



RUHR

ECONOMIC PAPERS

Christoph Hanck

Jan Prüser

House Prices and Interest Rates – Bayesian Evidence from Germany

Imprint

Ruhr Economic Papers

Published by

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI)
Hohenzollernstr. 1-3, 45128 Essen, Germany

Editors

Prof. Dr. Thomas K. Bauer
RUB, Department of Economics, Empirical Economics
Phone: +49 (0) 234/3 22 83 41, e-mail: thomas.bauer@rub.de

Prof. Dr. Wolfgang Leininger
Technische Universität Dortmund, Department of Economic and Social Sciences
Economics – Microeconomics
Phone: +49 (0) 231/7 55-3297, e-mail: W.Leininger@wiso.uni-dortmund.de

Prof. Dr. Volker Clausen
University of Duisburg-Essen, Department of Economics
International Economics
Phone: +49 (0) 201/1 83-3655, e-mail: vclausen@vwl.uni-due.de

Prof. Dr. Roland Döhrn, Prof. Dr. Manuel Frondel, Prof. Dr. Jochen Kluve
RWI, Phone: +49 (0) 201/81 49-213, e-mail: presse@rwi-essen.de

Editorial Office

Sabine Weiler
RWI, Phone: +49 (0) 201/81 49-213, e-mail: sabine.weiler@rwi-essen.de

Ruhr Economic Papers #620

Responsible Editor: Volker Clausen

All rights reserved. Bochum, Dortmund, Duisburg, Essen, Germany, 2016

ISSN 1864-4872 (online) – ISBN 978-3-86788-722-9

The working papers published in the Series constitute work in progress circulated to stimulate discussion and critical comments. Views expressed represent exclusively the authors' own opinions and do not necessarily reflect those of the editors.

Ruhr Economic Papers #620

Christoph Hanck and Jan Prüser

House Prices and Interest Rates – Bayesian Evidence from Germany



Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über:
<http://dnb.d-nb.de> abrufbar.

<http://dx.doi.org/10.4419/86788722>
ISSN 1864-4872 (online)
ISBN 978-3-86788-722-9

Christoph Hanck and Jan Prüser¹

House Prices and Interest Rates – Bayesian Evidence from Germany

Abstract

This study uses a Bayesian VAR to demonstrate that the recent house price boom in Germany can be explained by falling interest rates and that higher interest rates are likely sufficient to stop the increase of German house prices. The latter suggests a potential drawback of the current monetary policy of the ECB. The BVAR's prior information shrinks the model parameters towards a parsimonious benchmark. We provide a simulation study to compare the frequentist properties of two useful strategies to select the informativeness of the prior. The study reveals that prior information helps to obtain more precise estimates of impulse response functions in small samples. To choose relevant control variables, we use a new Bayesian variable selection approach by Ding and Karlsson (2014). In addition to impulse responses and variance decompositions, we use a Bayesian conditional forecast to test the hypothetical effect of an increase of interest rates on house prices. This approach has the crucial advantage that it is invariant to the ordering of the variables.

JEL Classification: C11, C32, C53, E37, E43

Keywords: Bayesian VAR; shrinkage; house prices

June 2016

¹ Christoph Hanck, UDE; Jan Prüser, RGS Econ and UDE. - We thank Torsten Schmidt for valuable comments. Hanck gratefully acknowledges the support of the German Research Foundation (DFG) through the project HA 6766/2-1. - All correspondence to: Jan Prüser, RWI, Hohenzollernstraße 1-3, 45128 Essen, Germany, e-mail: jan.prueser@rgs-econ.de

1. Introduction

The housing crises in the United States (US) and in Spain have forcefully demonstrated that real estate price fluctuations have a substantial impact on financial stability and real economic activity. After almost two decades of stagnation, German house prices increase at an accelerated rate since 2010, implying similar macroeconomic risks. One likely supporting factor is Germany’s robust economic recovery after the crisis. Furthermore, the expansionary monetary policy of the European Central Bank (ECB) has lead to historically low interest rates. The resulting favorable lending conditions and investors searching for a safe haven investing in German real estate may also have contributed to the increase. Low interest rates might allow for “unsustainable” house price growth whereas a more restrictive interest rate policy might prevent borrowers from bidding up house prices, resulting in a boom. How successfully interest rates can curb a boom depends on how responsive house prices are to interest rates. Our main findings are that the increase in house prices can be better explained by falling interest rates than by other fundamentals of the economy and that increasing interest rates likely are sufficient to stop the increase of house prices. These results are interesting for the recent debate whether central banks should use interest rate policy to bring down house prices, see e.g. Jorda et al. (2015). Furthermore, the results suggest a potential drawback of the current monetary policy of the ECB. As evidenced by the development of bubbles in countries like Spain and Ireland, an overly expansionary monetary policy is not suitable for all member countries of a monetary union.

To account for the interrelations over time between house prices, interest rates and other macroeconomic variables a vector autoregressive model (VAR) is a useful choice. However, in small samples the rich parametrization of VAR models may come at the cost of overfitting the data, possibly leading to imprecise inference and inaccurate forecasts. To avoid such overfitting we estimate a Bayesian VAR (BVAR). The BVAR can use prior information to shrink the model parameters towards a parsimonious benchmark, potentially leading to more precise estimates. While overly strong shrinkage does not let the data “speak”, too little shrinkage does not avoid the problem of overfitting. This raises the question as to how to select the appropriate amount of shrinkage. In many areas, researchers do not have any prior information on how to select the amount of shrinkage. Many existing studies therefore use benchmark values for the prior information, such as Sims and Zha (2006). This, however, does not necessarily yield the optimal amount of shrinkage for all relevant cases. Especially in cases like in our empirical application, where the sample size is small, the choice of values for the prior information has a crucial influence on the posterior. We investigate two useful strategies recently proposed in the literature, which make the choice of values for the prior information more “objective” and may lead to a more appropriate amount of shrinkage for the empirical case at hand. The first one, related to a recent proposal of Giannone et al. (2015), selects the amount of shrinkage by maximizing the marginal likelihood. The second selects it by minimizing out-of-sample forecast errors (see Robertson and Tallman, 1999).¹ We are the first, to the

¹A third strategy would be to control for overfitting by choosing the amount of shrinkage that yields a

best of our knowledge, to conduct a simulation study to compare the frequentist properties of the two approaches. Our results suggest that selecting the amount of shrinkage by maximizing the marginal likelihood outperforms the approach of selecting the amount of shrinkage by minimizing out-of-sample forecast errors in a mean squared error (MSE) sense, plausibly because it makes more efficient use of the data by using the full sample. Moreover, the simulation study reveals that prior information can help to obtain more precise estimates of impulse response functions in small samples.

The insights of the simulation study are then used to select the amount of shrinkage in the empirical analysis of the link between interest rates and house prices. Economic theory, while suggesting that many economic variables may have an impact on real estate prices, provides no clear guidance as to which variables are to be included in a model explaining real estate prices. Where variable selection in previous VAR studies is to some extent arbitrary, we are among the first to use a new Bayesian approach of variable selection by Ding and Karlsson (2014) to choose the relevant control variables. We find that it delivers plausible results, which are in line with previous studies. This sheds light on which variables besides the interest rate are important for explaining German house prices. In addition to impulse responses and variance decompositions, we use a Bayesian conditional forecast to test the hypothetical effect of an increase of interest rates on house prices. This approach has the advantage that it shows the likely implications of an increase in one variable on other variables and is invariant to the ordering of the variables.

The remainder of this paper is organized as follows. Section 2 lays out our empirical framework. Section 3 describes the setups for the simulation and presents findings. Section 4 describes the data and discusses the empirical findings. The last section concludes.

2. Methodology

2.1. Bayesian Vector Autoregression

It has recently become popular to use prior information in VARs to overcome the problem of overfitting (e.g., Koop and Korobilis, 2010). Since Sims' (1980) critique of the "credible restrictions" used by large macroeconomic models, VAR models have become the workhorse tool for macroeconomic forecasting and also a popular tool for policy analysis. In comparison to theory-based Dynamic Stochastic General Equilibrium models (see, e.g., Milani, 2012), VAR models impose fewer restrictions on the data and thereby allow the data to speak more freely.

Following Karlsson (2013), the VAR model with m variables can be written as

desired in-sample fit, which seems to be promising for large BVARs. See e.g. Bańbura et al. (2010).

$$\begin{aligned}\mathbf{y}'_t &= \sum_{i=1}^p \mathbf{y}'_{t-i} \mathbf{A}_i + \mathbf{x}'_t \mathbf{C} + \mathbf{u}'_t \\ &= \mathbf{z}'_t \mathbf{\Gamma} + \mathbf{u}'_t,\end{aligned}\tag{1}$$

with \mathbf{y}_t a $m \times 1$ vector containing the variables at time t , \mathbf{A}_i is a $m \times m$ matrix of parameters, \mathbf{x}_t a vector of d deterministic variables at time t , \mathbf{C} is a $d \times m$ parameter matrix, $\mathbf{z}'_t = (\mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p}, \mathbf{x}'_t)$ a $1 \times k$ vector with $k = mp + d$, $\mathbf{\Gamma} = (\mathbf{A}'_1, \dots, \mathbf{A}'_p, \mathbf{C})'$ a $k \times m$ matrix and $\mathbf{u}_t \sim N(\mathbf{0}, \mathbf{\Psi})$ are normally distributed errors.

Due to their rich parametrization, VAR models typically provide a good fit to the data. From a frequentist perspective, such “overfitting” may lead to imprecise inference and a high variance for out-of-sample forecasts. Litterman’s (1979) widely used “Minnesota prior” provides a potential remedy for this problem. The Minnesota prior embodies a set of prior beliefs which shrink the parameters towards a stylized representation of non-stationary macroeconomic time series data, leading to potentially more precise estimates. This provides a compromise between overfitting and using “incredible restrictions”.

The model (1) can be written more compactly by stacking the data as a multivariate regression model (see, e.g., Zellner, 1971, for a detailed discussion of such models)

$$\mathbf{Y} = \mathbf{Z}\mathbf{\Gamma} + \mathbf{U},\tag{2}$$

where $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_T)'$ and \mathbf{U} are $T \times m$ matrices with T the number of observations minus the number of lags p and $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_T)'$ is a $T \times k$ matrix. Under the normality assumption the likelihood is

$$f(\mathbf{Y}|\mathbf{\Gamma}, \mathbf{\Psi}) = 2\pi^{-mT/2} |\mathbf{\Psi}|^{-T/2} \exp \left\{ -\frac{1}{2} \text{tr} [\mathbf{\Psi}^{-1}(\mathbf{Y} - \mathbf{Z}\mathbf{\Gamma})'(\mathbf{Y} - \mathbf{Z}\mathbf{\Gamma})] \right\}.\tag{3}$$

In order to shrink the parameters, we next specify the prior beliefs for the parameters $\mathbf{\Gamma}$ and $\mathbf{\Psi}$.

2.1.1. The Normal-Wishart Prior

The Normal-Wishart prior is a natural conjugate prior for normal multivariate regressions, allowing for computationally convenient system estimation. To see conjugacy it is helpful to rewrite the likelihood function in the form of a Normal-Wishart distribution when considered as a function of $\mathbf{\Psi}$ and $\mathbf{\Gamma}$. Define the OLS estimate $\hat{\mathbf{\Gamma}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$ and $\mathbf{S} = (\mathbf{Y} - \mathbf{Z}\hat{\mathbf{\Gamma}})'(\mathbf{Y} - \mathbf{Z}\hat{\mathbf{\Gamma}})$. Adding and subtracting $\mathbf{Z}\hat{\mathbf{\Gamma}}$ in the exponential in (3) yields

$$\begin{aligned}f(\mathbf{Y}|\mathbf{\Gamma}, \mathbf{\Psi}) &\propto |\mathbf{\Psi}|^{-T/2} \exp \left\{ -\frac{1}{2} \text{tr} [\mathbf{\Psi}^{-1}\mathbf{S}] \right\} \\ &\times \exp \left\{ -\frac{1}{2} \text{tr} [\mathbf{\Psi}^{-1}(\mathbf{\Gamma} - \hat{\mathbf{\Gamma}})' \mathbf{Z}' \mathbf{Z} (\mathbf{\Gamma} - \hat{\mathbf{\Gamma}})] \right\}.\end{aligned}\tag{4}$$

The prior for $\mathbf{\Gamma}$ and $\mathbf{\Psi}$ can be specified in form of a conditional matrix-variate normal

distribution² (MN), where $\mathbf{X} \sim MN(\mathbf{M}, \mathbf{Q}, \mathbf{P})$ iff $\text{vec}(\mathbf{X}) \sim N(\text{vec}(\mathbf{M}), \mathbf{Q} \otimes \mathbf{P})$ and an inverse Wishart distribution $iW(\mathbf{\Sigma}, v)$, where $\mathbf{\Sigma}$ is the scale matrix and v denotes the degrees of freedom,

$$\mathbf{\Gamma} | \mathbf{\Psi} \sim MN(\underline{\mathbf{\Gamma}}, \mathbf{\Psi}, \underline{\mathbf{\Omega}}), \quad (5)$$

$$\mathbf{\Psi} \sim iW(\underline{\mathbf{S}}, \underline{v}). \quad (6)$$

The posterior using the Normal-Wishart prior can be shown to be

$$\begin{aligned} f(\mathbf{\Gamma}, \mathbf{\Psi} | \mathbf{Y}) &\propto |\mathbf{\Psi}|^{(\bar{v}+m+k+1)/2} \\ &\times \exp \left\{ -\frac{1}{2} \text{tr} \left[\mathbf{\Psi}^{-1} (\mathbf{\Gamma} - \bar{\mathbf{\Gamma}})' \bar{\mathbf{\Omega}}^{-1} (\mathbf{\Gamma} - \bar{\mathbf{\Gamma}}) \right] \right\} \\ &\times \exp \left\{ -\frac{1}{2} \text{tr} \left[\mathbf{\Psi}^{-1} \bar{\mathbf{S}} \right] \right\}, \end{aligned} \quad (7)$$

where $\bar{\mathbf{\Omega}} = (\underline{\mathbf{\Omega}}^{-1} + \mathbf{Z}'\mathbf{Z})^{-1}$, $\bar{\mathbf{\Gamma}} = \bar{\mathbf{\Omega}}(\underline{\mathbf{\Omega}}^{-1}\underline{\mathbf{\Gamma}} + \mathbf{Z}'\mathbf{Z}\hat{\mathbf{\Gamma}})$, $\bar{v} = T + \underline{v}$ and $\bar{\mathbf{S}} = \underline{\mathbf{S}} + \mathbf{S} + (\underline{\mathbf{\Gamma}} - \hat{\mathbf{\Gamma}})'(\underline{\mathbf{\Omega}} + (\mathbf{Z}'\mathbf{Z})^{-1})^{-1}(\underline{\mathbf{\Gamma}} - \hat{\mathbf{\Gamma}})$. And by the definition of the Normal-Wishart distribution we obtain the conditional and marginal posterior distributions

$$\mathbf{\Gamma} | \mathbf{Y}, \mathbf{\Psi} \sim MN(\bar{\mathbf{\Gamma}}, \mathbf{\Psi}, \bar{\mathbf{\Omega}}) \quad (8)$$

and

$$\mathbf{\Psi} | \mathbf{Y} \sim iW(\bar{\mathbf{S}}, \bar{v}). \quad (9)$$

This shows that the mean of the conditional posterior distribution $\bar{\mathbf{\Gamma}}$ is a weighted average of $\underline{\mathbf{\Gamma}}$ and $\hat{\mathbf{\Gamma}}$. The weights are given by the inverse of the variance-covariance matrix of the prior, $\underline{\mathbf{\Omega}}^{-1}$, and by $\mathbf{Z}'\mathbf{Z}$. Overall, smaller values of $\underline{\mathbf{\Omega}}$ imply that the prior receives a higher weight relative to $\hat{\mathbf{\Gamma}}$. Hence, if the posterior mean is not to be strongly influenced by the prior, one assigns a high variance to the prior.

2.1.2. The Prior Beliefs

Litterman's prior formulation is based on the belief that many macroeconomic variables are well characterized by unit-root processes. The prior mean $\underline{\mathbf{\Gamma}}$ is then set to

$$\underline{\mathbf{\Gamma}}_{ij} = \begin{cases} 1, & \text{first own lag, } i = j \\ 0, & i \neq j \end{cases}. \quad (10)$$

The specification of the prior variances for the Normal-Wishart prior is a bit more tricky.

²An excellent summary of matrix-variate distributions can be found in the appendix of Karlsson (2013).

Integrating Ψ out of $MN(\underline{\Gamma}, \Psi, \underline{\Omega})$ shows that the marginal prior distribution of Γ is matrixvariate t with $km \times km$ variance-covariance matrix $V(\gamma) = \frac{1}{\underline{v}-m-1} \underline{\mathbf{S}} \otimes \underline{\Omega}$, where $\gamma = \text{vec}(\Gamma)$. This implies that the variance-covariance matrix of one equation must be proportional to the variance-covariance matrix of another equation.³ Therefore, it is not possible to use the prior standard deviations as in the original Minnesota prior, which apply a harder shrinkage on lags of variables other than the dependent variable. Instead, we use the standard prior employed in Bayesian VARs due to Sims and Zha (1998). With $V(\gamma_j) = \frac{\tilde{s}_j}{\underline{v}-m-1} \underline{\Omega}$, where γ_j with $j = 1, \dots, m$ is a $k \times 1$ vector such that $\gamma = (\gamma'_1, \dots, \gamma'_m)'$, the diagonal elements of $\underline{\Omega}$ are set to

$$\omega_{ii} = \begin{cases} (\lambda_0 \lambda_1)^2 / (l^{\lambda_3} s_r)^2, & \text{for lag } l \text{ of variable } r, i = (l-1)m + r \\ (\lambda_0 \lambda_4)^2, & \text{for the constant, } i = mp + 1 \\ (\lambda_0 \lambda_5)^2, & \text{for the deterministic variables, } i = mp + 2, \dots, k \end{cases} \quad (11)$$

and $\tilde{s}_j = (\underline{v} - m - 1)s_j^2$ to approximate the variances of the Minnesota prior. Thus, the prior parameter matrix of the inverse Wishart is

$$\underline{\mathbf{S}} = (\underline{v} - m - 1) \text{diag} \left(\left(\frac{s_1}{\lambda_0} \right)^2, \dots, \left(\frac{s_m}{\lambda_0} \right)^2 \right). \quad (12)$$

To ensure that the prior variance exists \underline{v} is set to $m + 2$ (Kadiyala and Karlsson, 1997). The terms s_1^2 to s_m^2 are OLS residual variances of a univariate autoregression of order \tilde{p} for each series, used to correct for the different scales of the series. The λ_s are hyperparameters set by the researcher, which control the tightness of the prior. The hyperparameter λ_0 controls the overall scale of the prior variance-covariance matrix relative to the estimated covariance scale estimated from the univariate autoregressive OLS models, λ_1 controls how tightly the model complies to the random walk prior, λ_3 controls the degree to which coefficients on lags higher than one are likely to be zero and λ_4 and λ_5 control the degree to which coefficients of deterministic variables are likely to be zero (Canova, 2007).

In addition to the Minnesota prior, the Bayesian literature uses other priors to introduce beliefs about unit roots and cointegration. Rather than using knife-edge tests for unit roots or cointegration, such probabilistic statements avoid the need for pre-testing, which may produce mistaken inferences about the trend properties of the time series. In the following, two priors are introduced. These proceed by adding dummy observations to the data, consistent with Theil's mixed estimation method (Theil and Goldberger, 1960).

The sum of coefficients prior proposed by Doan et al. (1984) expresses the prior belief that the sum of coefficients on own lags is one and the sum of coefficients on the lags of each of the other variables is zero, i.e., $\sum_i \mathbf{A}_i = \mathbf{I}_m$, or similarly the belief that the recent

³Note that other priors, e.g., the independent-Wishart prior, do not share this possibly restrictive feature. However, more flexible priors imply a higher computational burden. Hence, we follow the preferred choice of Kadiyala and Karlsson (1997), the normal-Wishart prior.

average of lagged values of a variable $\bar{y}_{0i} = \frac{1}{p} \sum_{n=1-p}^0 y_{n,i}$ is likely to be a good forecast at the beginning of the sample. The prior beliefs are introduced by adding m rows of dummy observations to \mathbf{Y} and \mathbf{Z} . Let $y_{i,j}$ with $i, j = 1, \dots, m$ represent the elements of the first m rows of \mathbf{Y} and $z_{i,s}$ with $i = 1, \dots, m; s = 1, \dots, k$ represent the elements of the first m rows of \mathbf{Z} . Then

$$y_{i,j} = \begin{cases} \mu_5 \bar{y}_{0i}, & i = j \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

and

$$z_{i,s} = \begin{cases} \mu_5 \bar{y}_{0i}, & i = j, s \leq mp \\ 0, & \text{otherwise} \end{cases}. \quad (14)$$

When $\mu_5 \rightarrow \infty$ the model can be expressed entirely in terms of first differences.

The dummy initial observation prior introduced by Sims (1993) on the other hand allows for the possibility of cointegration between the variables. It adds one additional row to the system. Let now $y_{i,m+1}$ with $i = 1, \dots, m$ represent the elements of row $m+1$ of \mathbf{Y} and $z_{s,m+1}$ with $s = 1, \dots, k$ represent the elements of row $m+1$ of \mathbf{Z} . The additional row can then be written as

$$y_{i,m+1} = \mu_6 \bar{y}_{0i} \quad (15)$$

and

$$z_{s,m+1} = \begin{cases} \mu_6 \bar{y}_{0i}, & s \leq mp \\ \mu_6, & s = mp + 1, \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

with

$$i = \begin{cases} s, & s \leq m \\ s - m, & m < s \leq 2m \\ \vdots & \\ s - m(p-1), & m(p-1) < s \leq mp \end{cases}. \quad (17)$$

As $\mu_6 \rightarrow \infty$ the prior implies that the variables in the model are either all stationary or that the system is characterized by the presence of an unspecified number of unit roots without drift. Hence, cointegration is not ruled out in this limit. Taken together, the two prior allow to favour unit roots and cointegration (Sims and Zha, 1998).

2.2. Hyperparameter and Variable Selection

We now discuss the choice of hyperparameters of the prior. One possibility to select the hyperparameters is to evaluate the forecast performance of the model over a range of hyperparameters as suggested by Robertson and Tallman (1999). Examples for this approach include Summers (2001), Brandt and Freeman (2002), Wright (2009) and Litterman (1986). This approach is attractive as a good forecasting model likely is a model which does not overfit the data. However, it needs to be determined which forecast horizon and how many forecasts should be chosen to select the hyperparameters.

A second approach, related to Giannone et al. (2015), chooses the hyperparameters so as to maximize the marginal likelihood. The log-marginal likelihood of a given model M_i can be decomposed into the sum of one-step-ahead predictive scores

$$\ln m(\mathbf{Y}|\Xi, M_i) = \sum_{t=1}^T \ln m(\mathbf{y}_t|\{\mathbf{y}_s\}_{s=0}^{t-1}, \Xi, M_i), \quad (18)$$

with $m(\mathbf{Y}|\Xi, M_i) = \int f(\mathbf{Y}|\boldsymbol{\theta}_i, \Xi, M_i)\pi(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i$ for a parameter vector $\boldsymbol{\theta}_i$, a given model M_i , $\Xi = (\lambda_0, \lambda_1, \lambda_3, \lambda_4, \lambda_5, \mu_5, \mu_6)$. Whenever the distribution assigns a low density to the observation \mathbf{y}_t the predictive score is small. Thus, maximizing the marginal likelihood corresponds to maximizing the sum of one-step-ahead predictive scores, which are related to the forecasting ability of the model (Geweke et al., 2011). Estimating the hyperparameters by maximizing the marginal likelihood is an Empirical Bayes method, which has a frequentist flavor.⁴ At the same time, the marginal likelihood plays an important role for Bayesian model comparison. Given two competing models M_1 and M_2 , the posterior probabilities for each model are

$$\pi(M_i|\Xi, \mathbf{Y}) = \frac{m(\mathbf{Y}|\Xi, M_i)\pi(M_i)}{m(\mathbf{Y}|\Xi, M_2)\pi(M_1) + m(\mathbf{Y}|\Xi, M_2)\pi(M_2)}. \quad (19)$$

The posterior odds then are

$$\frac{\pi(M_1|\Xi, \mathbf{Y})}{\pi(M_2|\Xi, \mathbf{Y})} = \frac{m(\mathbf{Y}|\Xi, M_1)}{m(\mathbf{Y}|\Xi, M_2)} \times \frac{\pi(M_1)}{\pi(M_2)}. \quad (20)$$

The model choice can be based on the posterior odds, but it is more common to use the Bayes factors $BF_{1,2} = m(\mathbf{Y}|\Xi, M_1)/m(\mathbf{Y}|\Xi, M_2)$ directly for model comparison.

For the Normal-Wishart prior the marginal likelihood exists as a closed-form expression. It can be shown that the marginal likelihood has a matricvariate-t distribution $Mt(\mathbf{M}, \mathbf{P}, \mathbf{Q}, v)$, where \mathbf{M} is the mean matrix, \mathbf{Q} and \mathbf{P} are positive definite symmetric

⁴This is seen critically by pure Bayesians, as Carlin and Louis (2000) note: “Strictly speaking, empirical estimation of the prior is a violation of Bayesian philosophy: the subsequent prior-to-posterior updating ...‘would use the data twice’ (first in the prior, and again in the likelihood). The resulting inferences would thus be ‘overconfident’.” For a classical overview of the advantages and limitations of empirical Bayes methods from a Bayesian perspective, see Berger (1985).

scale matrices and v is the degree of freedom,

$$\mathbf{Y} \sim Mt(\mathbf{Z}\underline{\mathbf{\Gamma}}, (\mathbf{I}_T + \mathbf{Z}\underline{\mathbf{\Omega}}\mathbf{Z}')^{-1}, \underline{\mathbf{S}}, v). \quad (21)$$

As shown by Giannone et al. (2015) (21) can be rewritten into an expression reflecting the common trade-off between in-sample fit and model complexity of information criteria, emphasizing why maximizing the marginal likelihood may be expected to lead to a useful amount of shrinkage.⁵ Examples of choosing the hyperparameters by maximizing the marginal likelihood include Carriero et al. (2015), Brandt and Freeman (2009) and Deryugiana and Ponomarenko (2013).

What is more, the “marginalized” marginal likelihood can be used for variable selection: since multivariate likelihoods are not comparable when different dependent variables are included in the system, Ding and Karlsson (2014) propose marginalizing out the variables that are not of primary interest and then using the marginalized marginal likelihood for model selection. Their approach is based on the fit of a core subset of variables of interest that are always included in the model. Hence, one single model can be selected according to the marginalized marginal likelihood out of different VAR models which all include the variables of interest in addition to different combinations of other potentially relevant variables. Ding and Karlsson (2014) demonstrate in a simulation study that the marginalized marginal likelihood provides a sharp discrimination between models and variables, even in small samples with 100 observations.

For the Normal-Wishart prior there exists a closed-form expression for the marginalized marginal likelihood. Let w.l.o.g. $\mathbf{P} = (\mathbf{I}_q, \mathbf{0}_{q \times (m-q)})'$ be a $m \times q$ selection matrix such that $\mathbf{Y}_1 = \mathbf{Y}\mathbf{P}$, the matrix of the variables of interest. Then, for $\underline{\mathbf{\Gamma}}_1 = \underline{\mathbf{\Gamma}}\mathbf{P}$ and $\underline{\mathbf{S}}_1 = \mathbf{P}'\underline{\mathbf{S}}\mathbf{P}$ the marginalized marginal likelihood is

$$\mathbf{Y}_1 \sim Mt(\mathbf{Z}\underline{\mathbf{\Gamma}}_1, (\mathbf{I}_T + \mathbf{Z}\underline{\mathbf{\Omega}}\mathbf{Z}')^{-1}, \underline{\mathbf{S}}_1, v - m + q). \quad (22)$$

2.3. Conditional Forecasts

Conditional forecasts are a useful tool for policy analysis. Such forecasts have the advantage that they are invariant against the ordering of the variables, if (as in our case) the structural shocks are exactly identified. For different fixed paths of one variable the researcher can investigate the consequences for the other variables in the VAR system. Waggoner and Zha (1999) provide a framework for calculating the conditional forecast distribution using a Gibbs sampling algorithm. To illustrate the idea, note that at time T (1) (with a constant \mathbf{c} as the only deterministic variable for simplicity) can be written as

$$\mathbf{y}'_T = \sum_{i=1}^p \mathbf{y}'_{T-i} \mathbf{A}_i + \mathbf{c} + \boldsymbol{\epsilon}'_T \mathbf{A}_0^{-1}, \quad (23)$$

⁵Somewhat simplified relative to Giannone et al. (2015), we do not simulate the posterior mode of the distribution of the hyperparameters, but numerically maximize (21). This substantially reduces the computational burden, see footnote 7.

with $(\mathbf{A}_0^{-1})' \mathbf{A}_0^{-1} = \mathbf{\Psi}$ and ϵ'_t being the uncorrelated structural shocks, with $E(\epsilon_t \epsilon'_t) = \mathbf{I}$. Iterating the system h steps forward yields

$$\mathbf{y}'_{T+h} = \mathbf{c} \mathbf{K}_{h-1} + \sum_{i=1}^p \mathbf{y}'_{T+1-i} \mathbf{N}_i(h) + \sum_{j=1}^h \epsilon'_{T+j} \mathbf{M}_{h-j}, \quad (24)$$

where

$$\mathbf{K}_0 = \mathbf{I}, \quad (25)$$

$$\mathbf{K}_l = \mathbf{I} + \sum_{j=1}^l \mathbf{K}_{l-j} \mathbf{A}_j, \quad l = 1, 2, \dots; \quad (26)$$

$$\mathbf{N}_i(1) = \mathbf{A}_i, \quad i = 1, \dots, p; \quad (27)$$

$$\mathbf{N}_i(h) = \sum_{j=1}^{h-1} \mathbf{N}_i(h-j) \mathbf{A}_j + \mathbf{A}_{h+i-1}, \quad i = 1, 2, \dots, p, \quad h = 2, 3, \dots; \quad (28)$$

$$\mathbf{M}_0 = \mathbf{A}_0^{-1}, \quad (29)$$

$$\mathbf{M}_l = \sum_{j=1}^l \mathbf{M}_{l-j} \mathbf{A}_j, \quad l = 1, 2, \dots; \quad (30)$$

with the convention that $\mathbf{A}_j = \mathbf{0}$ for $j > p$. The future values \mathbf{y}'_{T+h} in (24) depend on a systematic part and structural shocks. By writing (24) as

$$\mathbf{y}'_{T+h} - \mathbf{c} \mathbf{K}_{h-1} - \sum_{i=1}^p \mathbf{y}'_{T+1-i} \mathbf{N}_i(h) = \sum_{j=1}^h \epsilon'_{T+j} \mathbf{M}_{h-j}, \quad (31)$$

one can see that restrictions placed on the future path of at least one of the variables in \mathbf{y}_t implies restrictions on the future shocks to each variable in the system. These constraints on future innovations are expressed as

$$\mathbf{r} = \mathbf{R} \boldsymbol{\epsilon}, \quad (32)$$

where \mathbf{r} is a $q \times 1$ vector containing the values for the constrained variables minus the unconditional forecast of the constrained variables, $\boldsymbol{\epsilon}$ is a $g \times 1$ vector containing the constrained future innovations and \mathbf{R} is a $q \times g$ matrix containing the impulse response of the constrained variables to the structural shock ϵ_t at horizon $1, 2, \dots, h$, where h is the maximum forecast horizon, q is the total number of conditions, $g = hm$ is the total number of future innovations and $q \leq g$. Imposing conditions on a variable in such a way has the advantage that the variable itself is still treated as endogenous over the forecast period. However, solving (32) for $\boldsymbol{\epsilon}$ generally yields infinitely many solutions. Doan et al. (1984) show that the unique solution that satisfies the constraints and minimizes the sum of constrained future innovations $\boldsymbol{\epsilon}' \boldsymbol{\epsilon}$ is given by

$$\hat{\boldsymbol{\epsilon}} = \mathbf{R}'(\mathbf{R}\mathbf{R}')^{-1} \mathbf{r}. \quad (33)$$

The conditional forecasts are then calculated by inserting $\hat{\epsilon}$ into (24). The Gibbs sampling algorithm of Waggoner and Zha (1999) simulates the distribution of the conditional forecast and accounts for both parameter uncertainty and the structural shocks that are constrained for the conditional forecasts.

3. Simulation Study

3.1. Simulation Design

The simulation study compares different ways to choose the prior hyperparameters, to study which performs better in estimating impulse response functions in small samples, also relative to standard OLS-based estimates. We employ the following two data generating processes (DGPs)

$$\mathbf{y}'_t = \mathbf{y}'_{t-1} \begin{pmatrix} 1 & 0.015 & -0.006 \\ 0.005 & 1 & 0.036 \\ -0.044 & -0.05 & 1 \end{pmatrix} + \mathbf{y}'_{t-2} \begin{pmatrix} -0.04 & -0.005 & -0.0261 \\ -0.035 & -0.061 & -0.011 \\ 0.028 & -0.0074 & 0.005 \end{pmatrix} + \mathbf{u}'_t \quad (34)$$

and

$$\begin{aligned} \mathbf{y}'_t = & \mathbf{y}'_{t-1} \begin{pmatrix} 1 & 0.015 & -0.006 \\ 0.005 & 1 & 0.036 \\ -0.044 & -0.05 & 1 \end{pmatrix} + \mathbf{y}'_{t-2} \begin{pmatrix} -0.04 & -0.005 & -0.0261 \\ -0.035 & -0.061 & -0.011 \\ 0.028 & -0.0074 & 0.005 \end{pmatrix} \\ & + \mathbf{y}'_{t-3} \begin{pmatrix} 0.001 & 0.062 & -0.004 \\ 0.01 & -0.029 & -0.004 \\ -0.01 & -0.003 & 0.03 \end{pmatrix} + \mathbf{y}'_{t-4} \begin{pmatrix} -0.001 & -0.062 & 0.0274 \\ 0.009 & 0.006 & 0.0311 \\ 0.46 & 0.022 & -0.01 \end{pmatrix} \\ & + \mathbf{y}'_{t-5} \begin{pmatrix} 0.001 & -0.0002 & 0.0004 \\ 0.002 & -0.0002 & -0.00031 \\ 0 & 0.0005 & -0.0001 \end{pmatrix} + \mathbf{y}'_{t-6} \begin{pmatrix} -0.0009 & 0.0005 & -0.0004 \\ -0.0004 & 0.0001 & 0.00021 \\ 0 & -0.0001 & 0.0003 \end{pmatrix} \\ & + \mathbf{u}'_t, \end{aligned} \quad (35)$$

with $\mathbf{u}_t \sim N(\mathbf{0}, \mathbf{I})$ or t-distributed (for details, see below).

Figures 1 and 2 plot a realization for 1000 observations of the two DGPs using normally distributed errors. The long horizon reveals that none of the series is explosive over time. (The largest eigenvalue for DGPa is 0.984 and the largest eigenvalue for DGPb is 0.983.) Typically, datasets in applied work are however much smaller. If one only observes a small stretch of the process some of the series might appear explosive, as it is the case for German house prices (see Figures 7 or 11). This is particularly apparent for the second DGP where y_1 is much more volatile than y_2 and y_3 . For example, y_1 could be representative of the interest rate, which is much more volatile than the consumer price index.

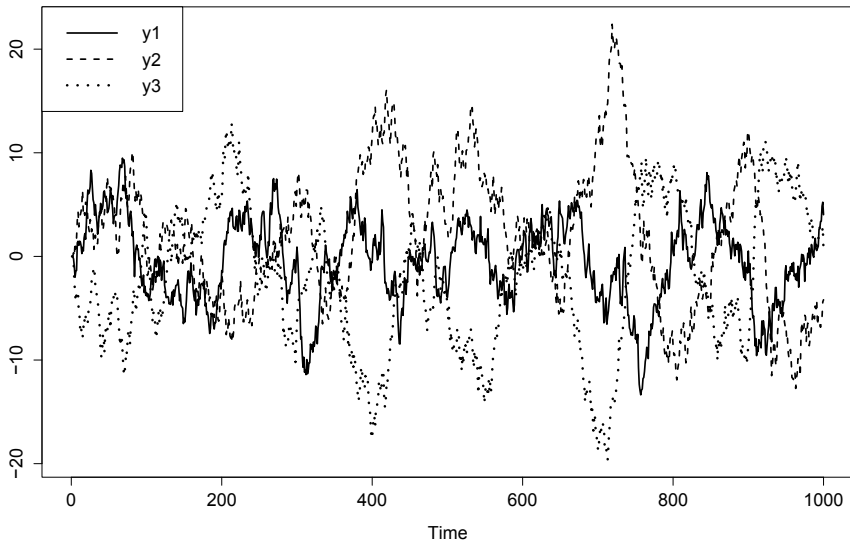


Figure 1: DGPa

In the simulation, only 96 observations of the series will be used for estimation, which is the also the number of observations used in the empirical empirical analysis in Section 4. We assess the performance of different ways of choosing the prior hyperparameters by comparing the estimated impulse response functions⁶ with the true impulse response functions. In total, four different approaches are compared. In the first approach, the hyperparameters are set to small numbers to obtain a tight prior. This provides us with a benchmark of how much a “high” amount of shrinkage can help to obtain more precise estimates. The second approach selects the hyperparameters by maximizing the marginal likelihood, as described in Section 2.3. The third and fourth approaches select the hyperparameters by minimizing the MSE of out of sample forecasts. This is done by minimizing the average relative MSE (AMSE) of all variables

$$\overline{AMSE} = \frac{1}{m} \sum_{i=1}^m \frac{MSE_i}{MSE_i(RW)}, \quad (36)$$

where m is the number of variables used in the model and the effect of different scales on

⁶The impulse response functions from the BVAR are estimated by taking the median of 1000 impulse response functions drawn. The Cholesky decomposition is used, as we also use it in the empirical part.

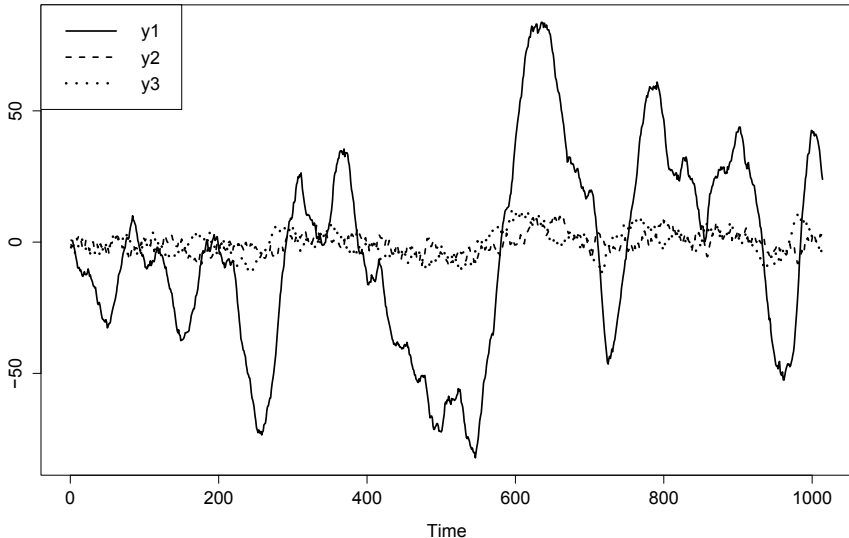


Figure 2: DGPb

time series is removed by the division over the MSE from a random walk (RW) forecast for each variable. Since it is in general not clear which forecast horizon is to be used for the evaluation of the forecast performance, we consider two different ways. The third approach leaves out the last 12 observations for the forecast evaluation of 1-to-12 period ahead forecasts. For the estimation of the impulse responses all observations are used. The fourth approach is similar, but instead of using 1-to-12 period ahead forecasts, uses 24 one period ahead forecasts. Each approach is applied to both DGPs and for each DGP five setups (explained below) are used. We thus consider $4 \cdot 2 \cdot 5 = 40$ simulations in total. Each simulation is based on 1000 replications.

The first setup for DGPa uses two lags for estimation. The second setup also uses two lags but has t -distributed errors with four degrees of freedom. In the third setup twelve lags are used for estimation with t -distributed errors. In the fourth and fifth setup two irrelevant additional variables are added. One follows a random walk with normally distributed error terms. The other can be described as $y_{5t} = y_{5t-1} + 0.3y_{2t-1} + v_t$ with $v_t \sim N(0, 1)$. In both setups 4 and 5 the errors of the VAR (i.e. y_{1t} to y_{3t}) are t -distributed with four degrees of freedom. In the fourth setup twelve lags are used, while in the fifth setup two lags are used. Table 1 provides a summary of the different setups.⁷

⁷A single setup with 1000 replications is already fairly computationally demanding. For the simulations,

Table 1: Simulation Setups

Setup	Number of Lags	Distribution	Additional Variables
1	2/6	normal	no
2	2/6	t-distributed	no
3	12/12	t-distributed	no
4	12/12	t-distributed	yes
5	2/6	t-distributed	yes

The table shows the number of lags used for estimation, the distribution of the error terms and if irrelevant additional variables are added to the VAR in each setup. For the t-distribution four degrees of freedom are used. Furthermore, the first number in the second column corresponds to DGPa and the second number corresponds to DGPb.

The hyperparameters for the first approach are set to $\lambda_0 = 1$, $\lambda_1 = 0.1$, $\lambda_3 = 0.1$, $\lambda_4 = 2$, $\mu_5 = 0$ and $\mu_6 = 0$. When the lag length for estimation is higher than the true lag length λ_3 is set to 3. For the other approaches the hyperparameters are chosen out of all combinations of $\lambda_0 \in \{0.1, 0.5, 0.9\}$, $\lambda_1 \in \{0.1, 0.5, 0.9\}$, $\lambda_3 \in \{0.1, 1, 2, 4\}$, $\lambda_4 \in \{0.1, 1, 2, 3\}$, $\mu_5 \in \{0, 0.1, 1\}$ and $\mu_6 \in \{0, 0.1, 1\}$. Although we know that all eigenvalues of the two DGPs lie inside the unit circle, we allow for values different from zero for μ_5 and μ_6 , as a researcher only observing a small sketch might do the same. Further note that we do not need to choose values for λ_5 , as no deterministic variables are included in the model.

The different setups aim to mimic realistic scenarios. In real world applications residuals typically have higher kurtosis than three and thus violate the normality assumption. The error terms are hence mostly drawn from a t-distribution. Moreover, the VAR is estimated with more lags and variables than necessary: in real world applications the researcher does not know which variables belong into the VAR model and thus might add irrelevant control variables to avoid an omitted variable bias. These irrelevant variables then increase the variance of the estimates. We account for this in setup 4 and 5. The same is true for the lag length, with researchers fitting too many lags to avoid biased estimates. For example, Litterman (1979) suggested to add as many lags as computationally feasible, with a prior distribution that incorporates the belief that higher lags are more likely to be close to zero. We account for this in setup 3 and 4.

3.2. Simulation Results

We present results in different ways. First, to investigate the bias of the BVAR and OLS-VAR (VAR for short) the mean of all estimated impulse response functions (for the BVAR this is the median of 1000 drawn impulse response functions) are compared

we used the HPC cluster Peregrine at the Rijksuniversiteit Groningen.

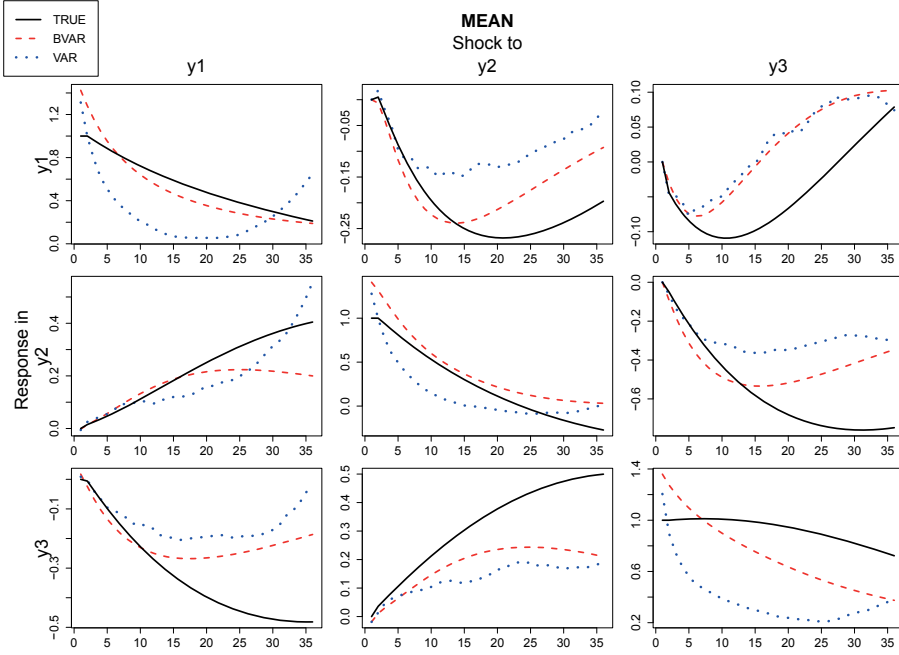


Figure 3: Comparison between the mean of the estimated impulse responses of the BVAR and VAR with the true impulse responses for DGPa and the fourth setup. The choice of hyperparameters is based on the marginal likelihood.

with the true impulse response functions. Second, the mean squared error (MSE) for the VAR and BVAR is calculated. The ratio of the MSE of the VAR over the MSE of the BVAR is used to compare their performance. Values above one indicate that the MSE of the BVAR is lower than the MSE of the VAR. But neither the mean nor the MSE are robust to outliers. Hence, results could potentially be strongly influenced by only a few number of extreme estimates in some of the replications. Third, quantiles of the estimated impulse responses may therefore provide a better impression of their distribution.

As an example, Figure 3 reports the mean of the estimated impulse responses of the BVAR and VAR for DGPa and the fourth setup with hyperparameters chosen based on the marginal likelihood. The responses of both models replicate the shape of the true responses with a bias. But, the bias when an informative prior is used is generally smaller than the bias of the VAR. Figure 4 shows the 0.05- and 0.95-quantiles of the estimated impulse responses for both models. The BVAR clearly dominates the VAR, being much closer to the true impulse responses. This result is qualitatively similar for the other simulations. The BVAR is more robust than the VAR against overspecifying the lag length or/and adding irrelevant variables.

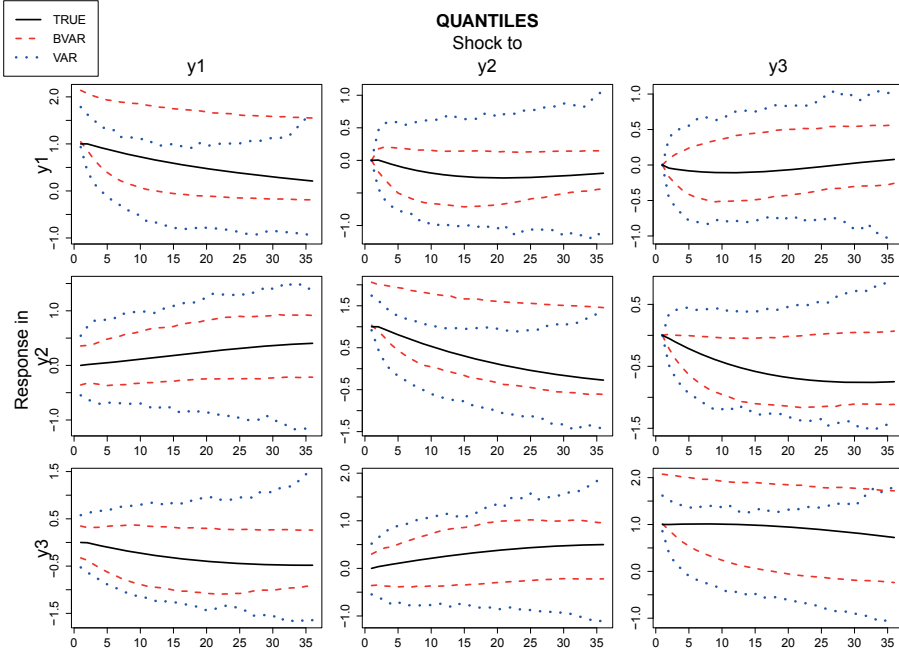


Figure 4: Comparison between the 0.05 and 0.95 quantiles of the estimated impulse responses of the BVAR and VAR with the true impulse responses for DGP_a and the fourth setup. The choice of hyperparameters is based on the marginal likelihood. Note that the true impulse responses here are the same as in Figure 3, but the scale of the y-axis is different.

To investigate the performance of the four different approaches we compare the MSE ratios. Figure 5 and Figures A.1-A.4 report the MSE ratio for each approach and all simulations of DGP_a for each impulse response function. The results for DGP_b are similar. With only a few exceptions the MSE of the approach based on the marginal likelihood is better than the frequentist VAR and works better with normally distributed errors, but is robust to t-distributed errors. Furthermore, in most cases the marginal likelihood outperforms the hyperparameter choice based on the forecast performance. The hyperparameter choice based on the forecast performance sometimes even performs quite poorly, producing an MSE much higher than the frequentist VAR. This is particularly the case for responses to own shocks. The approach of using 24 one step ahead forecasts works better compared to the approach of using one 12 step ahead forecast, indicating that in small samples a selection of the hyperparameters according to the out-of-sample performance may suffer from using only a small fraction of the sample observations. Instead, the marginal likelihood makes more efficient use of the data by using the full sample. This

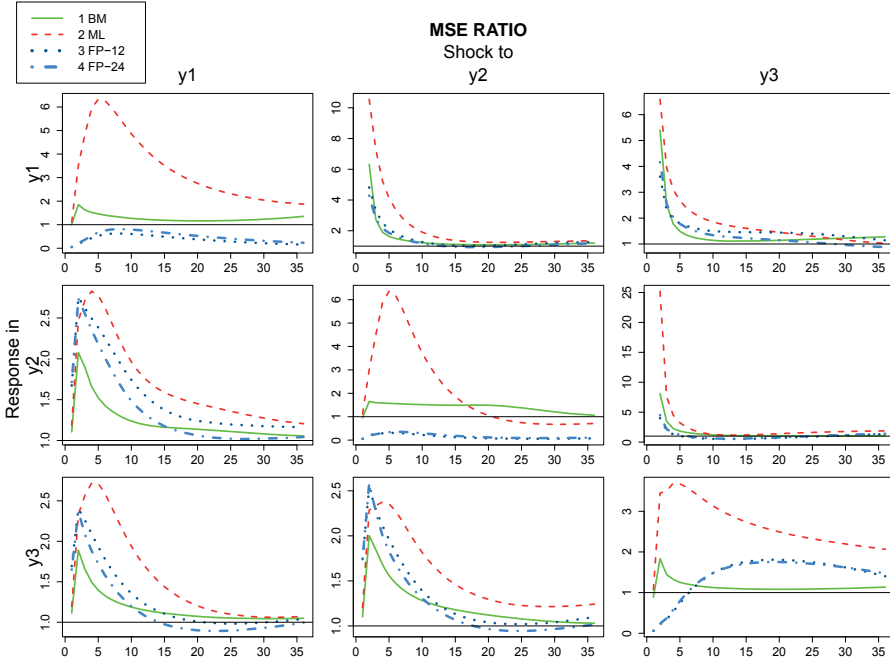


Figure 5: Comparison between the four different approaches for DGPa and setup 1. The hyperparameter choice in the first approach is based on the benchmark (BM), the second is based on the marginal likelihood (ML), the third is based on the forecast performance (FP) of 1-to-12 period ahead forecasts and the last is based on the forecast performance of 24 one period ahead forecasts. For each approach the ratio of the MSE, of the VAR, over the MSE, of the BVAR, is plotted for each impulse response function. Hence, values higher than one imply that the MSE of the BVAR is smaller than that of the VAR.

result is consistent with Ding and Karlsson (2014). Their simulation study shows that the marginalized marginal likelihood provides a sharper discrimination between models and variables than the predictive likelihood, as the marginalized marginal likelihood uses the full sample and the predictive likelihood only uses a smaller fraction of the data.

However, the simulation results, as usual, depend on the specification of the DGP. For example, the first own lags of both DGPs are one. This is not implausible for real data, but matches the prior beliefs. One can argue that a DGP which matches the prior beliefs too closely gives the BVAR an unfair advantage over the VAR. In real world applications prior beliefs are often more inaccurate. To address this concern we run additional simulations for both DGPs and each setup, where first own lags are set to 0.9 instead of 1. The hyperparameters are selected according to the marginal likelihood. The results of this

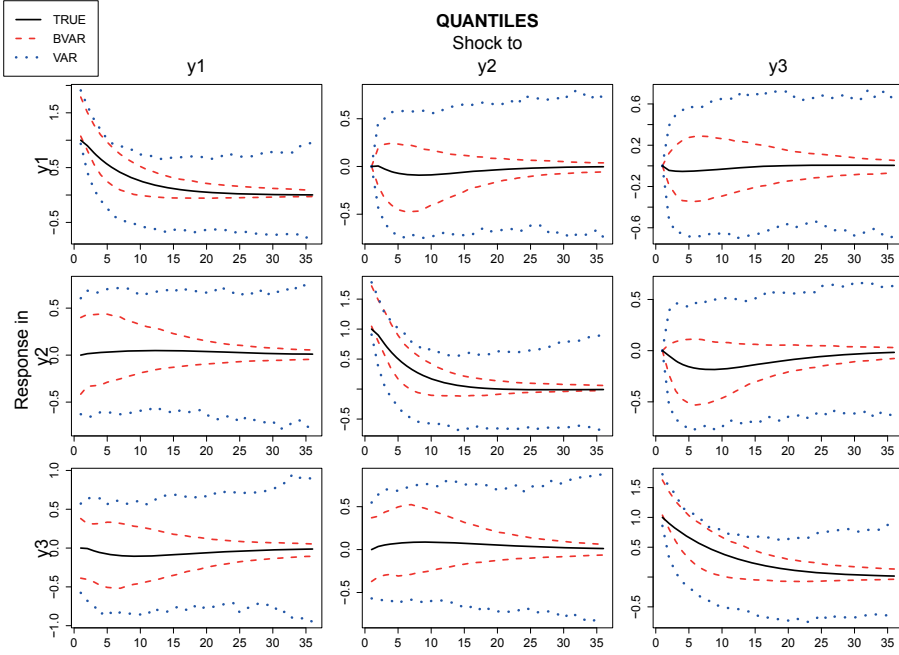


Figure 6: Comparison between the 0.05 and 0.95 quantiles of the estimated impulse responses of the BVAR and VAR with the true impulse responses for DGP_a where first own lags are set to 0.9 instead of 1 and the fourth setup. The choice of hyperparameters is based on the marginal likelihood.

simulation are consistent with the previous results. As an example, Figure 6 shows the 0.05 and 0.95 quantiles of the estimated impulse responses for both models. Thus, prior information centered at “false” parameter values may still help to produce more precise estimates.

4. House prices and interest rates

This section uses a BVAR to investigate the link between house prices and interest rates. We first consider a set of important control variables used in the literature. Out of these potential variables Section 4.2 selects the control variables via the marginalized marginal likelihood. Based on the selected variables, Section 4.3 finally presents the results of the impulse response analysis, variance decomposition and counterfactual analysis. Our main findings are that the increase in house prices can be better explained by falling interest rates than by other fundamentals of the economy and that increasing interest rates likely

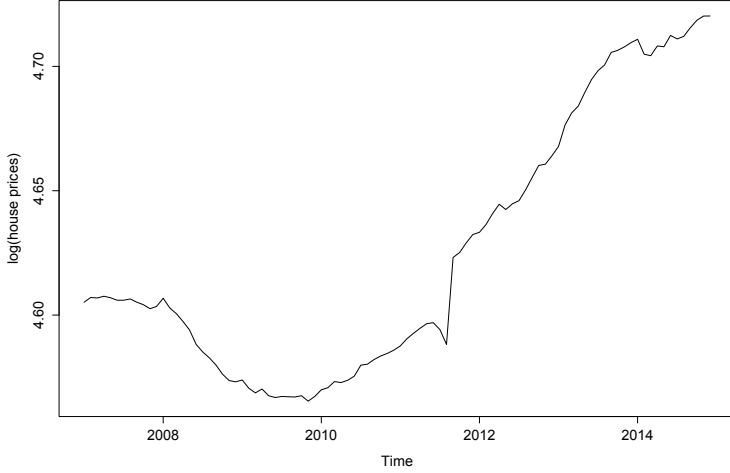


Figure 7: Log-house price indices

are sufficient to stop the increase of house prices.

4.1. Theoretical Considerations about the Data

We use the real estate price indices for existing houses by the German real estate internet platform ImmobilienScout24, Germany's largest real estate internet platform. Existing houses are much more important than newly built houses, accounting for the majority of the market. The indices are measured with monthly frequency starting in 2007 and are based on offer prices placed on ImmobilienScout24 (Bauer et al., 2013). Figure 7 shows the time series of log house prices indices.

After the financial crisis house prices initially dropped, but have increased fairly rapidly since 2010. Long term interest rates in Germany decreased from 4.02% to 0.59% during the same period, see Figure 8. Economic theory provides different explanations for why decreasing interest rates may lead to a higher demand for real estate projects. Decreasing interest rates lower the costs of financing real estate, making it more profitable to invest in real estate than in bonds. If lower interest rates are also associated with higher inflation in the future, investing in real estate can protect the wealth of economic agents. Since real estate projects are typically long term investments, their development is more likely to be linked to the development of long term interest rates than to short term interest rate.⁸ Existing literature however mostly uses short term interest rates, being interested

⁸This is confirmed by comparing the fit of the house prices with the marginalized marginal likelihood between a BVAR including short term interest rates besides house prices and a BVAR including long

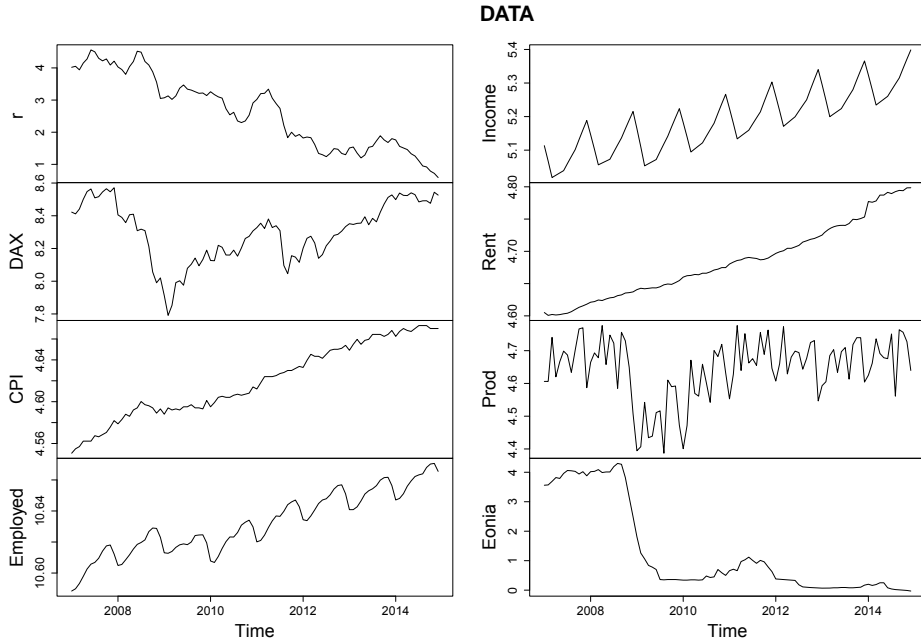


Figure 8: Time series overview.

in the direct effect of monetary policy on house prices.

Several studies analyze the link between central banks' monetary policy decisions and house prices. Examples include Jarociński and Smets (2008), Gopolicy and Hofmann (2008), Musso et al. (2011), Setzer and Greiber (2007) and Demary (2009). All studies use gross domestic product (GDP), the consumer price index (CPI) or both as control variables in their VAR models. Moreover, all find evidence that monetary policy influences house prices. A recent study for Switzerland by Berlemann and Freese (2013) analyzes the link between interest rates, house prices and stock prices. Other studies focus on the effect of other macroeconomic variables on house prices. Baffoe-Bonnie (1998) finds evidence that employment changes can explain real estate cycles of house demand and the number of houses sold. The fluctuations in employment cause a change in incomes and the change in incomes has an effect on the demand for houses. Furthermore, Bharat and Zan (2002) and McQuinn and O'Reilly (2007) find evidence of a long-run cointegration relationship between house prices, income and interest rates. In contrast to these studies we focus our attention specifically on Germany and investigate the time period after the financial crisis.

term interests besides house prices.

Table 2: Variable Selection first stage

	p	λ_0	λ_1	λ_3	λ_4	λ_5	μ_5	μ_6	AMSE	mse	mml
DAX	4	0.5	0.5	1	1	0.9	1	1	0.9	9.4	370.38
Production	4	0.9	0.9	2	1	0.9	1	1	0.7	9.3	373.25
Income	4	0.9	0.9	2	1	0.9	1	1	0.6	9.5	372.82
CPI	4	0.5	0.9	2	1	0.9	1	1	0.7	9.6	371.23
Eonia	4	0.5	0.5	1	1	0.9	1	1	1.1	9.6	370.15
Rent	4	0.5	0.9	2	2	0.9	1	1	0.7	8.5	372.42
Employment	4	0.9	0.9	2	1	0.9	1	1	0.6	9.5	372.93

The table shows the marginalized marginal likelihood (mml) expressed in logs for house prices, as described in (22). In each row the corresponding variable is added to the model in addition to r and house prices. AMSE gives the average MSE as described in (36) and mse gives the mean squared error just for house prices, measured in 1.0E-06 units. Both are based on 24 one step ahead forecasts for the observations ranging from 2013 to 2014.

We further consider rents, also taken from ImmobilienScout24. Moreover, the following time series taken from the databases of the Bundesbank (www.bundesbank.de) and the Statistical Data Warehouse of the ECB (<http://sdw.ecb.europa.eu/home.do>): r , the nominal secondary market yield of German government bonds with maturities of close to ten years, CPI , the German consumer price index, DAX , the German stock exchange index, $Eonia$, the effective overnight interest rate computed as a weighted average of all overnight unsecured lending transactions in the interbank market in Euro, $Employed$, the number of employed people in Germany, $Prod$, the industry production at constant prices in Germany and $Income$, the net income in Germany. For income we use a linear transformation to transform the series from quarterly to monthly frequency. All other time series are measured in monthly frequency ranging from 2007:01 to 2015:12. We reserve the data from 2015 for the out-of-sample performance assessment and use the remaining $T = 96$ observations for estimation. The time series are not seasonally adjusted and, except for the interest rates, all variables are expressed in logs. Figure 8 plots the series.

4.2. Empirical Variable Selection

We first compute the marginalized marginal likelihood to choose relevant control variables. We estimate trivariate BVARs for all potentially useful variables considered in Section 4.1. Long term interest rates and house prices are always included. Then, the variable which delivers the highest marginalized marginal likelihood for house prices is retained. The hyperparameters are chosen by maximizing the marginal likelihood, as the simulation has shown that this approach outperforms the approach of selecting the prior hyperparameters based on their forecast performance. The lag length is selected together with the hyperparameters by simultaneously maximizing the marginal likelihood w.r.t. both.⁹ The

⁹We choose the lag length according to the marginal likelihood to avoid a overly long lag length, as the simulation in Ding and Karlsson (2014) shows that an overly long lag length lowers the discriminative

Table 3: Variable Selection second stage

	p	λ_0	λ_1	λ_3	λ_4	λ_5	μ_5	μ_6	AMSE	mse	mml
DAX	5	0.5	0.9	1	1	0.9	1	1	0.8	9.7	362.45
Income	5	0.9	0.9	1	1	0.9	1	1	0.5	9.8	361.57
CPI	5	0.9	0.9	2	1	0.9	1	1	0.6	9.6	367.78
Eonia	5	0.5	0.5	1	1	0.9	1	1	1.0	9.7	365.12
Rent	5	0.5	0.9	2	1	0.9	1	1	0.7	8.6	367.49
Employment	5	0.9	0.9	1	1	0.9	1	1	0.5	9.8	361.94

The table shows the marginalized marginal likelihood (mml) expressed in logs for house prices, as described in (22). Production, house prices and r are always included in the BVAR and in each row the corresponding variable is added to the model. AMSE gives the average MSE as described in equation (36) and mse gives the mean squared error just for house prices, measured in 1.0E-06 units. Both are based on 24 one step ahead forecasts for the observations ranging from 2013 to 2014.

hyperparameters and lag lengths are chosen out of all combinations of $\bar{p} \in \{1, 2, \dots, 12\}$, $\lambda_0 \in \{0.1, 0.5, 0.9\}$, $\lambda_1 \in \{0.1, 0.5, 0.9\}$, $\lambda_3 \in \{0.1, 1, 2, 4\}$, $\lambda_4 \in \{1, 2, 3\}$, $\mu_5 \in \{0, 0.1, 1\}$ and $\mu_6 \in \{0, 0.1, 1\}$.

To account for seasonal effects in the time series a set of monthly dummies is added to the model. Moreover, one dummy variable for the month 2011:09 is added to account for a potential outlier in the time series house prices, see Figure 7. When rent is included one more dummy variable is added for the month 2014:02, see Figure 8. However, neither dummy has an important influence on our results. The hyperparameter λ_5 is set to 0.9 to ensure that $\lambda_5 < \lambda_4$. Otherwise, the variation in the endogenous variables will be over-explained by the deterministic variables relative to the endogenous variables (Brandt and Freeman, 2006).

Table 2 shows the average MSE as described in (36), the MSE just for house prices and the log marginalized marginal likelihood for house prices, in addition to the selected prior hyperparameters and lag length. The mean squared errors are calculated based on 24 one-step-ahead out-of-sample forecasts. Values over one mean that the forecast performance of the model is on average worse than the forecast of a random walk. Except for the Eonia all values are lower than one, indicating no serious issue of overfitting.

The values for the marginalized marginal likelihood, as described in (22), are lowest for the two financial variables. This seems reasonable as private investors and house owners are not directly involved in the inter-bank market and as the stock market in Germany is less important for private investors compared with countries like the US. The marginalized marginal likelihood is highest for the three fundamental variables production, income

power of the marginalized marginal likelihood.

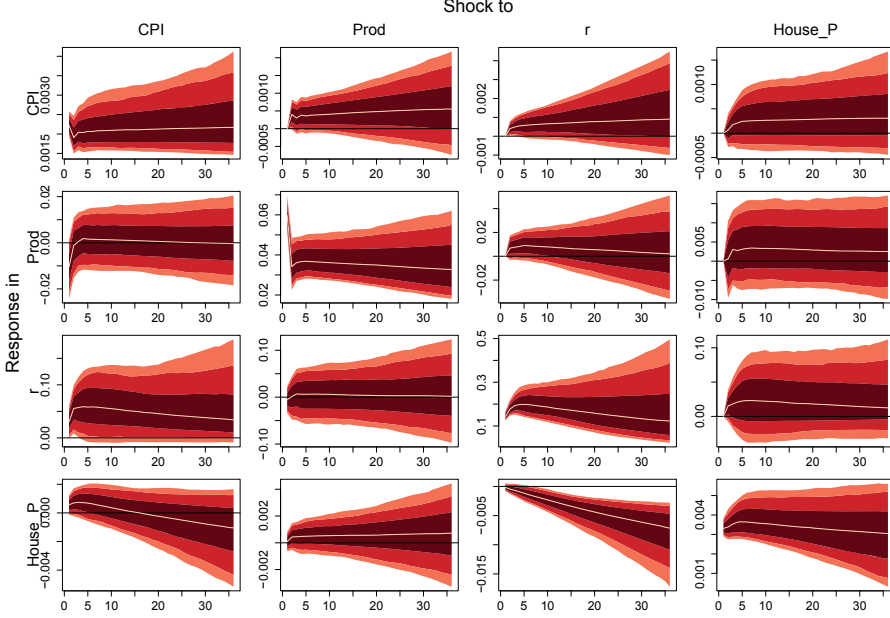


Figure 9: Median impulse response functions with 68%, 90% and 95% error bands. Note that *CPI* is ordered first, *Prod* second, *r* third and house prices last.

and employment. Not surprisingly, the value is similar for these three variables as they contain similar information about the economy. Industrial production has the highest marginalized marginal likelihood and also delivers the best out of sample forecasting performance of the fundamental variables and hence is added to the BVAR model.

Table 3 shows that given that production is now additionally included in the model, the marginalized marginal likelihood is lowest for income and employment as these do not provide much additional information.¹⁰ The model with *CPI* now has the highest marginalized marginal likelihood. Note, however, that the marginalized marginal likelihood is now a bit lower and the forecast performance for house prices has become slightly worse, compared to the trivariate BVARs from the first stage variable selection. Nevertheless, from an economic perspective *CPI* provide new information for the model. As the simulation results have shown that the BVAR is robust to adding irrelevant variables, we use *CPI* as additional control variable. The robustness checks in Sec. 4.4. will however demonstrate that our results are not driven by adding this variable. Furthermore, by including these variable, we will be able to highlight/show the importance of interest

¹⁰It is also possible to consider the fit of different pairs of variables. However, we believe that this is not important here, as many of the variables contain similar information.

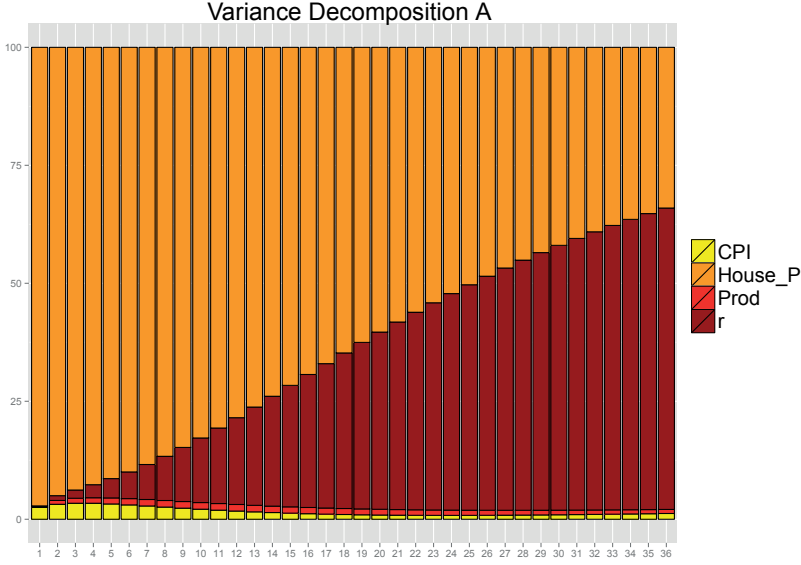


Figure 10: The figure shows the forecast error variance decomposition in percent for house prices. Note that *CPI* is ordered first, *Prod* second, *r* third and house prices last.

rates relative to other variables for the development of house prices.

4.3. Empirical Results

4.3.1. Innovation Accounting

We now turn to the discussion of the estimated impulse responses of the BVAR model including house prices, *CPI*, interest rates and production. Following the monetary policy literature (see Christiano et al., 1999) the ordering of the variable is as follows: *CPI* is ordered first, production second, interest rates third and house prices last. Table B.2 shows the contemporaneous correlation between the residuals. The correlation between the residuals is low with exception for *CPI*, whose residuals show some correlation with the other residuals. Thus, we estimate the model again with *CPI* ordered last as the only change. The estimated impulse responses are robust to the change in the ordering.

We now discuss the impulse responses shown in Figure 9. As expected, a positive shock to long term interest rates has a negative effect on house prices. Furthermore, the effect increases over time. Both structural shocks to *CPI* and production are likely to have a positive effect on house prices, though zero is mostly contained in the Bayesian credibility

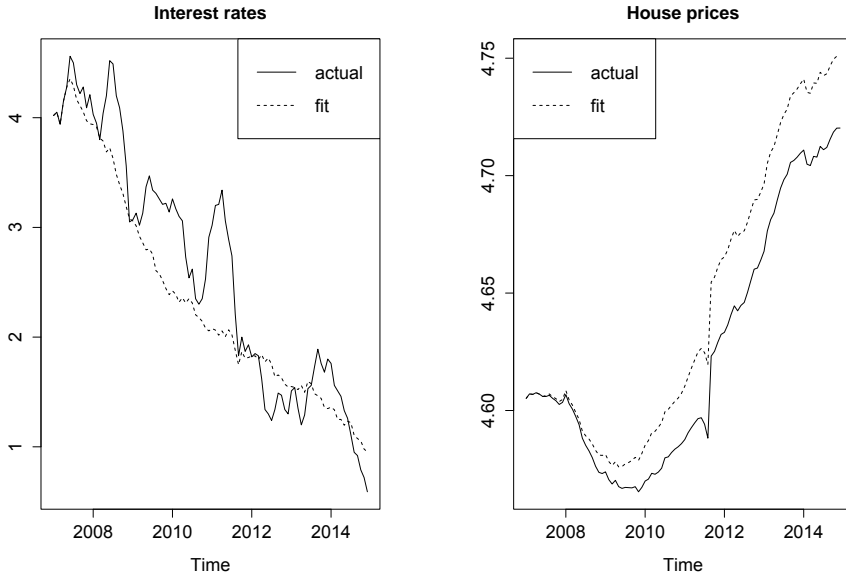


Figure 11: The figure shows the actual and the fitted path of the interest rate and house prices in Germany, by setting all interest rates shocks to zero.

interval. Furthermore, a shock to interest rates likely has a positive effect on *CPI*. This can be explained by via the Fisher Parity, according to which a nominal interest rate shock may be linked to increasing expected inflation. Increasing expected inflation can also explain the positive response of *Prod* to an interest rate shock: when agents expect price increases, they bring forward their purchases. However, the associated uncertainty with a response of *CPI* and *Prod* to an interest rate shock is quite high and to the extent to which the results also apply to more tranquil periods is questionable. In order to quantify the relative importance of the different shocks, Figure 10 reports the proportion that each shock contributes to the forecast error variance of house prices over a 36 month horizon, based on the mean posterior parameters. The results show that the shocks to the long term interest rates account for an increasing amount of the variation in house prices over time, stressing the importance of interest rates for house prices in the long run. Hence, the impulse responses and the variance decomposition indicate that the falling interest rates are a decisive factor for increasing house prices in Germany.

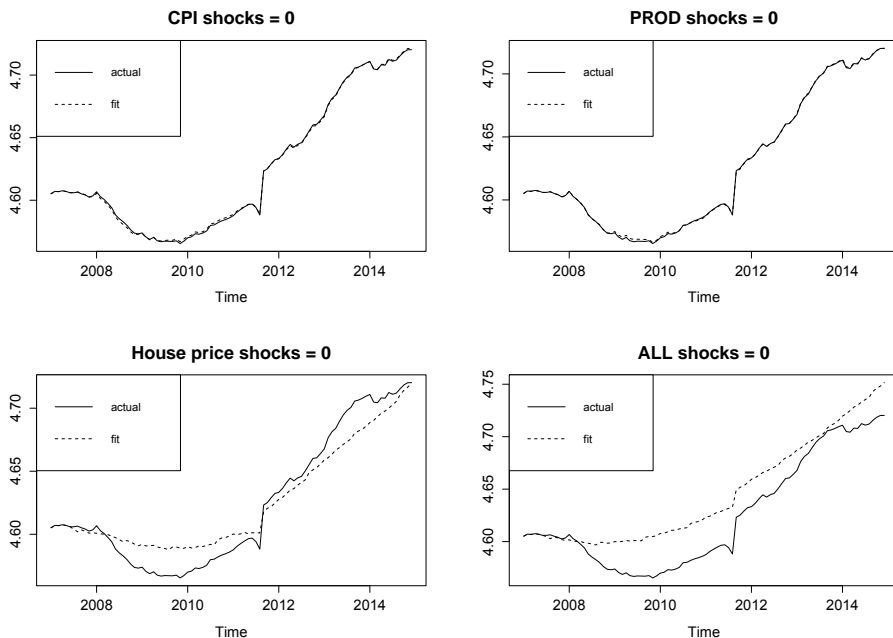


Figure 12: The figure shows the actual and the fitted path of house prices in Germany, by setting either all individual shocks (residuals) of the corresponding variable to zero or all shocks in the VAR system.

4.3.2. Counterfactual Analysis

To gain a better understanding of the development of house prices, we perform an in-sample counterfactual simulation conditional on the first p observations. We shall turn to the out-of-sample counterfactual analysis as described in Section 2.3. afterwards. First we simulate the VAR system by setting all coefficients to their mean values and adding in each period the estimated residuals. This produces a perfect fit. But setting all residuals in one equation to zero allows us to investigate the impact of the shocks of this variable on the development of house prices. Figure 11 shows the likely development of interest rates and house prices had there been no shocks to the interest rate equation. House prices would have been higher in the entire sample since 2008, as the left panel reveals that interest rate shocks had mostly a positive effect on the interest rate. However, this cannot entirely explain the decrease of the house prices from 2008 to 2010. Figure 12 shows the likely development of house prices to Germany if no shocks in the *CPI*, *Prod*, house prices equations and no shock in all equations had occurred. The shocks in the *CPI* and *Prod* equations had almost no impact on house prices, while the shocks in the house price equation can explain a large part of the decreasing house prices from 2008 to 2010.

If no shocks had occurred house prices would not have decreased. Hence, the decrease of house prices is likely related to negative shocks in the housing market, possibly related to the housing market crises in the US and Spain.

We now analyze how house prices might develop if the interest rate returned to its initial value of 4%, possibly due to a more restrictive monetary policy. It is not possible to answer this question by using the in-sample counterfactual analyses. Instead, we use a conditional forecast (described in Sec. 2.3.) to investigate the consequences of an increase in interest rates to 4% for the year 2015, comparing it with the results of an unconditional forecast. The unconditional forecast accounts for both parameter and forecast uncertainty. Figure 13 shows that house prices are predicted to continue to increase, in line with the actual path of house prices. If, however, interest rates were to rapidly go up to 4%, house prices are predicted to decrease.¹¹

4.4. Robustness Checks

In order to show that our results are not driven by adding *CPI* to the model (which lowers the marginalized marginal likelihood of house prices), we estimate the model that delivers the highest marginalized marginal likelihood for house prices (the model that only includes *Prod* as additional control variable), based on the hyperparameters and lag length from Table 2. The results of the impulse response functions are reported in Figure A.5, the results of the counterfactual analysis are reported in Figure A.6 and the results of the variance decomposition are reported in Figure A.7. Taken together, the results are similar to the results of the BVAR including *CPI*.

For the BVAR model including house prices, *CPI*, interest rates and *Prod* the chosen hyperparameters and lag length (see Table 3) are not sufficient to remove all autocorrelation in the residuals in all equations. In order to have statistically well-behaved residuals the hyperparameter λ_3 is switched from 2 to 1. To show that the residuals do not show any misspecification in the likelihood function due to serial correlation or ARCH effects, Table B.1 in the appendix shows the results for the Ljung-Box test for the squared and unsquared residuals. But changing the hyperparameter λ_3 from 2 to 1 leads to a change of the average MSE from 0.64 to 0.71. However, the results are robust to changing the hyperparameter λ_3 from 2 to 1. Furthermore, the results are robust to using a lag length of twelve.

To further test the robustness of the results, we include *Rent* instead of *CPI* in the model. The estimation is based on the hyperparameters and lag length from Table 3. Again the results are similar to results of the benchmark model including *CPI*. Finally, we repeat the analysis of innovation accounting for different control variables. We estimate a BVAR for each of the variables in Table 2 which were not already included in the BVAR, in

¹¹It is also possible to assume a more realistic smooth increase of the interest rates, which however leads to the same conclusion.

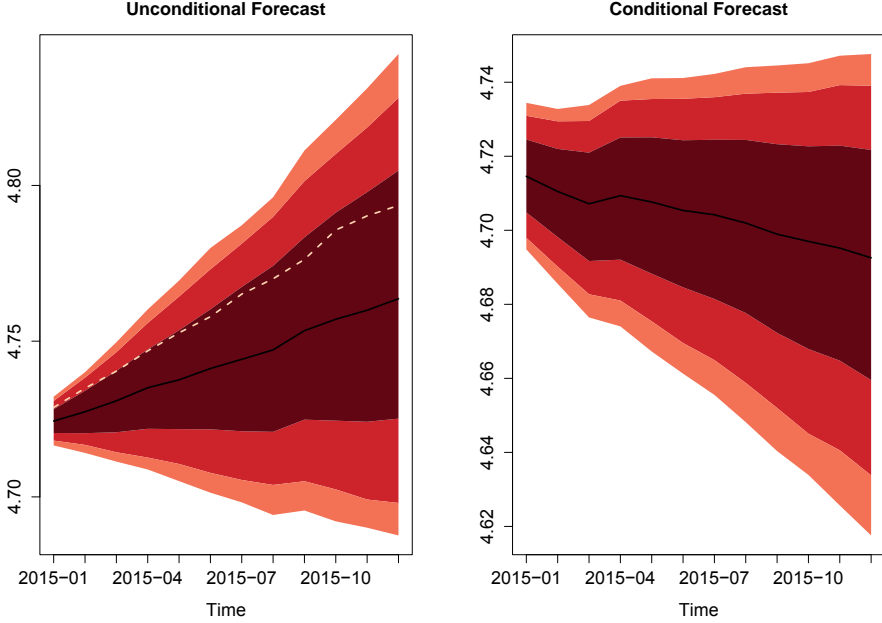


Figure 13: Conditional and unconditional forecasts with 68%, 90% and 95% error bands for house prices in Germany. The conditional forecast is based on the assumption that interest rates are at 4% over the forecast horizon and the dotted line represents the actual values. Furthermore, the variables production and *CPI* are included in the model.

addition to interest rates and house prices. The hyperparameters and lag length for each model are taken from Table 2. The results remain robust to these changes.

5. Conclusion

We study whether prior information helps to obtain more precise VAR estimates of impulse response functions in small samples. Moreover, we analyze different approaches of selecting the hyperparameters, which are compared via simulation. Building on the insights from the simulation, we empirically investigate the recent link between interest rates and house prices in Germany. The simulation study shows that the use of prior information to shrink the model parameters does indeed help to obtain sharper estimates of impulse responses. In addition, selecting hyperparameters by maximizing the marginal likelihood typically leads to more precise estimates than selecting the hyperparameters by minimizing out of sample forecast errors.

The empirical analysis reveals that interest rates play an important role in the recent development of house prices in Germany, as a large part of the variation in house prices can be explained by shocks to interest rates. The persistent negative response of house prices to interest rate shocks indicates that falling interest rates have substantially contributed to the sudden increase in house prices. It cannot even be ruled out that these have caused the increase in connection with Germany's robust economic recovery and positive expectations from agents, see an de Meulen et al. (2014). A counterfactual analysis shows that a permanent increase of interest rates to 4% would be sufficient to stop the increase of house prices. Overall, the results of the impulse response functions, variance decomposition and conditional forecasts indicate that the increase in house prices can be better explained by falling interest rates than by other fundamental values of the economy.

We therefore provide evidence for the recent turbulent period on the German real estate market that suggests a strong relationship between house prices and interest rates. Of course, the extent to which the results also apply to more tranquil periods is an open issue.

Future research using a panel VAR for European countries would allow to use more observations in estimation and to investigate possible spillover effects between countries.

References

- an de Meulen, P., Micheli, M., and Schmidt, T. (2014). Forecasting real estate prices in Germany: the role of consumer confidence. *Journal of Property Research*, 31(3):244–263.
- Baffoe-Bonnie, J. (1998). The dynamic impact of macroeconomic aggregates on housing prices and stock of house: a national and regional analysis. *The Journal of Real Estate Finance and Economics*, 17(2):179–197.
- Bañbura, M., Goammpe, D., and Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, 25(1):71–92.
- Bauer, T., Feuerschütte, S., Kiefer, M., an de Meulen, P., Micheli, M., Schmidt, T., and Wilke, L. H. (2013). A hedonic real estate price index for Germany based on internet data: 2007-2011. *AStA Wirtschafts und Sozialstaisches Archiv*, 7(1):5–30.
- Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. Springer series in statistics. Springer, New York, Berlin, Heidelberg.
- Berlemann, M. and Freese, J. (2013). Monetary policy and real estate prices: a disaggregated analysis for Switzerland. *International Economics and Economic Policy*, 10(4):469–490.
- Bharat, B. and Zan, Y. (2002). House prices and housing investment in Sweden and the UK: econometric analysis for the period 1970–1998. *Review of Urban & Regional Development Studies*, 14(2):189–216.
- Brandt, T. P. and Freeman, J. R. (2002). Moving mountains: Bayesian forecasting as policy evaluation. *Meeting of the Midwest Political Science Association, Chicago, Il.*
- Brandt, T. P. and Freeman, J. R. (2006). Advances in Bayesian time series modeling and the study of politics: theory testing, forecasting and policy analysis. *Political Analysis*, 14(1):1–36.
- Brandt, T. P. and Freeman, J. R. (2009). Modeling macro-political dynamics. *Political Analysis*, 17(2):113–142.
- Canova, F. (2007). *Methods for applied macroeconometric research*. Princeton University Press.
- Carlin, B. P. and Louis, T. A. (2000). *Bayes and empirical Bayes methods for data analysis*. FL: Chapman & Hall/CRC, 2nd ed edition.
- Carriero, A., Clark, T., and Marcellio, M. (2015). Bayesian VARs: specification choice and forecast accuracy. *Journal of Applied Econometrics*, 30(1):46–73.
- Christiano, L., Eichenbaum, M., and Evans, C. L. (1999). Monetary policy shocks: what have we learned and to what end? *Handbook of macroeconomics*, 1:65–148.

- Demary, M. (2009). The link between output, inflation, monetary policy and housing price dynamics. *MPRA Working Paper*.
- Deryugiana, E. B. and Ponomarenko, A. (2013). A large Bayesian vector autoregression model for russia. *BOFIT Discussion Paper*.
- Ding, S. and Karlsson, S. (2014). Model averaging and variable selection in VAR models. *manuscript*.
- Doan, T., Littermann, R. B., and Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, 3(1):1–100.
- Geweke, J., Koop, G., and Van Dijk, H. (2011). *The oxford handbook of Bayesian econometrics*. Oxford University Press.
- Giannone, D., Lenza, M., and Primiceri, G. E. (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics*, 97(2):436–451.
- Gopolicy, C. and Hofmann, B. (2008). House prices, money, credit and the macroeconomy. *Oxford Reviews of Economic Policy*, 24(1):180–205.
- Jarociński, M. and Smets, F. (2008). House prices and the stance of monetary policy. *ECB Working Paper*.
- Jorda, O., Schularich, M., and Taylor, A. M. (2015). Interest rates and house prices: pill or poison? *FRBSF Economic Letter*. Federal Reserve Bank of San Francisco.
- Kadiyala, K. R. and Karlsson, S. (1997). Numerical methods for estimation and inference in Bayesian VAR-models. *Journal of Applied Econometrics*, 12(2):99–132.
- Karlsson, S. (2013). Forecasting with Bayesian vector autoregressions. *Handbook of Economic Forecasting*, 2:791–897.
- Koop, G. and Korobilis, D. (2010). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics*.
- Litterman, R. B. (1979). Techniques of forecasting using vector autoregressions. *Working Paper, Federal Reserve Bank of Minneapolis*.
- Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions-five years of experience. *Journal of Business & Economic Statistics*, 4(1):25–38.
- McQuinn, K. and O’Reilly, G. (2007). A model of cross-country house prices. *Research Technical Paper*.
- Milani, F. (2012). The modeling of expectations in empirical DSGE models: a survey. *Advances in Econometrics*, 28:3–38.

- Musso, A., Neri, S., and Stracca, L. (2011). Housing, consumption and monetary policy: how different are the US and the euro area? *Journal of Banking & Finance*, 35(11):3019–3041.
- Robertson, J. and Tallman, E. (1999). Vector autoregressions: forecasting and reality. *Federal Reserve Bank of Atlanta Economic Review*, 84(1):4–18.
- Setzer, R. and Greiber, C. (2007). Money and housing: evidence for the euro area and the US. *Deutsche Bundesbank Discussion Paper*.
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1):1–48.
- Sims, C. A. (1993). A nine-variable probabalistic macroeconomic forecasting model. *Business Cycles, Indicators and Forecasting*, pages 179–212.
- Sims, C. A. and Zha, T. (1998). Bayesian methods for dynamic multivariate models. *International Economic Review*, 4(39):949–68.
- Sims, C. A. and Zha, T. (2006). Were there regime switches in US monetary policy? *The American Economic Review*, 96(1):54–81.
- Summers, P. (2001). Forecasting Australia’s economic performance during the Asian crisis. *International Journal of forecasting*, 17(3):499–515.
- Theil, H. and Goldberger, A. S. (1960). On pure and mixed statistical estimation in economics. *International Economic Review*, 2(1):65–78.
- Waggoner, D. F. and Zha, T. (1999). Conditional forecasts in dynamic multivariate models. *Review of Economics and Statistics*, 81(4):639–651.
- Wright, J. H. (2009). Forecasting US inflation by Bayesian model averaging. *Journal of Forecasting*, 28(4):131–144.
- Zellner, A. (1971). *An introduction to Bayesian inference in econometrics*. John Wiley & Sons.

Appendix A. Figures

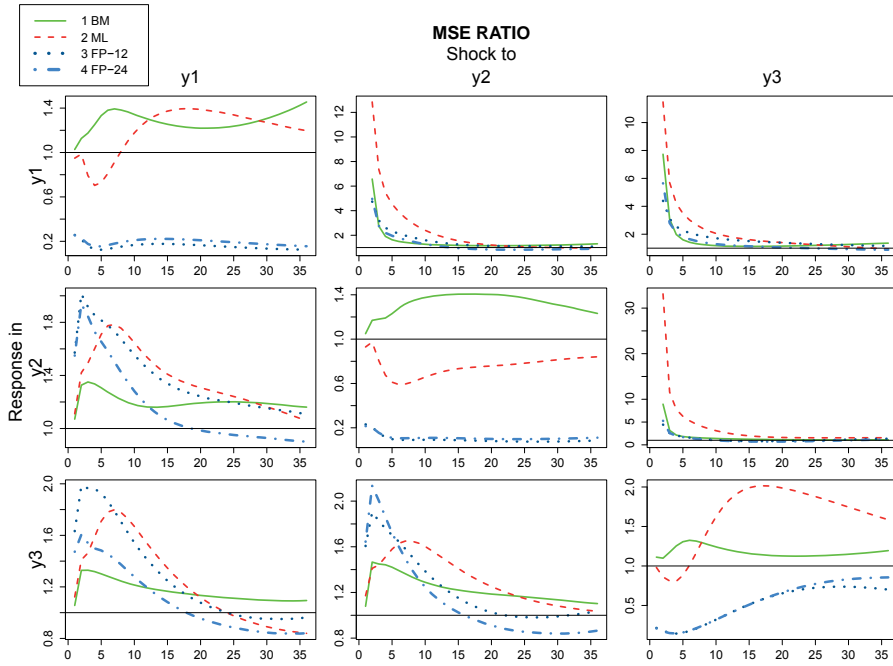


Figure A.1: Comparison between the four different approaches for DGPa and setup 2. For further explanation see Figure 5.

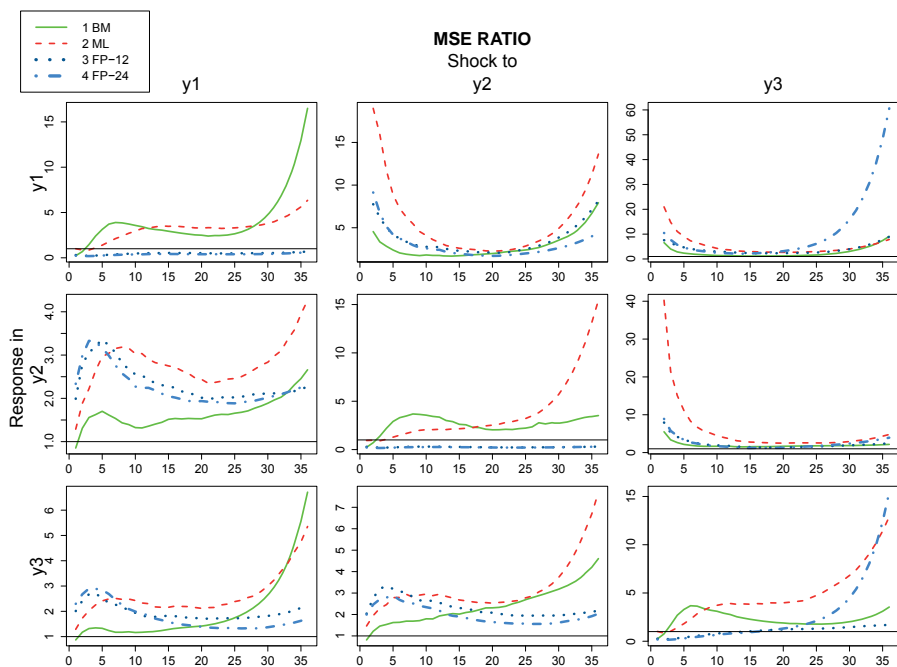


Figure A.2: Comparison between the four different approaches for DGPa and setup 3. For further explanation see Figure 5.

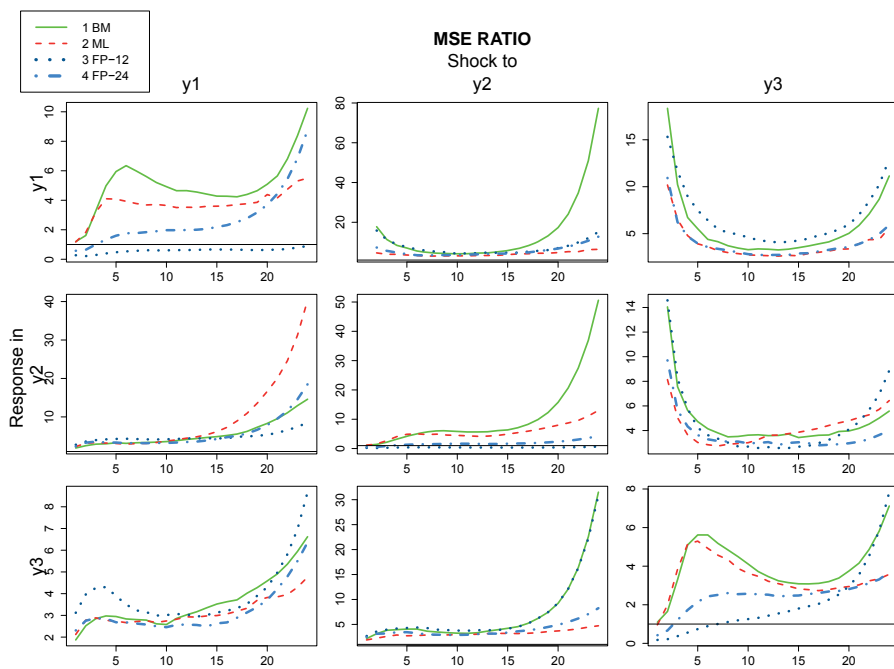


Figure A.3: Comparison between the four different approaches for DGPa and setup 4. For further explanation see Figure 5.

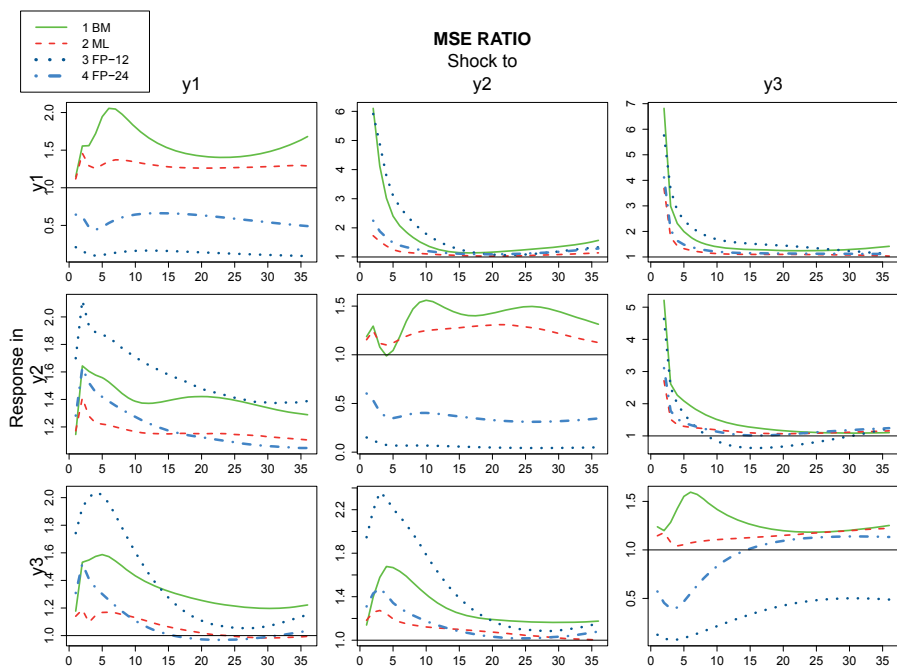


Figure A.4: Comparison between the four different approaches for DGPa and setup 5. For further explanation see Figure 5.

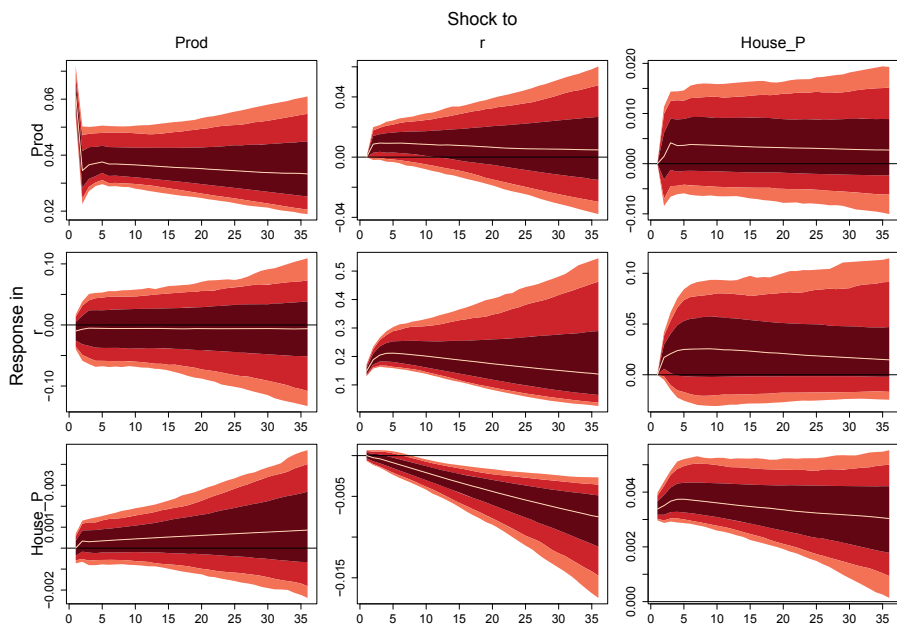


Figure A.5: Median impulse response functions with 68%, 90% and 95% error bands. Note that *Prod* is ordered first, *r* second and house prices last.

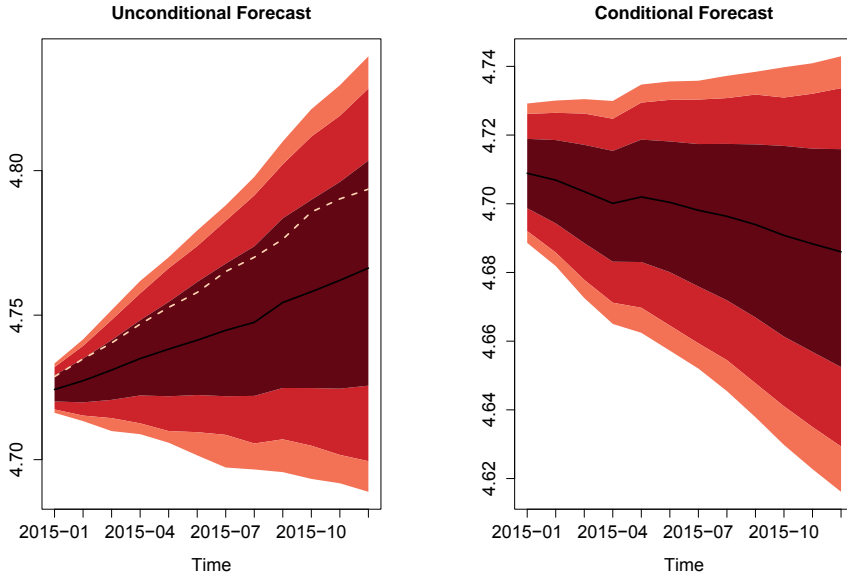


Figure A.6: Conditional and unconditional forecasts with 68%, 90% and 95% error bands for house prices in Germany. The conditional forecast is based on the assumption that interest rates are at 4% over the forecast horizon and the dotted line represents the actual values. Furthermore the variable production is included in the model.

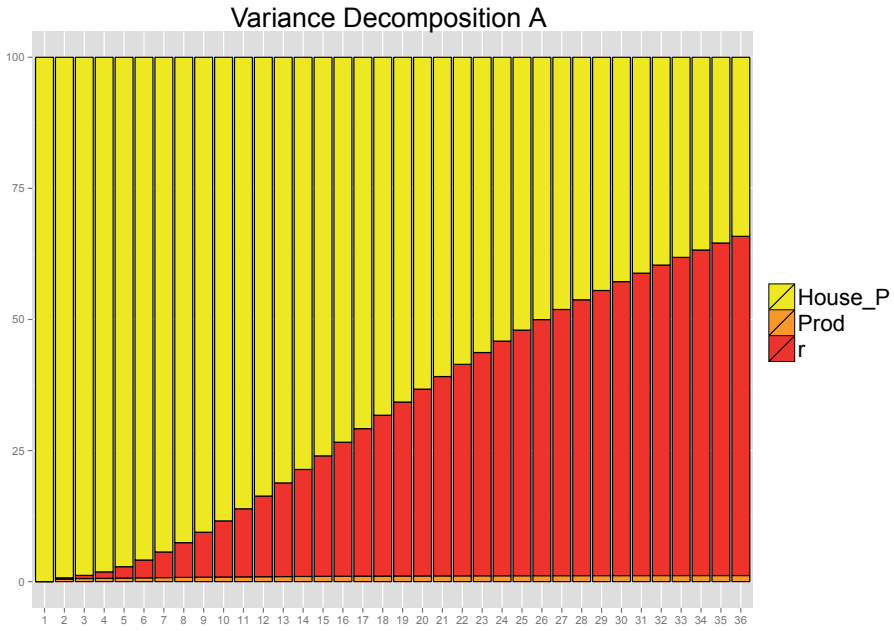


Figure A.7: The figure shows the forecast error variance decomposition in percent for house prices. Note that *Prod* is ordered first, *r* second and house prices last.

Appendix B. Tables

Table B.1: Residual Diagnostics

Lag Order	CPI	PROD	r	House prices
Autocorrelation				
1	0.9304085	0.5385309	0.5544260	0.7888553
2	0.6044125	0.4560631	0.6689359	0.9250233
3	0.5907237	0.1139220	0.8157002	0.9113061
4	0.6666178	0.1010447	0.8352627	0.8833816
Arch				
1	0.5238383	0.6821993	0.8314608	0.5940535
2	0.6090765	0.1569768	0.7811688	0.8565438
3	0.7369050	0.2934287	0.8740507	0.9307898
4	0.5550961	0.3250750	0.9140744	0.9751791

The table shows the p-values for autocorrelation and arch Ljung-Box tests.

Table B.2: Contemporaneous correlation between the residuals

	CPI	Prod	r	House prices
CPI	1.00	-0.11	0.25	0.18
Prod	-0.11	1.00	-0.04	0.01
r	0.25	-0.04	1.00	0.00
House_P	0.18	0.01	0.00	1.00