as residuals from a system of simultaneous equations for p and q if the sample $\{q_t, p_t\}_{t=1}^T$ is available to us. Since macroeconomic data often exhibit serial correlation, it is advantageous in the macroeconometric context to include lags of q_t and p_t in the equation system. Indeed, since the seminal article of Sims (1980), effects of macroeconomic shocks have been investigated principally by means of structural vector autoregressive (SVAR) models that are an important special case of dynamic simultaneous equation models.

1 https://www.ft.com/content/6723f69c-09a4-11e7-ac5a-903b21361b43, last accessed on 01.06.2017.
2 www.dailymail.co.uk/wires/ap/article-4234890/Merkel-Germany-raise-defense-spending-slowly.html, last accessed on 01.06.2017.
3 https://www.bloomberg.com/news/articles/2017-04-19/china-rolls-out-fresh-tax-cuts-in-bid-to-support-economic-growth, last accessed on 01.06.2017.

In this thesis I present results on identification in SVAR models. The supply and demand example above can serve as an intuitive illustration of the identification problem. It is well-known that the observed data points (q_t, p_t) are the results of market clearing. That is, they represent an equilibrium relationship. If a macroeconomic shock hits the system, and pushes (q_t, p_t) to the observed data point $(q_t', p_t') > (q_t, p_t)$, then, without additional knowledge, we cannot find out which curve has shifted. If, however, we postulate, e.g., that the demand curve is downward sloping, whereas the supply curve is upward sloping, then a basic supply-demand analysis makes it possible to decide, whether it was a supply shock or a demand shock (or both) that hit the system. Such additional assumptions are termed "structural." The central question of structural identification research is under what structural assumptions are the objects under scrutiny (e.g., demand shocks) identifiable.

The thesis proceeds as follows: in the next section of the present chapter, I demonstrate precisely the identification problem in structural vector autoregressions. Then I highlight recent selected advances in SVAR identification research in order to put the summary of main results into a wider context. In the summary of main results I explicitly list the findings in the thesis that are new to the literature. The main body of the thesis is made up of three chapters, corresponding to three independent and self-contained research papers. The full (bibliographic) details of these papers are listed in the last section of the present chapter. The thesis is concluded by a chapter in which I present an outlook on future research based on the findings in the contributing chapters.

1.1 THE IDENTIFICATION PROBLEM

The focus of this thesis is the identification problem in dynamic simultaneous equation systems of special form: structural vector autoregressions.[4]

A vector autoregressive process $\{y_t\}$ of order 1 is the stationary solution of the stochastic vector difference equation

$$y_t = B_1 y_{t-1} + u_t, \quad t \in \mathbb{Z}, \tag{1}$$

where $y_t \in \mathbb{R}^K$, $0 \neq B_1 \in \mathbb{R}^{K \times K}$, and $\{u_t\}$ is a white noise process with mean zero and positive definite variance-covariance matrix $\Sigma_u \in \mathbb{R}^{K \times K}$. The qualifier "structural" refers to the following modification of (1):

$$A_0 y_t = A_1 y_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}, \tag{2}$$

where $A_1 \in \mathbb{R}^{K \times K}$, $\{\varepsilon_t\}$ is a white noise process with mean zero and the identity variance-covariance matrix I_K. The two crucial elements of SVARs are 1.) the non-singular $A_0 \in \mathbb{R}^{K \times K}$ structural matrix that describes the contemporaneous relationship between the K variables in y_t, 2.) the structural innovations ε_t whose elements are uncorrelated or independent for any t.[5] The representation (1) can be obtained from (2) by multiplication with A_0^{-1}, the structural impact matrix. In this case $u_t = A_0^{-1}\varepsilon_t$, and $\Sigma_u = A_0^{-1}A_0^{-1\prime}$.

4 The definitions and observations that follow are standard. Hence, in this section I do not give explicit references. Further details can be found in, e.g. Brockwell and Davis (1991) or Lütkepohl (2005).

5 If $\{\varepsilon_t\}$ is Gaussian, then uncorrelatedness and independence coincide. Some contributions on SVARs, however, have focused on explicitly assuming that $\{\varepsilon_t\}$ is non-Gaussian, but independent, see Herwartz (2016), and Lanne, Meitz, and Saikkonen (2017).

One of the main objectives of SVAR analysis is to investigate the reaction of elements of y_t to an innovation in ε_t. In the stationary, causal case the VMA(∞) representation of y_t is written as

$$y_t = \sum_{j=0}^{\infty} \Phi_j u_{t-j} = \sum_{j=0}^{\infty} \Theta_j \varepsilon_{t-j}, \quad \Phi_0 = I_K, \tag{3}$$

where $\Phi_j = B_1^j$ for $j \geq 1$, and $\Theta_j = \Phi_j A_0^{-1}\ \forall j$, the structural impulse response coefficient matrix. Under the assumptions stated above, as long as the ordinary least squares estimator for $\widehat{B}_1$ exists, it is also unique, the reduced form is identified, and the parameters B_1, Φ_j, and Σ_u can be consistently estimated. We are now ready to state the structural identification problem in econometrics that has two aspects:

1. The decomposition of $\widehat{\Sigma}_u = \widehat{A_0^{-1} A_0^{-1\prime}}$ into $\widehat{A_0^{-1}}\widehat{A_0^{-1}}'$ is not unique without additional assumptions. Thus, A_0^{-1} (or A_0) cannot be consistently estimated without these additional, so-called structural assumptions.

2. The labeling of elements of ε_t as particular economic shocks (supply shock, monetary policy shock, etc.) requires economically founded interpretation, often in form of structural assumptions.

In the early history of SVAR research there was a genuine interplay between these two aspects. Economic (labeling) assumptions provided restrictions that made the matrix decomposition unique. And, vice versa, assumptions that make the matrix decomposition unique were interpreted also economically. As a typical example, cosider the Cholesky decomposition $\widehat{L}$ of $\widehat{\Sigma}_u$ that yields a unique triangular $\widehat{A_0^{-1}} = \widehat{L}$. The triangular structure implies which shock has an effect on y_t at the time of the arrival of the shock. For instance, it has been usually assumed that the monetary policy shock does not affect the real variables instantaneously. Thus, a monetary policy shock is an element of ε_t that is ordered after the real variables. We remark here, without further elaboration, that any structural impact matrix A_0^{-1} can be consistently estimated as $\widehat{L}Q$, for an appropriate orthogonal rotation matrix $Q'Q = QQ' = I_K$ that depends on the true A_0^{-1}. The structural identification problem can then be viewed as *"looking for Q."*

The preceding paragraphs assumed that the underlying data generating process (DGP) is, in fact, a vector autoregressive process of finite order. If the true DGP is, e.g., a VARMA process, then we can consistently estimate with a finite order VAR the correct structural innovations as long as the model is fundamental. Let $\{x_t\}_{t\in\mathbb{Z}} \in L^2(\Omega, \mathcal{F}, P)$ on some given probability space $(\Omega, \mathcal{F}, P)$, and consider $\mathcal{H}_t^x := \text{span}\{x_{t-k} : k \geq 0\}$, the closed linear span generated by x_t and its past values. The causal VARMA process $\{y_t\}$ with white noise innovations $\{\varepsilon_t\}$ is fundamental if and only if $\mathcal{H}_t^y = \mathcal{H}_t^\varepsilon$ for all t. Note, that for the reduced form innovations $\{u_t\}$ from a VAR fitted on y_t it holds by construction, that $\mathcal{H}_t^y = \mathcal{H}_t^u\ \forall t$. If the data generating VARMA process is not fundamental, then $\mathcal{H}_t^\varepsilon \supsetneq \mathcal{H}_t^y = \mathcal{H}_t^u$ for some t. That is, we cannot recover the correct structural innovations from the innovations of a finite order VAR. Thus, in my view, fundamentalness is the third (often neglected) aspect of the structural identification problem.

1.2 RECENT ADVANCES IN SVAR IDENTIFICATION RESEARCH

The identification problem has been extensively studied for the past 50 years, not only in the context of VARs, but also, in general, for simultaneous equation systems. Sufficient assumptions for the solution of the identification problem have been treated in, e.g., Hannan and Deistler (1988), or Lütkepohl (2005). There is an abundance of approaches in the econometric literature that aim to solve the identification problem. The only comprehensive book covering recent developments in SVAR analysis is still to appear at the Cambridge University Press (Kilian and Lütkepohl, to appear). This indicates that, despite the classical nature of the problem, research on identification in SVARs is still an active area.

In order to place the chapters of this thesis into a wider context, I selectively highlight three avenues of research, and some of the questions that I perceive as important open issues. Precise references to each of these can be found in the subsequent chapters.

First, sign restrictions on certain entries of the structural impulse response matrices Θ_j have become increasingly popular in the literature. The reason for this is, that agreeing on the assumption that a certain economic shock has negative (or positive) effect on an economic variable for the first several periods after the impact of the shock is less controversial than claiming that that effect is, e.g., zero in the impact period. However, it is important to note that sign restrictions do not point identify the structural parameters, since they only *restrict* the set of Q rotation matrices that are compatible with the sign restrictions. Further, sign restrictions, and in general set identifying restrictions may allow for structural models several of which are highly plausible from an economic viewpoint. It is, in my view, an advantage of this avenue of research. Note, however, that the set of structural parameters (and thus impulse responses) can be both theoretically and empirically quite large. Thus, there remain important open questions about, first, how to analyze the set of impulse responses, second, how to shrink the set further, and, third, how to discriminate between models compatible with sign restrictions.

Second, a recent development in the SVAR literature is the concept of proxy SVARs. Proxy SVARs use variables external to the VAR system (proxies) in order to solve the identification problem in a data-oriented way. In doing so, the researcher does not have to postulate structural economic assumptions, but rather the assumed exogeneity properties of the external variables provide enough additional moment conditions to identify the structural parameters. The proxy variables that are most often used in the SVAR context are benchmark policy shock measures that have been estimated outside of the VAR system, and for whom the postulated exogeneity properties (are assumed to) hold. In light of the previous paragraph, it might be that the researcher's aim is not exact identification. Then the question is, how can the proxies that produced credible results in the proxy SVAR context be used to aid identification in a milder, set-identifying context?

Third, some very recent contributions turned the focus of identification not on the structural parameters or the implied impulse responses, but on the identified shocks themselves. This is an equivalent approach insofar as to each structural form innovation ε_t corresponds an A_0 such that $\varepsilon_t = A_0 u_t$. Restrictions on the series $\{\varepsilon_t\}$ that are based on a priori economic arguments, established empirical facts, etc., can be beneficial in set-identifying the structural parameters. Since this approach is compatible with sign restrictions, or general set identifying restrictions, it also fits into the context of "shrinking the identified set" in order to discover meaningful empirical results.

As I highlight in the next section, the following chapters contribute to, and extend the ideas and questions of contemporary SVAR research outlined above.

1.3 summary of main contributions

In light of recent advances in SVAR research, this thesis makes the following contributions:

In Chapter 2, we introduce new sign restrictions to the fiscal policy literature that are based solely on the fundamental assumption of constant returns to scale (CRS) in the aggregate production function. We argue, that earlier classical results in the fiscal SVAR literature are not compatible with the CRS production function assumption, and we investigate the implications of explicitly imposing that assumption. Besides our restrictions being new to the literature, they are also methodologically new: we constrain the *relative* signs and magnitudes of certain impulse responses. Our empirical results, contrary to classical results, point towards contractionary effects of positive government spending shocks for a fiscal VAR model on quarterly US data. This finding opens up a debate about either the validity of empirical results established in the literature to date, or the validity of the CRS production function assumption.

In Chapter 3, we investigate the effect of including more (forward-looking) information in classical monetary policy SVAR models for monthly US data. To this end we augment a classical monetary policy SVAR with the federal fund futures series that arguably captures market expectations regarding policy, and Granger causes several variables in the specification. Besides futures-augmentation, we also estimate a FAVAR. We test explicitly, and establish that non-fundamentalness is not a problem in any of the empirical specifications. We contrast the estimated monetary policy shocks to two monetary policy benchmark shocks and conclude that information augmentation does not necessarily yield shocks that are more correlated with the benchmark measures. The empirical conclusions regarding the effects of monetary policy shocks are very similar among all specifications. The main finding of this chapter is, thus, that information-augmentation is not necessary from a methodological point of view, and the gains from using information-augmented models are negligible.

In Chapter 4, we develop two arguments new to the literature parallel to each other. First, our empirical motivation is to investigate the effects of monetary policy shocks on asset prices. To this end we augment standard monetary policy VARs with an asset price index, and use established sign and zero restrictions as structural identifying assumptions. To the best of our knowledge, these assumptions have not been utilized in the monetary policy – asset prices context. Second, we contrast the identified monetary policy shock estimates to an existing monetary policy shock benchmark measure. We propose to restrict attention only to those structural models that yield shocks highly correlated with the benchmark measure. We argue that such an analysis is highly successful in discovering empirical conclusions hidden by the usual practice of analyzing set-identified impulse responses. Thus, our methodological contributions challenge the interpretation of frequentist set-identified (sign- and zero-restricted) analyses in the empirical SVAR literature. Ultimately, we uncover negative but near-zero asset price responses to monetary policy shocks, coupled with mildly positive output responses.

1.4 CONTRIBUTING MATERIAL

The present thesis is comprised of three independent and self-contained research papers, two of which have been published in working paper form before.[6] The full details are:

1. Ludger Linnemann, Gábor B. Uhrin, and Martin Wagner (2016): "Government Spending Shocks and Labor Productivity," Discussion paper Nr. 9/2016, SFB 823.
2. Gábor B. Uhrin, Martin Wagner, and Uroš Herman (2017): "Monetary Policy Shocks and the Effect of the Information Set," Unpublished manuscript.
3. Gábor B. Uhrin and Helmut Herwartz (2016): "Monetary Policy Shocks, Set-Identifying Restrictions, and Asset prices: A Benchmarking Approach for Analyzing Set-Identified Models," Cege discussion papers Nr. 295.

The first item has been presented by GU as a contributed talk at the Conference of the European Economic Association in Genève, August 2016, and by MW at the Economics Research Seminar at the University of Graz. The third item has been presented by GU as an invited talk at the DIW Berlin Macroeconomics and Econometrics Seminar, in November 2016; as a contributed talk at the European Meeting of the Econometric Society in Genève, August 2016; and as a contributed talk at the Summer Workshop of the Institute of Economics of the Hungarian Academy of Sciences in Budapest, August 2016. It is, at the time of this writing, under review at the Journal of Monetary Economics. An earlier version of the third item has been circulated under the title: "Monetary Policy Shocks, Sign Restrictions, and Asset Prices: A Novel Approach for Analyzing Sign Restricted Models."

As the three papers were written independently of each other in the sequence 1, 3, 2, some parts and arguments may coincide, especially with regards to standard definitions and descriptions of standard methods. The individual papers' bibliographies have been combined and they appear at the end of this thesis. Due to submission prescriptions of certain journals, the original papers contain all figures and tables at the end. In this thesis all the figures and tables are displayed where they were intended to be displayed.

The research work culminating in the papers above has been supported by the German Research Foundation (DFG). Research papers 1 and 2 were written in the context of subproject A4 of the Collaborative Research Center (SFB) 823: "Statistical modelling of nonlinear dynamic processes." Research papers 2 and 3 benefited from the DFG Project "Macroeconomic fundamentals of asset prices: State dependence and implications for the conduct of monetary policy" (HE 2188/8-1). The financial support is gratefully acknowledged.

6 The publication of these materials was done in accordance with the regulations for the doctoral degree ("Promotionsordnung").

2

GOVERNMENT SPENDING SHOCKS AND LABOR PRODUCTIVITY

Ludger Linnemann, Gábor B. Uhrin, and Martin Wagner

Abstract. A central question in the empirical fiscal policy literature is the magnitude, in fact even the sign, of the fiscal multiplier. Standard identification schemes for fiscal VAR models typically imply positive output as well as labor productivity responses to expansionary government spending shocks. The standard macro assumption of decreasing returns to labor, however, implies that expansionary government spending shocks should lead to increasing output and hours, but to decreasing labor productivity. To potentially reconcile theory and empirical analysis we impose, amongst other sign restrictions, opposite signs of the impulse responses of output and labor productivity to government spending shocks in eight- to ten-variable VAR models, estimated on quarterly US data. Doing so leads to contractionary effects of positive government spending shocks. This potentially surprising finding is robust to the inclusion of variable capital utilization rates and total factor productivity.

2.1 INTRODUCTION

There is a large empirical literature (starting with Blanchard and Perotti (2002)) that uses structural VAR models to estimate the effects of shocks to government spending on the business cycle. A particular focus of this literature is on the identification of the fiscal multiplier, i.e., the effect of changes in government spending on aggregate output. Empirically, most studies find that an unexpected increase in government spending raises real output for at least a number of quarters, though the exact size of the multiplier is controversial. The central problem for the empirical fiscal policy literature is, of course, the problem of identification of exogenous changes in government spending. There is no consensus in the literature concerning which set of identifying restrictions should be used to disentangle government spending shocks from other shocks that affect cyclical variations in macroeconomic data.

In this paper, we propose to use the response of (hourly) labor productivity to help identify government spending shocks. The basic idea is straightforward: Consider the fiscal transmission mechanism that is embedded in most current DSGE models. If the government unexpectedly increases its spending, the resulting intertemporal tax burden imparts a negative wealth effect on households, which consequently expand their labor supply. Since the capital stock is predetermined in the short run, under a standard constant returns to scale aggregate production function there are decreasing returns to labor. As a consequence, the fiscal expansion should be associated with rising hours and output, but with decreasing hourly productivity.[1] Based on this stylized observation, we propose to use the restriction that output and labor productivity should respond with opposite signs as one of the identification conditions in a sign restricted VAR model.

We first review some popular alternative ways of identifying government shocks, namely that of Blanchard and Perotti (2002) who rely on a recursive ordering where government spending is assumed to be exogenous within the quarter; and the one proposed by Ramey (2011) who additionally controls for anticipation effects by estimating responses to the innovations to her narrative measure of the present discounted value of expected military expenditures. We demonstrate that either of these approaches implies an increase in labor productivity after a positive government spending shock in quarterly US macroeconomic data, opposite to the theoretical expectation based on the standard view of the fiscal transmission mechanism. Utilizing the above considerations on the relation between output and productivity responses to government spending shocks we estimate several variants of sign restricted VAR models for US quarterly macroeconomic time series. Sign restrictions have been used earlier in the literature on fiscal policy effects, e.g., by Mountford and Uhlig (2009) or Pappa (2009). The distinctive feature of our approach is the use of a sign restriction invoking the response of labor productivity that forces the estimated government spending shock responses to be compatible with the existence of an aggregate production function with constant returns to scale. In particular, we identify a government spending shock through the restrictions that the resultant impulse responses lead to positive comovement between government spending and public deficits, positive comovement between hours and output, and negative comovement between output (or hours) and labor productivity.

Using these restrictions, we find that the median target impulse response, as defined in detail in appendix 2.A, of private (non-farm business) output to a positive shock to government spending is negative. Since negative output reactions to government spending increases are in obvious contradiction to the consensus in the previous empirical literature, we undertake various robustness checks. In particular, we allow for cyclical capital utilization, and also include a measure of total factor productivity. The basic result remains: as soon as we impose that productivity and output have to comove negatively after government spending shocks, the median target impulse response implies a negative output reaction. Bootstrap confidence bands around the median target impulse response indicate that this negative response is statistically significantly different from zero for several periods.

Note that there are two possible interpretations of our result: First, it could be the case that government spending shocks do indeed have negative short run consequences for output and hours. In this case, one would have to assume that other identification schemes leading to the opposite result tend to confound the fluctuations due to government spending shocks with those due to other disturbances, e.g. technology shocks. Second, the transmission of government spending shocks needs to be analyzed in a setting featuring increasing returns to scale, since the data do not appear to be compatible with the combination of positive output effects of government spending and a constant returns to scale production function.

1 We consider the alternative possibility that government spending is productive in the sense of immediately shifting the aggregate production function unlikely for reasons discussed in section 3.2.

The empirical result that government spending seems to increase labor productivity is, of course, related to a finding emphasized earlier in the literature, viz., that positive government spending shocks appear to have a positive effect on the real wage rate (e.g., Perotti (2007), Monacelli and Perotti (2008)). With decreasing returns to labor, the real wage is, from a theory perspective, expected to fall if a government spending shock induces increasing labor supply. However, several authors, e.g., Hall (2009), Monacelli and Perotti (2008), or Ravn, Schmitt-Grohé, and Uribe (2012), have pointed out that higher wages may be compatible with higher employment if the price-marginal cost markup that imperfectly competitive firms charge declines in response to higher government spending. The point emphasized in the present paper is that even if declining markups make rising employment compatible with higher real wages, the increase in labor productivity that is also present in the data can still not be explained. Put differently, whatever the behavior of the markup is, it does not contribute to solving the question how sizably more output can be produced, following a government spending shock, with labor input changing only weakly.

Methodologically, we essentially use sign restrictions to impose a log-linear approximation to a standard neoclassical production function on the impulse responses. We propose to view this method as a combination of the a-theoretical nature of VAR modelling with a structural assumption concerning an aggregate production function underlying the US economy, whilst leaving all other equations unrestricted. This approach is similar in spirit to Arias, Caldara, and Rubio-Ramírez (2015), who use sign (and zero) restrictions to constrain impulse responses in a monetary VAR model such that they are compatible with a plausible central bank reaction function. Whereas Arias, Caldara, and Rubio-Ramírez (2015) require impulse responses to a monetary policy shock to reproduce a standard monetary policy rule, we impose a standard production function on the impulse responses to distinguish demand side disturbances, like government spending shocks, from supply side shifts in the production function itself. In both instances, the idea is to use only the structural information from relatively uncontroversial parts of a macroeconomic model that is implicitly thought of as the data generating process.

The paper proceeds as follows. In section 2.2, we discuss the sign restrictions that are used for identification of government spending shocks in more detail. In section 2.3, we first demonstrate the tendency for procyclical productivity responses under the Blanchard and Perotti (2002) and Ramey (2011) identifications of government spending shocks. We then discuss possible interpretations and present our own results based on sign restrictions. Finally, we show the central result to be robust to the inclusion of cyclical capacity utilization and total factor productivity. When including both additional variables we combine sign restrictions with standard short run (point) restrictions. Section 2.4 concludes. Two appendices follow the main text. Appendix 2.A presents some details of the econometric approach and appendix 2.B contains some further results.

2.2 GOVERNMENT SPENDING SHOCKS AND LABOR PRODUCTIVITY

Our main goal is to distinguish empirically between the effects of government spending shocks and of productivity shocks on the private business sector. To this end, we start by assuming that private (i.e., non-farm business) sector output Y_t is generated by a constant returns to scale production function that is standard in macroeconomics, i.e.,

$$Y_t = F(Z_t, H_t, S_t), \tag{4}$$

where Z_t is unobservable technology, H_t is labor input (measured in hours worked in the non-farm business sector), and S_t are the services derived from the installed capital stock. We concentrate on a log-linear approximation to this production function, where log-deviations from the balanced growth path are denoted by lower case letters. The log-linear representation of the production function is:

$$y_t = z_t + ah_t + (1-a)s_t, \tag{5}$$

with $a \in (0,1)$. This representation is exact in the special case that the production function is Cobb-Douglas, whereas for more general functional forms it is a first order approximation. The parameter $a \in (0,1)$ is the production elasticity of labor input, which in the Cobb-Douglas case is equal to the share of labor in total output. For other constant returns to scale production functions, that do not imply constancy of the labor share, the parameter a can also assume other values in the interval between zero and one. Macroeconomic models typically calibrate values for a in the range from 0.6 to 0.7.

Now consider estimating a VAR model containing (among others) the variables from above. Then, following any shock hitting the economy, the estimated impulse responses of output, technology, hours worked and capital services should, to a first order approximation at least, be related to each other as the variables in (5). In the following we will repeatedly compare relations between impulse response functions of VAR models and log-linearized structural economic relations.

We use this idea to disentangle government spending shocks from other shocks, in particular from technology shocks. If in period t a shock that does not change technology occurs, then $z_t = 0$ holds in this period and the impulse responses hence fulfill:

$$y_t - h_t = (a-1)h_t + (1-a)s_t. \tag{6}$$

However, capital services are typically not directly observable. We consider two alternative specifications to deal with this problem. The first assumes that capital services s_t are equal to the stock of installed capital (or are a fixed proportion of it), and the second assumes that capital services are given by the product of a time variable utilization rate and the capital stock. We present the first specification in the current section, and defer the discussion of the second as a robustness exercise to section 2.3.4.

If capital services are identical to the capital stock, then—since the capital stock is predetermined in the short run and slowly moving in response to shocks in general—their contribution can be neglected as long as the focus is on the economy's behavior in the immediate aftermath of a few quarters after a shock hits. Thus, the impact or short run effect of a non-technological shock on labor productivity is well approximated by:

$$y_t - h_t \approx (a-1)h_t, \tag{7}$$

since $s_t \approx 0$ on impact. Given the standard range of estimates of $a \in [0.6, 0.7]$, this implies that in the short run, if a non-technological shock increases hours worked by one percent, labor productivity should decline by between -2.5 to -3.3 percent. In the limiting case where $a \to 1$, the effect on labor productivity vanishes. Importantly, however, it cannot be positive for any value of a that implies decreasing or constant returns to labor in production.

While the exact value of a is unknown in general, (7) is nonetheless useful as the basis for identifying government spending shocks based on the signs of impulse responses. In particular, suppose we have estimates of a reduced form VAR model, and consider a particular candidate orthogonalization of the residuals in order to identify structural government spending shocks. Denote the impulse responses for the candidate orthogonalization at horizon $j \geq 0$ to a government spending shock f_t by a tilde over variables (e.g., $\widetilde{y}_j = \partial \log Y_{t+j} / \partial f_t$). Our maintained hypothesis is that government spending does not have a direct effect on technology (see below for further discussion of this point) and that the capital stock is predetermined in the short run. Therefore, a structural government spending shock should produce impulse responses that are compatible with (7) with $a \in (0,1)$ and that, hence, need to have the following properties:

(i) $\widetilde{y}_j$ and $\widetilde{h}_j$ have the same sign;

(ii) $\widetilde{y}_j$ and $\widetilde{y}_j - \widetilde{h}_j$ have opposite signs.

Since these properties of impulse responses can be expected to be present, in the short run, after any type of non-technological (or demand side) shock that leaves total factor productivity unchanged, we need a further restriction to ensure that the particular demand side shock we identify is indeed a government spending shock. Therefore, letting $\widetilde{g}_j$ and $\widetilde{d}_j$ denote the impulse responses at horizon j of government spending and the deficit, respectively, we add:

(iii) $\widetilde{g}_j$ and $\widetilde{d}_j$ have the same sign.

Below, we make use of these properties in the form of sign restrictions on the impulse responses of VAR models to identify government spending shocks. Restriction (i) requires that output and labor must comove positively, which is a basic requirement if a non-technological shock is considered and capital is predetermined in the short run. In this case labor is the only variable factor that can adjust in the short run to produce more or less output. Restriction (ii) is crucial for our approach. It imposes the decreasing returns to labor property following from a constant returns to scale production function with predetermined capital. Under non-technological shocks, output can only rise if measured labor productivity declines, such that we observe a positive response $\widetilde{y}_j$ only if $\widetilde{y}_j - \widetilde{h}_j$ declines at the same time, or vice versa. This restriction is pivotal in the present context, since it imposes the condition that a government spending shock is a pure demand side disturbance that does not shift the aggregate production function as, e.g., a technology shock would. Finally, condition (iii) serves to single out government spending shocks from other non-technological disturbances. It imposes that government spending shocks are at least partly deficit financed over the short run. This assumption is plausible in view of the political decision process, with spending changes rarely linked to specific tax changes required to finance them.

Note, importantly, that conditions (i) to (iii) neither constrain the signs of the reactions of output nor of hours worked to a government spending shock. It is only the *relation* between these two reactions that is restricted. The idea is that the basic notion of a demand side disturbance brought about by government spending changes imposes the required pattern of comovement between the impulse responses, as long as the data generating process is characterized by a constant returns to scale production function. It is left unrestricted, and hence decided by the data, whether this implies that output and hours increase while productivity decreases, or that output and hours decline while productivity rises.

In the next section, we proceed in three steps. First, in section 2.3.1 we review some popular identification schemes that have been used in the fiscal VAR literature to identify government spending shocks. We discuss whether the impulse response functions generated by these models are compatible with the theoretical requirements that characterize responses to government spending shocks as set out in conditions (i) to (iii) in section 2.3.2. Since the answer turns out to be negative, we proceed in section 2.3.3 by directly imposing conditions (i) to (iii) as the restrictions to identify government spending shocks via sign restrictions on VAR model impulse responses. Finally, in section 2.3.4 we investigate the robustness of the results with respect to allowing for variable capital utilization.

2.3 EMPIRICAL RESULTS

2.3.1 *Review of existing fiscal VAR model results*

We start off by reviewing standard findings of the empirical literature on the effects of government spending shocks. Given the above discussion, negative comovement between the impulse responses of output and labor productivity to government spending shocks should prevail. Consequently, the first question we ask is whether the available fiscal VAR model results are compatible with this restriction. The answer is no. In section 3.3 we therefore present results where we impose this negative comovement between the output and productivity responses to government spending shocks via sign restrictions.

All VAR models considered in this paper are estimated with quarterly US data from 1948q1 to 2013q4, which is the longest period over which all variables are available. The variables used in the baseline specification in this section are the logarithm of real government consumption and investment spending, $\log G_t$; the logarithm of real output in the non-farm business sector, $\log Y_t$; the logarithm of hourly labor productivity, $\log Y_t - \log H_t$, where H_t is hours worked in the non-farm business sector; the logarithm of real net taxes, $\log \tau_t$;[2] the nominal three months treasury bill rate, R_t; the inflation rate as measured by the annualized log change in the deflator of non-farm business output, π_t; the government deficit, D_t, defined as minus total government saving as a fraction of GDP; and the logarithm of real private nonresidential investment, $\log I_t$.

We have checked the robustness of our results by using, instead of τ_t as defined above, the Barro and Redlick (2011) measure of the average marginal tax rate, which is available only up to 2008q4 and thus requires using a shorter sample. The results do not change by

2 Here τ_t is defined as government current tax receipts plus contributions for government social insurance less government current transfer payments, deflated by the GDP implicit price deflator.

much, and therefore we use in our analysis the tax measure τ_t and the longer sample until 2013q4. The data on hours worked and the Barro–Redlick tax rate have been downloaded from Valerie Ramey's website, the other variables are obtained from the Federal Reserve Bank of St. Louis FRED database, except for private nonresidential investment, which is from the Bureau of Economic Analysis. Expressing the flow variables as per capita values by dividing through population does not change the results appreciably. To match the approach commonly used in the literature, all models also contain a constant as well as linear and quadratic time trends and are estimated with four lags of each endogenous variable.

Note that in all estimates below both output and hours, and thus productivity, are measured for the private (non-farm business) sector only. This seems important in the present context, because using economy-wide measures—such as real GDP and total hours worked—could be misleading. The reason for this is that GDP also contains the public sector output, which is difficult to measure and for which the existence of a standard production function is not necessarily guaranteed. Therefore, we only investigate the response of private output and private hourly productivity to government spending shocks. That being said, the results reported below only change very little if economy-wide GDP based measures for output and productivity are used instead of the non-farm business data, as we have ascertained by running this specification as another robustness check.

For comparison with our own results shown in the next subsection, as a first step we show the implications of three commonly used VAR identification methods for the response of labor productivity in the private non-farm business sector to a government spending shock. The first approach imposes Blanchard and Perotti's 2002 assumption that government spending does not react endogenously to the state of the economy within the quarter, but only with at least a one quarter lag. Thus, the government spending shock is in this setting identified by using the recursively orthogonalized residuals from a VAR model with the variables mentioned above with government spending ordered first. For brevity, this is called BP or recursive identification, henceforth. The BP approach has been criticized by Ramey (2011), who argues that the possible presence of anticipated changes in government expenditure invalidates the BP identifying assumption. If news of future rising expenditure arise, the private sector will respond before the econometrician actually observes an increase in measured spending. The resulting mismatch of timing could then lead to erroneous estimates of the shock responses. To overcome this problem, Ramey (2011) proposes the use of a narrative measure of the present discounted value of anticipated military spending to identify government spending shocks (orthogonal in addition to this variable). Therefore, the second approach shown below adds Ramey's 2011 variable for the present discounted value of expected future military expenditure as the first variable in the VAR model, and calculates an anticipated government shock as an orthogonalized innovation to this variable. This is called the Ramey identification for short. The third approach uses the same VAR model specification as the previous one, i.e., with the Ramey news variable ordered first and government spending ordered second, but considers a shock not to the anticipation variable, but to the spending variable itself. In this way, this identification can be seen as an attempt to capture an unanticipated spending shock while at the same time

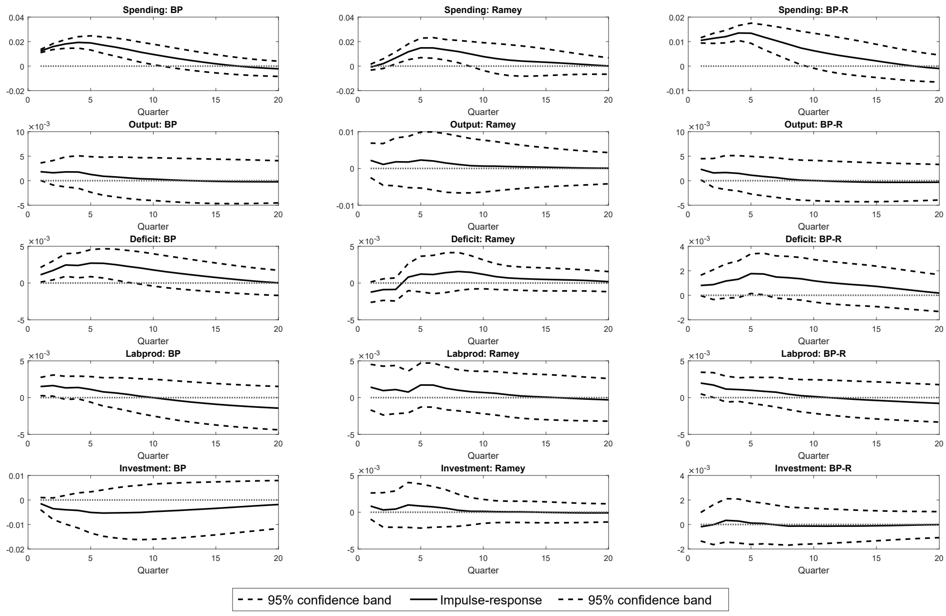


Figure 1: Impulse responses to government spending shock: BP, Ramey and BP-R identification schemes.

controlling for anticipation effects through the inclusion of Ramey's news variable, which continues to be ordered first. This third specification is abbreviated as BP-R below.

Figure 1 shows the results in terms of impulse responses to a one standard deviation shock to government spending or Ramey's 2011 news variable using these three identification schemes, along with ± 1.96 bootstrapped standard erros to capture symmetric 95% confidence bands. For brevity, only the responses of the most interesting variables for the question at hand are shown. The full set of impulse responses for all variables included in the VAR models is available upon request. In all identification schemes, a positive government spending shock raises private sector output (though only insignificantly so in the Ramey version), and the government deficit (though less clearly and with a lag in the Ramey specification). Most importantly for the present purpose, however, is the fact that under all identification schemes labor productivity (shown in the last but one row of figure 1) rises slightly. The increase in productivity is certainly not large, and in the Ramey case again not significantly different from zero. However, as argued above, if one believes that these models truly identify a government spending shock, then one expects a pronounced decrease in labor productivity.

In principle, it is possible that the increase in measured labor productivity is explained by the effect of a decline in hours worked on marginal productivity of labor. However, this does not seem to be the case. Replacing the productivity variable $\log Y_t - \log H_t$, used in the VAR models above, by the logarithm of hours worked, $\log H_t$, and re-estimating (leaving the rest of the VAR model unchanged) yields the estimated impulse responses of hours to a positive government spending shock in the three specifications shown in figure 2.

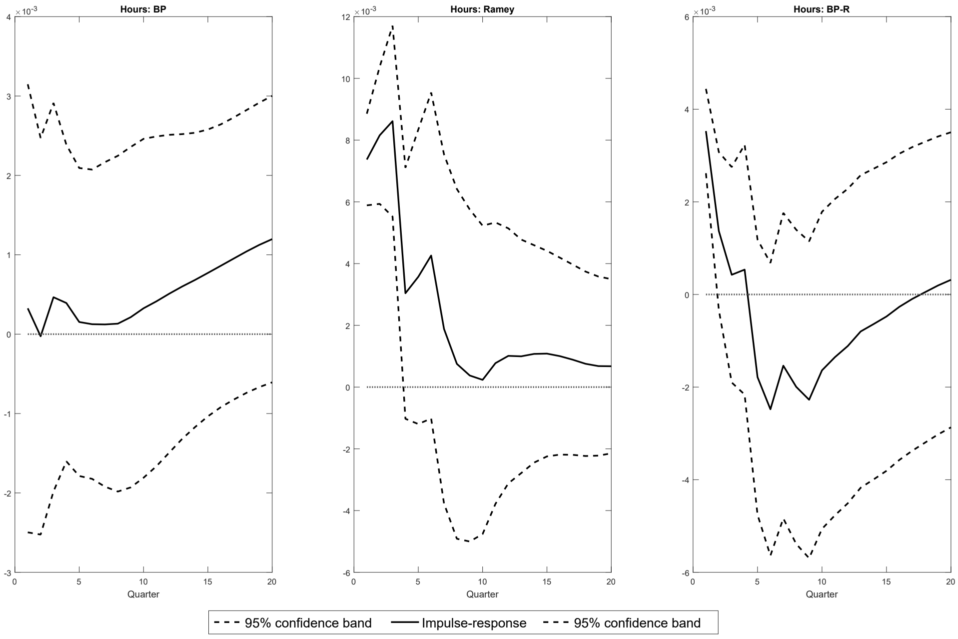


Figure 2: Impulse response of non-farm business hours to government spending shock: BP, Ramey and BP-R identification schemes.

In all cases, the response of hours appears to be close to zero or slightly positive, at least for the first couple of quarters after the shock, but not markedly negative. Thus, the behavior of hours does not seem to explain the estimated increase in productivity. Moreover, even if the hours response were indeed negative, this would raise the question how, in that case, a positive short run output response could be explained if the maintained assumption that these models correctly identify a purely non-technological government spending shock is correct. If technology does not change and the capital stock is predetermined in the short run, then rising output is associated with increases in hours worked (the possible caveat in the case that the output expansion is explained by a large concomitant increase in capital utilization is explored in section 2.3.4 below).

To sum up, the conclusion obtained so far from standard structural VAR model approaches is that output increases following a government spending expansion are difficult to explain without rising productivity. The very fact that the VAR model results point to productivity increases following rising government spending, casts doubt on their ability to identify a pure demand side innovation like a government spending shock. If the popular identification methods shown above truly identify government spending shocks, and if government spending shocks are truly non-technological in nature, one expects that impulse responses of output and hours have the same sign and are both of the opposite sign of the response of labor productivity. Yet, in the estimates it appears that output comoves positively with productivity, conditional on the identified shock, and weakly positively with hours. Thus, to the extent that these conventional identification schemes indeed succeed in isolating government spending shocks, one needs to explain how an increase in government spending is able to raise labor productivity.

2.3.2 *Discussion*

While in the recent literature the debate has revolved around estimating the magnitude of the effect of government spending shocks on output so far (the fiscal multiplier debate), the empirical evidence provided above highlights a different aspect: However large the output effects may be, they tend to derive not only from comparably large increases in hours or employment, but also from increases in labor productivity.

This poses an interesting challenge to our understanding of the fiscal transmission mechanism. The evidence given above seems incompatible with the usual view of the way government spending affects the economy, as it is embedded in most DSGE models. The standard transmission mechanism implies that an increase in government spending raises output because higher spending, through its associated tax burden, exerts a negative wealth effect on households. This gives households an incentive to reduce their consumption of leisure, which boosts labor supply such that output rises. Along a neoclassical production function with capital predetermined in the short run, this implies that decreasing returns to labor set in. Hence, a decrease in measured labor productivity results.

Three principally different reactions to the apparent conflict between theory and empirical evidence are conceivable. First, the standard view of the fiscal transmission mechanism needs to be augmented. If the positive labor productivity response is structural, one has to adjust theoretical models to accommodate it. Second, the identification methods discussed above tend to confound government spending shocks with other shocks, in particular with technological shocks that are known to raise productivity. A positive technology shock raises productivity and could be mistaken for a government spending shock in a recursive identification scheme, if the government immediately increases spending in response to the positive technological shock. Third, an increase in activity following a government spending shock triggers a rise in unmeasured factor utilization, in particular capital utilization. This might counteract decreasing returns to labor since the unobserved variable utilization rate of capital increases too. We discuss each of these possibilities in turn.

If one adopts the first view and maintains that the orthogonalizations applied in the VAR models shown above succeed in identifying structural government spending shocks, it could indeed be that the measured increase in labor productivity is structural. One possibility for this is that government spending is productive, in the sense of entering private sector production functions with a positive output elasticity. Higher government spending then shifts up the production functions of private firms and leads to a labor productivity increase. However, direct productivity effects of government spending most likely result from investment in public infrastructure. This, as a part of the economy's total capital stock, only changes slowly and therefore can be considered as predetermined in the short run following a spending boost.

Another possibility is that there are increasing returns to scale, and more stringently increasing returns to labor. In this case any increases in the scale of production, including those brought about by an increase in government spending, lower average costs and thus endogenously raise overall productivity. However, while this could, if the relevant effects are strong enough, also lead to a rise in measured labor productivity, one expects that (as also in the case of infrastructure effects from higher spending) private investment increases too, since private investors would attempt to take advantage of higher productivity. The impulse responses of investment are, however, not significantly different from zero in the three discussed structural VAR models. It is positive but not significantly different from zero in the specifications using the Ramey news variable, and negative (albeit not significantly different from zero) in the BP identification. Several other studies have also found negative investment responses to government spending shocks (e.g., Galí, López-Salido, and Vallés (2007)). Thus, the positive investment response that one expects if higher government spending truly increases productivity (either by shifting the production function by adding public capital, or by shifting the economy along an increasing returns to scale production function) does not seem to receive much empirical support.

Hence, we conclude that while we cannot strictly rule out the possibility that procyclical productivity is indeed a structural feature of government spending shocks, we consider the evidence in favor of this hypothesis to be weak. Note that this also rules out the possibility that labor productivity simply increases, because higher private investment raises the capital stock quickly enough. Even if there were a positive private investment response, this effect is expected to work intertemporally, with some delay because of the short run predetermined nature of the capital stock. The productivity response instead appears to be immediate.

In sum, this leaves us with either the second or the third view, namely that the non-negative productivity response either follows from failure to identify and disentangle government spending shocks from technological shocks with the methods employed above, or that it is the result of unaccounted increases of capacity utilization. The following two sections are dedicated to our attempt to distinguish between these possibilities.

2.3.3 *Results with sign restrictions imposed*

In this subsection, we present the results when we impose the discussed sign restrictions on the impulse responses from VAR models. We impose restrictions (i) to (iii) introduced above (positive comovement of output and hours, negative comovement of output and productivity, positive comovement of government spending and the budget deficit) on the impulse responses of the VAR model to identify government spending shocks. The crucial restriction is (ii), which has to be fulfilled by responses to demand side shocks like government spending shocks, but not by responses to technology shocks. In this way, the sign restrictions are used to separate government spending shocks, whose effects we want to analyze, from technology shocks.

The estimated VAR model contains essentially the same variables as discussed in the preceding subsection. The difference is that we include output and hours separately in order to be able to constrain their impulse response relation. Thus, the following variables are included $\log G_t$, $\log Y_t$, $\log H_t$, $\log \tau_t$, R_t, π_t, D_t, $\log I_t$. Furthermore, we again include four lags, a constant and linear and quadratic time trends. We implement the sign restrictions following the methodology outlined in Rubio-Ramírez, Waggoner, and Zha (2010). In brief

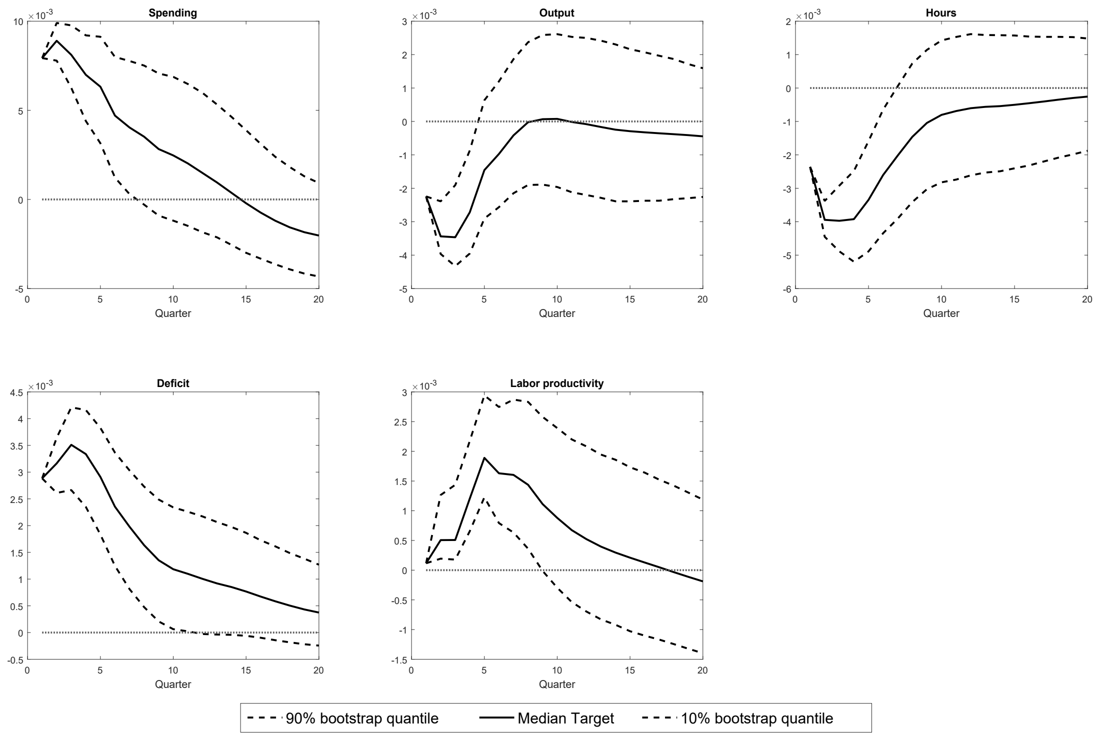


Figure 3: Median target impulse responses to government spending shock identified through restrictions (i) – (iii).

(for more details see appendix A), we randomly draw orthogonal matrices to rotate the so-called structural impact matrix until we have 5000 models in which the impulse response patterns to a government spending shock match restrictions (i) – (iii) over a horizon of four quarters. Robustness checks show that imposing the restrictions only for one or two quarters following a shock does not change the conclusions. From the responses fulfilling the sign restrictions, we calculate the median target (MT, henceforth) impulse response as advocated by Fry and Pagan (2011). The MT impulse response is the impulse response that is closest in a (variance weighted) squared distance sense to the (pointwise) median curve of the 5000 impulse responses satisfying the sign restrictions. To allow for inference, we quantify the uncertainty around the estimated MT responses by a bootstrap method described in appendix 2.A, and use this to construct 90 percent confidence bands that are depicted in the figures below as dashed lines.

Figure 3 shows the estimated effects of a positive impulse in government spending. In terms of the median target responses, a positive government spending shock is associated with an increase in the deficit and labor productivity, but with an initial decrease in output and hours worked. Note that while spending and the deficit have been restricted to be positive, output and hours are unrestricted. Only their relation is restricted by (i) and (ii) given above. The median target effect of government spending shocks on output is significantly negative for several periods.

The result that government spending expansions are associated with negative output and hours responses is, of course, surprising. As mentioned above, a large number of previous studies—using different identification assumptions—finds that positive government spending shocks are associated with short run increases in output. Hence, it is crucial to understand why our results differ markedly in this respect. The reason is, of course, that we restrict the relation between the responses of labor productivity by our restriction (ii) to be in accord with our view of the consequences of demand shocks, i.e., negative comovement between the impulse responses of output and productivity. In other words, if labor productivity rises when a government spending shock has occurred, this must have been due to the increase in the marginal product of labor. This increasing marginal product of labor is implied by the decline in hours, and thus in output. Hence, by using restriction (ii) we in a sense force the data to decide whether, conditional on a government spending increase, either an increase in output and hours with lower productivity, or a decrease in output and hours associated with a rise in productivity is more likely. The results shown in figure 3 indicate that the data appear to favor the latter possibility.

The results in figure 3 allow for different possible interpretations. One possible conclusion is that previous estimates that find a positive response of output to government spending increases (like those summarized in the preceding subsections) fail to disentangle government spending shocks from other confounding disturbances, like technology shocks. Our estimates, in contrast, explicitly rule out the influence of shocks that shift the short run production function and thus could be seen as identifying the pure demand side effects of government spending shocks.

It is important to stress that the results shown in figure 3, as well as in figures 4 and 5 to be discussed later, display the median target response. The corresponding figures 6 to 8 in appendix 2.B show the range of the sign-restricted impulse responses as generated by our simulation approach. The results from the appendix show that the largest part of the impulse responses has qualitatively the same shape as the median target impulse that we focus on, since the pointwise median curve over all impulse responses throughout is close to the median target impulse response. The figures in appendix 2.B, however, also show that there are feasible sign-restricted impulse responses with the opposite implications regarding the effects of government spending shocks on output and hours worked. There is no statistical way of discriminating between these different feasible orthogonalizations, as they are all observationally equivalent to the estimated reduced form VAR model. In the literature it is customary to focus on either the pointwise median curve (not itself an impulse response function) or the median target impulse response to capture the main tendency in the data. Both lead to very similar conclusions in our case. Second, and more problematically, we have thus far assumed that labor is the only variable factor of production that can adjust in the short run. This is debatable when the amount of services derived from the capital stock varies over the business cycle, as is implied by many theoretical models with variable capital utilization. We thus turn to an enlarged model where we allow for utilization changes in the following section.

2.3.4 *Robustness checks: variable capital utilization and total factor productivity*

So far, we have assumed that capital services s_t are identical (or proportional) to the capital stock k_t. This is a useful simplification, because the capital stock is predetermined in the short run, and moves only slowly even over the medium run. Hence, under this assumption it is possible to abstract from changes in capital services, at least for the small number of time periods for which sign restrictions are imposed on impulse responses. However, it is indeed likely that capital services are more variable than the capital stock itself, if the utilization rate of the latter is time varying. The question thus arises in how far our results are robust to allowing for variable capital utilization.

Time varying capital utilization is found to be an important feature of business cycles in several recent papers, e.g., Justiniano, Primiceri, and Tambalotti (2010). Empirically, Fernald (2014) provides a measure of the change in utilization that he computes based on the methodology described in Basu, Fernald, and Kimball (2006).[3] In the following, since the other variables in our VAR models are in log-levels as well, we use his measure of utilization change and integrate it (from a starting value of one) to obtain the level of utilization U_t (which is then taken to logarithms in the empirical model), and allow the services of capital to depend on it through $S_t = U_t K_t$, where K_t is the stock of installed capital.

Allowing for variable capital utilization, the log-linearly approximated production function thus reads as:

$$y_t = z_t + a h_t + (1-a)(u_t + k_t). \tag{8}$$

Under non-technological shocks, i.e., with $z_t = 0$, and upon neglecting movements in the capital stock which continues to be predetermined, it follows that measured labor productivity is approximately given by:

$$y_t - h_t \approx (1-a)(u_t - h_t). \tag{9}$$

Hence, given $a \in (0,1)$, labor productivity rises in response to a non-technological shock only if utilization u_t increases more strongly than hours worked h_t. Thus, with variable capital utilization our previous restriction (ii), which requires output and productivity to have opposite signs, may be too restrictive.

We thus extend the VAR model of the previous section with the logarithm of the level of utilization as an additional variable. The variables used are thus $\log G_t$, $\log Y_t$, $\log H_t$, $\log \tau_t$, R_t, π_t, D_t, $\log I_t$, $\log U_t$. Using the same notation as in section 2, let $\widetilde{u}_j$ denote the impulse response at horizon j of $\log U_t$ to a government spending shock. In terms of identification restrictions on the impulse responses, we replace the sign restriction (ii) by a new sign restriction (iv):

(iv) The difference of the impulse responses $\widetilde{y}_j$ and $\widetilde{h}_j$ has the same sign as the difference of the impulse responses $\widetilde{u}_j$ and $\widetilde{h}_j$.

3 The data are available at John Fernald's web site http://www.frbsf.org/economic-research/economists/john-fernald/.

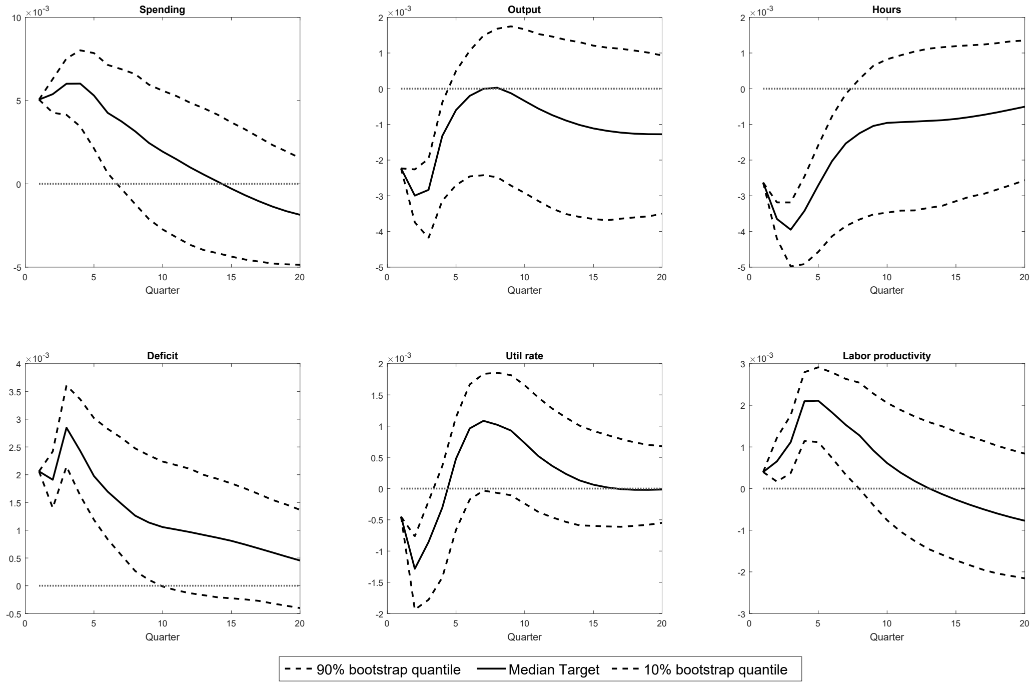


Figure 4: Median target impulse responses to government spending shock identified through restrictions (i), (iii) and (iv).

Figure 4 shows the MT impulse responses of the utilization-augmented VAR, with the government spending shock identified using sign restrictions (i), (iii), and (iv). A positive shock to government spending appears to trigger a strong negative adjustment of utilization in the short run. Note that the utilization response itself is not sign-restricted by (iv), but only its relation to the output and hours responses. As can be seen by a comparison with the results shown in figure 3 above, the other responses do not change qualitatively compared to the case without time varying utilization, although the magnitudes and the persistence of the responses differs. In particular, the median target output and hours reactions are still negative in the short run, even though less strongly so, since the decrease in utilization picks up part of the variation. Capital utilization itself is, as theoretically expected, procyclical, which in the current context means that it declines alongside output. Since utilization declines by less than hours, labor productivity rises by implication.

Thus, allowing for time varying utilization does not change the basic conclusion reached above that imposing constant returns to scale (in hours and in utilized capital, given the predetermined capital stock) leads, in the short run, to a negative median target response of output and hours to a government spending shock, accompanied by a positive productivity response.

Recall that the main purpose of the restrictions we use is to help disentangle government spending shocks from other shocks that directly shift the production function. Therefore, it might be useful, as a further robustness check, to control directly for a measure of technology in the VAR model. It is well known that standard measures of total factor productivity

(TFP) that are based on the classic Solow residual contain a component that is endogenous to the business cycle. The reason is that with procyclical utilization and possibly imperfect competition, the productive contribution of the input factors is larger than their income shares, with the latter commonly used as production elasticities in the computation of Solow residuals. Fernald (2014) also provides a corrected TFP measure that takes account of these effects, based on a methodology to purge spurious cyclicality due to utilization changes and markups expounded in Basu, Fernald, and Kimball (2006). Thus, his utilization corrected TFP series is likely to be a better proxy for exogenous shocks to the aggregate production function, and hence an appropriate control variable for us. Since his measure is in growth rates, we integrate it from a starting value of one to get the variable TFP_t, and use its log-level as an additional variable in the VAR model.

After a government spending shock that has no impact on technology, the impulse response at horizon j of the total factor productivity measure to this government spending shock, $\widetilde{tfp}_j$, should be zero, at least in the vicinity of the shock impact at short horizons j. We thus impose, as an additional identifying restriction, the exact zero-at-impact restriction:

(v) The impulse response $\widetilde{tfp}_j$ does not change on impact under government spending shocks.

This model version thus mixes the sign restrictions (i), (iii), and (iv) with the exact zero restriction (v). The implementation is based on the methodology set out in Arias, Rubio-Ramírez, and Waggoner (2014) described in appendix 2.A. Figure 5 shows the median target impulse responses of the VAR model with the variables $\log G_t$, $\log Y_t$, $\log H_t$, $\log \tau_t$, R_t, π_t, D_t, $\log I_t$, $\log U_t$, $\log TFP_t$.

The MT impulse responses shown in figure 5 show some differences compared to those previously discussed. In particular, while output and hours still decline in the short run, the size of the negative response is somewhat mitigated. To the extent that the inclusion of the $\log TFP$ variable succeeds in controlling for residual technological disturbances unaccounted for in the previous models, the estimates shown in figure 5 give a cleaner indication of the consequences of a government spending shock. Most clearly visible, the response of utilization now appears rather unclear, and statistically not significantly different from zero. This is also true for the TFP response itself, which is only constrained to be exactly zero in the impact period, and shows some endogenous but altogether insignificant variation thereafter. Labor productivity reacts less strongly than in the previous models, but the response is still positive. However, the main pattern found in the simpler models above still holds: Output and hours tend to decline for some periods following a government spending increase, whereas labor productivity rises slightly. We thus conclude that the central result presented in the previous subsection is robust to the consideration of both variable capital utilization and total factor productivity as additional control variables.

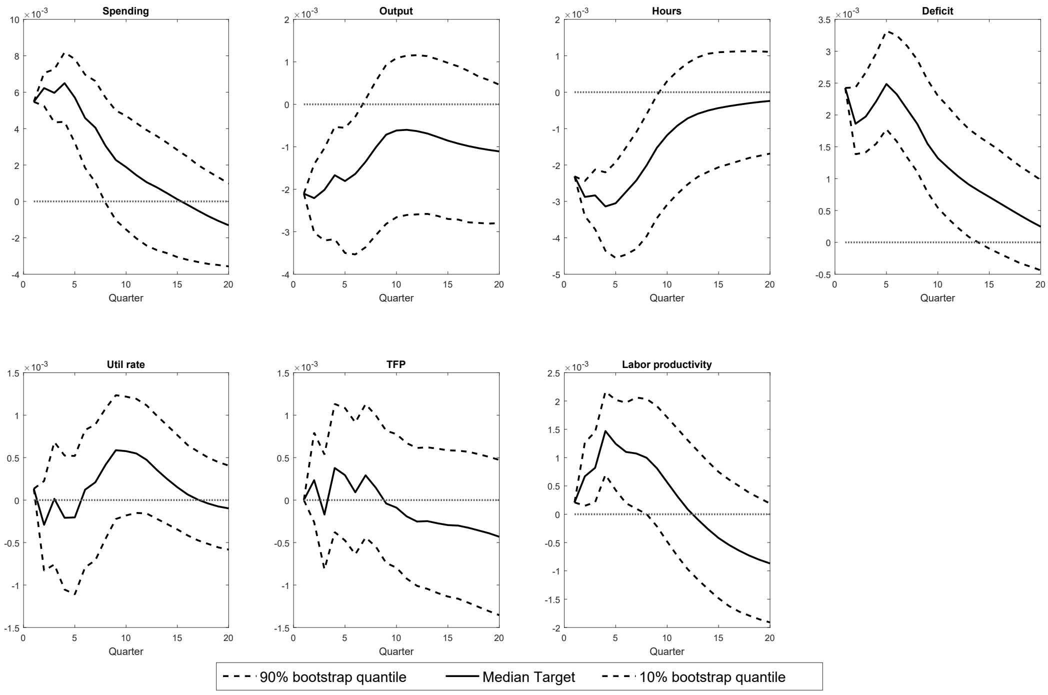


Figure 5: Median target impulse responses to government spending shock identified through restrictions (i), (iii), (iv) and (v).

2.4 CONCLUSIONS

Taking stock, the estimates presented above all highlight the central point: As soon as we impose the crucial requirement that the impulse responses of structural VAR models aimed at identifying government spending shocks exhibit behavior required to be consistent with a standard constant returns to scale aggregate production function, we find that the median target responses to a government spending increase imply short run declines in private sector output and hours, along with rising labor productivity. This result is robust to the inclusion of variable capacity utilization and to additionally including a measure of utilization-adjusted total factor productivity. Since the majority of previous studies has found positive output responses following government spending increases, the question arises, of course, how to interpret the results presented here.

Our results certainly cannot be taken to necessarily imply that other identification schemes that tend to find positive output responses to government spending increases are wrong. While there is the possibility that identification schemes that do not take into account the restrictions we impose on productivity behavior confound demand shocks deriving from government spending variations with technology shocks, we need to be cautious here for at least three reasons. First, sign restriction methods do not allow to exactly identify government spending shocks, but only the set of admissible model impulse responses given the restrictions. The range of admissible models includes impulse responses for output and hours of both signs. However, as demonstrated by the median target impulse responses shown above (and by the figures in appendix B), the majority of admissible impulse responses points towards a negative reaction of these variables, when forced to have a negative correlation of the responses of output and labor productivity in the short run.

Second, we use an estimated regressor as our variable capital utilization rate, which itself is not directly observable. Hence, although the measure is carefully constructed by Fernald (2014), there might still be unaccounted residual variation in true capital utilization that is not captured in the measured variable. As a consequence, observed labor productivity behavior could still be misleading and not fully capture movements in the marginal product of labor.

Finally, it might be that the main identifying restriction we use, namely constant returns to scale in the aggregate production function, does not hold empirically. In this case, the negative impulse response correlation between output and utilization-adjusted productivity need not hold. This would invalidate our central identification assumption. While this is possible in principle, we note that a large majority of business cycle models assumes the standard assumption of constant returns to scale. Allowing for increasing returns to scale requires an altogether rethinking of the fiscal transmission process in such models. The distinction between these possibilities is arguably an important topic of future research. At this point, we conclude that the data seem to imply that shocks to government spending either have negative output consequences, or if they have not, then this can, in our view, only be explained through the existence of aggregate increasing returns to scale.

2.A ECONOMETRIC DETAILS

The VAR model

We describe the employed econometric model in this section for the sake of completeness and, mainly, to fix notation. A k-variable structural VAR model for x_t is given by:

$$A_0 x_t = \mu + C_1 t + C_2 t^2 + A_1 x_{t-1} + A_2 x_{t-2} + \cdots + A_p x_{t-p} + \varepsilon_t, \tag{10}$$

where $x_t \in \mathbb{R}^k$, $\varepsilon_t \sim WN(0, I_k)$, $A_0, \ldots, A_p \in \mathbb{R}^{k\times k}$, $\mu, C_1, C_2 \in \mathbb{R}^k$. A_0, the so-called *structural matrix*, is assumed to be non-singular. In order to define a unique lag length p we assume $A_p \neq 0$, in our application $p = 4$. The corresponding reduced form, obtained from (34) by pre-multiplication with A_0^{-1}, is given by:

$$x_t = \nu + D_1 t + D_2 t^2 + B_1 x_{t-1} + \cdots + B_p x_{t-p} + u_t, \tag{11}$$

with $B_i = A_0^{-1} A_i$, $i = 1, \ldots, p$, $D_j = A_0^{-1} C_j$, $j = 1, 2$, $\nu = A_0^{-1}\mu$, and $A_0^{-1}\varepsilon_t = u_t \sim WN(0, \Sigma_u)$, with consequently $\Sigma_u = A_0^{-1} A_0^{-1\prime}$.

Denoting with $B(z) = I_k - B_1 z - \cdots - B_p z^p$, we impose the *causality assumption*, $\det(B(z)) \neq 0 \ \forall \ |z| \leq 1$. Under this assumption, the errors u_t correspond to the one-step prediction errors from the Wold decomposition, i.e., we obtain an infinite order moving average representation of the form:

$$x_t = \tilde{\nu} + \tilde{D}_1 t + \tilde{D}_2 t^2 + \sum_{j=0}^{\infty} \Phi_j u_{t-j} \tag{12}$$

$$= \tilde{\nu} + \tilde{D}_1 t + \tilde{D}_2 t^2 + \sum_{j=0}^{\infty} \Phi_j A_0^{-1} \varepsilon_{t-j} \tag{13}$$

The (r, s)-element of $\Theta_j = \Phi_j A_0^{-1}$ describes the change of variable r to a unit increase of $\varepsilon_{t,s}$ after j-periods, i.e., at horizon j. In the main text we use the short-hand notation $\widetilde{m}_j$ for this, with m denoting an element of the vector of variables x_t, since we are throughout only interested in the effects of government spending shocks, i.e., for one particular s only.

Identification schemes

As is well-known, the structural form (10) is not identified, for a detailed discussion see Hannan and Deistler (1988). The literature provides a large array of approaches to point or set identification of structural VAR models. In our paper we employ the following ones:

(1) Recursive identification: The structural matrix, A_0 is lower (upper) triangular, or, equivalently, A_0^{-1}, the *structural impact matrix*, is lower (upper) triangular.

(2) Sign restrictions: $\Theta_j^{(r,s)}$ is restricted to be either nonnegative or nonpositive for some combinations of (r, s, j), $r, z \in \{1, \ldots, k\}$, $j \in \mathbb{N}_0$.

(3) Zero and sign restrictions: $\Theta_{\bar{j}}^{(\bar{r},\bar{s})} = 0$ for some $(\bar{r}, \bar{s}, \bar{j})$, and some $\Theta_j^{(r,s)}$ are sign restricted as defined in the previous item (2). Note already here that in our paper we only consider zero-at-impact restrictions, i.e., $\bar{j} = 0$.

The recursive identification scheme yields an exactly identified structural VAR. It is important to note, however, that sign restrictions and the mixture of zero and sign restrictions, in general do *not* yield exactly identified structural forms. The identified set for the impulse responses is, thus, in general non-singleton for the second and third cases.[4]

Impulse response functions

The reduced form parameters are estimated by ordinary least squares, resulting in $\widehat{B}_1, \ldots, \widehat{B}_p$ and $\widehat{\Sigma}_u$. Thus, an estimate of the reduced form impulse response sequence $(\hat{\Phi}_j)_{j\geq 0}$ follows immediately. The approach to obtain structural impulse responses differs across cases (1) to (3). In every case, however, the starting point is the identity $\Sigma_u = A_0^{-1}A_0^{-1\prime}$ and the available estimate $\widehat{\Sigma}_u$.

For recursive identification consider the (unique) Cholesky decomposition of $\widehat{\Sigma}_u = \widehat{L}\widehat{L}'$, with $\widehat{L}$ lower triangular (and positive elements along the diagonal), and set $\widehat{A}_0^{-1} = \widehat{L}$.

Now, observe that for any unitary matrix $Q \in \mathbb{R}^{k\times k}$, with $QQ' = Q'Q = I_k$, it holds that

$$\widehat{\Sigma}_u = \widehat{L}QQ'\widehat{L}' = \widehat{L}_Q\widehat{L}_Q' \tag{14}$$

is another valid decomposition of $\widehat{\Sigma}_u$. Thus, the whole range of structural impulse responses consistent with the reduced form error variance matrix Σ_u is given by varying $Q \in \mathbb{R}^{k\times k}$ over all unitary matrices (for given Cholesky factor L). This is clearly not feasible and thus approximate solutions are required. Here we follow the approach of Rubio-Ramírez, Waggoner, and Zha (2010) to generate uniformly distributed Q-matrices:

1. Draw a matrix M with i.i.d. standard normal entries and perform the QR-decomposition of the matrix $M = QR$. Doing so, Q is unitary and has the uniform (or Haar) distribution.

2. Calculate the corresponding structural impulse response function $\{\widehat{\Theta}_j^Q\}_{j=0,\ldots,J} = \{\widehat{\Phi}_j\widehat{L}_Q\}_{j=0,\ldots,J}$ and verify whether the formulated sign restrictions are fulfilled. If so, keep $\left\{\widehat{\Theta}_j^Q\right\}_{j=0,\ldots,J}$, otherwise discard it.

3. Repeat these calculations until the set of retained structural impulse responses contains $n = 5000$ elements.

From the 5000 elements the median target impulse response function is then calculated as described below.

It remains to discuss the implementation of the combination of zero and sign restrictions, where we follow Arias, Rubio-Ramírez, and Waggoner (2014). As can be guessed by now, the solution consists of drawing random unitary matrices that imply that the resultant $\widehat{\Theta}_0^Q$ satisfies the required zero-at-impact restrictions in addition to the formulated sign restrictions. We describe the approach here only for our specific application, in which TFP is ordered last in the VAR model where we combine zero and sign restrictions:

4 Note that imposing *too many* sign restrictions can reduce the set of feasible impulse responses to the empty set. The same is, a fortiori, true for the combination of zero and sign restrictions.

1. Find a matrix $N_1 \in \mathbb{R}^{k\times(k-1)}$ with $N_1'N_1 = I_{k-1}$ such that $\widehat{L}_{[k,\bullet]}N_1 = 0$, with $\widehat{L}_{[k,\bullet]}$ denoting the k-th row of $\widehat{L}$.
2. Generate a vector $z \in \mathbb{R}^k$ with i.i.d. standard normally distributed entries and form the vector:
$$q = \frac{1}{||[N_1\ 0_{k\times 1}]z||}[N_1\ 0_{k\times 1}]z, \tag{15}$$
i.e., project the vector z on the space spanned by N_1 and normalize it to unit length.
3. Find a matrix $N_2 \in \mathbb{R}^{k\times(k-1)}$ with $N_2'N_2 = I_{k-1}$ such that $q'N_2 = 0$.
4. Draw a matrix $M \in \mathbb{R}^{(k-1)\times(k-1)}$ with i.i.d. standard normal entries and calculate the QR decomposition of N_2M, i.e.,
$$N_2M = [\tilde{Q}_1\ \ \tilde{Q}_2]\begin{bmatrix} R_1 \\ 0 \end{bmatrix}, \tag{16}$$
with $\tilde{Q}_1 \in \mathbb{R}^{k\times(k-1)}$.
5. Form the matrix $Q^+ = [q\ \tilde{Q}_1]$ and calculate the corresponding structural impulse response function $\{\widehat{\Theta}_j^{Q^+}\}_{j=0,\dots,J} = \{\widehat{\Phi}_j\widehat{L}_{Q^+}\}_{j=0,\dots,J}$, with $\widehat{L}_{Q^+} = \widehat{L}Q^+$, and verify whether the formulated sign restrictions are fulfilled. If so, keep $\left\{\widehat{\Theta}_j^{Q^+}\right\}_{j=0,\dots,J}$, otherwise discard it. Note that by construction, the zero-at-impact restriction on the structural impulse response of $\log TFP$ holds for all draws.
6. Repeat these calculations until the set of retained structural impulse responses contains $n = 5000$ elements.

In the discussion of results with sign restrictions we focus on the median target (MT) impulse response functions, compare Fry and Pagan (2011). The MT impulse response function is the element-wise closest impulse response function—out of the retained 5000 impulse responses—to the median curve, which itself is not an impulse response function corresponding to any of the structural models. Thus, we consider the set of structural impulse responses $\widehat{\Theta}^n = \{\widehat{\Theta}_j^n\}_{j=0,\dots,J}$ for $n = 1,\dots,5000$ and denote the (*element-wise*) median curve as $\widehat{\Theta}_{med} = \{\widehat{\Theta}_{j,med}\}_{j=0,\dots,J}$. The median target impulse response is defined as:
$$\widehat{\Theta}^{MT} = \text{argmin}_{n=1,\dots,5000}\sum_{r\in\mathcal{R}}\sum_{s\in\mathcal{S}}\frac{1}{\widehat{V}_{r,s}}\sum_{j\in\mathcal{J}}\left(\widehat{\Theta}_{j,n}^{(r,s)} - \widehat{\Theta}_{j,med}^{(r,s)}\right)^2, \tag{17}$$
with $\mathcal{R},\mathcal{S} \subseteq \{1,\dots,k\}$ and $\mathcal{J} \subseteq \{0,\dots,J\}$. $\widehat{V}_{r,s}$ is a measure of variability of the set of sign-restricted impulse responses for variable r and shock s. Starting with $\widehat{\text{Var}}(\widehat{\Theta}_{j,n}^{(r,s)}) = \frac{1}{5000}\sum_{n=1}^{5000}(\widehat{\Theta}_{j,n}^{(r,s)} - \overline{\widehat{\Theta}_{j,n}^{(r,s)}})^2$, with $\overline{\widehat{\Theta}_{j,n}^{(r,s)}} = \frac{1}{5000}\sum_{n=1}^{5000}\widehat{\Theta}_{j,n}^{(r,s)}$, we use two variability measures $\widehat{V}_{r,s}$:
$$\widehat{V}_{r,s}^{max} = \max_{j\in\mathcal{J}}\widehat{\text{Var}}(\widehat{\Theta}_{j,n}^{(r,s)}) \tag{18}$$
$$\widehat{V}_{r,s}^{avg} = \frac{1}{|\mathcal{J}|}\sum_{j\in\mathcal{J}}\widehat{\text{Var}}(\widehat{\Theta}_{j,n}^{(r,s)}), \tag{19}$$

with $|\mathcal{J}|$ denoting the cardinality of $\mathcal{J}$. In our application the results do not differ markedly when using either the maximum or the average variation measure. The results in the paper are based on the average measure $\widehat{V}_{r,s}^{avg}$.

Note that the general formulation above, with the index sets $\mathcal{R}, \mathcal{S}$ and $\mathcal{J}$ allows to calculate the distances for any combination of variables, shocks and horizons deemed important for the econometric analysis at hand.[5] In relation to our application we consider only on the impulse responses to the government spending shock, i.e., $\mathcal{S} = \{1\}$ and the restricted impulse responses. Thus, for the three specifications considered, we have $\mathcal{R} = \{1, 2, 5, 7\}$ (baseline specification), $\{1, 2, 5, 7, 9\}$ (utilization rate augmented specification) or $\{1, 2, 5, 7, 9, 10\}$ (utilization rate and TFP augmented specification). The horizons considered are $\mathcal{J} = \{0, 1, 2, 3\}$, with the results robust to choosing only one or two quarters.

Inference on impulse response functions

The confidence bands for the recursive identification scheme are obtained using the bootstrap algorithm proposed in Kilian (1998), which is based on a preliminary (simulation based) bias correction step. The 5000 bootstrap samples are then drawn using bias corrected parameter estimates.

Some more care has to be taken into account when bootstrapping the median target solution. The median target structural impulse response function by construction depends upon $\widehat{B}_1, \ldots, \widehat{B}_p$ as well as $\widehat{L}_{Q^{MT}} = \widehat{L}Q^{MT}$, with Q^{MT} denoting the rotation matrix corresponding to the minimizer of (17). Thus, resampling data from the reduced form model has to be combined with the structural decomposition given by $\widehat{L}_{Q^{MT}}$, which is done by a modification of the previous algorithm:

1. As in the standard case, generate a bootstrap sample, $x_1^*, \ldots, x_T^*$ using the Kilian (1998) bootstrap, i.e., bias corrected parameter estimates.

2. Estimate the parameters of the VAR model using x_t^*, resulting in parameter estimates $\widehat{B}_1^*, \ldots, \widehat{B}_p^*$. Calculate the structural impulse response function using these parameter estimates and the *original* $\widehat{L}_{Q^{MT}}$.

3. Verify whether the impulse response function from the previous item, $\{\widehat{\Theta}_j^{Q^{MT}*}\}_{j=0,\ldots,J}$, satisfies the formulated sign restrictions. If it does, keep it, otherwise discard it.

4. Repeat the above steps until 1000 impulse responses are retained and calculate pointwise bootstrap confidence bands as usual from these 1000 impulse responses.

2.B FURTHER RESULTS

In Sections 3.3–3.4 in the main text we present the median target impulse responses as summaries or typical representatives of the set of sign-restricted impulse responses. In the following we augment this information by additionally plotting the element-wise median curve as well as the element-wise 10-th and 90-th quantile curves. As already mentioned,

5 Clearly, the difference can also be calculated with any other quantile or the mean as target.

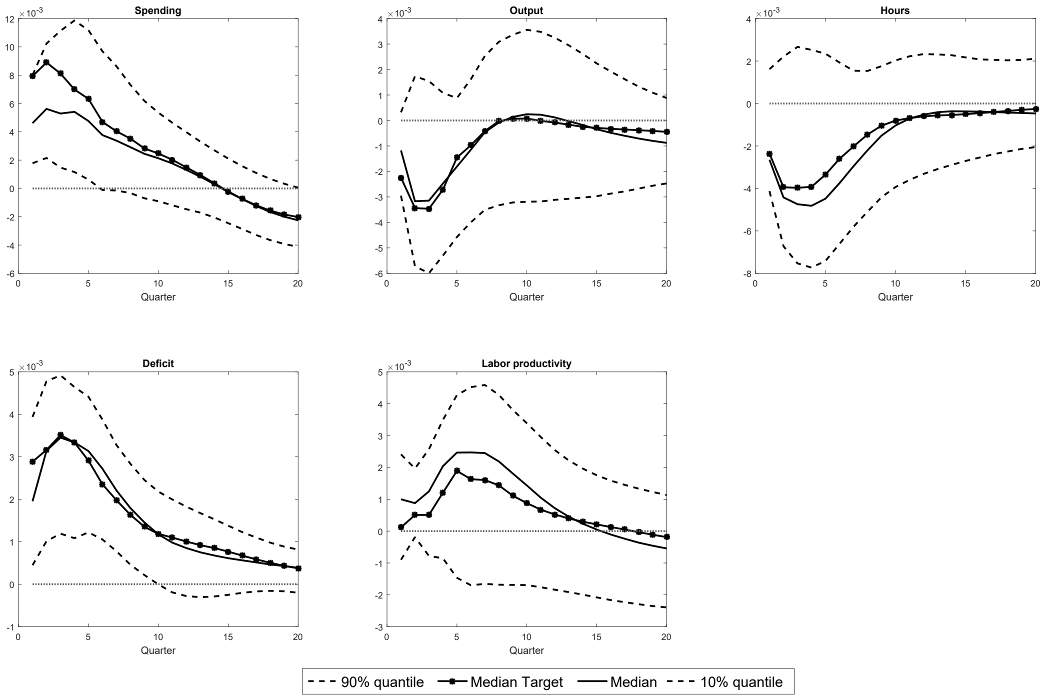


Figure 6: Set of sign-restricted impulse responses to government spending shock identified through restrictions (i) – (iii).

the median curve, and similarly the quantile curves, do not themselves correspond to structural models. This is, since for any variable, shock and horizon the median or quantile may—and in general will—correspond to a different structural model.

The structure of the following plots is the same as that of figures 3 – 5 in the main text. Figure 6 corresponds to the specification shown in figure 3, figure 7 to the specification of figure 4 and figure 8 to the specification of figure 5. The figures show that the feasible set of sign-restricted impulse responses includes elements with both positive and negative responses of output, hours and labor productivity to a government spending shock. The figures also show, however, that the main tendency points towards the direction discussed in the main text and represented by the median target impulse responses. There is only one small noticeable difference: the median target response of the utilization rate shown in figure 5 shows a positive albeit not significant value in the first period, whereas the median curve (shown in figure 8) starts off negatively.

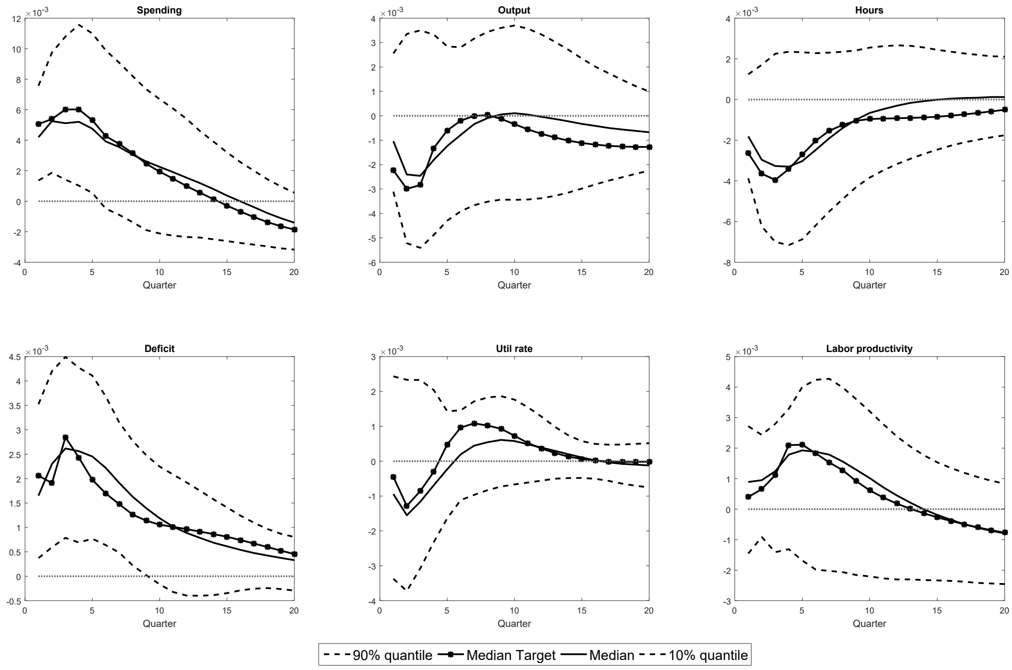


Figure 7: Set of sign-restricted impulse responses to government spending shock identified through restrictions (*i*), (*iii*) and (*iv*).

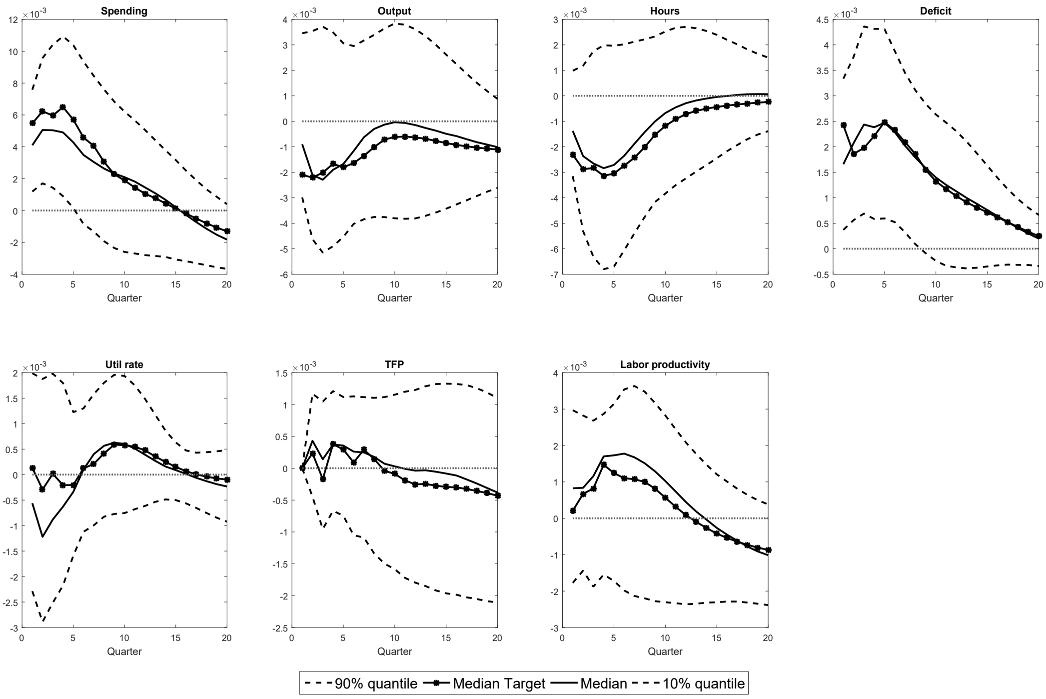


Figure 8: Set of point- and sign-restricted impulse responses to government spending shock identified through restrictions (i), (iii), (iv) and (v).

3 MONETARY POLICY SHOCKS AND THE EFFECTS OF THE INFORMATION SET

Gábor B. Uhrin and Martin Wagner, with contributions by Uroš Herman

Abstract. Including more (forward-looking) information in monetary policy SVARs is generally considered to be a good practice. In this paper we revisit the classical monetary policy SVAR of Bernanke and Mihov (1998) for monthly US data from 1989:01 to 2007:11. We augment the baseline VAR with more information in two ways. First, we add the federal funds futures series to the specification, second, we estimate a factor-augmented VAR (FAVAR) similar to Bernanke, Boivin, and Eliasz (2005). We first argue that the federal funds futures series is a reasonable variable to include in the present empirical specification. We explicitly test the fundamentalness of the structural shocks, and establish that they are fundamental even with respect to the variables in the baseline VAR. We compare the estimated monetary policy shocks to two benchmark monetary policy shock series, and conclude that information-augmentation does not necessarily lead to monetary policy shocks more highly correlated with the benchmark measures. By means of impulse response analysis and counterfactual analysis we establish that monetary policy shocks do not contribute much to the evolution of other (real) variables. We also find, however, that information-augmentation can help mitigating the price puzzle.

3.1 INTRODUCTION

One of the central questions of the empirical monetary policy literature is how to identify and estimate exogenous monetary policy shocks. Since Sims (1980), a large empirical literature has been using structural vector autoregressions (SVARs) to this end. Monetary policy SVARs, however, are generally prone to the problem of foresight, that may render the empirically estimated monetary policy shocks not exogenous. If the forward-looking information incorporating agents' foresight is missing from the empirical model, then the conclusions based on the SVAR estimates are unreliable. In particular, missing (forward-looking) information may imply the non-fundamentalness of the structural innovations. While foresight problems have been recognized early in the literature (Sims, 1992; Rudebusch, 1998), there is a recent surge of contributions that suggest ways to test and correct for non-fundamentalness in empirical models (Chen, Choi, and Escanciano, 2017; Forni and Gambetti, 2014; Giannone and Reichlin, 2006).

In this paper we revisit the classical monetary policy SVAR model of Bernanke and Mihov (1998) for monthly US data from 1989 to 2007. We explicitly test whether the estimated monetary policy VAR model is non-fundamental. Then we evaluate the adequateness of two conceivable answers to problems caused by insufficient information. First, we augment the baseline specification with a variable that is argued to capture market information on policy expectations: the federal funds futures. This series is available since 1989. Second, we estimate a factor augmented vector autoregression (FAVAR), as in Bernanke, Boivin, and Eliasz (2005). As proponents of FAVAR and dynamic factor models argue, first, the factors that are estimated from 113 time series are likely to capture and condense information that is not necessarily contained in few-variable VARs, and, second, factor models are generically expected to be fundamental (Anderson and Deistler, 2008; Alessi, Barigozzi, and Capasso, 2011).

Our second motivation to include forward looking information stems from the following consideration: the estimated monetary policy shocks will, by construction, be orthogonal to the expectations contained in these informational variables. Thus, we can a priori expect the monetary policy shocks to be closer to being "pure", exogenous monetary policy shocks.

Our empirical results suggest that non-fundamentalness does, in fact, *not* appear in our baseline empirical specification. While federal funds futures augmentation and factor augmentation seem a priori sensible, we conclude that the resulting monetary policy shocks are not necessarily "better" than those obtained from the baseline classical monetary policy SVAR. In particular, the estimated monetary policy shocks' correlation with the (extended) Romer and Romer (2004) measure is lower for the augmented specifications than for the baseline specification. Impulse response analysis and counterfactual simulations suggest that empirical conclusions based on the classical monetary policy SVAR are very similar to those based on the information augmented specifications.

The paper proceeds as follows: In Section 3.2 we motivate the need to address non-fundamentalness in the monetary policy modelling context, and we detail possible answers to the problem. In Section 3.3 we describe the econometric models, structural identifying assumptions, and the idea behind the monetary policy benchmark measures of Kuttner (2001) and Romer and Romer (2004). Section 3.4 contains our empirical results including testing for non-fundamentalness, comparing estimated monetary policy shocks to the benchmark measures, impulse response analysis, and counterfactual simulations. Finally, Section 3.5 concludes.

3.2 INFORMATION AND FUNDAMENTALNESS

Consider the following K-dimensional infinite order moving average (MA(∞)) process

$$y_t = \sum_{j=0}^{\infty} \Theta_j \varepsilon_t, \qquad t \in \mathbb{Z} \tag{20}$$

where $y_t \in \mathbb{R}^K$, $\{\varepsilon_t\} \sim WN(0, I_K)$, is a sequence of white noise *structural innovations*; $\Theta_j \in \mathbb{R}^{K \times K}$ are the structural impulse response matrices. Throughout this paper we assume that the relationship between y_t and ε_t given in Equation (20) indeed exists. That is, we assume that $\{\varepsilon_t\}$ is a causal function of $\{y_t\}$, or, in short, ε_t is y_t-causal.[1] In a traditional

1 For this notion of causality, as a relation between two processes $\{\varepsilon_t\}$ and $\{y_t\}$, see further Brockwell and Davis (1991, p. 418).

structural vector autoregressive (SVAR) analysis one is interested in recovering $\{\varepsilon_t\}$ and $\{\Theta_j\}$. This is typically achieved by fitting a p-order vector autoregression on y_t, recovering the *reduced form residuals* $\{\hat{u}_t\}$, their variance-covariance matrix $\widehat{\Sigma}_u$, and the corresponding reduced form impulse response matrices $\{\widehat{\Phi}_j\}$. The sequence $\{\hat{\varepsilon}_t\}$ is then estimated as, e.g., $\hat{\varepsilon}_t = \widehat{A}_0 \hat{u}_t$, where the estimate of the *structural matrix*, $\widehat{A}_0$, is a rotation of the lower triangular Cholesky decomposition $\widehat{L}$ of $\widehat{\Sigma}_u$. That is, $\widehat{A}_0 = (\widehat{L}Q)^{-1}$ for some Q such that $Q'Q = QQ' = I$. If the assumptions on A_0 correspond to some economic intuition, then the *structural shocks* are consistently estimated with this procedure.

For any stochastic process $\{x_t\}_{t\in\mathbb{Z}} \in L^2(\Omega, \mathcal{F}, P)$ for some given probability space $(\Omega, \mathcal{F}, P)$, consider $\mathcal{H}_t^x := \text{span}\{x_{t-k} : k \geq 0\}$, the closed linear span generated by x_t and its past values. There are four such sets of interest to us. The sets $\mathcal{H}_t^y$ and $\mathcal{H}_t^\varepsilon$, and the sets $\mathcal{H}_t^u$ and $\mathcal{H}_t^{u^{(y)}}$. Here the series $\{u_t^{(y)}\}$ is the Wold innovation series corresponding to a VAR(p) model imposed on y_t.[2] If $u_t^{(y)}$ is y_t-causal, then by construction $\mathcal{H}_t^{u^{(y)}} = \mathcal{H}_t^y \; \forall t$. Further, the equality $\mathcal{H}_t^\varepsilon = \mathcal{H}_t^u \; \forall t$ holds because u_t is a linear combination of the elements in ε_t. The discussion in the preceding paragraph implicitly assumed a more demanding equality, viz., $\mathcal{H}_t^{u^{(y)}} = \mathcal{H}_t^u \; \forall t$. The y_t-causal process $\{\varepsilon_t\}$ is called *y_t-fundamental* if $\mathcal{H}_t^y = \mathcal{H}_t^\varepsilon$ for all t.[3] That is, the information contained in y_t and its past values corresponds to the information contained in ε_t and its past values. If $\mathcal{H}_t^{u^{(y)}} = \mathcal{H}_t^u \; \forall t$ holds, then ε_t is y_t-fundamental, and fitting a VAR recovers the correct structural innovations—at least up to an orthogonal rotation.

Fundamentalness is very closely related to invertibility.[4] The process $\{\varepsilon_t\}$ is an *invertible function* of the process $\{y_t\}$ if there is a sequence of matrices $\{\Pi_j\}$ with absolutely summable components such that[5]

$$\varepsilon_t = \sum_{j=0}^{\infty} \Pi_j y_{t-j} \,. \tag{21}$$

If $\{\varepsilon_t\}$ is an invertible function of $\{y_t\}$, then $\{\varepsilon_t\}$ is y_t-fundamental.[6] The reverse statement is generally not true, and a fundamental process need not be invertible. However, this happens only in borderline cases, c.f., the discussion in Alessi, Barigozzi, and Capasso (2011). This case is detectable (Tsay, 1993), and it can be argued, that it is empirically rare (Watson, 1986). In light of this, we treat fundamentalness and invertibility interchangably

2 Note, that the true data generating process may not be a finite order vector autoregression. In this case fitting a finite order VAR will only yield an approximation, see the discussion in Lütkepohl (2005, Chapter 15).

3 Causality implies $H_t^y \subseteq H_t^\varepsilon$. Fundamentalness alone means the reverse: $H_t^y \supseteq H_t^\varepsilon$. Both conditions imply the equality between the two sets.

4 We note here that there is a "tower of Babel" situation with respect to terminology concerning invertibility and fundamentalness. First, our definition of fundamentalness originates, to our knowledge, in Rozanov (1967, pp. 56-57). Second, Brockwell and Davis (1991) views fundamentalness as "extended invertibility", and does not give weight to explicitly distinguishing between invertibility and fundamentalness. Thus, fundamentalness as a word does not appear in Brockwell and Davis (1991). Third, the miniphase assumption of Hannan and Deistler (1988) is equivalent to fundemantalness. Invertibility is, in turn, equivalent to the strict miniphase assumption. To complicate matters further, the minimum phase of Rosenblatt (2000) is, in fact, invertibility. In the recent econometrics literature the term "fundamentalness" is what appears most frequently, hence we settle on this term.

5 See Brockwell and Davis (1991, p. 418).

6 This is an immediate consequence of Theorem 3.1.2 and Proposition 4.4.1 on pages 86 and 127, respectively, in Brockwell and Davis (1991). The multivariate analogue follows from the respective multivariate statements.

in this paper. We note also that our empirical results below support the hypothesis of invertibility, hence, fundamentalness. Thus, the discussion of the borderline case where the two concepts are not equivalent is not necessary.

Non-fundamentalness can arise for several, related reasons that are equally prevalent in econometric research. First, it can be the result of an omitted variables problem (Lütkepohl, 2012; Kilian and Lütkepohl, to appear; Sims, 2012), where key variables are missing from the estimated VAR. Second, the underlying data generating process, motivated by, e.g., an economic model might analytically imply non-fundamental innovations. A particular example of such a model is the growth model with fiscal foresight in Leeper, Walker, and Yang (2013), where economic agents have (unbiased) expectations about future tax rates at the time t of their decision making. Since the tax rate change is implemented at time $t' > t$, the agents—having *foreseen* the policy change—base their decisions between t and t' already upon the new tax rate arriving at t'. The recent Handbook of Macroeconomics chapter of Ramey (2016) devotes a separate section to the problem of foresight, which can imply non-fundamentalness. There are two sides to this problem. On the one hand, as discussed above, economic agents might foresee policy actions. On the other hand, the policy makers might observe more information on the state of the economy than economic actors. In both cases, the investigated structural shock might include the policy maker's endogenous response to expectations about the future of key macroeconomic variables. Besides fiscal foresight, these problems are all prevalent in macroeconometric research on monetary policy. An early critique recognizing the issue explicitly is Rudebusch (1998), but already Sims (1980, p. 7) suggests that expectations regarding policy actions may be crucial to the question of identification. Furthermore, the motivating examples in the seminal Bernanke, Boivin, and Eliasz (2005) article proposing the use of factor augmented VARs (FAVARs) seem to implicitly refer to the problem of foresight and non-fundamentalness. Ramey (2016, Section 3.3.2) contains an up-to-date and accessible overview about the problem of foresight in empirical monetary policy research.

Once we have established that non-fundamentalness might be present theoretically, the question is if it does indeed appear empirically for the problem at hand. Several contributions on testing non-fundamentalness appeared recently in the literature. Forni and Gambetti (2014) and Beaudry, Fève, Guay, and Portier (2015) propose extending the VAR information set with factors extracted from a large dataset and then testing whether the additional variables contribute significantly to forecasts of the original variables. A critique of this approach is given by Canova and Sahneh (2016). In our paper we, however, utilize the non-fundamentalness test of Sahneh (2015) that relies on the result, that if the y_t-causal innovations $\{\varepsilon_t\}$ are serially independent and non-Gaussian, then they form a martingale difference sequence (MDS) if and only if they are an invertible function of $\{y_t\}$.[7] That is, if $\{\varepsilon_t\}$ is not a martingale difference sequence, then it is not an invertible function of $\{y_t\}$. The key insight is, that under the Gaussian case, fundamental and non-fundamental representations are observationally equivalent. Under non-Gaussianity, however, fundamental and non-fundamental representations generally imply different moment structures.

7 This statement can also be found in the similar contribution of Chen, Choi, and Escanciano (2017), who, however, assume that the innovations are iid. This is more restrictive than the approach of Sahneh (2015). Both contributions ultimately stem from the results in Rosenblatt (2000, Section 5.4).

As Chen, Choi, and Escanciano (2017) argue, if the residuals are indeed not Gaussian, then this approach is more fruitful, since the residual-based tests for non-fundamentalness have power against alternatives that are not contained in the information set spanned by the larger dataset available to the researcher. The point that the researcher has the sufficient amount of data available is the implicit assumption made in the approach of, e.g., Forni and Gambetti (2014).

An exogenous policy shock has to be fundamental with respect to the information contained in the empirical model with variables y_t in order to reach sensible conlcusions based on such a model. For a y_t-causal ε_t non-fundamentalness means that $H_t^y \subsetneq H_t^\varepsilon$. That is, the information contained in y_t and its past values is less than what ε_t and its past values contain. Sims (2012) argues that (non-)fundamentalness is not an "either/or problem" by comparing, in a Monte Carlo experiment based on a DSGE model, the impulse responses to non-fundamental shocks with those of the impulse responses to fundamental shocks. According to his results, the biases stemming from non-fundamentalness may be moderate. The more comprehensive study of Leeper, Walker, and Yang (2013), on the other hand, discovers more serious distortions in the empirical analysis if non-fundamentalness is not taken into account. We side with the latter conclusions, adding that we believe that the underlying data generating process is, in principle, unknowable. Thus, quantifying such a "bias" or "distortion" emprically is, at best, a challenging problem—if not impossible. Therefore, in our view it is better to remain on the "safe side" and ask for the structural shocks to be fundamental with respect to the variables contained in the empirical model.

In order to overcome the problem of non-fundamentalness, Giannone and Reichlin (2006) propose to simply extend the information set of the (S)VAR.[8] This can be achieved either by simply augmenting the VAR with additional variables that contribute significantly to the forecasting ability of the VAR system, or by including principal components obtained from a large dataset. The former approach has been taken by, e.g., Ramey (2011) in the context of fiscal policy. In the context of monetary policy, the latter approach also resonates in the motivation of Bernanke, Boivin, and Eliasz (2005), who propose factor augmented vector autoregressions as useful empirical tools for monetary policy. The factor-augmented approach is theoretically strengthened by the finding that dynamic factor models are generically fundamental (Alessi, Barigozzi, and Capasso, 2011; Forni, Giannone, Lippi, and Reichlin, 2009). Or, put differently, a tall transfer function—typical of dynamic factor models—implies a pure AR representation, thus, by construction, fundamentalness (Anderson and

8 Other authors consider turning the non-fundamental representation into a fundamental one by using so-called Blaschke matrices. As Lütkepohl (2012) and Kilian and Lütkepohl (to appear) convincingly argue, however, it is more fruitful to view the non-fundamentalness problem as an omitted variable problem and try to ameliorate it as such. This is the avenue that we pursue in this paper.

Deistler, 2008).[9] That is, in a factor-augmented context, non-fundamentalness is an unlikely problem.[10]

In this paper we explore the implications of both directions to correct for the problem of non-fundamentalness. First, we augment the VAR specification with an arguably forward-looking variable that contains (market) expectations regarding policy, the federal funds futures. As it turns out, the federal funds futures Granger-causes two out of six variables in our VAR (industrial production and the federal funds rate) at a significance level of 0.01. We also test, whether the federal funds rate Granger causes the rest of the variables in the VAR jointly, employing the multivariate out-of-sample Granger causality test of Gelper and Croux (2007). Further, we contrast the results obtained from this augmentation to the benchmark non-augmented model and a FAVAR model similar to Bernanke, Boivin, and Eliasz (2005).

3.3 MODELS OF MONETARY POLICY AND IDENTIFYING ASSUMPTIONS

The empirical monetary policy literature that utilizes SVARs has employed a variety of specifications and identifying assumptions, see the comprehensive surveys of Christiano, Eichenbaum, and Evans (1999) and Ramey (2016). In the present paper we use the specification of the classical paper of Bernanke and Mihov (1998), since their aim in the structural analysis is to estimate a measure of the stance of monetary policy. That is, they explicitly concentrate on the monetary policy shock series in their analysis. Our approach is closest in spirit to this classical approach. Thus, we consider a standard monetary policy VAR of order p with the following variables: output, prices, commodity prices, and a set of monetary policy instruments: the federal funds rate (ffr), non-borrowed reserves (nbr) and total reserves (tr). The identification schemes of Bernanke and Mihov (1998), in short BM, are based on two ideas. First, the reduced form innovations corresponding to the monetary policy instruments (in short, policy block) are jointly orthogonal to the reduced form innovations corresponding to the non-policy variables. Second, the structural identification within the policy block is based on the following set of equations:

$$u_{\mathrm{tr},t} = -\alpha u_{\mathrm{ffr},t} + \varepsilon_{dt} \tag{22}$$

$$u_{\mathrm{br},t} = \beta u_{\mathrm{ffr},t} + \varepsilon_{bt} \tag{23}$$

$$u_{\mathrm{nbr},t} = \phi_d \varepsilon_{dt} + \phi_b \varepsilon_{bt} + \varepsilon_{st}. \tag{24}$$

9 The papers Anderson and Deistler (2008, 2009) prove this statement precisely. The starting point of these papers is an $n \times q$ rational transfer function matrix $W(z)$ with minimal realization $\{A, B, C, D\}$ of dimension m. The matrices A, B, C, D are similar to those in Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson (2007). If B and C have q columns and n rows, respectively, where $n > q$, and if the entries of A, B, C, D assume generic values, then $W(z)$ has no finite or infinite zeros (Anderson and Deistler, 2008, Proposition 1, p. 285). The situation described in the previous statement corresponds to that encountered in dynamic factor models, viz., strictly more outputs than inputs. The zero-free property of $W(z)$ implies stability and (strict) miniphase as in Hannan and Deistler (1988). Stability is equivalent to our notion of causality. Then, as discussed in a previous footnote, these conditions together imply fundamentalness. Note, that a precise treatment of the terms used in the present footnote can be found in Hannan and Deistler (1988).

10 On the precise connection between dynamic factor models and FAVARs see Stock and Watson (2016).

where $\varepsilon_t^{\text{pol}} \equiv (\varepsilon_{dt}, \varepsilon_{bt}, \varepsilon_{st})'$ are the structural innovations, and $u_t^{\text{pol}} \equiv (u_{\text{tr},t}, u_{\text{nbr},t}, u_{\text{ffr},t})'$ are the reduced form innovations.[11] Of particular interest to us is $\{\varepsilon_{st}\}$, the monetary policy shock series.

We can write the above system of equations in a more compact way as:

$$A_0^{\text{pol}}(\alpha, \beta, \phi_d, \phi_b) u_t^{\text{pol}} = \varepsilon_t^{\text{pol}}, \tag{25}$$

where A_0^{pol} is a 3-by-3 matrix, whose values depend on the parameters $\alpha, \beta, \phi_d, \phi_b$.[12] Based on economic arguments, certain parametric assumptions can be put on A_0^{pol} depending on the economic model being considered. The structural models put forth in Bernanke and Mihov (1998) are contained in Table 1.

Table 1: Identifying restrictions in Bernanke and Mihov (1998).

	Structural model				
	BM-JI	BM-BR	BM-FFR	BM-NBR	BM-STR
α	0	$\times$	$\times$	$\times$	0
β	$\times$	$\times$	$\times$	$\times$	$\times$
ϕ_d	$\times$	1	1	0	$\times$
ϕ_b	$\times$	α/β	-1	0	0

Notes: The symbol $\times$ means that the parameter is left unrestricted. The mnemonics stand for: JI – just identified, BR – borrowed reserves, FFR – federal funds rate, NBR – non-borrowed reserves, STR – Strongin, as in Strongin (1995).

In addition to a standard monetary policy SVAR, we also estimate monetary policy shocks by means of a factor augmented VAR (FAVAR) model. Similar to Bernanke, Boivin, and Eliasz (2005), the model is described by the following equations:

$$X_t = \Lambda^f F_t + \Lambda^y Y_t + e_t \tag{26}$$

$$\Phi(L) \begin{bmatrix} F_t \\ Y_t \end{bmatrix} = \nu + u_t, \tag{27}$$

$F_t \in \mathbb{R}^r$, $Y_t \in \mathbb{R}^K$, $X_t \in \mathbb{R}^M$, and $\Phi(L) \equiv I - \Phi_1 L - \cdots - \Phi_p L^p$, an autoregressive lag polynomial of order p, and $e_t \sim WN(0, \Sigma_e)$ and $v_t \sim WN(0, \Sigma_u)$ are assumed to be independent of each other for all t.[13] We refer to Equation (26) as the *observation equation*, and to Equation (27) as the *state equation*. We estimate the *unobservable factors* F_t as principal components from the standardized dataset X_t that excludes the *observable factors* Y_t.

The identification scheme for recovering the structural innovations ε_t from the reduced form innovations u_t involves two considerations. First, the identification scheme is recursive, with the monetary policy instrument ordered last. Second, notice that the monetary policy instrument might be contained in the space spanned by the estimated factors $\hat{F}_t$. If

11 Note, that nbr is the difference between tr and br.

12 A_0^{pol} is the lower-right 3-by-3 submatrix of the structural matrix A_0.

13 Note, that the autoregressive lag order p need not be the same for the BM models and the FAVAR.

this is the case, the recursive identification assumption is invalid, since then the identified monetary policy shock is not guaranteed to be orthogonal to (the space spanned by) the unobserved factors. To overcome this problem, Bernanke, Boivin, and Eliasz (2005) uses transformed factors $\tilde{F}_t$ in order to remove this dependence from the space spanned by $\hat{F}_t$. Therefore, in effect, instead of (27), we estimate the VAR:[14]

$$\tilde{\Phi}(L)\begin{bmatrix}\tilde{F}_t\\ Y_t\end{bmatrix} = \tilde{\nu} + \tilde{u}_t, \tag{28}$$

where $\tilde{F}_t \in \mathbb{R}^r$ are linear transformations of $\hat{F}_t$ that are, based on economic grounds, assumed to be orthogonal to Y_t.[15]After estimating $\hat{\tilde{u}}_t$ as the ordinary least squares residuals from (28), a Cholesky decomposition of $\hat{\Sigma}_u$ yields the structural impact matrix and the structural innovations $\hat{\varepsilon}_t = \hat{A}_0\hat{\tilde{u}}_t$. The monetary policy shock is the last element in ε_t, and, as such it is also last in the recursive order imposed by a Cholesky decomposition. We assume that the VAR in (28) is stationary and causal. Then the reduced form impulse responses for the variables in (28) are calculated recursively as coefficients of the moving average representation of $(\tilde{F}_t, Y_t)'$:

$$\begin{bmatrix}\tilde{F}_t\\ Y_t\end{bmatrix} = \sum_{j=0}^{\infty}\Psi_j u_{t-j}. \tag{29}$$

One advantage of the factor augmentation is that we can calculate impulse responses to shocks $\tilde{u}_t$ of variables in X_t. This is achieved by estimating the *factor loadings* $\Lambda = (\Lambda^f, \Lambda^y)$ from (26) by OLS and calculating the impulse response coefficients as $\hat{\Lambda}\widehat{\Psi}_j$.

3.3.1 *Monetary policy shock benchmark measures*

A monetary policy shock is an unexpected, exogenous innovation to the equation guiding monetary policy, e.g., a Taylor-type rule. There are two avenues in the literature that aim to construct monetary policy shock measures not assuming a VAR system. First, the narrative-based monetary policy shock series of Romer and Romer (2004) identifies changes in the federal funds rate around the meetings of the Federal Open Market Committee (FOMC), and exogenizes these with respect to the forecasts about the real economy available to the FOMC directly prior to the meeting. In particular, the monetary policy shock series are constructed as the estimated residuals from the following regression (Romer and Romer, 2004, Eq. 1, p. 1061):

$$\begin{aligned}\Delta ff_m &= \alpha + \beta ffb_m + \sum_{i=-1}^{2}\gamma_i\widetilde{\Delta y}_{mi} + \\ &+ \sum_{i=-1}^{2}\lambda_i(\widetilde{\Delta y}_{mi} - \widetilde{\Delta y}_{m-1,i}) + \sum_{i=-1}^{2}\phi_i\tilde{\pi}_{mi} + \sum_{i=-1}^{2}\theta_i(\tilde{\pi}_{mi} - \tilde{\pi}_{m-1,i}) + \rho\tilde{u}_{m0} + \nu_m,\end{aligned} \tag{30}$$

14 As usual in the FAVAR literature, we assume in this second step that the variables $\tilde{F}_t$ are given.

15 This orthogonalization is achieved by obtaining $\tilde{F}_t$ in a way such that it is contained in the space spanned by principal components extracted from so-called "slow-moving" variables, span$\{\hat{F}_t^{(s)}\}$. These variables are assumed to not react to the policy variables contemporaneously. This assumption is justified economically in Bernanke, Boivin, and Eliasz (2005), and is almost always employed in the FAVAR literature. Note, that span$\{\hat{F}_t^{(s)}\} \subseteq$ span$\{\hat{F}_t\}$.

where Δff_m is the change in the intended federal funds rate at the FOMC meeting m, ffb_m is the intended federal funds rate before any changes decided on meeting m, and $\widetilde{\Delta y}_{mi}$, $\tilde{\pi}_{mi}$, $\tilde{u}_{m0}$ are the forecasts of real output growth, inflation and unemployment, respectively, for quarter i at the time of meeting m.[16] Arguably, this method results in a monetary policy shock series $\{\hat{v}_m\}$ that is exogenous to the FOMC's information about future developments in the economy. As there can be several meetings, or no meetings at all in any given month, $\{\hat{v}_m\}$ is converted to the monthly monetary policy shock series $\{\hat{\varepsilon}_t^{\mathrm{rr}}\}$ by averaging over the respective month. In the following we refer to this series as the RR measure.

The second type of methodology, exemplified by Kuttner (2001), investigates the changes in the spot price of the federal funds futures in a window around the federal funds rate announcement. The federal funds futures is a contract on the average federal funds rate's level at the time of the contract's expiry. The argument here is that any endogenous change in the federal funds rate should be incorporated in the market price of the federal funds futures. Any change that is above, or below this expected change is interpreted as an unexpected, exogenous change in the federal funds rate. More precisely, following the notation of Thornton (2014, Equations (3)–(6)), the expectation of the average of the effective funds rate through the current month on day d of a month with s days is

$$fff_d^0 = \frac{1}{s}\left\{ (d-1)\Big(\sum_{i=1}^{d-1} ff_i\Big) + (s-d+1)E_t\Big[\sum_{i=d}^{s} ff_i\Big] \right\}, \tag{31}$$

where fff_d^0 is the current-month federal funds futures rate, and ff_d is the effective federal funds rate. The monetary policy surprise Δff_d^u can then be calculated as

$$\Delta ff_d^u = \frac{s}{s-d}(fff_d^0 - fff_{d-1}^0). \tag{32}$$

The intuition is the following: if the market expects the federal funds rate to change on day d, but not later during the month, then Δff_d^u would be zero. If it is not, then the change is unexpected from the market's perspective. The measure can be calculated for any day, except for the first and last day of the month. Taking this into account, Kuttner (2001) proposes to calculate

$$\Delta ff_d^u = (fff_d^1 - fff_{d-1}^1), \tag{33}$$

for the last three days of the month, where fff_d^1 is the federal funds futures rate of the futures contract for the *next* month (one-month-ahead rate). For the first day of the month, fff_{d-1}^0 in (32) is replaced with the one-month-ahead rate on the *last day* of the *previous* month. Calculating Δff_d^u for all days on which a federal funds rate announcement took place, and averaging it within months yields a monthly monetary policy surprise measure $\{\hat{\varepsilon}_t^{\mathrm{u}}\}$. In the following we use the terms Kuttner measure and surprise measure to refer to this series.

For any monetary policy SVAR, we can estimate a monetary policy shock series as the identified structural form innovations $\hat{\varepsilon}_t^{\mathrm{mp}}$ that is an element of $\hat{\varepsilon}_t$ from section 3.2. The

16 As an example: Assume that the meeting takes place on the 17th of May, 2017. This is the 287-th meeting in the dataset. Then for this meeting $m = 287$. The index $i = 0$ denotes the current (2nd) quarter forecast from the Greenbook forecast prepared for meeting m. Similarly, $i = -1$ is the forecast for the first quarter, etc.

monthly measure $\hat{\varepsilon}_t^{\mathrm{rr}}$ represents monetary policy shocks that are exogenous with respect to the information of the central bank. On the other hand, $\{\hat{\varepsilon}_t^{\mathrm{u}}\}$ is a measure that is derived from the market's bet on the federal funds futures. Hence, it arguably contains the economic actors best information on the (future) stance of monetary policy.[17] In light of the discussion in section 3.2, we postulate that a monetary policy shock estimated from a SVAR should, ideally, be similar to both of the benchmark monetary policy shock series above. In the context of this paper we describe similarity by the empirical correlation between the benchmark measure and the estimated monetary policy shock. Thus, in the following we evaluate the various monetary policy models normatively by examining how strongly the implied monetary policy shock series correlate to the narrative-, or market-based benchmark monetary policy shock measures.

3.4 EMPIRICAL RESULTS

All SVAR and FAVAR models are estimated with monthly US data from January 1989 to November 2007. Our data source is an updated Stock–Watson dataset based on series from the FRED database of the St. Louis Fed. During the initial drafts of this manuscript the FRED-MD dataset (McCracken and Ng, 2016) was not yet available to us, but our data set is almost identical to the FRED-MD. The federal funds futures series are obtained from Bloomberg. The starting date, January 1989, depends on the fact that January 2, 1989 was the first (week)day of trading federal funds futures for a full month. We end the sample in November 2007 for two reasons: First, after the 2008 crisis the practice of monetary policy has changed notably compared to that of the previous decades, thus we do not believe that a model estimated on pre-crisis data would be valid for post-crisis monetary policy analysis. Second, including December 2007 in the data yields reduced form parameter estimates in the VAR model with a maximal eigenvalue greater than 1 for the companion form matrix, i.e., an explosive solution. Our sample size is, thus, $T = 227$.

The BM models are estimated with the following variable ordering: industrial production (IP), consumer price index (CPI), commodity price index (COMM), total reserves (TR), non-borrowed reserves (NBR), and federal funds rate (FFR).[18] The FFR, TR, and NBR are considered in levels; the IP, CPI, and COMM are in logarithms. In the subsequent analysis, we also augment the VAR specification with the federal funds futures (FUT) variable, considered in levels, and ordered first in the VAR. Note, that the ordering of FUT *within* the non-policy block (see section 3.3) is not consequential to the identification. We use $p = 12$ lags in the VAR specification due to the fact that we have monthly data available, and this is consistent with the empirical monetary policy literature using monthly US data (cf., Uhlig (2005); Bernanke and Mihov (1998)). We include a constant in our specification. We estimate the reduced form VAR parameters and innovations by ordinary least squares. The structural parameters are estimated by generalized method of moments, as described in de-

17 As Piazzesi and Swanson (2008) point out, the Kuttner measure is also robust to time-varying risk premia that could potentially invalidate the expectations hypothesis for the federal funds futures series. This is due to the short-window transformation (Δ) in the Kuttner measure.

18 Bernanke and Mihov (1998) uses an approximated monthly GDP and GDP deflator series instead of IP and CPI, respectively. We have also employed this specification, and the results did not change markedly. We decided to retain the IP and CPI variables since they are available to us as monthly data by default.

tail in Bernanke and Mihov (1998). We obtain confidence bands for the impulse responses using the bias-corrected bootstrap of Kilian (1998).

For estimating the FAVAR model one has to decide on the set of observable factors, Y_t. As a baseline specification, following Bernanke, Boivin, and Eliasz (2005), we include only the federal funds rate in Y_t. Similar to Bernanke, Boivin, and Eliasz (2005), two other specifications for Y_t are as follows. First, non-borrowed reserves, total reserves, and federal funds rate. Second, industrial production, consumer price index, and federal funds rate. Note, that in all specifications, the federal funds rate is ordered last in Y_t. We estimate the unobserved factors in the observation equation (Eq. (26)) as principal components from X_t *not* containing the observable factors. We extract the first $r = 3$ principal components. The lag order for the state equation (Eq. (27)) is $p = 13$, in accordance with Bernanke, Boivin, and Eliasz (2005). We estimate the state equation by ordinary least squares, including a constant. The confidence bands for the impulse responses are obtained following the procedure in Bernanke, Boivin, and Eliasz (2005). It is ultimately based on the Kilian (1998) bootstrap.

In order to estimate the principal components, a stationary panel X_t is necessary. Thus, we apply the stationarity corrections reported in Bernanke, Boivin, and Eliasz (2005) to the 113 macroeconomic time series in our dataset.[19] In particular, IP, CPI, TR and NBR are in differenced logarithms in the panel X_t, whereas the FFR is in levels. The fact, that the variables in the BM specifications are transformed differently than in the FAVAR specifications might raise the suspicion, that the estimation results are not directly comparable. We note the following points regarding this issue. First, the monetary policy shock is ultimately an innovation to the FFR equation in either the BM-VAR model, or in the state equation of the FAVAR. The FFR is in levels, the economic model described by Equations (22) – (24) is specified in levels, and the policy block variables in the BM-VAR are, accordingly, in levels. Thus, the monetary policy shock series is an innovation series "in levels". Second, there is no loss of information through the stationarity transformations in X_t, as there are also several variables that are left untransformed. Third, the shocks estimated in the FAVAR specification and in the BM specifications are monetary policy shock candidates in their own right. Our focus in this paper is the information content of the "vehicle" used to estimate these monetary policy shocks. In light of these points, we can see that there is no impediment to comparing the monetary policy shocks from the FAVAR specification with those obtained from the BM specifications.[20]

3.4.1 *Additional information and fundamentalness*

The federal funds futures arguably contains forward-looking market information about monetary policy actions, and federal funds futures outperform several other variables in predicting changes in the federal funds rate up until six months in the future (Gürkaynak,

19 We also conduct stationarity tests and establish that the transformations indeed result in stationary variables.

20 As a robustness check, we have also estimated the BM models with the same transformations as used for the FAVAR models. Since the statistical results and conclusions in the following sections did not change markedly, we decided to remain within the spirit of Bernanke and Mihov (1998).

Sack, and Swanson, 2007).[21] Thus, a priori, it is a useful variable to include in the empirical specification for two reasons. First, it can correct for possible non-fundamentalness in the empirical model. Second, in any case, including the federal funds futures renders the estimated monetary policy shock orthogonal to the information that is contained in this variable. Thus, we can a priori expect our monetary policy shock estimates to be closer to the "pure", unexpected monetary policy shocks.

Following the recommendations of Giannone and Reichlin (2006) and Forni and Gambetti (2014), we test empirically the following hypotheses. First, whether the federal funds futures Granger-causes any of the variables in the non-augmented VAR. Second, whether the federal funds futures has forecasting power for the non-augmented VAR as a whole, controlling for past information in the non-augmented VAR (multivariate Granger causality). The former, textbook Granger causality test (see, e.g., Lütkepohl, 2005, pp. 44–45) rejects the null hypothesis that the futures series does not Granger cause the federal funds rate and the industrial production at 1% significance level. For the rest of the variables the causality test does not reject the null hypothesis of no Granger causality at 1% significance level. The lag length for the test statistic was selected based on information criteria in each of the preceding cases. For testing multivariate Granger causality, we implemented the multivariate out-of-sample Granger causality test of Gelper and Croux (2007).[22] Based on 5000 bootstrap replications, the test's p-value is approximately zero for lag orders smaller than or equal to the lag order in the BM VAR specification, i.e., $1, \dots, 12$. This indicates the rejection of the null hypothesis of no Granger causality: the federal funds futures can significantly contribute to joint forecasts of the variables in the baseline VAR. Both of these results empirically support our claim that the federal funds futures can be a useful source of information mitigating the problem of non-fundamentalness—should the problem be present.

As discussed in the previous section, non-fundamentalness can only be detected from the error terms u_t if their joint distribution is not Gaussian, see also Chen, Choi, and Escanciano (2017). We apply the Lomnicki–Jarque–Bera (LJB, Lütkepohl, 2005, p. 175) test for multivariate normality on the reduced form residuals of the baseline, and the fed funds futures augmented specification, respectively. The p-value is numerically zero in both cases and, thus, the null hypothesis of joint normality is rejected at all significance levels. This result is robust to all reasonable lag length specifications.

Besides non-Gaussianity, we have to test for serial independence of the reduced form innovations u_t in order to apply the invertibility test of Sahneh (2015). In Table 2 we report the p-values of the multivariate serial independece test of Kojadinovic and Yan (2011), implemented in the R-package `copula`.[23] The null hypothesis is serial independence, and there are two types of test statistics. CvM is a copula-based Cramér–von Mises-type test statistic, similar to Genest and Rémillard (2004), and Möbius (M) is based on the Möbius decomposition thereof. The maximal lag order l is specified so that the potential dependence

21 Gürkaynak, Sack, and Swanson (2007) contrast the federal funds futures' performance in measuring monetary policy expectations with term federal funds loans, term eurodollar deposits, eurodollar futures, treasury bills, and commercial paper.

22 There are three test statistics suggested in Gelper and Croux (2007). Based on the paper's recommendation, we implement the test statistic termed 'Reg', defined in the first display on page 3322 of the paper.

23 We use version 0.999-16 of the `copula` package, and our R version is 3.2.4 Revised ("Very Secure Dishes").

Table 2: Tests for serial independence of the reduced form innovations.

		p-value		
Maximal lag	VAR specification	CvM	M–Fisher	M–Tippett
2	With futures	0.9126	0.8806	0.7827
	Without futures	0.7298	0.6109	0.5649
6	With futures	0.7300	0.4481	0.2652
	Without futures	0.4271	0.3881	0.2173
12	With futures	0.1264	0.7458	0.8297
	Without futures	0.1024	0.5639	0.6778

Notes: p-values of the tests of the null hypothesis of serial independence for multivariate time series of Kojadinovic and Yan (2011). Based on 1000 bootstrap replications. CvM is the test statistic in Equation (6), on p. 352, *ibid*. The statistic behind the M(öbius) columns is $M_{A,n}$ on p. 355, *ibid*. The Fisher and Tippett methods for obtaining p-values for M are detailed on p. 357, *ibid*. Maximal lag lengths are the assumed embedding dimensions, see p. 348, *ibid*.

within the l-blocks $(u_t, \ldots, u_{t+l})$ is taken into account. There are $2^{l-1} - 1$ p-values obtained with the M method, and the authors propose to combine them in two ways: according to Fisher (1932) and according to Tippett (1931). We report the p-values obtained from both methods. We specify $l = 1, \ldots, 12$, and report the results for $l \in \{2, 6, 12\}$.[24] The p-values are obtained from 1000 bootstrap replications, and the bootstrap procedure has been shown to be consistent in Kojadinovic and Yan (2011). The results imply that we cannot reject the null hypothesis of serial independence at reasonable significance levels.

Table 3: Invertibility test for the reduced form monetary policy VAR model.

	Baseline VAR			Futures augmented VAR		
$\bar{h}$	Parzen	Daniell	Quadratic	Parzen	Daniell	Quadratic
2	0.7593	0.7337	0.7315	0.6269	0.5586	0.5910
5	0.8198	0.7977	0.7956	0.6793	0.6741	0.6707
10	0.8625	0.8385	0.8408	0.6920	0.7003	0.6940

Notes: p-values of Sahneh's (2015) test for the null hypothesis of invertibility. There are three kernels (Parzen, Daniell, and Quadratic), and $\bar{h}$ are the initial bandwidths needed for the computation of the test statistic.

In light of these test results, table 3 contains the p-values of Sahneh's (2015) test for the null hypothesis of invertibility. The test statistic is based on the generalized spectral tests in Hong and Lee (2005). The computation of the test requires that one specifies a kernel and a bandwidth. For the calculation of a data-driven bandwidth it is necessary that we specify an initial bandwidth. We report our results for the Parzen, Daniell, and quadratic kernels and several initial bandwidths. The results show that the null hypothesis of u_t

24 Other specifications of l do not change the statistical conclusions. Estimation results are available upon request.

being a martingale difference sequence cannot be rejected for either model specifications for significance levels smaller than 55.86%. As a consequence, the investigated monetary policy VAR specifications, in fact, imply shocks that are fundamental regardless of whether we include forward-looking information or not.

One motivation for the inclusion of the federal funds futures in the VAR by its forward-looking nature, capturing market expectations. However, the results from this section imply that including more (forward-looking) information in the VAR analysis is not necessary from the fundamentalness point of view. Our second motivation is that by the inclusion of a more forward-looking variable, we can get "better" monetary policy shock estimates that are closer to being pure, exogenous monetary policy shock. Thus, the question whether we *gain* any additional empirical insight through federal funds futures augmentation in the empirical analysis remains. We explore this issue in the following sections.

It is important to note, that the test of non-fundamentalness is applied to the *reduced form* residuals; the structural identification is, at this step, irrelevant. Thus, according to our results, any VAR model fitted to the same data with the same specifications is fundamental, regardless of the structural identification scheme. Since our data and specifications are standard in the monetary policy literature, we have reason to believe that our specific result about fundamentalness might carry over to several other contributions in the literature. That is, we have reason to believe that the literature on monetary policy SVARs by and large investigates shocks that are indeed fundamental.

3.4.2 *Investigating the monetary policy shocks*

While SVAR analysis usually focuses on impulse response analysis and forecast error variance decompositions, it is also important to investigate the identified monetary policy shock series $\{\hat{\varepsilon}_t^{\mathrm{m}p}\}$ alone. A correctly estimated and identified shock series should, ideally, resemble established empirical facts about monetary policy shocks. This principle could aid both identification and analysis, and could provide a useful "backtesting" of the estimation of the monetary policy shock series.[25]

In the following we report the empirical correlations between the identified monetary policy shocks from our various estimated structural models and, first, the monetary policy surprise measure proposed by Kuttner (2001), updated through 2007 by Uroš Herman for the purposes of the present project, second, the monetary policy shock series of Romer and Romer (2004), updated through December 2007 by Johannes Wieland for the handbook chapter of Ramey (2016). We argued earlier, that a correctly specified monetary policy shock should be highly correlated with both of these series: the market data based monetary policy surprise series, and the narrative-based monetary policy shock series.

In Table 4 we display the pair-wise empirical correlations between the identified monetary policy shock series and the Kuttner or RR benchmark measure for each respective BM structural model. We can conclude the following insights. First, the structural model that yields the highest correlation with the benchmark shocks is the BM-FFR model. The rest of the models notably underperform the BM-FFR model in terms of correlations. This

25 On how focusing on shocks per se can aid identification, see the recent contributions by Ludvigson, Ma, and Ng (2017), Rubio-Ramírez and Antolín-Díaz (2016), and Uhrin and Herwartz (2016).

Table 4: Correlations between benchmark measures and identified monetary policy shocks: BM specifications.

		Structural model				
Benchmark	VAR specification	BM-JI	BM-BR	BM-FFR	BM-NBR	BM-STR
Kuttner	With futures	0.1411	0.1475	0.4529	-0.1127	0.0937
	Without futures	0.1232	0.1396	0.4119	-0.1183	0.0247
RR	With futures	0.1218	0.0482	0.2212	0.0247	0.0890
	Without futures	0.1205	0.0377	0.2974	0.0190	0.0746

finding is in line with the findings of Bernanke and Mihov (1998), who conclude that the relevant monetary policy instrument (structural model) is indeed the federal funds rate (BM-FFR). Second, augmenting the VAR with the federal funds futures typically yields monetary policy shocks that are better correlated with the benchmark measures. A notable, and puzzling exception is the FFR model, where (a) the futures augmentation increases the correlation with the Kuttner measure, but (b) the futures augmentation *decreases* the correlation with the RR monetary policy shock measure.

Let us take a closer look at these two points. Finding (a) need not be surprising. As we have seen earlier, the Kuttner measure $\{\hat{\varepsilon}_t^u\}$ is calculated directly from the federal funds futures. If the correct structural model describing monetary policy is the BM-FFR model, then augmenting a VAR with federal funds futures should, at least intuitively, imply monetary policy shocks closer to the surprise benchmark measure. Finding (b) implies a more convincing point. First, the best performing structural model seems to be the BM-FFR model. Thus, even if in the other models the futures augmentation yields improvements, we can argue that these models do not describe monetary policy adequately. Second, for the BM-FFR model the futures augmentation decreases the correlation between the RR benchmark measure and the identified monetary policy shock. The RR measure, in contrast to the Kuttner measure, is not a direct derivative of the futures series. Thus, the result based on the RR series seems normatively more compelling. Since we have established earlier that the shocks obtained from the baseline model are fundamental, we conclude, based on the points described above, that extending the VAR information set with a forward-looking variable is not necessarily beneficial.

Table 5: Correlations between benchmark measures and identified monetary policy shocks: BBE specifications.

	Benchmark	
FAVAR observed factor	Kuttner	RR
FFR	0.3822	0.2717
NBR, TR, FFR	0.3179	0.2243
IP, CPI, FFR	-0.0394	0.0673

Further support towards this conclusion can be found in Table 5. In this table we report the pair-wise empirical correlations between the identified monetary policy shock series and the Kuttner or RR measure for each specification of observable factors in Equation (26). The FAVAR specification that yields the highest correlation with the benchmark shocks is the one where the observable factor is solely the federal funds rate. However, the highest correlation among the FAVAR specifications, 0.3822 (or 0.2717), is notably lower than the highest correlation from the BM models even *without additional information*: 0.4119 (or 0.2974). That is, the identified monetary policy shocks from BBE underperform the identified monetary policy shocks for the simpler VAR models of Bernanke and Mihov (1998) in terms of correlations with benchmark measures. It is, on the other hand, particularly surprising that the futures augmented BM-FFR model performs worse than the FAVAR-FFR model compared to the RR benchmark.

Similar, or higher (> 0.39) correlations with the original Romer and Romer (2004) series, running from 1969 to 1996, have been reported in Coibion (2012) and Uhrin and Herwartz (2016) for monetary policy VAR specifications similar to the VAR in the present paper. We note that in these papers the structural assumptions and sample periods are distinct from ours. Nevertheless, there seems to be weak, but growing evidence for the observation that with a simple monetary policy SVAR model one can indeed recover shocks that are well correlated with existing benchmark measures.

There is one last insight that we can conclude from correlation analysis between distinct monetary policy shocks. A FAVAR arguably contains more information than a smaller dimensional VAR. Further, the federal funds futures—being spot rates on a futures market—arguably contains the market's information on macroeconomic developments. Thus, a priori one might expect that augmenting the BM VAR models with the federal funds futures would yield monetary policy shocks more similar to the monetary policy shocks from the FAVAR model. However, the opposite can be true. Table 6 reports the 30 pair-wise correlations between the monetary policy shock obtained from a FAVAR specification and the monetary policy shock obtained from the BM VAR specification.[26] As the table shows, in our benchmark BM specification (BM-FFR) the federal funds futures augmentation yields monetary policy shocks that are less correlated with the monetary policy shocks obtained from any of the FAVAR models. The same is true for the BM-NBR model, whereas in the other models the futures augmentation yield shocks that are correlated more strongly with the FAVAR monetary policy shocks. All of the latter models, however, imply correlations that are notably weaker than for the BM-FFR case.

3.4.3 *Structural impulse response analysis*

The innovations obtained from the baseline monetary policy SVAR model are fundamental, and so are the shocks obtained from the same VAR augmented with federal funds futures. Thus, by a structural analysis of the SVAR estimates we can safely draw conclusions about the underlying monetary policy shocks. In the previous section we argued that the most adequate structural model describing monetary policy is the BM-FFR model. Hence, in the

26 There are 30 pair-wise correlations, since we have 2×5 BM VAR models and 3 FAVAR specifications.

Table 6: Correlations between identified monetary policy shocks from BM and FAVAR specifications.

		Structural model				
FAVAR specification	VAR specification	BM-JI	BM-BR	BM-FFR	BM-NBR	BM-STR
FFR	With futures	0.0959	0.1234	0.6500	-0.1518	0.0235
	Without futures	0.0608	0.0581	0.7574	-0.1190	-0.0828
NBR, TR, FFR	With futures	0.1659	0.0833	0.6446	-0.0351	0.0799
	Without futures	0.1234	-0.0096	0.7547	0.0184	0.0119
IP, CPI, FFR	With futures	0.0196	-0.0124	0.1026	0.0200	0.0021
	Without futures	-0.0196	-0.0573	0.1029	0.0392	-0.0226

case of the BM models, we report the structural impulse responses for the BM-FFR model only.[27]

Figure 9 shows the VAR impulse responses to a 0.25 basis point increase in the federal funds rate. The monetary policy shock does not have a contractionary effect on output for the first several periods. In fact, the 95% confidence bands include zero for all considered horizons. This result is in contrast with the original results of Bernanke and Mihov (1998), who report a clear contractionary effect. Uhlig (2005), on the other hand, reports a non-contractionary and non-significant effect of monetary policy shocks. Recent evidence on post-1990 samples, however reports results similar to ours, cf., Ramey (2016, Figure 3.3). The impulse responses show an emergence of a mild price puzzle: the response of prices to a positive monetary policy shock is clearly positive for the first several periods. This trend, however gets reversed quickly.

The futures augmented VAR (Figure 10) provides a clearer picture in two respects. First, the price puzzle seems to be completely eliminated. Second, the output response starts out mildly positively, and then turns negative. This shape is similar to the shapes found in the most recent empirical monetary policy literature, see, in particular, Gertler and Karadi (2015), Ramey (2016), and Uhrin and Herwartz (2016). The futures series rises clearly in response to a monetary policy shock, and then comoves with the response of the federal funds rate. Even though the price puzzle does not appear in the (consumer) prices, the commodity prices markedly increase. This is in contrast with the baseline, non-augmented VAR, where the commodity price response to a positive monetary policy shock does not significantly deviate from zero.

Figure 11 displays the impulse responses obtained from the FAVAR-FFR specification. We report the impulse responses of only those variables that are contained in the baseline BM specification. With the exception of the federal funds rate, the variables are in logarithms and first differences, and transforming back to logged levels (by a cumulative sum) achieves smooth impulse responses. The behavior of the federal funds rate response is similar to the BM specifications in the first several periods. Later, however, the impulse response steadily declines towards zero. The responses of output and prices are essentially zero initially. The

27 Results on the other model variants are available upon request.

Figure 9: Impulse responses and bootstrap confidence bands for the baseline VAR specification: BM-FFR model. Monetary policy shock.

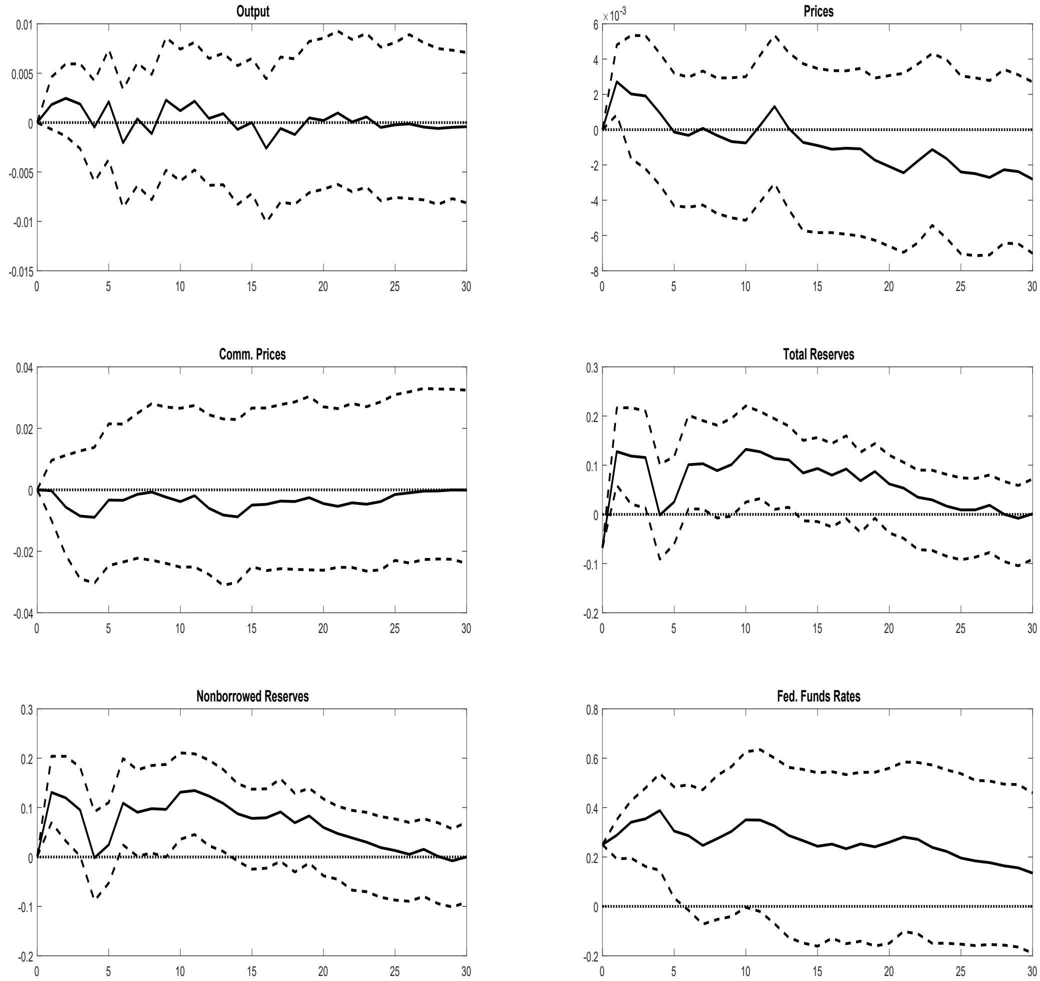


Figure 10: Impulse responses and bootstrap confidence bands for the futures augmented VAR specification: BM-FFR model. Monetary policy shock.

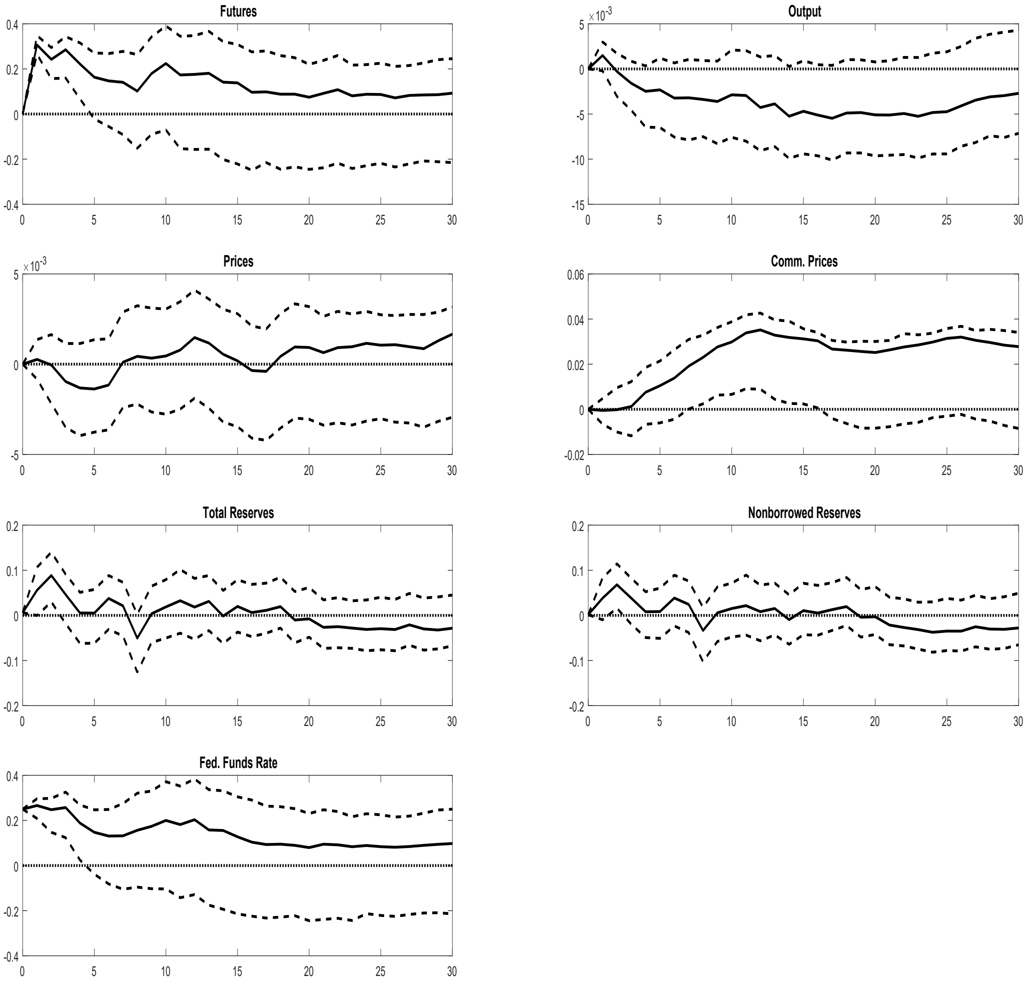
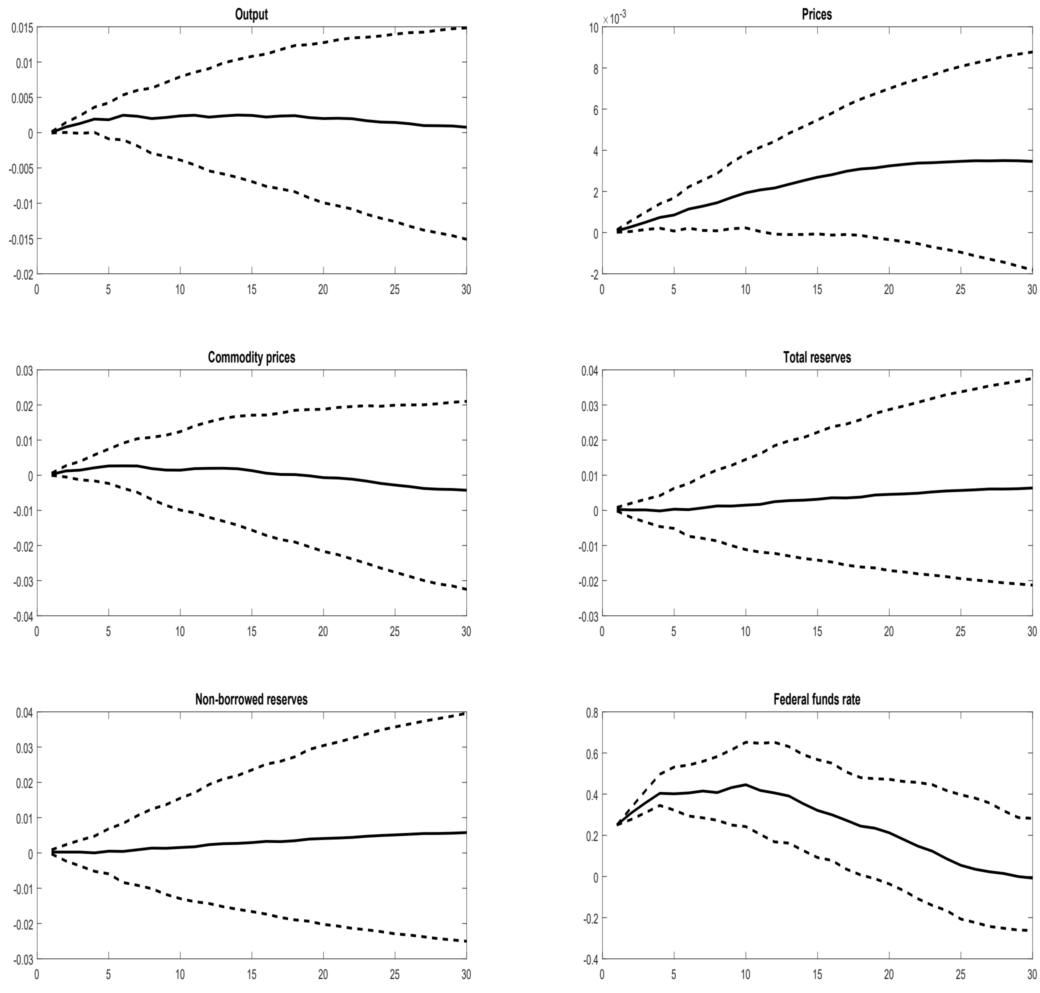


Figure 11: Cumulative impulse responses and bootstrap confidence bands for the FAVAR-FFR specification. Monetary policy shock.



confidence bands of the prices response indicate the plausibility of the price puzzle in this specification. However, on the long run, both output and prices become statistically not significantly different from zero.

The results of this section indicate, that including more (market-based) information in the monetary VAR can lead to the mitigation of the price puzzle. This finding is perfectly in line with Sims (1992), who argued that the price puzzle is a result of monetary VARs not containing sufficient information.[28] Bernanke, Boivin, and Eliasz (2005) use this insight as a motivation for their proposal of factor-augmentation. However, they also point out that the existence or non-existence of the price puzzle is not necessarily related to the inclusion of more information—a point further strengthened recently by Ramey (2016).[29] Rather, it can be the case that the monetary policy shocks are incorrect estimates of the true monetary policy shocks, and whether we discover a price puzzle empirically is only a statistical question. Our results above indicate that the shocks obtained from the baseline VAR are fundamental. Further, our preferred model in terms of precision of monetary policy shock estimation is the baseline VAR, and, in particular the BM-FFR structural model. Hence, we conjecture that the non-existence of the price puzzle in the augmented specifications is not the result of more information, but a statistical artefact of the estimation of a less parsimonious model.

3.4.4 *Counterfactual simulations*

Early empirical evidence based on SVAR studies showed that the effects of monetary policy shocks on other (real) variables are relatively small. In contrast, Romer and Romer (2004), Gertler and Karadi (2015) discovered much larger effects of monetary policy. Coibion (2012) reconciles the results of Romer and Romer (2004) with the earlier results, suggesting that the true effects are likely to lie in between those found by Romer and Romer (2004) and the previous consensus in the literature.

How much different are the effects of monetary policy shocks that are obtained from information augmented models? In the following we investigate this question through counterfactual simulations where only the monetary policy shocks drive the fluctuations of key variables. More precisely, we estimate a counterfactual $\hat{y}_t^*$ for each t in the sample using the reduced form parameter estimates, and assuming that the counterfactual shock series in the autoregressive equation is, e.g., $\hat{\varepsilon}_t^* = (\mathbf{0}, \hat{\varepsilon}_t^{\mathrm{mp}})'$. As in the previous section, we focus on, and report our results only for the BM-FFR model and the FAVAR-FFR specification.

Figure 12 displays the actual (dotted blue line) and the counterfactual (solid black line) yearly growth rates of industrial production, consumer price index, and commodity price index, respectively from 1990:01 to 2007:11. In the left column the counterfactual series are generated with only the monetary policy shocks from the baseline VAR having effect on the economy. In the right column the counterfactual series are generated with only the

28 The monetary authority might endogenously react to an inflationary pressure, thereby mitigating the future inflation. If we would be able to include in our empirical model all the information available to the monetary authority, we could establish that, *ceteris paribus*, the monetary contraction does *not* increase the price level (Sims, 1992, p. 988–989).

29 The argument of Bernanke, Boivin, and Eliasz can be found in the working paper version: Bernanke, Boivin, and Eliasz (2004, pp. 18).

monetary policy shocks from the futures-augmented VAR having effect on the economy. As we can see, the monetary policy shocks do not contribute much to the developments in the economy. The shocks from the futures-augmented specification, however, contribute even less, except for the episodes around 2001 and 2007 in the commodity price series. A similar insight can be gained from taking a look at counterfactual series where the monetary policy shocks are not present—the situation depicted in Figure 13.[30] As a benchmark, we can also contrast the counterfactual evolution of the variables when no shocks are driving the economy. Figure 14 displays this case. The pure autoregressive evolution of the two specifications is quite similar, verifying that the observed differences in the previous pictures were indeed caused by the shocks and not the different atoregression parameter estimates.

In the case of the FAVAR specification, there are two sources of uncertainty for those variables that are not contained in the state equation. First, the (identified) shocks from the state equation, second, the idiosyncratic shocks of the observation equation. It is, therefore, useful to investigate, first, how much the shocks to the observation equation contribute to the evolution of the time series, and, second, how much does the monetary policy shock contribute to the evolution of the series in the observation equation. The first row of Figure 15 displays the counterfactual scenario where the macroeconomic time series (in differenced logarithms) have only been affected by shocks in the state equation. We can see, that the disturbance terms in the observation equation have a sizable influence on the variables over and above the shocks from the state equation. This fact is more pronounced for the commodity price index, and the price series. However, the state equation does seem to drive the evolution of the output reasonably well. In the second row of Figure 15 we have set the monetary policy shocks to zero. If we contrast the second row with the first row, we can conclude that the contribution of the monetary policy shocks is moderate. This finding is further supported by the figures in the third row where only the monetary policy shock is affecting the equation system. While the monetary policy shock has a moderate influence on output, its influence on the prices and commodity prices is negligible.

We can conclude from the counterfactual analysis the following. The identified monetary policy shocks have small influence on the evolution of output, prices and commodity prices, and this is regardless of whether we augment the VAR specification with more information. The only sizable difference between the baseline and futures-augmented specification can be observed in the case of commodity prices, where the futures-augmented monetary policy shocks have a greater influence on the macroeconomic series in comparison with the baseline shocks. In the case of the FAVAR, the shocks to the unobserved factors contribute sizably to the evolution of output, but less so to that of prices and commodity prices. In contrast, the monetary policy shock's influence on these variables is only moderate.

30 Here $\hat{\varepsilon}_t^* = (\hat{\varepsilon}_t^-, 0)$, where $\hat{\varepsilon}_t^-$ is the estimated structural innovation excluding the monetary policy shock.

Figure 12: The contribution of monetary policy shocks to the fluctuations of macroeconomic variables.

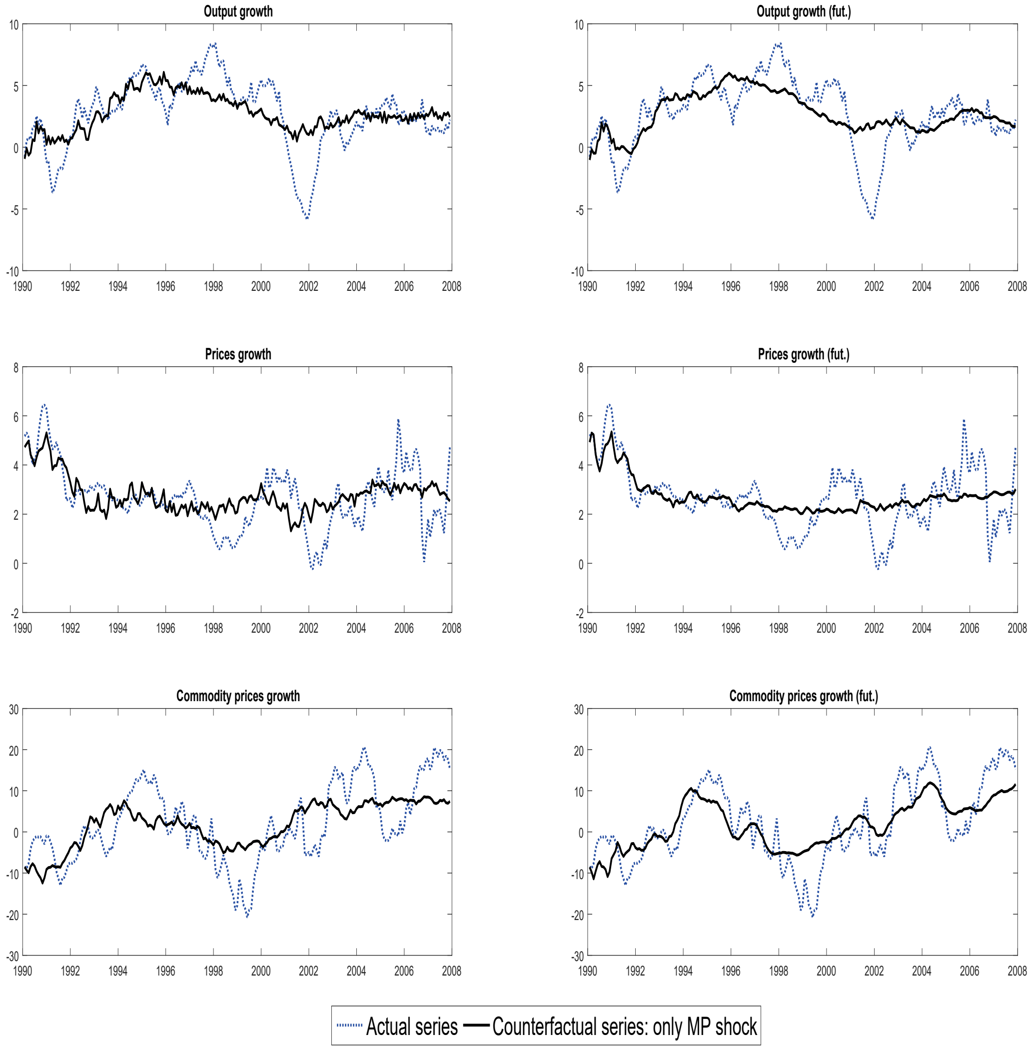


Figure 13: The evolution of macroeconomic variables without the monetary policy shocks.

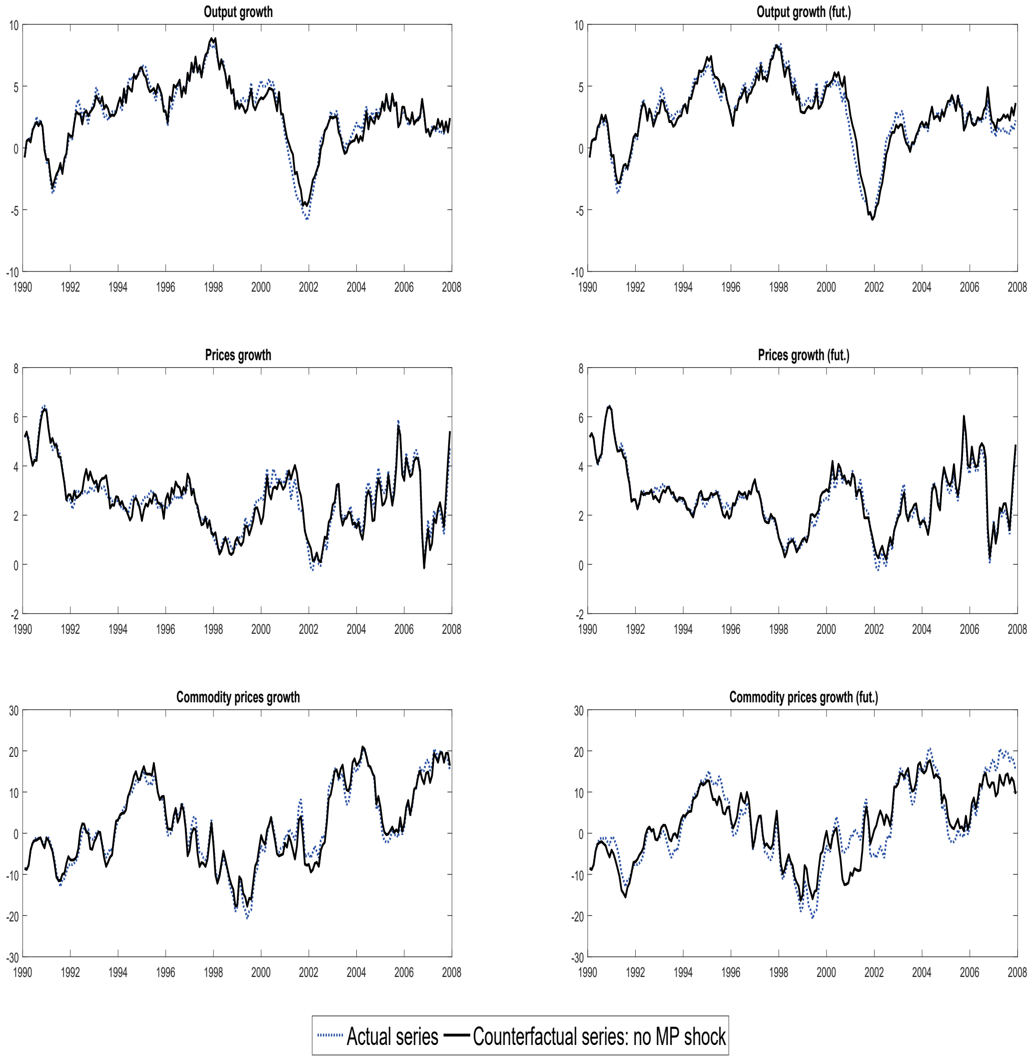


Figure 14: The evolution of macroeconomic variables without shocks.

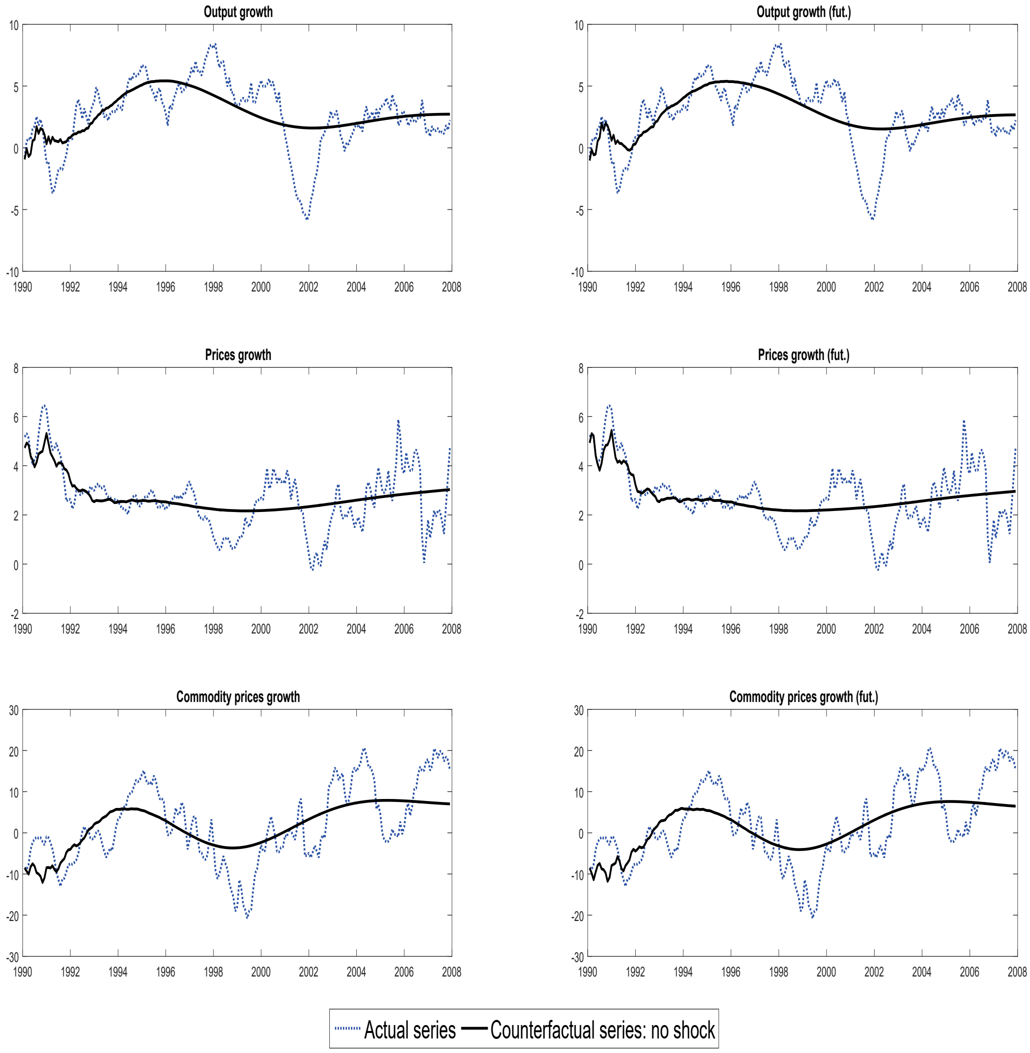
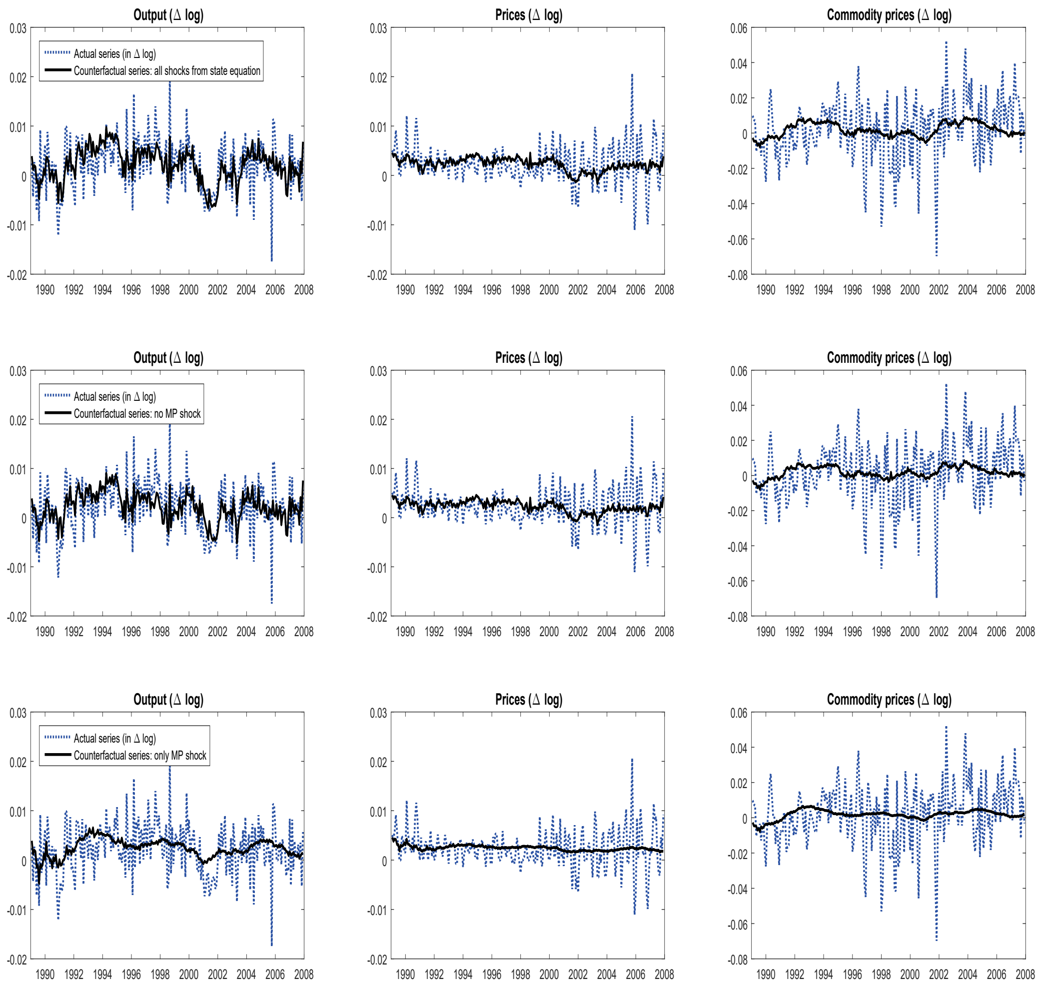


Figure 15: The contribution of the state equation shocks to the macroeconomic variables.



3.5 CONCLUSIONS

The inclusion of more forward looking information in structural VAR (SVAR) analyses is generally considered to be a good practice. It can achieve at least two goals. First, it can mitigate foresight problems, and, ultimately, render non-fundamental shocks in an empirical specification fundamental. We view fundamentalness of the structural shocks with respect to the variables in the empirical specification as a minimum requirement. Second, theoretically, the inclusion of variables containing expectations in the empirical model renders, by construction, the monetary policy shocks orthogonal to these expectations. Thus it seems a priori more likely that the thus obtained monetary policy shocks are closer to being "pure", exogenous monetary policy shocks.

In the present paper we investigate the classical monetary policy SVAR of Bernanke and Mihov (1998) for monthly US data running from 1989:01 to 2007:11. We augment the baseline VAR with more information in two ways. First, we add the federal funds futures series to the specification, second, we estimate a factor-augmented VAR (FAVAR) similar to Bernanke, Boivin, and Eliasz (2005).

We first argue that the federal funds futures series is a reasonable variable to include in monetary policy VARs. It arguably contains the market's expectation about monetary policy, Granger causes several of the key variables in the VAR, and the VAR as a whole. We view this finding alone as a useful insight for empirical monetary policy analysis. Then we test statistically whether the shocks from our estimated specification (with or without futures) are fundamental. To this end, we establish that the reduced form innovations are jointly non-Gaussian and serially independent. The null hypothesis of fundamentalness cannot be rejected for either specifications for all reasonable significance levels. Thus, we proceed to investigate whether information-augmentation yields monetary policy shock and impulse response estimates that are "better" than the baseline estimates. To this end, we compare the (identified) monetary policy shock estimates to the updated benchmark series of Romer and Romer (2004) and Kuttner (2001). According to our results, the preferred classical monetary policy SVAR model without the futures augmentation yields monetary policy shock estimates that are better correlated with both benchmark shocks than the monetary policy shocks obtained from a FAVAR, and not necessarily less correlated with the benchmark measures than the monetary policy shocks obtained from the futures-augmented specification. The price puzzle of the baseline specification is mitigated by factor and futures augmentation. We demonstrate that, aside from the ameliorated price puzzle, the empirical conclusions based on the classical monetary policy VAR do not differ from the conclusions from the information-augmented specifications: monetary policy shocks have negligible effects on the real economy. This finding is corroborates results in the literature on monetary policy shocks estimated based on post-1990 data, cf., Ramey (2016).

The reduced form VAR innovations are non-Gaussian, serially independent, and they form a martingale difference sequence. Thus, the implied structural shocks are fundamental. Furthermore, our VAR specification is standard in the monetary policy literature. While most of our messages are specific to the structural assumptions of the S(FA)VARs that we investigate, these observations on the reduced form innovations is expected to hold more generally. The result that the reduced form innovations are non-Gaussian opens the stage for extending the scope of investigation towards the recent SVAR literature on non-Gaussian monetary policy SVARs, see Lanne, Meitz, and Saikkonen (2017) and Herwartz (2016).[31]

31 Note, that we have only verified that the joint distribution of the residuals is not Gaussian. What is necessary for non-Gaussian SVARs is more: at most one marginal distribution needs to be Gaussian. This can be tested by, e.g., using the R-package `ICtest`.

4

MONETARY POLICY SHOCKS, SET-IDENTIFYING RESTRICTIONS, AND ASSET PRICES: A BENCHMARKING APPROACH FOR ANALYZING SET-IDENTIFIED MODELS

Gábor B. Uhrin and Helmut Herwartz

Abstract. A central question for monetary policy is how asset prices respond to a monetary policy shock. We provide evidence on this issue by augmenting a monetary SVAR for US data with an asset price index, using set-identifying structural restrictions. The impulse responses show a positive asset price response to a contractionary monetary policy shock. The resulting monetary policy shocks correlate weakly with the Romer and Romer (2004) (RR) shocks, which matters greatly when analyzing impulse responses. Considering only models with shocks highly correlated with the RR series uncovers a negative, but near-zero response of asset prices.

4.1 INTRODUCTION

The financial crisis of 2008–2009 has stirred up the debate on the conduct of monetary policy all around the world. One of the questions that came into the focus of the discussion is the extent to which monetary policy should react to developments in asset markets. Should central banks "lean against the wind" and try to mitigate turbulences in asset markets through raising interest rates, or should they rather concentrate solely on stabilizing the output gap and the inflation?[1] Arguing for either of these positions raises the need to quantify the (contemporaneous) effects of monetary policy actions on asset prices.

Starting with Sims (1980), such empirical questions have often been investigated by means of (structural) vector autoregressive (VAR) models. The crucial problem of identifying exogenous, unanticipated monetary policy shocks has been addressed in several studies that aimed to quantify the effects of monetary policy on, for instance, real output. Most of the classical procedures developed in these studies have been applied in a monetary policy – asset price context. As a particular exception, the agnostic sign-restriction approach exemplified by Uhlig (2005) has not been employed yet to explore the linkage between monetary policy shocks and asset prices.

Parallel to the SVAR literature, alternative approaches to identify monetary policy shocks have also been proposed. A major contribution has been put forth by Romer and Romer (2004), henceforth RR, who combined narrative evidence with statistical methods to construct a monetary policy shock series free of endogeneity and anticipation effects.

In the present paper we make several, related, contributions. First, we augment the VAR specification of Uhlig (2005) with the S&P 500 Composite Index, and estimate the model on

1 A concise summary of these debates can be found, e.g., in Assenmacher-Wesche and Gerlach (2010).

monthly US data from 1959 January to 2007 December. We use two set identifying restrictions to identify monetary policy shocks and examine the effects of these shocks on asset prices. The first restrictions are the sign restrictions of Uhlig (2005) (Scheme I), and the second set of restrictions are the zero and sign restrictions on the structural matrix A_0 put forth recently by Arias, Caldara, and Rubio-Ramírez (2015) (Scheme II). According to our results, the SVAR impulse responses point towards a mildly positive asset price response to an increase in the monetary policy instrument. This result is puzzling in light of earlier literature. Second, we argue that the resulting identified monetary policy shocks correlate only weakly with the monetary policy shock series of RR. We show that this finding matters greatly when analyzing (structural) impulse responses. In particular, we make the following observations: *i.*) the majority of admissible models yield impulse responses that vary widely in their shapes and impact magnitudes; *ii.*) this ambiguity affects those variables most whose responses are left agnostic by the identification scheme; *iii.*) models that are highly correlated with the RR shocks yield clearly shaped and less ambiguous impulse responses. Thus, third, we propose to restrict attention to those specifications that yield monetary policy shocks highly correlated with the RR series. We show that impulse response analysis of these models leads to more robust and reliable conclusions. Ultimately, we find evidence of: 1.) asset prices responding mildly negatively (in Scheme I), or ambiguously (in Scheme II) to a positive monetary policy shock, 2.) a mildly positive output response to what is understood to be a "contractionary" monetary policy shock. The former findings are contrary to our first results, but in line with conclusions of earlier studies. The latter finding is contrary to the baseline results obtained recently by Arias, Caldara, and Rubio-Ramírez (2015). Thus, we also conclude that comparing structurally (set-) identified shocks to a benchmark series can uncover by default hidden, but relevant and robust empirical conclusions. As a result, our methodological contribution complements the concerns of Kilian and Murphy (2012) regarding the interpretation of results from set-identified SVARs, and can be a useful empirical strategy when the identified set is not sufficiently narrow for sharp empirical conclusions. In fact, the benchmarking approach that we put forth can be considered as a step towards a frequentist parallel of the most likely models of Inoue and Kilian (2013).

The paper proceeds as follows: In Section 4.2 we provide an overview of existing results in identifying monetary policy shocks and their effects on asset prices. In Section 4.3 we detail the econometric model and the structural identifying assumptions. In Section 4.4 we present our baseline results,. In Section 4.5 we analyze the identified monetary policy shock series and compare them with the Romer and Romer (2004) series. In Section 4.6 we re-investigate our baseline results concentrating only on a certain subset of admissible models. Section 4.7 provides a discussion, some further results and robustness checks. Finally, Section 4.8 concludes.

4.2 MONETARY POLICY SHOCKS AND ASSET PRICES

While the crucial empirical problem in characterizing effects of monetary policy shocks is identifying exogenous, unanticipated changes in monetary policy, there seems to be no consensus in the literature on the identifying assumptions to use. Ramey (2016) provides a critical review of several identifying assumptions and argues that previous results based on distinct identifying assumptions cannot easily be reconciled, especially in longer, more recent samples.

Since distinct identifying assumptions may generate distinct results, the lack of consensus also applies to the empirical question: what are the effects of monetary policy shocks on asset prices? Compared with the literature on quantifying the effects of monetary policy on real variables, the empirical literature on monetary policy and asset prices is relatively small-scale. While the literature generally concludes that asset prices react negatively to an exogenous increase in the monetary policy instrument, the magnitude, the timing and the persistence of this negative reaction varies greatly across studies.

Earlier papers that use a recursive identification scheme, including Patelis (1997), Thorbecke (1997), Neri (2004), find that an increase in the monetary policy instrument leads to a small decrease in the stock prices. Bjørnland and Leitemo (2009) criticize the use of recursive identification schemes. Applying short and long run restrictions, they find large and persistent negative effects. More recently, Lanne, Meitz, and Saikkonen (2017) assume a non-Gaussian SVAR and confirm the findings of Bjørnland and Leitemo (2009) in rejecting the recursive identification scheme, and finding a significant instantaneous negative effect that, however, dies out quickly. In contrast, utilizing changes in the heteroskedasticity structure of the error term, Rigobon and Sack (2004) and Lütkepohl and Netšunajev (2014) find smaller, but relatively persistent negative effects. In a time-varying SVAR, Galí and Gambetti (2015) find negative short run effects that quickly turn into positive after impact especially in the 1980s and 1990s. Following an event-study approach around the monetary policy decision changes, Bernanke and Kuttner (2005) uncover that a 25 basis point cut in the federal funds rate leads, on average, to a 1% increase in asset prices.

As the above list of contributions indicate, a wide variety of approaches to SVAR analysis have been applied in the monetary policy – asset prices context. Notable exceptions are the usage of sign restrictions as proposed by, e.g., Uhlig (2005), and sign and zero restrictions advocated by Arias, Caldara, and Rubio-Ramírez (2015). We aim to fill this gap in the present paper, and we argue in the next section that using these restrictions as identifying assumptions in the context of our empirical question has several advantages over other identification schemes.

4.3 IDENTIFYING MONETARY POLICY SHOCKS WITH SIGN AND ZERO RESTRICTIONS

We consider the following K-dimensional structural VAR,

$$A_0 y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \varepsilon_t, \tag{34}$$

where $y_t \in \mathbb{R}^K$, $\varepsilon_t \sim WN(0, I_K)$, $A_0, \ldots, A_p \in \mathbb{R}^{K \times K}$, and A_0, what we call the *structural matrix*, is assumed to be non-singular. In order to define a unique lag length we assume that $A_p \neq 0$. In the above equation ε_t is the vector of *structural innovations*. The corresponding, estimable reduced form is

$$y_t = B_1 y_{t-1} + \cdots + B_p y_{t-p} + u_t, \tag{35}$$

with $B_i = A_0^{-1} A_i$, $i = 1, \ldots, p$. For u_t, the vector of *reduced form innovations* the following holds: $A_0^{-1} \varepsilon_t = u_t \sim WN(0, \Sigma_u)$. That is, the vector of structural innovations is a linear combination of the vector of reduced form innovations. Writing $B(z) = I_K - B_1 z - \cdots - B_p z^p$, we assume that the reduced form is causal, that is, $\det(B(z)) \neq 0 \ \forall |z| \leq 1$. Then, the moving average representation of y_t exists and is given by (Brockwell and Davis, 1991, Th. 11.3.1, p. 418):

$$y_t = \sum_{j=0}^{\infty} \Phi_j u_{t-j} = \sum_{j=0}^{\infty} \Theta_j \varepsilon_{t-j}, \quad \Phi_0 = I_K, \tag{36}$$

where element (i,k) of the coefficient $\Theta_j = \Phi_j A_0^{-1}$ is interpreted as the reaction of the i-th variable on the k-th structural innovation at horizon j. We call A_0^{-1} the *structural impact matrix*, since $\Theta_0 = A_0^{-1}$.

In this paper we aim to identify only one particular structural form innovation, the monetary policy shock, ε_t^{mp}, that is an element of the vector ε_t. We use sign restrictions on the impulse responses and zero restrictions on the structural matrix A_0 as identifying assumptions. Zero and sign restrictions on the structural matrix A_0 are straightforward: $A_0^{(i,k)}$, the (i,k)-th element of A_0 is restricted to be zero, positive, or negative. Sign restrictions on the impulse responses can be formulated as follows: $\Theta_j^{(i,k)}$, the (i,k)-th element of Θ_j, is restricted to be either negative or positive for some a priori selected combinations of (i,k,j), $i,k \in \{1,\ldots,K\}$, $j \in \mathbb{N}_0$. Note, that an insufficient amount of zero restrictions on A_0, or sign restrictions in general cannot point identify the structural parameters $A_0, \ldots, A_p$.[2]

Using sign restrictions on impulse responses for several periods to identify monetary policy shocks has been first proposed by Uhlig (2005).[3] Somewhat surprisingly, we are not aware of any attempt to utilize sign restrictions in a monetary policy – asset prices context. The use of zero restrictions on the structural matrix to restrict the systematic component of monetary policy in the SVAR has been put forth recently by Arias, Caldara, and Rubio-Ramírez (2015), and we are not aware of any research employing this identification to our empirical question. While the employed sign and zero restrictions cannot, in general, point-identify a structural VAR model, or a structural shock, using set identification has two important advantages in our view.[4]

First, sign restrictions by construction avoid the problem of deciding upon the exact recursive ordering of shocks. As Bjørnland and Leitemo (2009) pointed out, it is important to allow for the possibility of the monetary policy shocks contemporaneously affecting asset prices *and vice versa*—a view supported by theoretical models of, e.g., Castelnuovo and Nisticò (2010). A simple recursive identification scheme necessarily excludes one of these possibilities. Further, this critique of the recursive identification schemes was also strengthened recently by Lanne, Meitz, and Saikkonen (2017), who assume non-Gaussian error terms, and statistically test and reject the adequacy of the recursive scheme.

Second, sign restrictions, and a small number of zero restrictions, on the other hand, are considered to be mild assumptions that are relatively easy to interpret, justify, and agree upon. If we are striving for exact identification together with allowing for non-recursivity,

2 For necessary and sufficient conditions for exact (point) identification see, for example, Rubio-Ramírez, Waggoner, and Zha (2010).

3 Similar contributions are Faust (1998) and Canova and De Nicoló (2002).

4 We refer to a shock being set identified if there are at least two parameter points in the structural parameter space that are observationally equivalent, i.e., lead to the same reduced form parameters. This terminology is also used by Arias, Caldara, and Rubio-Ramírez (2015).

or contemporaneous interdependence, then we have to argue for at least one additional restriction to achieve it. If, however, one is willing to give up on exact identification, then restrictions that were used to achieve exact identification may become harder to argue for.

Thus, in the following we use sign and zero restrictions to identify the monetary policy shock and investigate the effect of this identified shock to a stock price index variable. To this end, we augment a VAR specification similar to Uhlig (2005) with the S&P500 Composite Index and use, as our first assumption, the same identifying assumption as Uhlig (2005):

RESTRICTION SR1 A monetary policy shock's effects on the impulse responses of commodity prices, GDP deflator and non-borrowed reserves is non-positive, and on the impulse response of the federal funds rate is non-negative for the impact period and four periods after impact.

We call this restriction *Scheme I*. Besides being intuitively reasonable, these sign restrictions are also supported by New-Keynesian DSGE models under a wide set of parameter calibrations (Carlstrom, Fuerst, and Paustian, 2009). Note, that the original formulation by Uhlig (2005) requires the impulse responses of "prices" in general to be non-positive. While this assumption may also include asset prices, we prefer to remain agnostic about the signs of effects of monetary policy on asset prices, hence we do not constrain the response of asset prices to monetary policy shocks.

Arias, Caldara, and Rubio-Ramírez (2015) argue, however, that the sign restrictions of Uhlig (2005) imply parameter estimates that are incompatible with theoretical considerations about and empirical evidence on the systematic component of the monetary policy, the Taylor rule. Since monetary policy shocks are innovations to the Taylor rule, the identification of monetary policy shocks should be coupled with identifying the corresponding systematic monetary policy equation in the SVAR. This can be achieved by means of zero restrictions on the structural matrix. Following Arias, Caldara, and Rubio-Ramírez (2015), we use the following zero and sign restrictions on A_0 to identify monetary policy shocks:

RESTRICTION ZR The federal funds rate only reacts contemporaneously to GDP, GDP deflator, commodity prices, and asset prices.

RESTRICTION SR2 The federal funds rate's contemporaneous reaction to GDP, and to the GDP deflator is positive.

We call these restrictions jointly *Scheme II*. These restrictions explicitly impose a Taylor-type rule on the federal funds rate equation of the SVAR consistent with empirical and theoretical evidence about the systematic component of monetary policy. In particular, restriction ZR implies that the contemporaneous reaction of the federal funds rate to non-borrowed reserves and total reserves is zero. The monetary policy shock is identified as the innovation corresponding to this correctly specified equation in the SVAR. Since we are interested in the response of the monetary policy instrument to asset prices, we allow the federal funds rate to react contemporaneously to asset prices. This leaves our identification agnostic in the asset price – monetary policy context. It is important to note that Restrictions

ZR and SR2 restrict the structural matrix A_0. Thus, in contrast to Restriction SR1, the impact period impulse response coefficient A_0^{-1} is restricted only indirectly.[5]

4.4 MONETARY POLICY SHOCKS AND ASSET PRICES

First we investigate the effects of monetary policy on asset prices in a structural VAR similar to Uhlig (2005). The VAR is estimated with monthly US data from 1959:01 to 2007:12. The seven variables used in the specification are: Real GDP, GDP deflator, commodity price index, stock price index, federal funds rate, non-borrowed reserves, total reserves. Monthly series for real GDP and the GDP deflator were interpolated as in Mönch and Uhlig (2005). Real GDP was interpolated using the industrial production index, while the GDP deflator was interpolated by means of consumer and producer price indices. The commodity price index is the Commodity Research Bureau's BLS spot index obtained from Thomson Reuters' Datastream and is determined as the monthly average of daily data. Monthly observations of the S&P 500 Composite Index were obtained from the FRED MD project website maintained by Michael W. McCracken (McCracken and Ng, 2016). For the empirical analysis, the values were deflated by the GDP deflator. The remaining variables were obtained from the St. Louis FRED database under the following names: GDPC1 (real GDP), INDPRO (industrial production), GDPDEF (GDP deflator), CPIAUSL (consumer price index), PPIFGS (producer price index), FEDFUNDS (federal funds rate), TOTRESNS (total reserves), and BOGNONBR (non-borrowed reserves).[6]

To facilitate comparability, we employ the same VAR specification as Uhlig (2005): the VAR contains $p = 12$ lags and does not include a constant or deterministic trend. The federal funds rate is considered in levels. All other variables are in logarithms and multiplied by 100. We estimate the VAR by OLS. In order to simulate the set of sign and zero restricted impulse responses, we use the algorithms proposed by Rubio-Ramírez, Waggoner, and Zha (2010) and Arias, Rubio-Ramírez, and Waggoner (2014). A detailed description of these algorithms can be found in Appendix 4.B. In short, we draw random orthogonal matrices, Q, to rotate the lower-triangular Cholesky decomposition $\widehat{A}_u^c$ of $\widehat{\Sigma}_u$, the estimated reduced form variance-covariance matrix. The rotation matrices Q are constructed in a systematic way so that the structural form parameters estimated using $\widehat{A}_0^{-1} = \widehat{A}_u^c Q$ satisfy the sign and zero restrictions (i.e., they are *admissible*). We repeat the random drawing procedure until we have 65000 admissible impulse responses. Each of these 65000 impulse responses corresponds to a distinct *admissible model*, $\widehat{A}_{0,s}^{-1} = \widehat{A}_u^c Q_s$, where $s = 1, \dots, 65000$ is the simulation index.[7] In the plots below we also report the median target (MT) impulse responses as advocated by Fry and Pagan (2011). The MT impulse response is the impulse response that is closest in terms of a standardized squared distance to the median of the set of admissible impulse responses.

5 Note, that a zero restriction in A_0 in general does not imply a zero restriction in A_0^{-1}.
6 Further details on the data and sources can be found in Appendix 4.A.
7 While the number 65000 is based on computational constraints, our results are robust to several dozen runs of the same procedure.

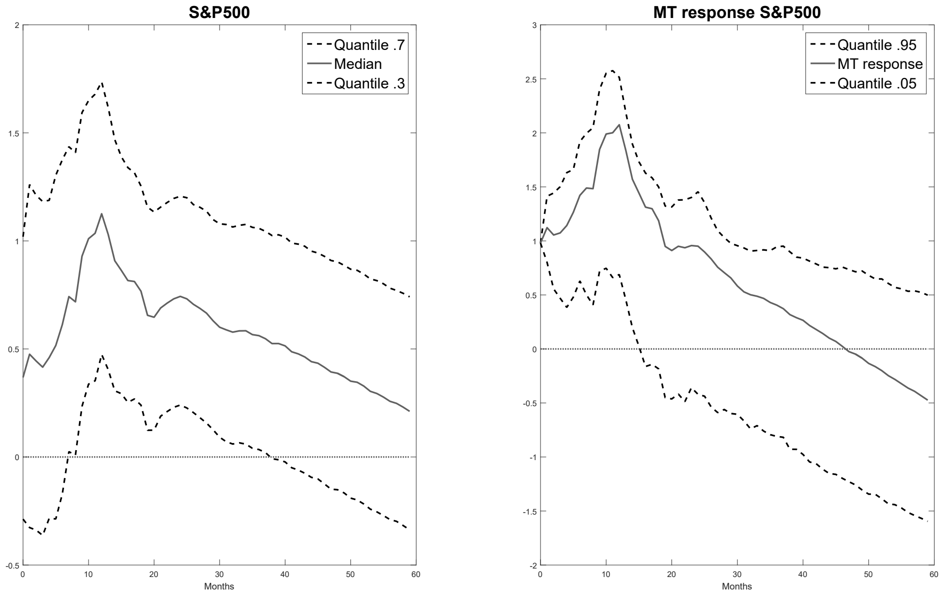


Figure 16: Sign restricted (left) and median target (right) impulse responses of the S&P 500 index for a one per cent increase in the federal funds rate. Identification Scheme I.

Sign restrictions on impulse responses

We identify a monetary policy shock first solely via Scheme I (Restriction SR1), our baseline restriction on the impulse responses. In Figure 16 we display the impulse response of the asset price index to a one per cent increase in the federal funds rate. The sub-figure on the left contains the pointwise median, as well as the pointwise 0.3 and 0.7 quantiles of the set of admissible impulse responses. In the sub-figure on the right we report the median target impulse response joint with a 90% bootstrap confidence band.[8]

Figure 17 visualizes the set of admissible impulse responses of the rest of the variables to a one per cent positive monetary policy shock (increase in the federal funds rate). The results are very similar to those presented in Uhlig (2005, Fig. 6., p. 397), thus we do not discuss them in detail. The only difference is the slight rising trend of the GDP deflator after the impact period. Figure 18 contains the MT impulse responses for the same variables as in Figure 17.

8 Details of the bootstrap procedure can be found in Appendix 4.B.

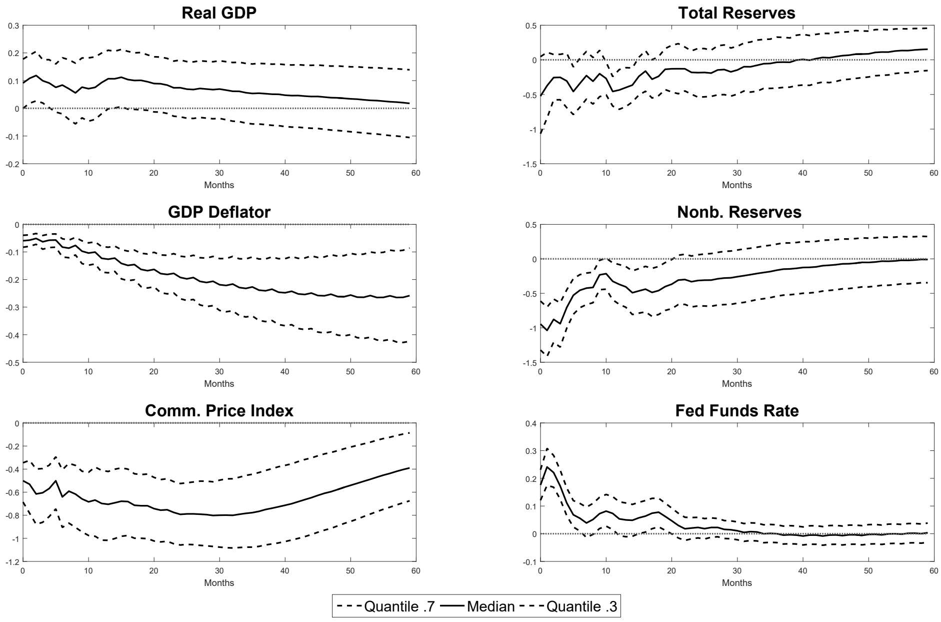


Figure 17: Sign restricted impulse responses for a one per cent increase in the federal funds rate. Identification Scheme I.

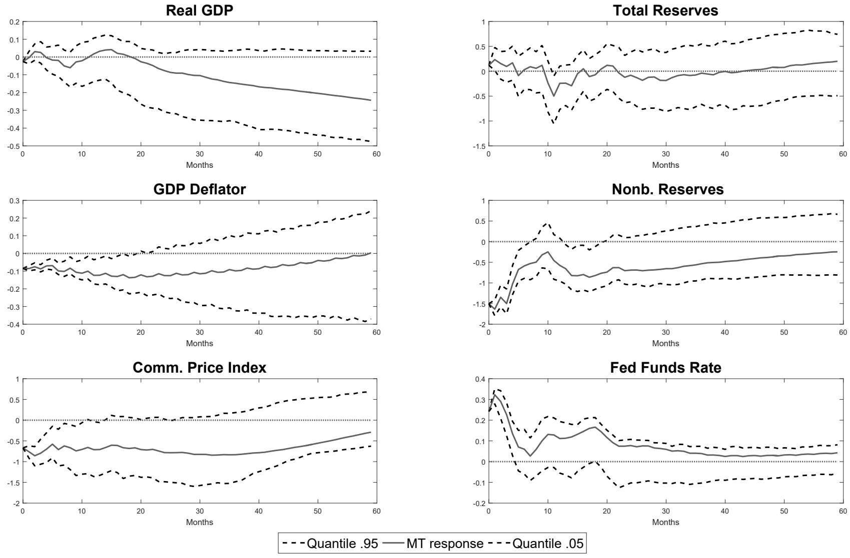


Figure 18: Median target impulse responses for a one per cent increase in the federal funds rate. Identification Scheme I.

The conclusion of our baseline analysis is that, for a one per cent increase in the federal funds rate, the set of admissible impulse responses of asset prices have more mass above the zero line. A similar result holds also for the response of GDP. While one can argue that these conclusions are ambiguous, as there are also admissible models that yield negative asset price responses, the MT impulse responses further point towards positive responses to a positive monetary policy shock.

Zero and sign restrictions on the structural matrix

The alternative identification scheme we employ is described by Scheme II (Restrictions ZR and SR2). These are restrictions on the structural matrix, A_0, and they are imposed jointly. The structure and interpretation of the following figures is similar to those in the previous subsection.

Figure 19 paints a more ambiguous picture than the baseline sign restriction specification: the median of the admissible impulse responses for asset prices starts at zero. While it turns positive in the short and medium run, and remains so later on, the admissible set does not leave the neighborhood of zero markedly. In the right-hand-side subfigure, the median target impulse response for asset prices, we can observe a similarly ambiguous pattern, even though the impulse response is significantly poisitive after the 10th month.

Figure 20 mostly corroborates the results of Arias, Caldara, and Rubio-Ramírez (2015, Figure 1., p. 12). Under the zero and sign restriction on the structural matrix the identified monetary policy shock has mostly a contractionary effect on the output. This is in contrast to our baseline results, and results by Uhlig (2005). Since the responses of the GDP deflator are mostly negative, the often discovered price puzzle does not seem to appear in this setup. This picture is further strengthened by the median target impulse responses in Figure 21.

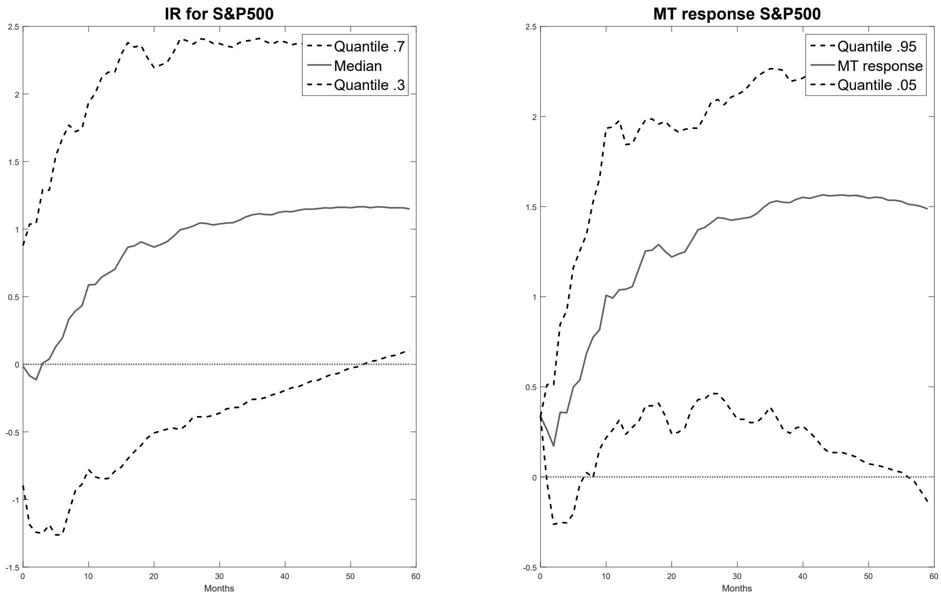


Figure 19: Sign restricted (left) and median target (right) impulse responses of the S&P 500 index for a one per cent increase in the federal funds rate. Identification Scheme II.

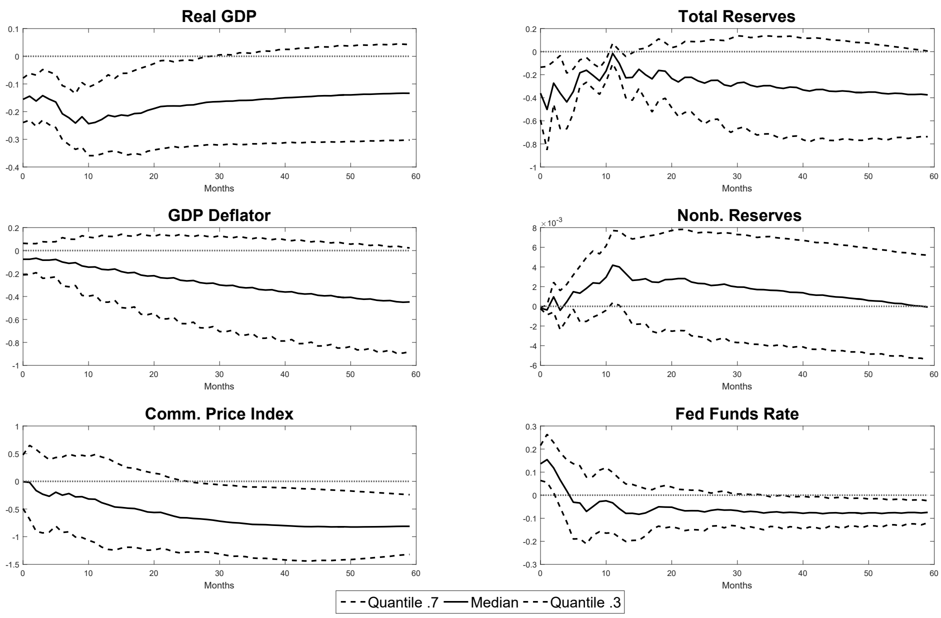


Figure 20: Sign restricted impulse responses for a one per cent increase in the federal funds rate. Identification Scheme II.

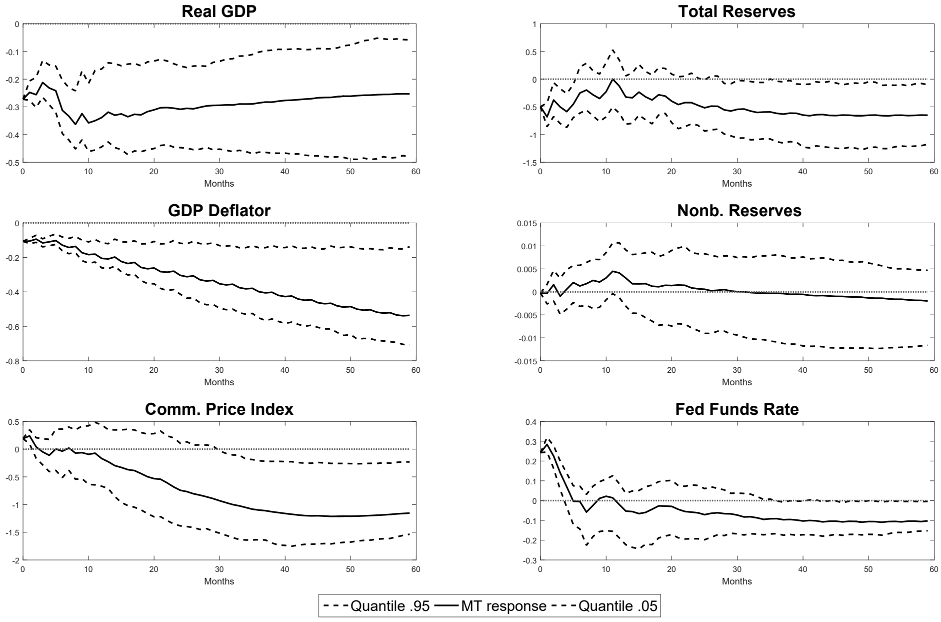


Figure 21: Median target impulse responses for a one per cent increase in the federal funds rate. Identification Scheme II.

4.5 examining the monetary policy shock

As we noted in the previous subsection each of the $s = 1, 2, \ldots, 65000$ impulse responses corresponds to a different admissible model $\widehat{A}_{0,s}^{-1}$. Similarly, to each $\widehat{A}_{0,s}^{-1}$ corresponds an identified monetary policy shock series $\{\hat{\varepsilon}_{ts}^{mp}\}_{t=1,\ldots,T}$ that is obtained from the reduced form residuals by the relation $\hat{\varepsilon}_t = \widehat{A}_0 \hat{u}_t$. In this section we investigate the identified monetary policy shocks by comparing the obtained series $\{\hat{\varepsilon}_{ts}^{mp}\}_{t=1,\ldots,T}$ for each s with the Romer and Romer (2004) series. In order to keep the argumentation compact, in the present section we report results using only the baseline identification restrictions, Scheme I. The following arguments, however, hold similarly for monetary policy shocks identified with the restrictions of Scheme II.[9]

Romer and Romer (2004) develop a monthly measure of monetary policy shocks for the period 1969–1996 that is based on the following methodology: the authors 1.) identify intended federal funds rate changes around meetings of the Federal Open Market Committee (FOMC) by combining narrative accounts of the FOMC meetings and the report of the manager of open market operations; 2.) regress the intended changes on the Fed's internal (so called "Greenbook") forecasts of inflation, real output growth and unemployment in

9 Further results on the latter case can be found in Appendix 4.C.

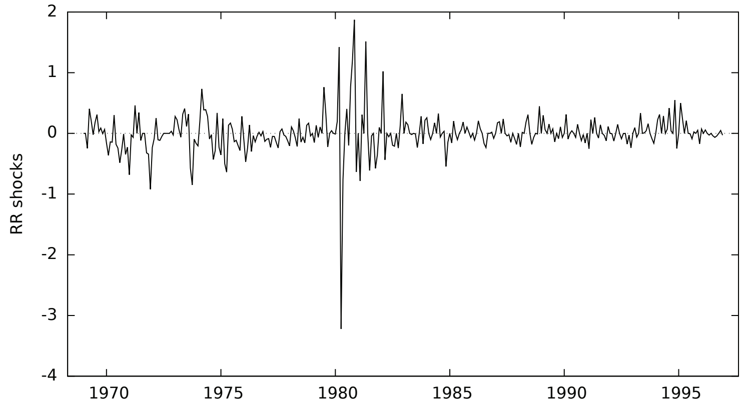


Figure 22: Romer and Romer (2004) shock series.

order to control for information about future developments in the economy. Specifically, the regression that they estimate is the following (Romer and Romer, 2004, Eq. 1, p. 1061):

$$\Delta ff_m = \alpha + \beta ffb_m + \sum_{i=-1}^{2} \gamma_i \widetilde{\Delta y}_{mi} + \\ + \sum_{i=-1}^{2} \lambda_i (\widetilde{\Delta y}_{mi} - \widetilde{\Delta y}_{m-1,i}) + \sum_{i=-1}^{2} \phi_i \tilde{\pi}_{mi} + \sum_{i=-1}^{2} \theta_i (\tilde{\pi}_{mi} - \tilde{\pi}_{m-1,i}) + \rho \tilde{u}_{m0} + \nu_m, \quad (37)$$

where Δff_m is the change in the intended federal funds rate at the FOMC meeting m, ffb_m is the intended federal funds rate before any changes decided on meeting m, and $\widetilde{\Delta y}_{mi}$, $\tilde{\pi}_{mi}$, $\tilde{u}_{m0}$ are the forecasts of real output growth, inflation and unemployment, respectively, for quarter i at the time of meeting m. The estimated residuals $\hat{\nu}_m$ represent unanticipated monetary policy shocks, and they are averaged over months to obtain the monthly series $\hat{\varepsilon}_t^{rr}$, the RR monetary policy shock series that runs from January 1969 to December 1996.[10] Figure 22 shows the RR series.

Romer and Romer (2004) carefully argue about the validity of the interpretation of their measure as monetary policy shocks. To our knowledge, only Coibion (2012) provides a critical examination of the implications of the RR shocks. The main objective of Coibion (2012) is to try to reconcile the surprisingly large influence of monetary policy shocks on, for example, output, with earlier similar studies. While he argues that the implications of Romer and Romer (2004) are not robust to, for example, excluding certain episodes in US central banking history, we do not read Coibion (2012) as an argument against the validity of interpreting the RR series as a "pure" monetary policy shock series. We, in fact, go further and argue that any identified structural monetary policy shock series obtained from, e.g., a SVAR analysis should resemble the RR series, $\hat{\varepsilon}_t^{rr}$. Further, Coibion (2012)'s analysis is based partly on alternative monetary policy shock series proposed in Coibion and Gorodnichenko (2011). These alternative shock series allow for *i.)* heteroskedasticity in the error term ν_m, *ii.)* time-varying coefficients in Equation (37). These modifications seem a priori sensible, however, using the monetary policy shock series from Coibion and

10 Note, that in any particular month there can be several FOMC meetings m or no meetings at all.

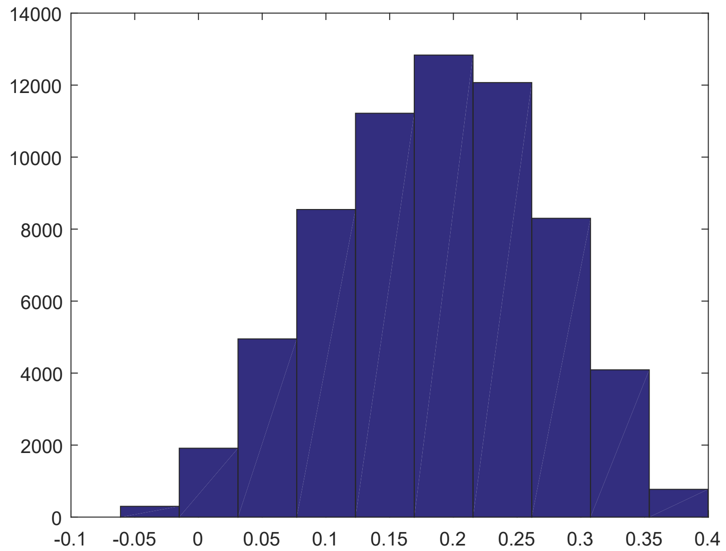


Figure 23: Histogram of correlations between $\hat{\varepsilon}_t^{mp}$ and $\hat{\varepsilon}_t^{rr}$. Scheme I.

Gorodnichenko (2011) does not lead to conclusions different from what we describe below.[11] Thus, in the following we view the RR series, as a benchmark monetary policy shock series.

How strongly do the identified monetary policy shocks from our baseline analysis resemble the benchmark monetary policy shocks? In order to answer this question, for each admissible monetary policy shock series $\hat{\varepsilon}_{ts}^{mp}$, with $s = 1, \dots, 65000$, we calculate the correlation $\rho_s = \widehat{\mathrm{Corr}}(\hat{\varepsilon}_{ts}^{mp}, \hat{\varepsilon}_t^{rr})$ on the subsample running from January 1969 to December 1996.[12] For simplicity of notation, and without loss of generality, assume that for $s < s'$ it holds that $\rho_s \leq \rho_{s'}$, that is, we index admissible models by their corresponding correlation coefficients.

Figure 23 contains the histogram of the 65000 obtained correlation coefficients ρ_s. As visual inspection immediately suggests, the correlations are mostly quite weak. Indeed, the average correlation is 0.1863, and the median is 0.1899. In the previous section we reported the median target impulse responses, and used these as further evidence for our results. However, the correlation corresponding to the median target model is 0.1621, i.e., lower than both the average and the median. This implies that at least half of the models have larger ρ_s values than the median target model.

What does it imply for the impulse response analysis if the identified monetary policy shocks are weakly correlated with the Romer and Romer shock series? In Figure 24 we report the impulse responses for each variable of the models $s = 1, \dots, 100$, i.e., those 100 models where ρ_s is the weakest. Figure 25 contains the impulse responses of models $s = 37450, \dots, 37550$, i.e., the 100 median models, and Figure 26 contains the impulse

11 Estimation results are available upon request.

12 Note, that, while for $\hat{\varepsilon}_{ts}^{mp}$ the t index runs from January 1960 to December 2007, the RR series, $\hat{\varepsilon}_t^{rr}$, is available only between January 1969 and December 1996.

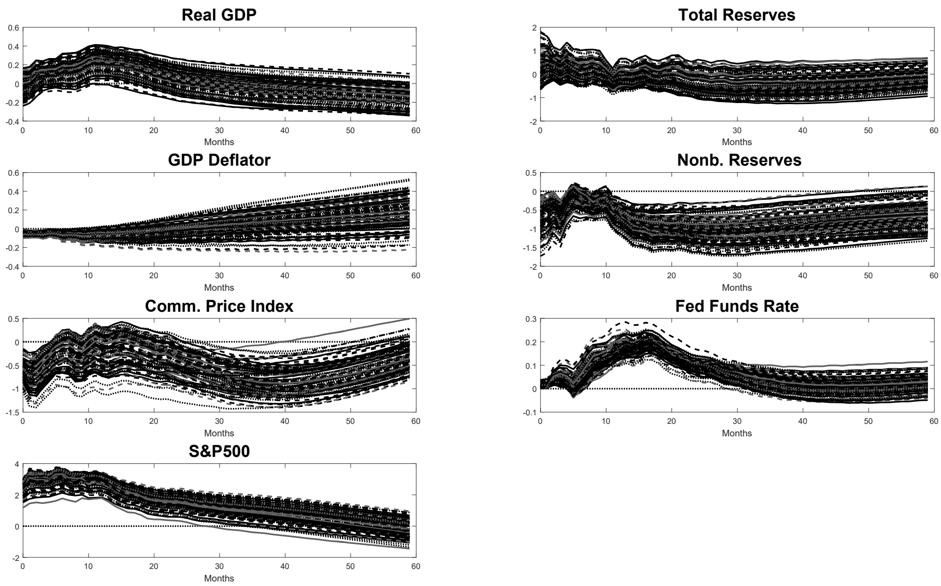


Figure 24: Impulse responses from models $s = 1, \ldots, 100$. Identification Scheme I.

responses from models $s = 64900, \ldots, 65000$—the 100 models with the highest correlations ρ_s.

The figures suggest that by concentrating on impulse responses from models with low correlations ρ_s, we might be led to notably different qualitative and quantitative conclusions in comparison with the impulse responses of models with high correlations ρ_s. In particular, Figure 24 supports the conclusion that asset prices react positively to a positive monetary policy shock, and they persistently remain so for several periods after impact. Similarly, GDP reacts rather ambiguously on impact, but there is a clear hump-shape suggesting a sluggish response to a monetary policy shock. In contrast, Figure 26 suggests that GDP reacts positively to an increase in the monetary policy instrument followed by a steady gradual decrease. Asset prices, on the other hand, respond mildly negatively on impact.

More surprisingly, according to Figure 25, models that yield shocks featuring an average correlation with the RR series hardly support any unambiguous empirical conclusion. Anything can happen as regards the shapes and magnitudes of impulse responses for *all* variables, and, *especially*, for those responses that are left unrestricted by the identification scheme. This finding is worth emphasizing for two reasons. First, the large majority of models are close to an average (or median) correlation level, cf., Figure 23. Second, note, that a central idea of sign restrictions, emphasized by Uhlig (2005), is to leave those variables' responses agnostic whose behavior is of key interest to the analysis at hand. Hence, if one randomly selects two admissible models, then they might, with high probability, lead to distinct conclusions especially with respect to key variables. Further, since the median target impulse response is inevitably tied to some "average" model, analyzing the median target alone offers inconclusive, or even misleading results.

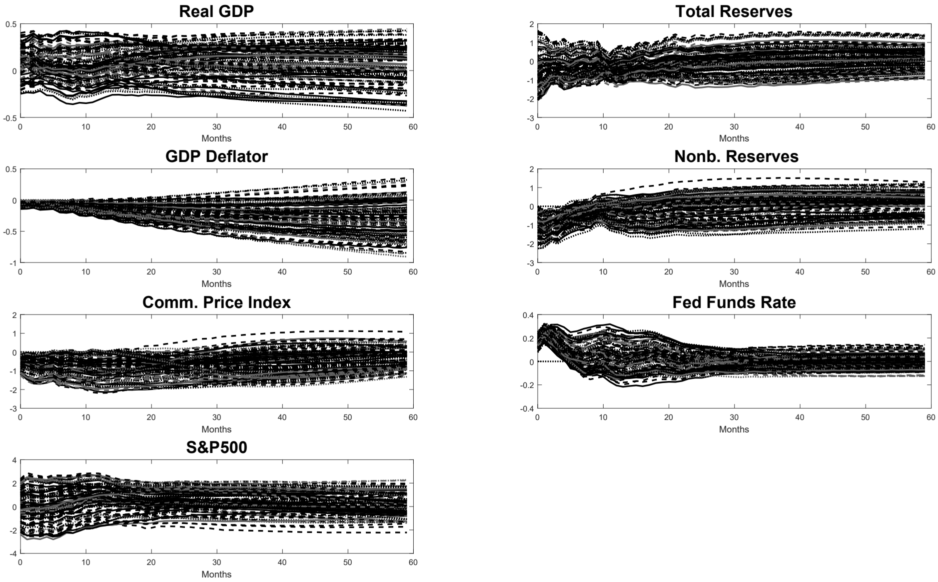


Figure 25: Impulse responses from models $s = 37450, \ldots, 37550$. Identification Scheme I.

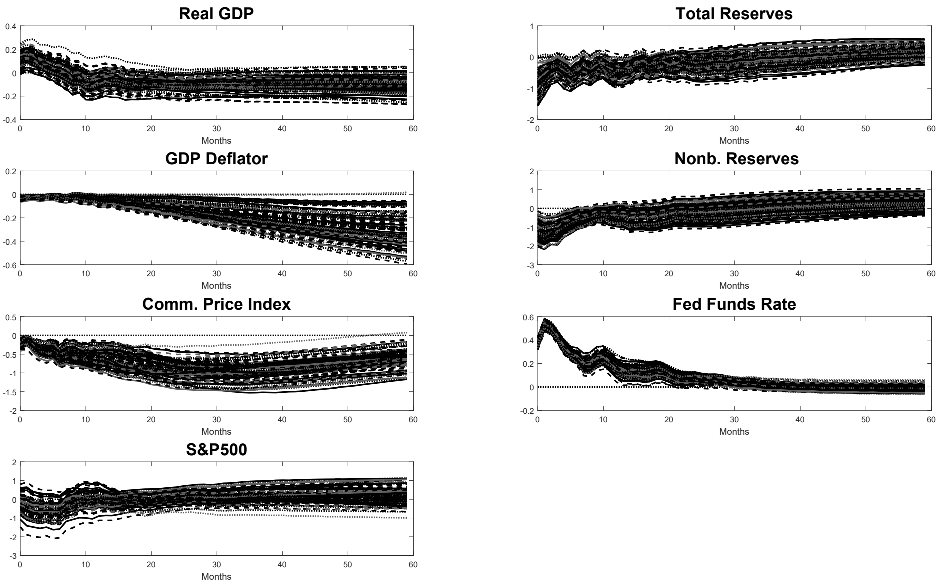


Figure 26: Impulse responses from models $s = 64900, \ldots, 65000$. Identification Scheme I.

In sum, we have argued in the previous paragraphs that the qualitative and quantitative features of the impulse response functions of models $s = 1, \ldots, 65000$ are closely linked to their implied correlation with the RR shocks, ρ_s. Models with similar *high* (or *low*) ρ_s values imply similar impulse responses. Models with notably different ρ_s values imply notably different impulse responses.

Thus, if one accepts that the identified monetary policy shocks should ideally be closely correlated with the RR shocks, one should put particular emphasis on analyzing the impulse responses displayed in Figure 26. Indeed, such highly correlated models grasp best what is implied by a "true" monetary policy shock. Thus, in the next section we reconsider our baseline results concentrating on those 100 models that yield monetary policy shocks that are closest to the RR series. In addition, we give special attention to model number 65000 showing the highest correlation with $\hat{\varepsilon}_t^{rr}$.

4.6 ANALYZING MODELS WITH THE HIGHEST CORRELATION

We have argued that most of the monetary policy shocks identified in our baseline model correlate only weakly with the Romer and Romer (2004) monetary policy shock series that we view as a benchmark series. In this section we reconsider the impulse responses concentrating only on $\mathcal{S}^{high} = \{ s : s \in [64900, 65000] \cap \mathbb{N} \}$, those 100 models that have the highest ρ_s, and $s^* = 65000$, the model with the highest ρ_s.

Sign restrictions on impulse responses

Figure 27 displays the minimal envelope, the maximal envelope and the median of impulse responses of models in $\mathcal{S}^{high}$ identified through Scheme I (Restriction SR1). In response to a positive monetary policy shock we find evidence for a positive, rising, but then quickly falling and in the end negative GDP response. This finding strengthens the result from the baseline analysis. Contrary to the baseline analysis, however, we also see evidence for a zero or mildly negative asset price response, which seems to reconcile results from the baseline analysis with existing previous studies. Note, that the hump shapes around months 7 and 10, respectively, are typical in the considered set $\mathcal{S}^{high}$, and this shape is also in line with impulse responses obtained earlier in the literature (e.g., Galí and Gambetti (2015)). The remaining variables behave similarly to the baseline analysis, but we get a notably sharper picture on how the GDP deflator, non-borrowed reserves and the federal funds rate behave in reaction to a positive monetary policy shock.

We also show in Figure 28 the impulse responses corresponding to model s^*, i.e., the model that yields identified monetary policy shock series that has the highest correlation with the RR series. Model s^* strengthens the results in the previous paragraph: in particular, asset prices react sizably negatively and they return only slowly to their starting value.

Figure 29 displays the (standardized) monetary policy shock series corresponding to model s^*, together with the (standardized) RR series.

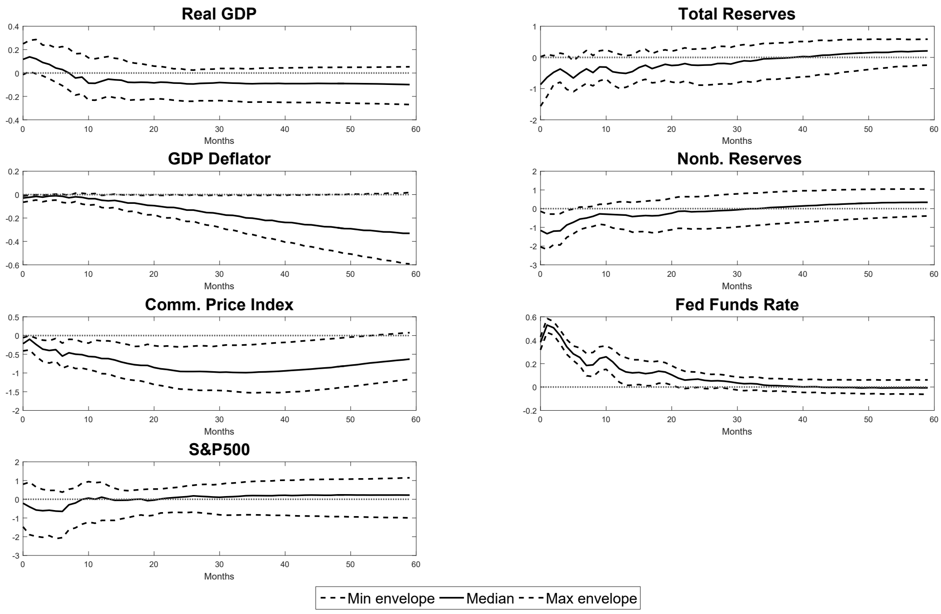


Figure 27: Range and median of models from $\mathcal{S}^{high}$. Identification Scheme I.

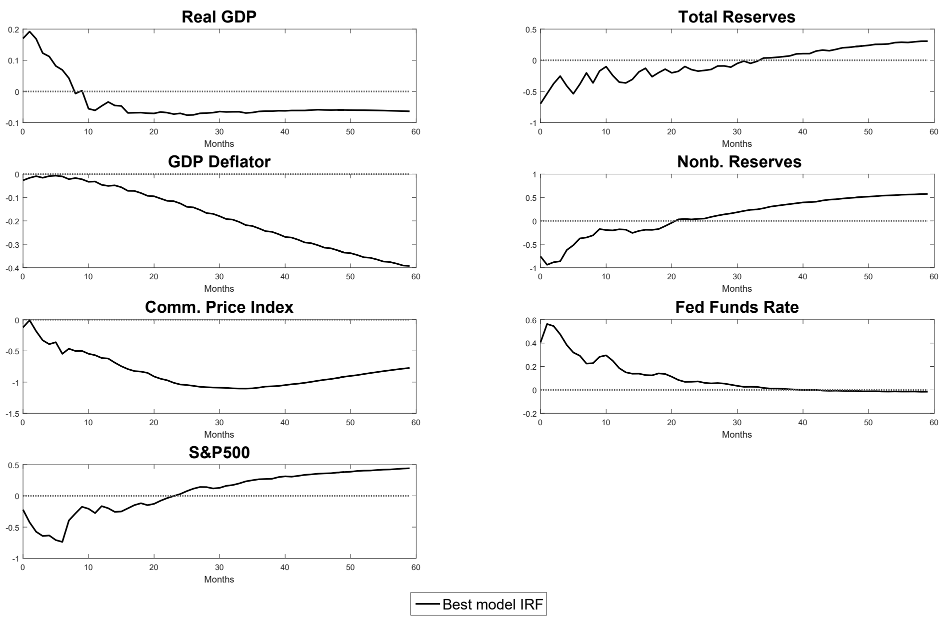


Figure 28: Impulse responses from model s^*. Identification Scheme I.

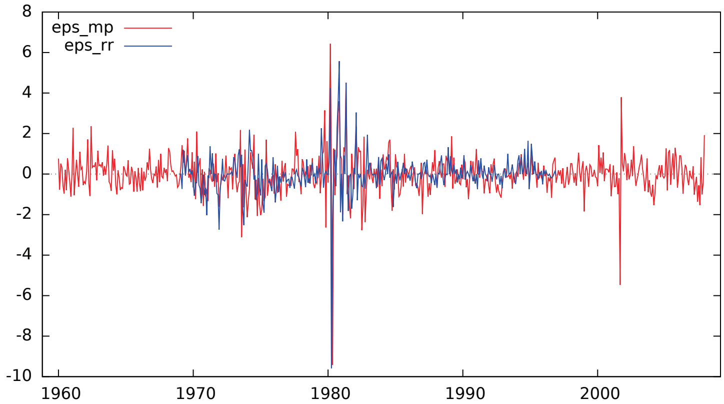


Figure 29: The series $\hat{\varepsilon}_{ts^*}^{mp}$ and $\hat{\varepsilon}_t^{rr}$, divided by their respective standard deviations. Scheme I.

As visual inspection of Figure 29 suggests, $\hat{\varepsilon}_{ts^*}^{mp}$ matches quite well with $\hat{\varepsilon}_t^{rr}$, especially in those times when $\hat{\varepsilon}_t^{rr}$ shows sizable swings, strengthening the argument to restrict attention to models highly correlated with the RR shocks.[13]

Zero and sign restrictions on the structural matrix

Figure 30 shows the impulse response of models in $\mathcal{S}^{high}$ determined from Scheme II (Restrictions ZR and SR2). A remarkable feature of the GDP response to a positive monetary policy shock is a qualitative similarity to the GDP response of Figure 27. This finding is worth noting especially in light of our Figure 20, and the arguments of Arias, Caldara, and Rubio-Ramírez (2015), where the main tendency of the GDP response is markedly negative.

Further, while in the baseline analysis of Scheme II we didn't find strong evidence for the price puzzle, based on Figure 30 we cannot claim that the existence of the price puzzle is not plausible. The asset price response is centered around zero. Thus, contrary to the results in the previous paragraph, we find no evidence of exogenous monetary policy shocks affecting asset prices. The rest of the responses exhibit high similarity to the baseline analysis, but concentrating only on hundred models leads to sharper conclusions.

In comparison to the figures describing $\mathcal{S}^{high}$ the impulse responses of model s^* in Scheme II are quite sensitive to the particular draw of the rotation matrix, therefore, we do not attempt to analyze the impulse responses this particular scheme. The (standardized) monetary policy shock series implied by Scheme II is shown in Figure 31, jointly with the (standardized) RR series. Similarly to the conclusions of the previous subsection, we can observe that the monetary policy shock series matches the RR series quite well especially in the high volatility phases.

13 Note, that an important caveat in interpreting results from model s^*, however, is that the model and the precise form of the impulse responses depend on the particular draw of the orthogonal matrix.

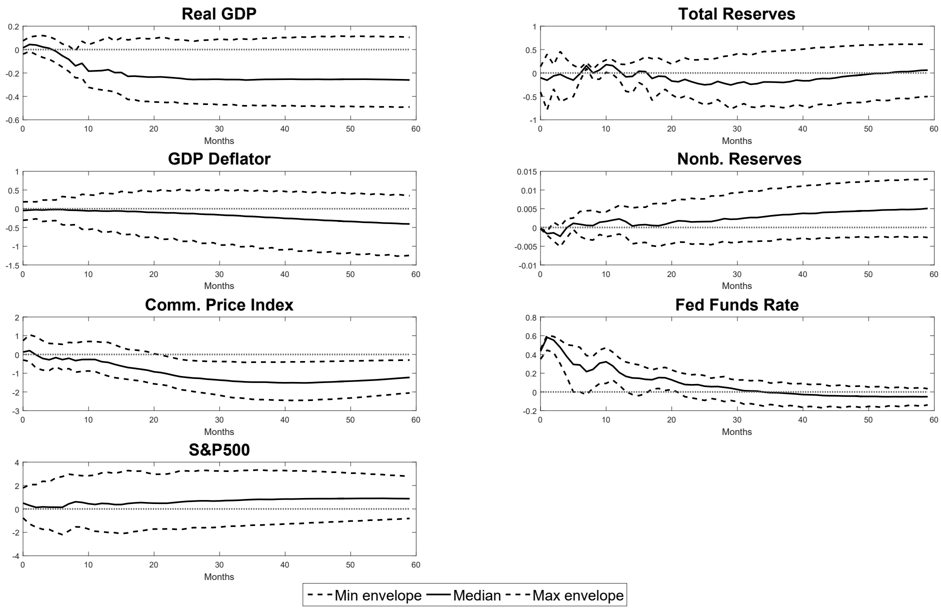


Figure 30: Range and median of models from $\mathcal{S}^{high}$. Identification Scheme II.

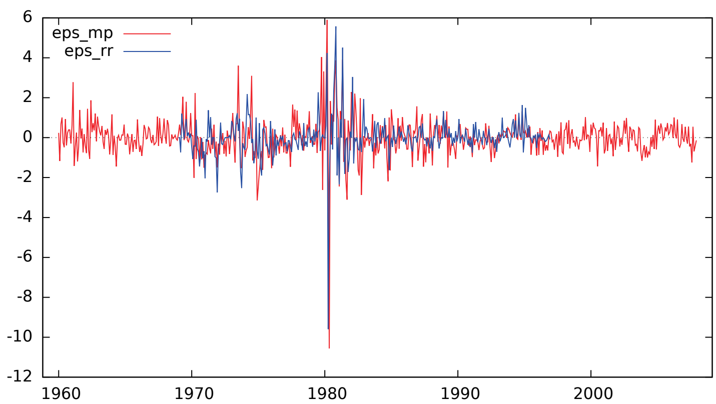


Figure 31: The series $\hat{\varepsilon}^{mp}_{ts^*}$ and $\hat{\varepsilon}^{rr}_{t}$, divided by their respective standard deviations. Scheme II.

4.7 DISCUSSION AND ROBUSTNESS

Our proposal of combining structural identifying assumptions with some benchmark series is similar in spirit to two different approaches proposed earlier in the literature. First, Faust, Swanson, and Wright (2004) identify monetary policy shocks by requiring the federal funds rate response to the policy shock to be equal to a certain benchmark response directly measured from federal funds futures data. Second, Mertens and Ravn (2013, 2014) utilize a narrative series as a proxy for the identified policy shock series and thereby provide additional identifying moment conditions. While both approaches solve the identification problem in a data-oriented way, their explicit aim is exact identification. In contrast, we believe that there can be several economic (structural) models compatible with the data. Sign restrictions, or non-exact identification in general, are adequate assumptions in line with this view as long as one carefully interprets the results.

From a methodological point of view, in the previous sections we argued that one should evaluate the effects of structural monetary policy shocks only within a subset of the set of *admissible* models. That is, in our analysis we started out by constraining the set of admissible models with set-identifying restrictions. An other possibility would be to simply generate several $\widehat{A}_0^{-1}$ matrices that do not satisfy any a priori structural assumptions and select the one that generates an estimated shock that has the highest correlation with the Romer and Romer (2004) series. However, there are several theoretical and empirical caveats for this approach. First, in the comparison exercise we might simply discover a high but spurious correlation between the benchmark and the estimated series. In particular, if the estimated series does not have a structural interpretation, then we cannot claim that the selected estimated series is, in fact, not a completely different shock. Second, the selected estimated series will highly depend on the particular random draw of $\widehat{A}_0^{-1}$. In our analysis it turns out that by not restricting the response of GDP deflator to be negative in the first several periods we immediately encounter the so-called price puzzle in a particularly severe form: the GDP deflator reacts positively to a monetary policy shock and its response remains positive in the long run. Further, by considering unrestricted models we could not significantly improve the "fit" to the RR shocks: the maximal correlation is around 0.42 compared with the approximately 0.39 of the set-identified specifications.

Our approach in this paper is explicitly frequentist. Thus, we cannot discriminate statistically between competing admissible models, and claiming that a particular admissible model is "most likely" (Inoue and Kilian, 2013) is not feasible. Nevertheless, our benchmarking approach extends the possibilities of empirical SVAR analysis in two directions. First, it restricts the set of models beyond what is achievable by the structural assumptions alone. This gives the possibility to sharply focus the evaluation of empirical and economic implications of the identified structural models. The empirical analysis then avoids the point made by Kilian and Murphy (2012) about the potential perils of analyzing summary statistics of the identified set when this set is too large. Second, while speaking of most likely models is not possible in a frequentist setting, if one accepts the postulated structural assumptions and the validity of the benchmark series, the models in $\mathcal{S}^{high}$ can be argued to be frequentist counterparts to the most likely models of Inoue and Kilian (2013).

We argued earlier that using the series proposed by Coibion and Gorodnichenko (2011) as the benchmark series instead of the RR series does not change any of our conclusions. This is despite the fact that Coibion (2012) arrives at distinct conclusions regarding the contribution of monetary policy shocks to fluctuations of real variables using these alternative narrative-based series. This might imply that our results are not driven by the "identifying power" of the benchmark shocks. However, if we use a completely uninformative simulated white noise series as benchmark series in place of the RR series, then the ordering of the models according to their ρ_s coefficients yield completely uninformative results: the impulse responses for models in $\mathcal{S}^{high}$ are similarly unstructured as the impulse responses for models in any other subsets of the ordering. This finding strengthens our results: investigating models relative to a benchmark indeed provides additional information.

Up until now we were silent about how more conventionally identified monetary policy shocks compare with the RR shock. As an example, using our baseline variable ordering[14] with a straight-forward recursive identification we obtained a correlation of $\rho = 0.3838$, which is very close to our best set-identified models. Values of similar magnitude for exact identification were reported also by Coibion (2012). These correlations are higher than the majority of correlations that we uncover in our analysis. On the one hand, this result might indicate that the applied sign restrictions need not be very successful assumptions to identify monetary policy shocks. However, the fact that models with the highest ρ_s have correlation around 0.40 implies that by a careful analysis of admissible models we can *improve* on other, especially exact identification procedures while at the same time settling on less restrictive identifying assumptions. We leave a comprehensive comparative evaluation of other exact identification schemes and empirical specifications for future research.[15]

4.8 CONCLUSIONS

How do asset prices respond to exogenous monetary policy shocks? We provide empirical results on this question. To this end, we augment the VAR specification of Uhlig (2005) with the S&P 500 Composite Index, and estimate the model on monthly US data from 1959 January to 2007 December. We use two identification schemes that result in set identification of the monetary policy shock. First, we use the sign restrictions put forth in Uhlig (2005) (Scheme I). Second, we utilize zero and sign restrictions on the structural matrix A_0 proposed by Arias, Caldara, and Rubio-Ramírez (2015) (Scheme II).

The SVAR impulse responses identified via Scheme I and Scheme II both point towards a mild positive asset price response to an increase in the monetary policy instrument. We argue that the resulting identified monetary policy shocks correlate only weakly with the monetary policy shock series of Romer and Romer (2004) that we view as a benchmark series for monetary policy shocks. We show that this finding matters greatly when analyzing (structural) impulse responses. In particular, the majority of admissible models yield impulse responses that vary widely in their shapes and impact magnitudes. Thus, we propose to restrict attention to those specifications that yield monetary policy shocks highly corre-

14 Real GDP, GDP deflator, commodity price index, stock price index, federal funds rate, non-borrowed reserves, total reserves.

15 Estimation results regarding the experiments in the above section are available upon request.

lated with the RR series, and we show that the impulse response analysis of these models leads to more robust and reliable conclusions.

Ultimately, we find evidence of asset prices responding mildly negatively (Scheme I) or with ambiguous sign (Scheme II) to a positive monetary policy shock. Besides the asset price response, we also uncover a mildly positive output response under both identification schemes when concentrating on models with the highest correlation with the Romer and Romer (2004) series. The expansionary effect of a "contractionary" monetary policy shock on output may seem surprising, Ramey (2016), however, points out that the consensus on "contractionary" monetary policy shocks indeed having contractionary effects easily disappears once one lifts the recursiveness identification assumption.

While the theme of comparing identified monetary policy shocks with the Romer and Romer (2004) series is quite specific to monetary policy applications on US data, our proposed approach of evaluating (set-) identified shocks against some benchmark series is more general and can be applied to a wide variety of empirical questions. As Kilian and Murphy (2012, p. 1186) point out: "If the set of admissible models remains large, the most useful exercise will be to search for the admissible model most favorable to each of the competing economic interpretations (...)". We believe that our approach complements and extends this advice, and is, thus, beneficial for empirical research. We have shown that comparing structurally (set-) identified shocks to a benchmark series can uncover by default hidden, but relevant and robust empirical conclusions.

4.A DATA

All data, except for the Romer and Romer (2004) series, is fully available to us from 1959 January to 2007 December. The data was gathered on 16.12.2015.

- ROMER AND ROMER MONETARY POLICY SHOCKS: The monthly series from 1/1/1969 to 12/1/1996 was obtained from Christina Romer's website: `http://eml.berkeley.edu/~cromer/#data`.

- REAL GDP: The monthly GDP was approximated with state-space methods using the quarterly GDP series GDPC1 and the monthly industrial production series INDPRO obtained from the FRED database. The interpolation method is described in Mönch and Uhlig (2005).

- GDP DEFLATOR: The monthly GDP deflator was approximated with state-space methods using the quarterly GDP deflator series (GDPDEF) and the monthly series CPIAUCSL (consumer price index for all urban consumers) and PPIFGS (producer price index for finished goods). All series were downloaded from the FRED database, and the interpolation method is described in Mönch and Uhlig (2005).

- COMMODITY PRICE INDEX: Daily data from the Commodity Research Bureau BLS spot index was obtained from Thomson Reuters' Datastream. Monthly observations were calculated as the averages of daily observations for each month.

- STOCK PRICE INDEX: Monthly observations of the S&P 500 composite index was obtained from the FRED MD project website (`https://research.stlouisfed.org/econ/mccracken/fred-databases/`) maintained by Michael W. McCracken. For the empirical analysis, the values were deflated by means of the GDP deflator.

- MEASURES OF MONETARY POLICY: Monthly series of the federal funds rate (FEDFUNDS), total reserves (TOTRESNS), and non-borrowed reserves (BOGNONBR) were obtained from the FRED database.

4.B ECONOMETRIC DETAILS

VAR model and impulse responses

Recall from the main text that we consider the following K-dimensional structural VAR,

$$A_0 y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \varepsilon_t, \tag{38}$$

where $y_t \in \mathbb{R}^K$, $\varepsilon_t \sim WN(0, I_K)$, $A_0, \ldots, A_p \in \mathbb{R}^{K \times K}$, and A_0, what we call the *structural matrix*, is assumed to be non-singular. In order to define a unique lag length we assume that $A_p \neq 0$. In the empirical application $K = 7$, and $p = 12$. In the above equation ε_t is the vector of *structural innovations*. The corresponding, estimable reduced form is

$$y_t = B_1 y_{t-1} + \cdots + B_p y_{t-p} + u_t, \tag{39}$$

with $B_i = A_0^{-1}A_i,\ i = 1, \dots, p$. For u_t, the vector of *reduced form innovations* the following holds: $A_0^{-1}\varepsilon_t = u_t \sim WN(0, \Sigma_u)$. Writing $B(z) = I_K - B_1 z - \cdots - B_p z^p$ we assume, that the reduced form is causal, that is, $\det(B(z)) \neq 0\ \forall |z| \leq 1$. Then the moving average representation of y_t exists and is given by (Brockwell and Davis, 1991, Th. 11.3.1, p. 418)

$$y_t = \sum_{j=0}^{\infty} \Phi_j u_{t-j} = \sum_{j=0}^{\infty} \Theta_j \varepsilon_{t-j}, \quad \Phi_0 = I_K, \tag{40}$$

where element (i, k) of the coefficient $\Theta_j = \Phi_j A_0^{-1}$ is interpreted as the reaction of the i-th variable on the k-th structural innovation at horizon j.

We estimate the parameters with ordinary least squares, obtaining $\widehat{B}_1, \dots, \widehat{B}_p$ and $\widehat{\Sigma}_u$. The corresponding estimate for the reduced form impulse response sequence $(\hat{\Phi}_j)_{j\geq 0}$ follows immediately. The starting point for estimating sign restricted impulse responses is the lower triangular Cholesky decomposition of $\widehat{\Sigma}_u = \widehat{A}_u^c \widehat{A}_u^{c\prime}$. Since $\widehat{\Sigma}_u$ is symmetric and positive definite, its Cholesky decomposition is unique (Meyer, 2000, p. 154). Note, that for any $K \times K$ orthogonal matrix Q with $Q'Q = QQ' = I_K$ it holds that

$$\widehat{\Sigma}_u = \widehat{A}_u^c Q Q' \widehat{A}_u^{c\prime}. \tag{41}$$

We are interested in finding those $\widehat{A}_u^c Q = \widehat{A}_0^{-1}(Q)$ matrices that imply structural form impulse response sequences $(\hat{\Theta}_j)_{j\geq 0} = (\hat{\Phi}_j \widehat{A}_0^{-1}(Q))_{j\geq 0}$ that satisfy the sign restrictions maintained in the main text. To this end we use the method proposed by Rubio-Ramírez, Waggoner, and Zha (2010). Let H be the length of the impulse response horizon that we wish to estimate, and let $J \leq H$ be the length of the horizon on which sign restrictions are imposed. Then the procedure can be described as follows:

1. Draw a matrix M with i.i.d. standard normal entries and perform the QR-decomposition of the matrix $M = QR$. The resulting matrix Q is orthogonal and has the uniform (or Haar) distribution on the group of orthogonal matrices.

2. Calculate the corresponding structural impulse response function $\{\widehat{\Theta}_j^Q\}_{j=0,\dots,H} = \{\hat{\Phi}_j \widehat{A}_0^{-1}(Q)\}_{j=0,\dots,H}$ and verify whether the formulated sign restrictions are fulfilled for $j = 1, \dots, J$. If so, keep $\left\{\widehat{\Theta}_j^Q\right\}_{j=0,\dots,H}$, otherwise discard it.

3. Repeat steps 1–2 until the set of retained structural impulse responses contains $S = 65000$ elements.

In the empirical application, we set $H = 60$, and $J = 4$ according to Restriction SR in the main text.

In order to simulate the set of impulse responses resulting from the sign and zero restrictions on A_0 we use the method proposed by Arias, Rubio-Ramírez, and Waggoner (2014). Let $\widehat{A}$ be $(\widehat{A}_u)^{-1}$, that is, the inverse of the Cholesky decomposition of $\widehat{\Sigma}_u$. Then, for our particular application, the algorithm can be described by the following steps:

1. Find a matrix $N_1 \in \mathbb{R}^{K\times(K-2)}$ with $N_1' N_1 = I_{K-2}$ such that $\widehat{A}_{[(K-1:\,K),\bullet]} N_1 = 0$, with $\widehat{A}_{[(K-1:\,K),\bullet]}$ denoting the $2 \times K$ matrix formed by the $(K-1)$-th and K-th rows of $\widehat{A}$.

2. Generate a vector $z \in \mathbb{R}^K$ with i.i.d. standard normally distributed entries and form the vector:
$$q = \frac{1}{||[N_1\ 0_{K\times 2}]z||}[N_1\ 0_{K\times 2}]z, \tag{42}$$
i.e., project the vector z on the space spanned by N_1 and normalize it to unit length.
3. Find a matrix $N_2 \in \mathbb{R}^{K\times(K-2)}$ with $N_2'N_2 = I_{K-2}$ such that $q'N_2 = 0$.
4. Draw a matrix $M \in \mathbb{R}^{(K-2)\times(K-2)}$ with i.i.d. standard normal entries and calculate the QR decomposition of N_2M, i.e.,
$$N_2M = [\tilde{Q}_1\ \ \tilde{Q}_2]\begin{bmatrix} R_1 \\ 0 \end{bmatrix}, \tag{43}$$
with $\tilde{Q}_1 \in \mathbb{R}^{K\times(K-2)}$.
5. Form the matrix $Q^+ = [q\ \tilde{Q}_1]$, calculate the corresponding structural matrix $\widehat{A}_0^{Q^+} = Q^{+\prime}\widehat{A}$, and verify whether the formulated sign restrictions are fulfilled. If so, keep $\widehat{A}_0^{Q^+}$, and the implied structural parameters, otherwise discard it. Note that by construction, the zero restrictions on the structural matrix hold for all draws.
6. Repeat these steps until the set of retained structural parameters contains $S = 65000$ elements.

Inference on the median target impulse response

The median target (MT) impulse response is the impulse response that is (element-wise) closest in terms of standardized squared distance to the pointwise median of the set of sign restricted impulse responses (termed here the *median curve*). Our implementation of the MT impulse response follows Fry and Pagan (2011). In particular, let $\widehat{\Theta}^s = \{\widehat{\Theta}_j^s\}_{j=0,\dots,H}$ be the set of structural impulse responses for $s = 1,\dots,S$ with $S = 65000$, estimated on horizons $0,\dots,H$. Denote the (element-wise) median curve as $\widehat{\Theta}_{med} = \{\widehat{\Theta}_{j,med}\}_{j=0,\dots,H}$. The median target impulse response is defined as:
$$\widehat{\Theta}^{MT} = \text{argmin}_{s=1,\dots,S} \sum_{r\in\mathcal{R}}\sum_{k\in\mathcal{K}}\sum_{j\in\mathcal{J}} \left(\frac{\widehat{\Theta}_{j,s}^{(r,k)} - \widehat{\Theta}_{j,med}^{(r,k)}}{\widehat{SD}_{r,k,j}}\right)^2, \tag{44}$$
with $\mathcal{R},\mathcal{K} \subseteq \{1,\dots,K\}$ and $\mathcal{J} \subseteq \{0,\dots,H\}$. Starting with $\overline{\widehat{\Theta}_{j,s}^{(r,k)}} = \frac{1}{S}\sum_{s=1}^S \widehat{\Theta}_{j,s}^{(r,k)}$, we can write $\widehat{SD}_{r,k,j} = \sqrt{\left(\frac{1}{S}\sum_{s=1}^S \widehat{\Theta}_{j,s}^{(r,k)} - \overline{\widehat{\Theta}_{j,s}^{(r,k)}}\right)^2}$. That is, $\widehat{SD}_{r,k,j}$ is, for each impulse response horizon j, each shock k, and each variable r the (pointwise) empirical standard deviation of the set of admissible impulse responses. In the empirical analysis we consider in Equation (44) only responses to the monetary policy shock, i.e., $\mathcal{K} = \{5\}$, and all the impulse responses, $\mathcal{R} = \{1,\dots,7\}$. The length of the estimated impulse response horizon is $H = 60$, and $\mathcal{J} = \{0,\dots,12\}$, i.e., impact period plus one year.

We denote by Q^{MT} the rotation that yields $\widehat{A}_0^{-1}(Q^{MT})$ corresponding to the median target model. While the median curve does not correspond to any particular structural model, it is possible to provide inference on the MT impulse response: an impulse response that corresponds to a well-defined, unique structural model.

Our bootstrap procedure for inference on the MT response follows Linnemann, Uhrin, and Wagner (2016). The algorithm can be described by the following steps:

1. Generate a bootstrap sample, $y_1^*, \ldots, y_T^*$ using the Kilian (1998) bootstrap.
2. Estimate the parameters of the VAR model using y_t^*, resulting in parameter estimates $\widehat{B}_1^*, \ldots, \widehat{B}_p^*$. Calculate the structural impulse response function using these parameter estimates and the *original* $\widehat{A}_0^{-1}(Q^{MT})$.
3. Verify whether the impulse response function from the previous item, $\{\widehat{\Theta}_j^{Q^{MT}*}\}_{j=0,\ldots,J}$, satisfies the formulated sign restrictions. If it does, keep it, otherwise discard it.
4. Repeat the above steps until 1000 impulse responses are retained and calculate pointwise bootstrap confidence bands as usual from these 1000 impulse responses.

4.C FURTHER RESULTS

In this Appendix we report results on comparison between the monetary policy shock series, and the Romer and Romer (2004) (RR) series , $\hat{\varepsilon}_t^{rr}$, where the monetary policy shocks are identified with Scheme II (Restrictions ZR and SR2). With a slight abuse of notation we denote the identified monetary policy shock series as $\hat{\varepsilon}_{ts}^{mp}$, similarly to the series obtained with identification Scheme I. In the context of the present Appendix no confusion should arise from this shorthand. The arguments of Section 4.5 remain valid, and are further strengthened by the evidence below.

Figure 32 displays the histogram of the 65000 obtained correlation coefficients $\rho_s = \widehat{\mathsf{Corr}}(\hat{\varepsilon}_{ts}^{mp}, \hat{\varepsilon}_t^{rr})$. As visual inspection suggests, the large majority of models are only midly correlated with the RR series. The average level of correlation is 0.1692, and the median is 0.1654. These values suggest, that the achievable correlation level using Scheme II are on average lower than those attained using Scheme I. However, the maximal correlation level (0.4) is very similar to the one obtained using Scheme I and the simple recusrive identification scheme, cf., Section 4.7). These three experiments suggest that there may be a cap on the achievable correlation level that is most likely influenced by the data and the model specification, but not the identification scheme.

Figures 33 – 35 display models with low ρ_s, medium ρ_s, and high ρ_s coefficients, respectively. As discussed in the main text, these figures show models that imply quite distinct impulse reponses both qualitatively and quantitatively. The lack of information content of Figure 34, displaying models with medium ρ_s coefficients, is even more pronounced than that of the corresponding figure in the main text, Figure 25.

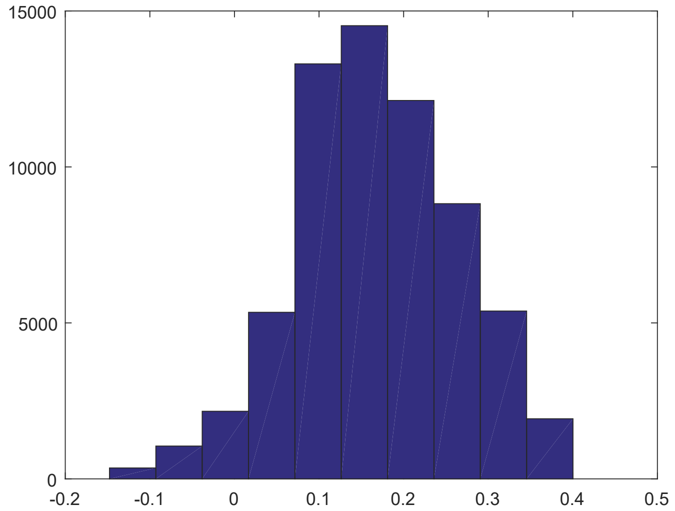


Figure 32: Histogram of correlations between $\hat{\varepsilon}_t^{mp}$ and $\hat{\varepsilon}_t^{rr}$. Scheme II.

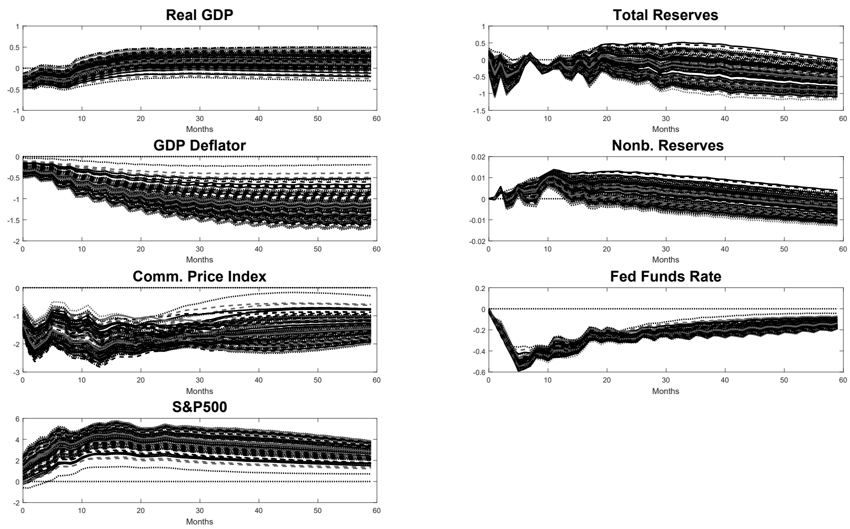


Figure 33: Impulse responses from models $s = 1, \ldots, 100$. Identification Scheme II.

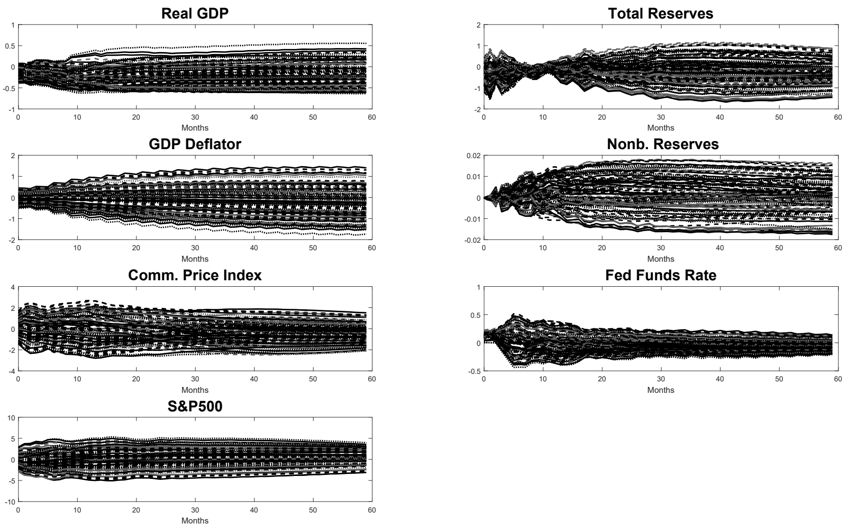


Figure 34: Impulse responses from models $s = 37450, \ldots, 37550$. Identification Scheme II.

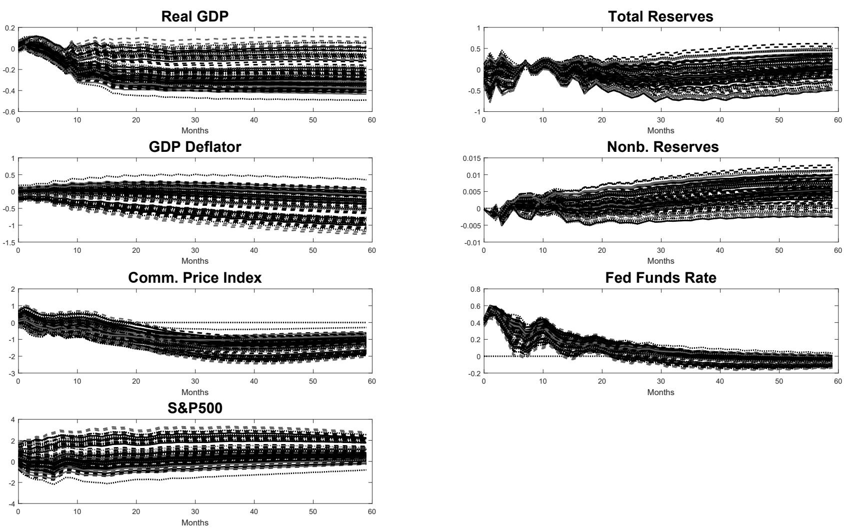


Figure 35: Impulse responses from models $s = 64900, \ldots, 65000$. Identification Scheme II.

5 CONCLUSIONS AND OUTLOOK

The contributing articles of the present thesis include self-contained conclusions and short abstracts. I have summarized the new scientific results in Section 1.3 of the Introduction chapter. In the present chapter I aim to highlight directions of future research based on the results of this thesis.

The focus of scrutinizing the identified monetary policy shock originated ultimately in the work resulting in Chapter 3, in cooperation with Martin Wagner. In this chapter we evaluated on a normative basis how "good" monetary policy shock estimates of a classical monetary policy SVAR are. Furthermore, we extended this evaluation to the classical SVAR augmented with a forward-looking variable, and a factor augmented VAR. According to the results, information-augmentation is neither methodologically necessary, nor explicitly beneficial. The simple, benchmark monetary policy model specification seems adequate. This result may be surprising in light of the recent surge of contributions discussing the problem of foresight in empirical studies, and the problem of non-fundamentalness. There are two explanations in my view. First, non-fundamentalness may indeed be not as prevalent in empirical monetary policy reserach as it is in empirical fiscal policy research. In order to evaluate this conjecture, there needs to be more empirical evidence. Second, it can be that the tools for detecting non-fundamentalness are not sufficiently developed, and the test results are not reliable. As the recent paper, of, e.g., Chen, Choi, and Escanciano (2017) shows, under non-Gaussianity the null hypothesis of fundamentalness is equivalent to the requirement that the reduced form innovations form a martingale difference sequence. This is ultimately the same as the null hypothesis of the specification testing literature that reads as follows. A regression model $y = f(x, \beta) + u$ for some fixed f is correctly specified if $E(u \mid x) = 0$ almost surely for some β_0 in the set of admissible coefficients β. See Bierens (2016) for further details. Thus, in my view, drawing on the ideas of the consistent specification testing literature seems to be a natural next step in developing tools to detect fundamentalness.[1]

Since the first discussions on the issue, I became increasingly convinced that the identified shocks themselves indeed deserve explicit attention that is often lacking in SVAR studies. These considerations, and intensive reflections on the nature of monetary policy shocks led to the idea of Chapter 4. The core benchmarking idea of this chapter turned out to be useful in the particular context of set identified monetary policy SVAR analysis. Through investigating only those structural models that imply monetary policy shocks highly correlated with benchmark measures we can discover by default hidden empirical conclusions. The method is easily adaptable to other empirical questions, and future research will investigate

1 Sahneh (2016) takes a first step in this direction. However, the proposed test ultimately falls short of being an extension of Bierens (1990) and Bierens and Ploberger (1997) to testing non-fundamentalness. This is despite the fact that the motivation explicitly stems from these papers.

if comparison with a benchmark shock is successful in shrinking the identified set also in other empirical specifications.

I have taken a first step in this direction, and the details are as follows. We have seen in Chapter 2 that the set of admissible impulse responses to the government spending shock is quite large, and the empirical conclusions based on them may not be unambiguous. In light of the results of Chapter 4, it seems natural to ask whether the benchmarking approach is successful in shrinking the identified set in this particular empirical model. In the context of fiscal policy, several candidate benchmark shocks exist in the literature. First, the military news series of Ramey (2011) can be viewed as a government spending shock series. Second, Romer and Romer (2010) construct narrative-based tax shock series using methodology similar to that of Romer and Romer (2004). I have used both of these benchmarks on the fiscal VAR studied in Chapter 2 explicitly identifying a tax shock via additional sign restrictions. However, I could not discover results different from what is described in the paper. My conjecture is, that this is because both candidate benchmark series are relatively low frequency in the sense that most of the elements of the series are zeros. Thus, benchmarking via investigating pair-wise *correlations* is not necessarily beneficial.

There are two potential next steps for the analyses carried out in this thesis. First, drawing on the ideas in Ludvigson, Ma, and Ng (2017) one can possibly further restrict the identified sets in Chapters 2 and 4. The key insight is to restrict certain episodes in the policy shock series. This amounts to postulating that the (fiscal or monetary) policy shock should or should not exceed a certain threshold for designated periods in the series. These restrictions could stem from stylized facts established from the benchmark policy shock series mentioned above. Second, it would be of independent interest to estimate a time varying fiscal SVAR with the sign restrictions designed for Chapter 2. The motivation for this step comes from the observation, that the impulse responses of Chapter 2 are on both sides of the zero line, but the shape of the bottom and top envelopes is quite similar. This suggests that there might be two shock propagation structures corresponding to, e.g., different episodes in the data. A time varying parameter estimation could explicitly verify whether this conjecture is correct.

The particularly negative result in Chapter 4 about the sign restricted median curve and the median target, in conjunction with the critiques in Kilian and Murphy (2012), points out that it is still an open issue how to adequately report and analyze set identified impulse responses. At a general level, the challenge of choosing between competing structural models in a frequentist setting remains still open for future research.

BIBLIOGRAPHY

Alessi, L., M. Barigozzi, and M. Capasso (2011): "Non-Fundamentalness in Structural Econometric Models: A Review," *International Statistical Review*, 79(1), 16–47.

Anderson, B. D. O., and M. Deistler (2008): "Properties of Zero-Free Transfer Function Matrices," *SICE Journal of Control, Measurement, and System Integration*, 1(4), 284–292.

Anderson, B. D. O., and M. Deistler (2009): "Properties of Zero-Free Spectral Matrices," *IEEE Transactions on Automatic Control*, 54(10), 2365–2375.

Arias, J. E., D. Caldara, and J. F. Rubio-Ramírez (2015): "The Systematic Component of Monetary Policy in SVARs: An Agnostic Identification Procedure," International Finance Discussion Papers nr. 1131, Board of Governors of the Federal Reserve System (U.S.).

Arias, J. E., J. F. Rubio-Ramírez, and D. F. Waggoner (2014): "Inference Based on SVARs Identified with Sign and Zero Restrictions: Theory and Applications," Working paper.

Assenmacher-Wesche, K., and S. Gerlach (2010): "Monetary Policy and Financial Imbalances: Facts and Fiction," *Economic Policy*, 25(63), 437–482.

Barro, R. J., and C. J. Redlick (2011): "Macroeconomic Effects from Government Purchases and Taxes," *The Quarterly Journal of Economics*, 126(1), 51–102.

Basu, S., J. G. Fernald, and M. S. Kimball (2006): "Are Technology Improvements Contractionary?," *American Economic Review*, 96(5), 1418–1448.

Beaudry, P., P. Fève, A. Guay, and F. Portier (2015): "When is Nonfundamentalness in VARs a Real Problem? An Application to News Shocks," NBER Working Papers 21466, National Bureau of Economic Research, Inc.

Bernanke, B., J. Boivin, and P. S. Eliasz (2005): "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach," *The Quarterly Journal of Economics*, 120(1), 387–422.

Bernanke, B. S., J. Boivin, and P. Eliasz (2004): "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach," NBER Working Papers nr. 10220, National Bureau of Economic Research.

Bernanke, B. S., and K. N. Kuttner (2005): "What Explains the Stock Market's Reaction to Federal Reserve Policy?," *The Journal of Finance*, 60(3), 1221–1257.

Bernanke, B. S., and I. Mihov (1998): "Measuring Monetary Policy," *The Quarterly Journal of Economics*, 113(3), 869–902.

Bierens, H. J. (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58(6), 1443–1458.

——— (2016): *Econometric Model Specification: Consistent Model Specification Tests and Semi-Nonparametric Modeling and Inference*. World Scientific Publishing Company.

Bierens, H. J., and W. Ploberger (1997): "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica*, 65(5), 1129–1151.

Bjørnland, H. C., and K. Leitemo (2009): "Identifying the Interdependence Between US Monetary Policy and the Stock Market," *Journal of Monetary Economics*, 56(2), 275–282.

Blanchard, O., and R. Perotti (2002): "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output," *The Quarterly Journal of Economics*, 117(4), 1329–1368.

Brockwell, P. J., and R. A. Davis (1991): *Time Series: Theory and Methods*. Springer, second edn.

Canova, F., and G. De Nicoló (2002): "Monetary Disturbances Matter for Business Fluctuations in the G-7," *Journal of Monetary Economics*, 49(6), 1131–1159.

Canova, F., and M. H. Sahneh (2016): "Are Small Scale VARs Useful for Business Cycle Analysis? Revisiting Non-Fundamentalness," Discussion Paper nr. 11041, Centre for Economic Policy Research.

Carlstrom, C. T., T. S. Fuerst, and M. Paustian (2009): "Monetary Policy Shocks, Choleski Identification, and DNK Models," *Journal of Monetary Economics*, 56(7), 1014–1021.

Castelnuovo, E., and S. Nisticò (2010): "Stock Market Conditions and Monetary Policy in a DSGE Model for the U. S.," *Journal of Economic Dynamics and Control*, 34(9), 1700–1731.

Chen, B., J. Choi, and J. C. Escanciano (2017): "Testing for Fundamental Vector Moving Average Representations," *Quantitative Economics*, 8(1), 149–180.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999): "Monetary Policy Shocks: What Have We Learned and to What End?," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and M. Woodford, vol. 1, chap. 2, pp. 65–148. Elsevier.

Coibion, O. (2012): "Are the Effects of Monetary Policy Shocks Big or Small?," *American Economic Journal: Macroeconomics*, 4(2), 1–32.

Coibion, O., and Y. Gorodnichenko (2011): "Monetary Policy, Trend Inflation, and the Great Moderation: An Alternative Interpretation," *American Economic Review*, 101(1), 341–70.

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

Faust, J., E. T. Swanson, and J. H. Wright (2004): "Identifying VARs Based on High Frequency Futures Data," *Journal of Monetary Economics*, 51(6), 1107–1131.

Fernald, J. G. (2014): "A Quarterly, Utilization-Adjusted Series on Total Factor Productivity," Working Paper Series nr. 2012-19, Federal Reserve Bank of San Francisco.

Fernández-Villaverde, J., J. F. Rubio-Ramírez, T. J. Sargent, and M. W. Watson (2007): "ABCs (and Ds) of Understanding VARs," *American Economic Review*, 97(3), 1021–1026.

Fisher, R. A. (1932): *Statistical Methods for Research Workers*. Olivier and Boyd.

Forni, M., and L. Gambetti (2014): "Sufficient Information in Structural VARs," *Journal of Monetary Economics*, 66(C), 124–136.

Forni, M., D. Giannone, M. Lippi, and L. Reichlin (2009): "Opening The Black Box: Structural Factor Models With Large Cross Sections," *Econometric Theory*, 25(5), 1319–1347.

Fry, R., and A. Pagan (2011): "Sign Restrictions in Structural Vector Autoregressions: A Critical Review," *Journal of Economic Literature*, 49(4), 938–60.

Galí, J., and L. Gambetti (2015): "The Effects of Monetary Policy on Stock Market Bubbles: Some Evidence," *American Economic Journal: Macroeconomics*, 7(1), 233–57.

Galí, J., J. D. López-Salido, and J. Vallés (2007): "Understanding the Effects of Government Spending on Consumption," *Journal of the European Economic Association*, 5(1), 227–270.

Gelper, S., and C. Croux (2007): "Multivariate Out-of-Sample Tests for Granger Causality," *Computational Statistics & Data Analysis*, 51(7), 3319–3329.

Genest, C., and B. Rémillard (2004): "Test of Independence and Randomness Based on the Empirical Copula Process," *Test*, 13(2), 335–369.

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

Giannone, D., and L. Reichlin (2006): "Does Information Help Recovering Structural Shocks from Past Observations?," *Journal of the European Economic Association*, 4(2-3), 455–465.

Gürkaynak, R. S., B. T. Sack, and E. P. Swanson (2007): "Market-Based Measures of Monetary Policy Expectations," *Journal of Business & Economic Statistics*, 25(2), 201–212.

Hall, R. E. (2009): "Reconciling Cyclical Movements in the Marginal Value of Time and the Marginal Product of Labor," *Journal of Political Economy*, 117(2), 281–323.

Hannan, E. J., and M. Deistler (1988): *The Statistical Theory of Linear Systems*. John Wiley & Sons.

Herwartz, H. (2016): "Structural VAR Modelling with Independent Shocks: An Analysis of Macroeconomic Dynamics in the Euro Area Based on a Novel Identification Scheme," Working paper.

Hong, Y., and Y.-J. Lee (2005): "Generalized Spectral Tests for Conditional Mean Models in Time Series with Conditional Heteroscedasticity of Unknown Form," *The Review of Economic Studies*, 72(2), 499–541.

Inoue, A., and L. Kilian (2013): "Inference on Impulse Response Functions in Structural VAR Models," *Journal of Econometrics*, 177(1), 1–13.

Justiniano, A., G. E. Primiceri, and A. Tambalotti (2010): "Investment Shocks and Business Cycles," *Journal of Monetary Economics*, 57(2), 132–145.

Kilian, L. (1998): "Small-Sample Confidence Intervals for Impulse Response Functions," *The Review of Economics and Statistics*, 80(2), 218–230.

Kilian, L., and H. Lütkepohl (to appear): *Structural Vector Autoregressive Analysis*. Cambridge University Press.

Kilian, L., and D. P. Murphy (2012): "Why Agnostic Sign Restrictions Are not Enough: Understanding the Dynamics of Oil Market VAR Models," *Journal of the European Economic Association*, 10(5), 1166–1188.

Kojadinovic, I., and J. Yan (2011): "Tests of Serial Independence for Continuous Multivariate Time Series Based on a Möbius Decomposition of the Independence Empirical Copula Process," *Annals of the Institute of Statistical Mathematics*, 63(2), 347–373.

Kuttner, K. N. (2001): "Monetary Policy Surprises and Interest Rates: Evidence From the Fed Funds Futures Market," *Journal of Monetary Economics*, 47(3), 523–544.

Lanne, M., M. Meitz, and P. Saikkonen (2017): "Identification and Estimation of non-Gaussian Structural Vector Autoregressions," *Journal of Econometrics*, 196(2), 288–304.

Leeper, E. M., T. B. Walker, and S.-C. S. Yang (2013): "Fiscal Foresight and Information Flows," *Econometrica*, 81(3), 1115–1145.

Linnemann, L., G. B. Uhrin, and M. Wagner (2016): "Government Spending Shocks and Labor Productivity," Discussion Paper nr. 9/2016, SFB 823.

Ludvigson, S. C., S. Ma, and S. Ng (2017): "Shock Restricted Structural Vector-Autoregressions," Working Paper.

Lütkepohl, H. (2005): *New Introduction to Multiple Time Series Analysis*. Springer.

——— (2012): "Fundamental Problems with Nonfundamental Shocks," Discussion Paper nr. 1230, DIW Berlin, German Institute for Economic Research.

Lütkepohl, H., and A. Netšunajev (2014): "Structural Vector Autoregressions with Smooth Transition in Variances: The Interaction between U. S. Monetary Policy and the Stock Market," Discussion Paper nr. 1388, DIW Berlin, German Institute for Economic Research.

McCracken, M. W., and S. Ng (2016): "FRED-MD: A Monthly Database for Macroeconomic Research," *Journal of Business & Economic Statistics*, 34(4), 574–589.

Mertens, K., and M. O. Ravn (2013): "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States," *American Economic Review*, 103(4), 1212–1247.

——— (2014): "A Reconciliation of SVAR and Narrative Estimates of Tax Multipliers," *Journal of Monetary Economics*, 68(S), S1–S19.

Meyer, C. D. (2000): *Matrix Analysis and Applied Linear Algebra*. Society for Industrial and Applied Mathematics.

Monacelli, T., and R. Perotti (2008): "Fiscal Policy, Wealth Effects, and Markups," NBER Working Papers nr. 14584, National Bureau of Economic Research, Inc.

Mönch, E., and H. Uhlig (2005): "Towards a Monthly Business Cycle Chronology for the Euro Area," *Journal of Business Cycle Measurement and Analysis*, 2005(1), 43–69.

Mountford, A., and H. Uhlig (2009): "What are the Effects of Fiscal Policy Shocks?," *Journal of Applied Econometrics*, 24(6), 960–992.

Neri, S. (2004): "Monetary Policy and Stock Prices: Theory and Evidence," Temi di discussione (Economic working papers) nr. 513, Bank of Italy, Economic Research and International Relations Area.

Pappa, E. (2009): "The Effects of Fiscal Shocks on Employment and the Real Wage," *International Economic Review*, 50(1), 217–244.

Patelis, A. D. (1997): "Stock Return Predictability and the Role of Monetary Policy," *The Journal of Finance*, 52(5), 1951–1972.

Perotti, R. (2007): "In Search of the Transmission Mechanism of Fiscal Policy," NBER Working Papers nr. 13143, National Bureau of Economic Research.

Piazzesi, M., and E. T. Swanson (2008): "Futures Prices as Risk-Adjusted Forecasts of Monetary Policy," *Journal of Monetary Economics*, 55(4), 677–691.

Ramey, V. A. (2011): "Identifying Government Spending Shocks: It's all in the Timing," *The Quarterly Journal of Economics*, 126(1), 1–50.

——— (2016): "Macroeconomic Shocks and Their Propagation," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and H. Uhlig, vol. 2, chap. 2, pp. 71–162. Elsevier.

Ravn, M. O., S. Schmitt-Grohé, and M. Uribe (2012): "Consumption, Government Spending, and the Real Exchange Rate," *Journal of Monetary Economics*, 59(3), 215–234.

Rigobon, R., and B. Sack (2004): "The Impact of Monetary Policy on Asset Prices," *Journal of Monetary Economics*, 51(8), 1553–1575.

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

——— (2010): "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks," *American Economic Review*, 100(3), 763–801.

Rosenblatt, M. (2000): *Gaussian and Non-Gaussian Linear Time Series and Random Fields*. Springer.

Rozanov, Y. A. (1967): *Stationary Random Processes*. Holden-Day.

Rubio-Ramírez, J. F., and J. Antolín-Díaz (2016): "Narrative Sign Restrictions for SVARs," Working Paper.

Rubio-Ramírez, J. F., D. F. Waggoner, and T. Zha (2010): "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference," *Review of Economic Studies*, 77(2), 665–696.

Rudebusch, G. D. (1998): "Do Measures of Monetary Policy in a VAR Make Sense?," *International Economic Review*, 39(4), 907–931.

Sahneh, M. H. (2015): "Are the Shocks Obtained from SVAR Fundamental?," Discussion Paper nr. 65126, MPRA.

Sahneh, M. H. (2016): "Testing for Non-Fundamentalness," Discussion Paper nr. 71924, MPRA.

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

——— (1992): "Interpreting the Macroeconomic Time Series Facts : The Effects of Monetary Policy," *European Economic Review*, 36(5), 975–1000.

Sims, E. R. (2012): "News, Non-Invertibility, and Structural VARs," in *DSGE Models in Macroeconomics: Estimation, Evaluation, and New Developments (Advances in Econometrics, Volume 28)*, ed. by N. Balke, F. Canova, F. Milani, and M. A. Wynne, pp. 81–135. Emerald Group Publishing.

Stock, J. H., and M. W. Watson (2016): "Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and H. Uhlig, vol. 2, chap. 8, pp. 415–525. Elsevier.

Strongin, S. (1995): "The Identification of Monetary Policy Disturbances Explaining the Liquidity Puzzle," *Journal of Monetary Economics*, 35(3), 463–497.

Thorbecke, W. (1997): "On Stock Market Returns and Monetary Policy," *The Journal of Finance*, 52(2), 635–654.

Thornton, D. L. (2014): "The Identification of the Response of Interest Rates to Monetary Policy Actions Using Market-Based Measures of Monetary Policy Shocks," *Oxford Economic Papers*, 66(1), 67.

Tippett, L. (1931): *The Method of Statistics*. Williams and Norgate.

Tsay, R. S. (1993): "Testing for Noninvertible Models with Applications," *Journal of Business & Economic Statistics*, 11(2), 225–233.

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

Uhrin, G. B., and H. Herwartz (2016): "Monetary Policy Shocks, Set-Identifying Restrictions, and Asset Prices: A Benchmarking Approach for Analyzing Set-Identified Models," CEGE Discussion Papers nr. 295, University of Göttingen.

Watson, M. W. (1986): "Vector Autoregressions and Cointegration," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. McFadden, vol. 4, chap. 47, pp. 2843–2915. Elsevier.