Essays in Modeling Fat Time Series Data using Bayesian Econometrics

Dissertation

zur

Erlangung des Grades Dr. rer. pol.
der Wirtschaftswissenschaftlichen Fakultät
der Universität Duisburg-Essen

vorgelegt von
Jan Prüser, Duisburg

Essen, September 2018

DOI: 10.17185/duepublico/70115
URN: urn:nbn:de:hbz:464-20190513-081931-2

Tag der mündlichen Prüfung: 16.04.2019
Erstgutachter: Prof. Dr. Christoph Hanck
Zweitgutachter: Prof. Dr. Ansgar Belke

Contents

Acknowledgments

I would like to thank my supervisor, Christoph Hanck, for all his support, for the freedom I was granted in my research and for his excellent comments on my thesis.

I am very grateful to Prof. Dr. Ansgar Belke for agreeing to be my second supervisor.

I wrote this thesis while being a student at the Ruhr Graduate School of Economics. I gratefully acknowledge the financial support of the Mercator Scholarship.

I also would like to thank Alexander Schlösser for a fruitful cooperation and for stimulating discussions.

Finally, I would like to thank all of my colleagues at the Ruhr Graduate School of Economics and at the department of Economic at the University of Duisburg-Essen. It was a great time.

Jan

1 Introduction

The major challenge facing empirical macroeconomics is how to successfully extract information from datasets to built models that are flexible enough to capture key data features without being overly flexible as to be seriously over-parametrized. Empirical researchers often find it necessary to allow for some sort of non-linearity and wish to include a large set of variables, but have only a few hundred observations. This raises problems for conventional methods of econometric inference. In order to address these problems, Bayesian methods turned out to be useful. The first purpose of this chapter is to provide the reader with an introduction to the Bayesian view with an overview of Bayesian theory (Section 1.1), computation (Section 1.2) and prior specification (Section 1.3). Section 1.4 gives an overview on how this thesis explores several aspects of Bayesian methods in modeling and forecasting macroeconomic and financial time series.

1.1 Bayesian Fundamentals

Bayesian econometrics is based on a few elementary rules in probability theory. These rules form the underlying principle on how an econometrician estimates model parameters, conducts inference, compares models and obtains predictions from a model. At the core is Bayes' Theorem, which states that

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(B|A)\,\mathrm{P}(A)}{\mathrm{P}(B)}, \tag{1.1}$$

for any random event A and B. In order to use this result for estimation and inference, assume that the random event B is a stochastic process y, which generates the observed data, and A is the parameter vector of interest θ. The Bayesian is interested in probability statements about the parameter given the observed data y and therefore treats the parameters in θ as random variables. Once we accept that unknown things like model coefficients, models and future events are random variables, the rest of the Bayesian approach is non-controversial as it follows the rules of probability. The treatment of θ as a random variable is controversial as frequentists view parameters as (unknown) constants. However, Bayesians econometrics is based on a subjective view of probability, which argues that our uncertainty about anything unknown can be expressed using the rules of probability. Taken it as given that econometrics involves learning about unknown coefficients in a statistical model, given the observed data and that the conditional probability of the unknown, given the known, is the best way of summarizing what we have learned, it is useful to write Bayes' theorem as

$$\pi(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{f(y)}, \tag{1.2}$$

where function $\pi(\theta|y)$ is the posterior density, $f(y|\theta)$ is the likelihood function, $f(y)$ is the marginal likelihood function and $\pi(\theta)$ is the prior density. When the researcher is only interested in the distribution of the parameter vector θ and not in the marginal likelihood $f(y)$, the denominator $f(y)$ can be dropped since it does not depend on θ and ensures that $\int \pi(\theta|y)\,\mathrm{d}\,\theta = 1$. Hence, the expression $f(y|\theta)\pi(\theta)$ does not integrate to one, but it has the same shape as $\pi(\theta|y)$. The resulting relationship

$$\pi(\theta|y) \propto f(y|\theta)\pi(\theta) \tag{1.3}$$

states that the posterior is proportional to the product of likelihood and prior. In the Bayesian philosophy the prior density summarizes what is known about θ before seeing the data. We update the a priori beliefs via the likelihood function. While the marginal likelihood is unimportant for the estimation of the parameters, it plays an important role for Bayesian model comparison. Given two competing models M_1 and M_2. The Bayes Theorem gives us the posterior probabilities for each model as

$$\pi(M_i|y) = \frac{m(y|M_i)\pi(M_i)}{m(y|M_2)\pi(M_1) + m(y|M_2)\pi(M_2)}, \tag{1.4}$$

with $m(y|M_i) = \int f(y|\theta_i, M_i)\pi(\theta_i)\,\mathrm{d}\,\theta_i$ and the posterior odds are given by

$$\frac{\pi(M_1|y)}{\pi(M_2|y)} = \frac{m(y|M_1)}{m(y|M_2)} \times \frac{\pi(M_1)}{\pi(M_2)}. \tag{1.5}$$

The model choice can be based on the posterior odds, but it is more common to use the Bayes factors $BF_{1,2} = m(y|M_1)/m(y|M_2)$ directly for model comparison. Furthermore, the posterior model probabilities give a theoretical justification for model averaging.

1.2 Bayesian Computation

The Bayesian is interested in techniques to compute the posterior distribution and whatever features of it are of interest.[1] These features typically include the expected value, the covariance matrix, univariate marginal densities of θ and their quantiles, and also functions of θ, denoted by $g(\theta)$. Such quantities can be expressed in general as

$$\mathrm{E}[g(\theta)|y] = \int g(\theta)p(\theta|y)d\theta, \tag{1.6}$$

where $p(\theta|y)$ is a proper density, i.e., $\int p(\theta|y)d\theta = 1$. For $g(\theta) = \theta$ we obtain the expected value of θ, for $g(\theta) = \theta\theta'$ we get the matrix of uncentered second moments and can use these to get the covariance matrix of θ. If the interest is in the posterior probability that the i-th element of θ lies in the interval (a, b), $g(\theta)$ can be defined as the indicator function that is equal to one if $\theta_i \in (a, b)$ and zero otherwise. Furthermore, for $g(\theta) = p(y_f|y, \theta)$, where y_f denotes out of sample observations to be predicted, we get the predictive density $p(y_f|y)$ of y_f.

For a few (simple) econometric models (for example the normal linear regression model), analytic results for the integral in (1.6) exist. However, in most econometric models the integral in (1.6) is not known analytically and numerical methods are required to compute them. One common technique to compute the integral as in (1.6) is Monte Carlo simulation. The basic idea of this tool is to simulate a large sample of θ which is distributed according to the posterior density $\pi(\theta|y)$. The sample of random draws with size R, $\{\theta^{(r)}\}_{r=1}^{R}$, can be used to consistently estimate $\mathrm{E}[g(\theta)|y]$ by the sample mean of the draws, i.e.

$$\frac{1}{R}\sum_{r=1}^{R} g(\theta^{(r)}) \xrightarrow{p} \mathrm{E}[g(\theta)|y] \tag{1.7}$$

as R goes to infinity. Since the researcher controls the simulation size, he can approximate the posterior mean arbitrarily well if he is patient enough.

The classic example, for which it is possible to obtain the posterior analytically, is the linear regression model under the normality assumption. The regression model with T observations and k regressors can be written as $y = X\beta + \epsilon$, where X is a $T \times k$ matrix of observations, β is the $k \times 1$ vector of coefficients and $\epsilon \sim N(0, \sigma^2 I)$. The likelihood function of the model is

[1] Note that this section draws heavily on Bauwens & Korobilis (2013).

$$Ł(\beta, \sigma^2; y, X) \propto (\sigma^2)^{-\frac{T}{2}} \exp\left[-\frac{(y - X\beta)'(y - X\beta)}{2\sigma^2}\right]. \tag{1.8}$$

Considering each regression coefficient and the logarithm of σ^2 as uniformly distributed on the real line corresponds to the noninformative prior $p(\beta, \sigma^2) \propto \frac{1}{\sigma^2}$.[2] Multiplying the prior with the likelihood according to formula (1.3) gives the posterior

$$p(\beta, \sigma^2|y, X) \propto (\sigma^2)^{-\frac{T+2}{2}} \exp\left[-\frac{(\beta - \hat{\beta})'X'X(\beta - \hat{\beta}) + s^2}{2\sigma^2}\right], \tag{1.9}$$

where $\hat{\beta} = (X'X)^{-1}X'y$ is the OLS estimator and $s^2 = (y - X\hat{\beta})'(y - X\hat{\beta})$ is the sum of squared OLS residuals. Integrating the above expression with respect to σ^2 yields the marginal posterior density for β, which is proper for $T > k$, given by

$$p(\beta|y, X) \propto \left[-\frac{(\beta - \hat{\beta})'X'X(\beta - \hat{\beta}) + s^2}{2\sigma^2}\right]^{-\frac{v+k}{2}}, \tag{1.10}$$

where $v = T - k$ is the degrees of freedom parameter. Formula (1.10) is the kernel of a multivariate t distribution and hence the posterior of β follows a multivariate t distribution. According to the properties of the t density, the posterior mean of β is given by $\hat{\beta}$ and its posterior covariance matrix is $s^2(X'X)^{-1}/(T - k - 2)$. Hence the posterior mean is equal to the OLS estimator (and the Maximum Likelihood estimator under the given assumptions) and the posterior covariance is close to the OLS estimator, which is given by $s^2(X'X)^{-1}/(T - k)$. The multivariate t distribution converges to the multivariate normal distribution as v tends to infinity. Consequently, for a sample size that is large relative to the number of regressors, the posterior in (1.10) can be well approximated by the $N(\hat{\beta}, s^2(X'X)^{-1}/(T-k))$ density.[3]

In order to obtain features of the posterior density $g(\beta)$ that are not available analytically (for example β_1/β_2), it is possible to generate draws from the posterior (which is feasible due to the known analytic form in the simple example above) and

[2] Note that if $p(\log \sigma^2) \propto 1$, then $p(\sigma^2) \propto 1/\sigma^2$.

[3] While this is an example of a model and prior which gives numerical results very similar from the frequentist and Bayesian perspective, it is worth to stress that the interpretation of the results is different. Under the Bayesian perspective, the unknown parameter β has a posterior density centered on $\hat{\beta}$ given for the observed unique sample, while from the frequentist perspective, the sampling distribution of $\hat{\beta}$ is centered on the unknown (true) value of β. Hence, the frequentist refers to a hypothetically experiment repeated infinitely many times of sampling the data from the population model, whereas the Bayesian just refers to the single sample that has been observed.

to compute $g(\beta)$ for each draw. The exact procedure is as follows:

1. Generate R draws $\{\theta^{(r)}\}_{r=1}^{R}$ of β from its posterior.
2. Calculate $g(\beta)$ for $r = 1, 2, \ldots, R$.
3. Use a kernel method (like a histogram) to obtain an estimate of the posterior density.

In most cases, multiplying the prior with the likelihood gives a mathematical expression of the posterior density which does not belong to any known family of densities or is not easy to simulate by direct sampling. Then, we cannot sample from the posterior directly and need to rely on other simulation methods. One simulation method, that turned out to be useful to sample from complex distributions, is called Markov chain Monte Carlo (MCMC). A Markov chain is defined by a stochastic process describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event. The term Markov property refers to the memoryless property of this process. More formally, a process $\{s_t\}, t = 1, 2, \ldots, T$, satisfies the Markov property if the probability of moving from one state s_t to the next s_{t+1} satisfies $P(s_{t+1}|s_t, s_{t-1}, \ldots, s_1) = P(s_{t+1}|s_t)$. If the variable is continuous, the above property holds using densities and P gets replaced by p. The MCMC algorithm involves a variable s_t with initial state s_0 and a transition density $p(s_{t+1}|s_t)$, which satisfies the Markov property. The basic idea behind MCMC is then to construct a Markov chain such that its limiting distribution is the target distribution (i.e. the posterior distribution). After the Markov chain has converged to the target distribution, the draws can be used to sample from the posterior. The Markov property implies that the draws from the posterior will not be independent. Nevertheless, due to ergodic theorems it is still valid to use these draws to estimate any functions of the parameters.

The most widely used MCMC sampling method is the Gibbs algorithm. In order to use the Gibbs sampler the parameter vector θ (of k elements) needs to be partitioned into b sub-vectors, with $b \leq k$, denoted by $\theta_{[i]}$, i.e. $\theta = (\theta'_{[1]}\theta'_{[2]} \ldots \theta'_{[b]})'$, such that the conditional posterior density $p(\theta_{[i]}|\theta_{-[i]}, y)$ can be directly simulated for each block. Then, a Gibbs sampler constructs a Markov chain using the full conditional distributions $p(\theta_{[i]}|\theta_{-[i]}, y)$ as the transition densities. Under certain regularity conditions, the limiting distribution of the Markov chain constructed in this way is the target distribution (i.e. the posterior distribution). In order to construct such a Markov chain and to simulate a sample of size R from the posterior $p(\theta|y)$ (after a burn in phase of R_0 draws), the algorithm proceeds as follows:

1. Choose an initial value $\theta_{-[1]}^{(0)}$ which belongs to the parameter space.

2. Set $r = 1$.
3. Draw successively

$$
\begin{aligned}
&\theta_{[1]}^{(r)} \text{ from } p(\theta_{[1]}|\theta_{-[1]}^{(r-1)}, y)\\
&\theta_{[2]}^{(r)} \text{ from } p(\theta_{[2]}|\theta_{[1]}^{(r)}, \theta_{[3]}^{(r-1)}, \dots, \theta_{[b]}^{(r-1)}, y)\\
&\vdots\\
&\theta_{[i]}^{(r)} \text{ from } p(\theta_{[i]}|\theta_{[1]}^{(r)}, \dots, \theta_{[i-1]}^{(r)}, \theta_{[i+1]}^{(r-1)}, \dots, \theta_{[b]}^{(r-1)}, y)\\
&\vdots\\
&\theta_{[b]}^{(r)} \text{ from } p(\theta_{[b]}|\theta_{-[b]}^{(r)}, y)
\end{aligned}
$$

4. Set $r = r + 1$ and go to step 3 unless $r > R_0 + R$.
5. Discard the first R_0 values of $\theta^{(r)}$. Use the last R draws to compute any function of interest.

If it is not possible to directly sample from the conditional posterior distributions, the Metropolis-Hastings (MH) algorithm can be applied. The MH algorithm can be used to sample from the joint posterior distribution directly. Though, it is more common to use it if it is not feasible to draw from one of the conditional posteriors in a Gibbs sampler scheme. Instead of directly sampling from this distribution, a MH algorithm can then be used to sample indirectly from it. Therefore it is called MH-with in Gibbs step. The basic idea of the MH algorithm is to take draws from a density that approximates the target (posterior) density. The drawn parameter values from the approximating density (called candidate density) are subject to an acceptance-rejection step to decide whether the drawn value (called candidate) is a draw of the posterior. If it is accepted, it is kept as a valid draw $\theta^{(r)}$ and if it is rejected, the previous accepted draw is accepted. Consequently, there will be sequences of identical draws, showing the dependence of the draws. In case the candidate density is identical to the target density, all draws will be accepted and the method corresponds to direct sampling of independent draws of the posterior. Hence, the candidate density should be constructed to be as close as possible to the target density. This, however, can be difficult in huge dimensions, which is the reason why the MH algorithm is typically used as an with-in-Gibbs step. The candidate density may depend on the last accepted draw and is denoted by $q(\theta|\theta^{(r-1)})$. The MH algorithm operates as follows:

1. Set $r = 1$ and choose an initial value $\theta^{(0)}$ which belongs to the parameter space.
2. Draw $\theta^{(cand)} \sim q(\theta|\theta^{(r-1)})$. Compute $\alpha = \min\left\{\frac{p(\theta^{(cand)}|y)}{p(\theta^{(r-1)}|y)}\frac{q(\theta^{(r-1)}|\theta^{(cand)})}{q(\theta^{(cand)}|\theta^{(r-1)})}, 1\right\}$.

3. Set $\theta^{(r)} = \theta^{(cand)}$ with probability α, and set $\theta^{(r)} = \theta^{(r-1)}$ with probability $1 - p$.
4. Set $r = r + 1$ and go to step 2 unless $r > R_0 + R$.
5. Discard the first R_0 values of $\theta^{(r)}$. Use the last R draws to compute any function of interest.

Step 3 can be implemented by drawing a random number U from a Uniform(0,1) density, and if $U < \alpha$ one sets $\theta^{(r)} = \theta^{(cand)}$ and otherwise $\theta^{(r)} = \theta^{(r-1)}$. The ratio in the min operator of step 2 is called the MH ratio. The probability of a candidate draw to be accepted depends on two factors. First, the probability is higher when the candidate draw is more likely than the previous draw under the posterior density and, second, the probability is higher when the candidate draw is less likely than the previous draw under the candidate density. The choice of the proposal density is crucial for the MH algorithm to work efficiently. In case the proposal density does not depend on the previous draw $\theta^{(r-1)}$, the proposal density produces independent candidate draws but is typically difficult to design to approximate the posterior distribution well. If the candidate density is symmetric in the sense that $q(\theta^{(r-1)}|\theta^{(cand)}) = q(\theta^{(cand)}|\theta^{(r-1)})$, the MH ratio simplifies. One way to design such a candidate density is the random walk MH algorithm. The random walk MH algorithm generates candidate values as $\theta^{(r-1)} + v$ where v is a draw from a distribution (typically normal) with zero mean and a variance matrix to be selected not overly small to allow the candidate draws to walk in the parameter space without staying overly close to the previous draws.

1.3 Prior Specification

The use of prior information is a controversial aspect of Bayesian methods and the need to specify a prior distribution is sometimes seen as a weakness of the Bayesian approach. However, the prior allows to formally incorporate additional information. This can be exploited to impose "structure" into the analysis. In case the researcher has no prior information, he can attend to use a non-informative prior.[4] As seen in the previous section, it is possible to use a uniform prior under which the mode of the posterior is equivalent to the maximum likelihood estimator. Or, if the researcher has very little information on which to base the prior, but has a large number of observations, he can select a portion of the sample as the training sample and estimate the prior from the training sample. However, in many cases prior information is available. For example, it is well known that VAR models are often overparametrized and that

[4] In this thesis a noninformative prior refers to a prior with a large variance and typically a mean of zero.

the most recent lags of a variable are expected to contain more information about the variable's current value than do earlier lags. One way frequentists typically respond to this is by using hypothesis tests to avoid an overly long lag length, using this information implicitly. Bayesians sometimes use this information explicitly by using the so-called Minnesota prior, due to Litterman (1986*a*), that formally incorporates this belief by shrinking earlier lags stronger than recent lags towards zero in order to avoid overfitting the data and still using a high number of lags. Hence, the prior can be interpreted as a form of regularization from the frequentist perspective and Litterman (1986*a*) argued from a largely frequentist viewpoint that the precision of estimates and forecasting performance can be improved by incorporating shrinkage in the form of a prior distribution on the parameters.

Shrinkage estimators and Bayesian estimators in general are attractive from an frequentist perspective because they (often) have better frequentist properties than frequentist estimators like the OLS or the maximum likelihood estimator. Results due to James & Stein (1961) show that the Maximum Likelihood estimator is inadmissible, but Bayesian estimators are admissible. Moreover, shrinkage estimators decrease the variance compared to the OLS estimator at the cost of introducing an estimation bias, which may lead, overall, to more precise estimates in a mean squared error sense. This is the reason why frequentists also consider shrinkage estimators to be useful. They derive them by adjusting the loss functions accordingly. Under a uniform prior the mode of the posterior corresponds to an frequentist estimate in a standard regression framework as shown in the previous section. The same can be shown for shrinkage estimators like the Lasso estimator. The Lasso estimator of β is a penalized least squares estimator where β is chosen to minimize:

$$(y - X\beta)'(y - X\beta) + \kappa \sum_{j=1}^{k} |\beta_j|, \tag{1.11}$$

where κ is a shrinkage parameter. The Lasso estimate of β is equivalent to Bayesian posterior modes if independent Laplace priors are placed on the elements of β (see Park & Casella (2008) for details).

For an applied researcher who wants to incorporate prior information (like the Minnesota prior to avoid overfitting), it may be a challenging task to determine the strength of the prior, i.e., its variance. Typically, the prior variance is designed in a way that it is a function of a few so called hyperparamters denoted by the vector λ, which have a similar effect as κ (assuming the prior mean is zero). The problem then boils down to selecting a few hyperparamters. Ideally, they should be selected to avoid overfitting without suppressing the signal in the data. One way to determine these hyperparamters is to use an empirical Bayes approach and estimate them from the data, for example by maximization of the marginal likelihood. The use of data-based information to specify the prior violates a philosophical basic premise of Bayesian methods. Nevertheless, empirical Bayes methods are becoming increasingly popular for researchers who are interested in objective tools that are useful in practice (see Berger (1985) for a discussion). Another way to estimate these hyperparameters which does not violate Bayesian principles and in addition accounts for uncertainty surrounding these hyperparameters, is to use hierarchical Bayes. Instead of assuming λ to be fixed, it is possible to assume that λ is also a random variable, and as such to also characterize it with some prior distribution. Since λ introduces additional variables to the model, the posterior becomes $\pi(\theta, \lambda|y)$ and not just $\pi(\theta|y)$. According to Bayes' rule we get

$$
\begin{aligned}
\pi(\theta, \lambda|y) &= \frac{f(y|\theta, \lambda)\pi(\theta, \lambda)}{f(y)}, \\
&= \frac{f(y|\theta, \lambda)}{f(y)} \frac{\pi(\theta, \lambda)}{\pi(\lambda)} \pi(\lambda), \\
&= \frac{f(y|\theta, \lambda)}{f(y)} \pi(\theta|\lambda)\pi(\lambda), \\
&\propto f(y|\theta, \lambda)\pi(\theta|\lambda)\pi(\lambda), \\
&\propto f(y|\theta)\pi(\theta|\lambda)\pi(\lambda).
\end{aligned}
$$

At the last stage, it is often convenient to use non-informative priors for $\pi(\lambda)$.

1.4 Contribution of this Thesis

This thesis explores several aspects of Bayesian methods for modeling and forecasting macroeconomic and financial time series. In Chapter 2, the interest lies on forecasting US inflation based on a Phillips curve augmented with additional predictors. While the original Phillips curve is likely to miss some important predictors, an augmented Phillips curve with too many predictors bears the risk of overfitting the data, leading to imprecise out-of-sample predictions. This raises the question of which predictors are relevant. However, the relevance of the predictors may change over time. In this case, only asking whether a variable is important or not is not addressing the right question. A researcher may not only be interested in assessing whether a variable is important, but rather when it is. The main contribution of this chapter is to address this issue by introducing a novel modeling approach called Markov Dimension

Switching (MDS). This approach allows for the calculation of time-varying variable inclusion probabilities to shed light on the question which variables are important in determining inflation at different times. Furthermore, the time-varying inclusion probabilities introduce a time-varying data-based shrinkage on the coefficients and thereby may avoid overfitting and hence can be a useful tool for forecasting. In a forecasting exercise, the MDS approach compares favorably to Bayesian variable selection which is unable to account for model change over time for one quarter and one year ahead inflation.

Chapter 3 extends the framework of dynamic model averaging (DMA) by introducing adaptive learning from model space. The conventional DMA approach consists of independently estimating K different time-varying parameter (TVP) models. In order to combine the different models, their individual forecasts are weighted by time-varying inclusion probabilities. This allows not only the marginal effect of each predictor to change over time but also allows the entire forecasting model to change over time. However, as each model is estimated independently, the information provided by the other models and the information provided from the time-varying inclusion probabilities is left unexploited in the process. In order to exploit the information in the estimation of the individual TVP models, this chapter proposes to not only average over the forecasts but, in addition, to also dynamically average over the time-varying parameters. This is done by approximating the mixture of individual posteriors with a single posterior in each period. By doing so, the information of all model posteriors is summarized in one single posterior, which is then used in the upcoming period as the prior for each of the individual models. This is attractive because it is often argued that pooling information is optimal relative to pooling forecasts; as the latter introduces an efficiency loss. The relevance of this extension is illustrated in three empirical examples involving forecasting US inflation, US consumption expenditures and forecasting of five major US exchange rate returns. In all applications adaptive learning from model space delivers improvements in out-of-sample forecasting performance.

Chapter 4 explores the use of Bayesian additive regression trees (BART) from the machine learning literature for forecasting with many predictors in a macroeconomic context. The BART model is attractive because it has variable selection built in and allows for non-linearities and interaction effects between predictors in a natural way. Government statistical agencies collect data on a wide range of macroeconomic variables, e.g. measures of output, capacity, employment and unemployment, prices, wages, housing, inventories, orders, stock prices, interest rates, exchange rates and monetary aggregates. Forecasting with many predictors provides the opportunity to exploit a much richer base of information. However, macroeconomic time series are typically rather short. A large number of predictors combined with only a small number of observations raises problems to conventional econometric methods. So far, factor models estimated with principal components and shrinkage methods like the Lasso estimator have been found to be useful to address these problems. The main contribution of this chapter is to provide a forecasting comparison between the BART, Lasso and Factor approaches using a large data set of over one hundred quarterly US macroeconomic time series. It turns out that BART outperforms the Lasso and Factor approach and therefore is a valuable addition to existing methods for handling large dimensional data sets in a macroeconomic context.

Chapter 5 is joint work with Alexander Schlösser. It studies the time-varying impact of Economic Policy Uncertainty (EPU) on the US Economy by using a VAR with time-varying coefficients. The coefficients are allowed to evolve gradually over time which allows to discover structural changes without imposing them a priori. The use of priors provides the opportunity to regularize the amount of time-variation. While an overly loose prior may result in overfitting, an overly tight prior may suppress possible time-variation, which is the main interest of this chapter. The strength of these priors depends on a small set of hyperparameters which are estimated using a new Bayesian approach, instead of using benchmark values typically used in other applications. Moreover, the VAR model is augmented with a few factors which capture information from a large data set without introducing a degrees of freedom problem. This allows to simultaneously investigate the impact of EPU on variables which represent real economic activity and on variables which mirror the activity on financial markets and turns out to be empirically important since EPU has an impact on a wide range of different variables. The VAR with time-varying coefficients reveals three different regimes which match the three major business cycles of the US economy, namely the Great Inflation, the Great Moderation and the Great Recession. This finding is in contrast to previous literature which typically imposes two regimes a priori.

Chapter 6 is also joint work with Alexander Schlösser. It exploits the econometric framework from Chapter 5 in order to investigate the effects of Economic Policy Uncertainty (EPU) on a wide range of macroeconomic variables for eleven European Monetary Union (EMU) countries. In particular, it allows to address the following three questions: First, has EPU potentially harmed growth in the Euro area since the introduction of the Euro or has the effect of EPU changed since the financial crisis? Put differently, has the transmission or the size of EPU shocks varied over time? Second, the theory of optimum currency areas suggests that asymmetric shocks undermine business cycle synchronization. Therefore, it is important to investigate whether all countries respond similarly to an EPU shock. Third, three major transmission channels have been identified in the theoretical literature. The first describes how

EPU affects investment, the second investigates the effect on private consumption and the third explains how EPU affects financial variables. This raises the question which channels are most important for each country and which variables are affected most by EPU shocks. In turns out that the transmission of EPU shocks is quite stable over time. Moreover, the model reveals that a group of fragile countries (GIIPS-countries) suffered most due to EPU shocks compared to a group of stable countries (Northern countries). Finally, it turns out that EPU shocks are transmitted through various channels, such as the real options, the financial and the precautionary savings channel and private investors and financial market participants react more sensitively than consumers to EPU shocks.

Chapter 7 is joint work with Christoph Hanck. It uses a Bayesian VAR in order to provide empirical evidence 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 VAR model is a useful choice to study the complex dynamic interrelationship among macroeconomic variables. However, macroeconomic times series are rather short (like in the empirical application of this chapter) and VARs are richly parametrized. This may lead to the risk of overfitting the data, possibly leading to imprecise inference and inaccurate forecasts. The Bayesian VAR can employ prior information to shrink the model parameter, potentially avoiding such overfitting. This chapter provides 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. The insights from this simulation study are used to select the informativeness of the prior in the empirical application. To choose relevant control variables, a new Bayesian variable selection approach by Ding and Karlsson (2014) is used. In addition to impulse responses and variance decompositions, a Bayesian conditional forecast to test the hypothetical effect of an increase of interest rates on house prices is employed.

2 Forecasting US Inflation using Markov Dimension Switching

This chapter considers Bayesian variable selection in the Phillips curve context by using the Bernoulli approach of Korobilis (2013*c*). The Bernoulli model, however, is unable to account for model change over time, which is important if the set of relevant predictors changes. To tackle this problem, this paper extends the Bernoulli model by introducing a novel modeling approach called Markov Dimension Switching (MDS). MDS allows the set of predictors to change over time. The MDS and Bernoulli model reveal that inflation expectations, the growth rate of the oil price and the Treasury bill rate are the most important variables for one quarter inflation. For one year inflation the unemployment rate, inflation expectations, the Treasury bill rate and the number of newly built houses turn out to be the most important variables. Furthermore, the relevant predictors exhibit a sizable degree of time variation for which the Bernoulli approach is not able to account, stressing the importance and benefit of the MDS approach. In a forecasting exercise the MDS model compares favorably to the Bernoulli model for one quarter and one year ahead inflation.

2.1 Introduction

The Phillips curve has served as an important tool in macroeconomics for explaining and forecasting inflation in the US over the past five decades. In the original Phillips curve, inflation depends on lags of inflation and the unemployment rate. In order to obtain a better understanding and potentially more precise forecasts, a large literature extends the Phillips curve with additional explanatory variables. Influential papers include Stock & Watson (1999), Atkeson & Ohanian (2001), Ang et al. (2007), Stock & Watson (2007) and Groen et al. (2013). Forecasting inflation is crucial, e.g., for central banks, but at the same time challenging. One difficulty arises from the problem of which additional variables to include in the Phillips curve. While the original Phillips curve is likely to miss some important predictors, an augmented Phillips curve with too many predictors bears the risk of overfitting the data, leading to imprecise out-of-sample predictions. This raises the question of which predictors are relevant. However, the relevance of the predictors may change over time. In this case, only asking whether a variable is important or not is not addressing the right question. A researcher may not be interested in assessing whether a variable is important, but rather when it is.

This chapter addresses the question of which predictor is relevant by following Korobilis (2013*c*) and considers Bayesian variable selection in the Phillips curve context. Korobilis (2013*c*) provides an algorithm for stochastic variable selection. The key idea is to introduce an indicator for each predictor, which determines if a variable is included in the model. Each indicator is drawn from a Bernoulli distribution in a Gibbs sampler scheme. By doing so, it is possible to calculate variable inclusion probabilities to assess the importance of single predictors in determining inflation. However, a potential drawback is that the set of indicators is assumed to be constant over time. Thus, the Bernoulli approach is unable to account for model change over time, which is desirable if the set of relevant predictors changes over time. The importance of changing predictors over time is documented by, inter alia, Stock & Watson (2010), who find that most predictors for inflation improve forecast performance only in some specific time periods. Therefore, it may be empirically important for predictors to change over time. Conventional hypothesis testing approaches designed for constant parameter models are also not capable to allow for this, as they only test whether a restriction holds for all time periods or never.[1] The main contribution of this chapter is to tackle this problem by introducing a novel modeling approach called Markov Dimension Switching (MDS). The MDS model can be seen as an extension of the Bernoulli model. In the MDS model each indicator follows a Markov-switching process and thus allows for changing predictors over time. Hence, this approach allows for the calculation of time-varying variable inclusion probabilities to shed light on the question which variables are important in determining inflation at different times.

The relevance of this extension is illustrated by using the Bernoulli and the MDS approach to assess the importance of the predictors for one quarter and one year inflation. Most important predictors for one quarter turn out to be inflation expectations, the percentage change of the oil price and the Treasury bill rate. The unemployment rate, inflation expectations, the Treasury bill rate and the number of newly built houses turn out to be the most important predictors for one year ahead inflation. The relevant variables show a sizable degree of time variation, which the

[1] Furthermore, the Bayesian methods used in this chapter have the advantage that they allow for a formal treatment of model uncertainty. Using hypothesis tests to select a parsimonious model ignores model uncertainty, as the selected model is assumed to be the one which generated the data. Treating one model as if it were the “true” model and ignoring the huge number of other potential models may be seen as problematic.

Bernoulli approach can not account for, highlighting the benefit and importance of the proposed MDS approach of this chapter. In particular MDS reveals that the relevance of inflation expectations, unemployment and house prices for the one year horizon changes abrupt over time, which would be difficult to capture for existing methods which assume a gradually change of the relevance of predictors. From an economic perspective it is particular interesting that the relevance of unemployment rate changes that rapidly as it has long been assumed that economic policymakers face a trade-off between unemployment and inflation. This result, however, suggests that this inverse relation might not be stable over time and that a break down of the Phillips curve may only be temporary. Furthermore, this chapter investigates the forecasting performance of both approaches. It turns out that the MDS approach exhibits a better forecasting performance than the Bernoulli approach for one quarter and one year ahead inflation. An additional finding is that the forecasting performance of the MDS approach is competitive in comparison with a range of other plausible approaches.

The remainder of this chapter is organized as follows. Section 2 lays out and discusses the econometric framework. Section 3 presents the empirical findings and the last section concludes.

2.2 Markov Dimension Switching

The Phillips curve serves as a starting point and motivation for many models that forecast inflation. In the original Phillips curve, inflation depends only on the unemployment rate and lags of inflation. Including additional predictors, as Stock & Watson (1999) among many others do, leads to the so-called generalized Phillips curve

$$\pi_{t+h} = \alpha + \sum_{j=0}^{p-1} \phi_j \pi_{t-j} + \boldsymbol{x_t} \boldsymbol{\beta} + \epsilon_{t+h}, \tag{2.1}$$

where $\boldsymbol{x_t}$ is a $1 \times q$ vector of exogenous predictors, $\pi_{t+h} = \log(P_{t+h}) - \log(P_t)$, P_t denotes the price level and $\epsilon_t \sim N(0, \sigma_t^2)$. The number of parameters may be large relative to the number of observations, as in many macroeconomic applications. Estimation of the Phillips curve in this case may cause imprecise estimation and overfitting (i.e., the model fits the noise in the data, rather than finding the pattern useful for forecasting). Both, imprecise estimation and overfitting translate into inaccurate out-of-sample predictions. Hence, it is important to identify the truly relevant predictors out of a set of many potentially relevant predictors. To do so, this chapter follows Korobilis (2013*c*) and considers Bayesian variable selection in the Phillips curve context by introducing $m = q + p + 1$ indicators $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)$. The model can now be written as

$$\pi_{t+h} = (\boldsymbol{z}_t \odot \boldsymbol{\gamma})\boldsymbol{\theta} + \epsilon_{t+h}, \tag{2.2}$$

where $\boldsymbol{z}_t = (1, \pi_t, \ldots, \pi_{t-p+1}, \boldsymbol{x}_t)$, $\boldsymbol{\theta} = (\alpha, \phi_0, \ldots, \phi_{p-1}, \boldsymbol{\beta}')'$ and $\odot$ denotes elementwise multiplication. Hence, if $\gamma_i = 1$, the ith variable is included in the model and if $\gamma_i = 0$, it is not. By sampling the indicators from their posterior, all 2^m possible variable combinations can be considered and estimated in a stochastic manner. A potential drawback, however, is that the indicators are constant over time. Thus, a predictor is either included or excluded from the model for all periods, which is undesirable if the set of predictors changes over time. To address this problem, this chapter introduces MDS to allow the indicator variables to change over time. In the MDS each indicator variable γ_i follows a first-order Markov-switching process $S_{i,t}$ and therefore $\boldsymbol{\gamma}$ now has a time index t:

$$\pi_{t+h} = (\boldsymbol{z}_t \odot \boldsymbol{\gamma}_t)\boldsymbol{\theta} + \epsilon_{t+h}, \tag{2.3}$$

where $\boldsymbol{\gamma}_t = (S_{1,t}, \ldots, S_{m,t})$. Each Markov switching process $S_{i,t}$ can take on the value one or zero and is characterized by a 2×2 transition matrix $\boldsymbol{\mu}_i$, where $\mu_{kj,i} = \Pr(S_{i,t+1} = j | S_{i,t} = k)$, $k = 0, 1$ and $j = 0, 1$.[2] If $S_{i,t} = 1$, the ith variable is included in the model at period t and if $S_{i,t} = 0$, it is not. Therefore, the means of the posterior draws of $S_{i,t}$ can be interpreted as a time-varying variable inclusion probability in this modeling context. Furthermore, note that keeping $\boldsymbol{\theta}$ constant does not imply that a certain variable has either an impact of zero or an impact given by $\boldsymbol{\theta}$. This is because the time-varying inclusion probabilities introduce a time-varying data-based shrinkage on the coefficients. Therefore, MDS may avoid overfitting and hence can be a useful tool for forecasting. In contrast, estimating $\boldsymbol{\theta}$ in a time-varying manner bears a high risk of overfitting and can empirically only poorly approximate changing predictors by allowing coefficients to be estimated as being approximately zero. Furthermore, models with time-varying parameters typically assume a gradually change in parameters and therefore are not well suited to capture abrupt changes in the relevance of predictors.

[2] The Markov mixture modeling approach allows that the probability of switching depends on the current state of the stochastic process, which is not the case for i.i.d. mixture models, but may be useful to model dependence over time and allows to formulate different prior beliefs about the frequency of dimension switching and the level of sparsity in the model, see section 2.1. The i.i.d. case is however nested as a special case of the Markov mixture approach.

2.2.1 Gibbs Sampler

This section describes the Gibbs Sampler, which allows to draw from the posterior distribution of the Bernoulli and the MDS model.

1. Sample $\boldsymbol{\theta}$ from the following density

$$\boldsymbol{\theta}|\boldsymbol{\gamma}_{1:T}, \boldsymbol{z}_{1:T}, \pi_{1+h:T+h}, \sigma^2_{1+h:T+h} \sim N(\overline{\boldsymbol{\theta}}, \overline{\boldsymbol{\Omega}}), \tag{2.4}$$

   with

$$\overline{\boldsymbol{\theta}} = \overline{\boldsymbol{\Omega}} \left(\boldsymbol{V}(\hat{\boldsymbol{\theta}}_{OLS})\hat{\boldsymbol{\theta}}_{OLS} + \sum_{t=1}^{T} (\boldsymbol{z}_t \odot \boldsymbol{\gamma}_t)' \sigma^{-2}_{t+h} \pi_{t+h} \right),$$

$$\overline{\boldsymbol{\Omega}} = \left(\boldsymbol{V}(\hat{\boldsymbol{\theta}}_{OLS}) + \sum_{t=1}^{T} (\boldsymbol{z}_t \odot \boldsymbol{\gamma}_t)' \sigma^{-2}_{t+h} (\boldsymbol{z}_t \odot \boldsymbol{\gamma}_t) \right)^{-1}.$$

   For the prior, the OLS estimate of the full model is used. When one variable is omitted from the model for the full sample period, the parameter of this predictor is drawn from the prior. In order to obtain reasonable draws in this case, the OLS estimate of the model seems to be a useful choice. Then the mean of the posterior of $\boldsymbol{\theta}$ is the weighted average of the OLS estimate of the full model and the OLS estimate using only a subset of the predictors. While the OLS estimate of the full model likely has a higher variance as it is likely to include irrelevant predictors, the OLS estimate based on the sparse data matrix is more likely to suffer from omitted variables bias. Hence, the posterior addresses the classic bias variance trade-off in a convenient way by placing weights on both estimates in a data-driven way.

2. Sample $\boldsymbol{\gamma}_t$:

   - If γ_i is constant, sample it from

$$\gamma_i|\gamma_{-i}, \pi_{1+h:T+h}, \boldsymbol{z}_{1:T}, \boldsymbol{\theta}, \sigma^2_{1+h:T+h} \sim \text{Bernoulli}\left(\frac{l_{1i}}{l_{1i} + l_{0i}}\right), \tag{2.5}$$

     with

$$l_{1i} = \exp\left(-\frac{1}{2}\sum_{t=1}^{T}\left(\frac{\pi_{t+h} - (\boldsymbol{z}_t \odot \boldsymbol{\gamma}_{[\gamma_i=1]})\boldsymbol{\theta}}{\sigma^2_{t+h}}\right)^2\right) p(\gamma_i = 1),$$

$$l_{0i} = \exp\left(-\frac{1}{2}\sum_{t=1}^{T}\left(\frac{\pi_{t+h} - (\boldsymbol{z}_t \odot \boldsymbol{\gamma}_{[\gamma_i=0]})\boldsymbol{\theta}}{\sigma^2_{t+h}}\right)^2\right) p(\gamma_i = 0),$$

where $p(\gamma_i = 1) = 0.5$.

- In the MDS model $S_{i,t}$ is sampled for $t = 1, \ldots, T$ conditioning on $\boldsymbol{\gamma}_{-i,1:T}$, $\pi_{1+h:T+h}$, $\boldsymbol{z}_{1:T}$, $\boldsymbol{\theta}$, $\sigma^2_{1+h:T+h}$ and the transition probabilities of the ith Markov process $\boldsymbol{\mu}_i$ using the algorithm of Chib (1996) (see Appendix B for details). The transition probabilities of the ith Markov process are drawn from a Beta distribution

$$\mu_{11,i}|S_{i,1:T} \sim \text{Beta}(u_{11} + n_{11}, u_{10} + n_{10}), \tag{2.6}$$

$$\mu_{00,i}|S_{i,1:T} \sim \text{Beta}(u_{00} + n_{00}, u_{01} + n_{01}), \tag{2.7}$$

where n_{jk} counts the number of transitions from state j to k and u_{jk} is the prior hyperparamter. Setting $u_{11} = u_{00} = u_{10} = u_{01} = 1$ corresponds to the uniform prior. The posterior is not sensitive to this prior choice if none of the four possible transitions is rare. However, it is also possible to use a more informative prior. For example a researcher may want to avoid a high frequency of regime changes and smooth the variable inclusion probability over time. Thus, once we are in a regime, i.e. a variable is excluded or included in the model, the regime should only be switched if there is a strong signal in the data. This prior belief can be implemented by setting $u_{11} = u_{00} = T$. Sparse models are typically known to forecast better than models with overly many variables. A stronger favor for sparse models would be achieved by setting only $u_{00} = T$. All three prior parametrizations, i.e. the uniform, the smooth and the sparse prior, are considered in the empirical part.

3. Sample σ_t^{-2}:
   - In the case of homoscedastic errors where $\sigma_t^2 = \sigma^2$, sample from the density

$$\sigma^{-2}|\boldsymbol{\theta}, \pi_{1+h:T+h}, \boldsymbol{z}_{1:T}, \boldsymbol{\gamma}_{1:T} \sim \text{Gamma}(a, b^{-1}), \tag{2.8}$$

   where $a = T + a_0$ and $b = b_0 + \sum_{t=1}^{T}(\pi_{t+h} - (\boldsymbol{z}_t \odot \boldsymbol{\gamma}_t)\boldsymbol{\theta})^2$.
   The hyperparameters a_0 and b_0 are set to zero.
   - In the case of heteroscedastic errors, sample conditioning on $\boldsymbol{\theta}$, $\pi_{1+h:T+h}$, $\boldsymbol{z}_{1:T}$, $\boldsymbol{\gamma}_{1:T}$ using the algorithm of Kim et al. (1998) by assuming that

$$\log(\sigma_t) = \log(\sigma_{t-1}) + \xi_t, \tag{2.9}$$

   where $\xi_t \sim N(0, \zeta)$ and ζ is sampled from

$$\zeta^{-1}|\sigma^2_{1+h:T+h} \sim \text{Gamma}(a, b^{-1}), \tag{2.10}$$

where $a = T + \kappa_1$ and $b = \kappa_2 + \sum_{t=1+h}^{T+h}(\log(\sigma_t) - \log(\sigma_{t-1}))^2$. The hyperparameters κ_1 and κ_2 are set to 3 and 0.0001.

2.2.2 Comparison with Existing Literature

A growing literature works with Bayesian priors in models with many parameters, which shrink some of the parameters towards zero to ensure parsimony. For example, Bańbura et al. (2010) find that shrinking parameters leads to improved forecasts in large VAR models. There is also an increasing number of papers applying shrinkage by using hierarchical priors, such as the lasso prior introduced by Park & Casella (2008). Hierarchical priors have the advantage that the priors introducing the shrinkage dependent on unknown parameters which are estimated from the data, resulting in data-driven shrinkage. For example, Korobilis (2013*b*) shows that hierarchical shrinkage is useful for macroeconomic forecasting using many predictors. In a Phillips curve context, Belmonte et al. (2014) use the lasso prior in a time-varying parameter (TVP) model. The lasso prior in their model automatically decides which parameter is time-varying, constant or shrunk towards zero. This approach may be well suited to model structural changes in the Phillips curve while avoiding overfitting.

Fewer papers deal with model change over time as opposed to parameter change (which empirically can only poorly approximate model change by allowing coefficients to be estimated as being approximately zero). Chan et al. (2012) consider dimension switching in a TVP framework using the algorithm of Gerlach et al. (2000). However, in their forecasting study, they only consider models with no predictors, a single predictor or all m predictors. In other words, $\boldsymbol{\gamma}$ can only take on $m + 2$ values and not 2^m as this would be computationally infeasible for the algorithm they used. To consider all variable combinations, dynamic model averaging (DMA) can be applied, using approximations in form of so called forgetting factors (sometimes also called discount factors) as proposed by Raftery et al. (2010). Koop & Korobilis (2012) find that DMA leads to substantial improvements in forecasting inflation over simple benchmark models and more sophisticated approaches. DMA assigns time-varying weights over the set of 2^m possible TVP models.[3]

In contrast to DMA or hierarchical shrinkage, the MDS model has the advantage that through the indicator variables the likelihood contains information about the relevance of every predictor at each point in time and thereby may lead to more efficient estimates. In the DMA approach each model is estimated independently and does not use the information of the time-varying weights. For example, at the

[3]In the empirical application only the intercept and the first lag are always included.

beginning of the sample, the most weight may be placed on models with only a few predictors and at the end of the sample more weight may be assigned to model with a large set of predictors. However, each individual model is estimated using the same set of predictors for the whole sample ignoring this information. It would be useful to take this information into account when estimating the parameters and this is exactly what the MDS model does. In the hierarchical shrinkage approach some parameters are shrunk towards zero (i.e., the corresponding variables are irrelevant), but this information is only contained in the prior and not in the likelihood function. Furthermore, this approach cannot account for model change over time, as it shrinks the parameters towards zero for all time periods or never.

Moreover, in contrast to DMA, the MDS model does not need approximations. It can easily be estimated using Gibbs sampling and thereby take full parameter uncertainty into account. Another potential drawback is that in the DMA approach all model combinations have to be estimated in a deterministic fashion, while MDS uses a stochastic search algorithm. The stochastic search is still feasible when the model space is too large to be assessed in a deterministic manner by visiting only the most probable models in a stochastic manner. Despite the potential advantages of MDS, the assumption of constant parameters may appear restrictive. However, this assumption is less restrictive than it seems, as the time-varying inclusion probabilities introduce a time-varying data-based shrinkage on the coefficients. Therefore, MDS addresses overfitting concerns and allows for model change over time.

2.3 Forecasting Inflation

2.3.1 Data

This study forecasts core inflation as measured by the Personal Consumption Expenditure (PCE) deflator for 1978Q2 through 2016Q4. The period from 1992Q1 to 2016Q4 is used to evaluate the out-of-sample forecast performance. A wide range of variables is considered as potential predictors, reflecting the major theoretical explanations of inflation as well as variables which have been found to be useful in forecasting inflation in other studies. The following predictors are used:

- DJIA: the percentage change in the Dow Jones Industrial Average.
- EMPLOY: the percentage change in employment.
- HSTARTS: the log of housing starts.
- INFEXP: University of Michigan survey of inflation expectations.

- MONEY: the percentage change in the money supply (M1).
- OIL: the percentage change of Spot Crude Oil Price: WTI
- PMI: the change in the Institute of Supply Management (Manufacturing): Purchasing Managers Composite Index.
- CONS: the percentage change in real personal consumption expenditures.
- GDP: the percentage change in real GDP.
- INV: the percentage change in Real Gross Private Domestic Investment (Residential)
- SPREAD: the spread between the ten year and three month Treasury bill rates.
- TBILL: three month Treasury bill (secondary market) rate.
- UNEMP: unemployment rate.
- CAPUT: the change in Capital Utilization (Manufacturing).

The variables are obtained from the "Real-Time Data Set for Macroeconomists" database of the Philadelphia Federal Reserve Bank and from the FRED database of the Federal Reserve Bank of St. Louis. All predictors are real time quarterly data so that all forecasts are made using versions of the variables available at the respective time. Furthermore, all data are seasonally adjusted if necessary. If not stated otherwise, all models considered in the next section include four lags of quarterly inflation as additional predictors. This is consistent with quarterly data.

2.3.2 Out-of-sample Results

In this section, the forecasting performance of the MDS model is investigated. In a first step, MDS and Bernoulli models are considered in which the first lag of inflation and the intercept are always included and all other variables are allowed to be omitted from the model. In order to assess whether the MDS or the Bernoulli approach is useful to avoid overfitting, their forecast performance is compared with an AR(1) model with intercept and a multiple regression model containing all variables. Furthermore, the uniform, the smooth and the sparse prior for the transition probabilities are compared. All these models are applied with a constant and a stochastic variance specification as described in the description of the Gibbs Sampler.

In a second step the forecasting performance of the MDS model is compared with two modeling approaches which have been found useful in inflation forecasting. These approaches are DMA proposed by Koop & Korobilis (2012) and the hierarchical shrinkage in TVP-models proposed by Belmonte et al. (2014). For DMA, three forgetting factors have to be set by the researcher. The first controls the amount of time variation in the coefficients, the second the amount of time variation of the volatility and the third controls the amount of time variation of the model probabilities (see Koop & Korobilis (2012) for details). Setting these forgetting factors to one leads to the special case of constant coefficients, constant variance and a constant model probabilities. Values close to one are typically used in the literature because of overfitting concerns. Koop & Korobilis (2012) set the hyperparamter for the variance to 0.98 and set the foregtting factors for the coefficients and model probabilities to either 0.95 or 0.99, which they find to deliver a favorable forecasting performance over simple benchmark regressions and more sophisticated approaches. Thus, this set of values is used to forecast inflation. Moreover, dynamic model selection (DMS) is considered next to DMA in the forecasting comparison. In the TVP-model with hierarchical shrinkage the specification of the hierarchical gamma prior is crucial, see Belmonte et al. (2014) for details. In the application the shape and scale parameter of the inverse gamma prior is set to 0.1 leading to a relatively non-informative prior. As a special case of this model, the lasso prior by Park & Casella (2008) in a regression model with constant coefficients is also considered using the same hierarchical inverse gamma prior. Furthermore, the last two models are estimated using the same two specifications for the variance as for the MDS models.

In order to evaluate the forecast performance, the root mean squared forecast error (RMSFE) and the mean absolute forecast error (MAFE) as standard forecast metrics are used. However, these only evaluate the point forecasts and ignore the remaining part of the predictive distribution. This is the reason why the predictive likelihood may be preferable to evaluate the forecast performance. The predictive likelihood is the predictive density for π_{t+h} (given data through time t) evaluated at the actual outcome. As a forecast metric it has the advantage of evaluating the forecast performance of the entire predictive density. Additionally, the predictive likelihood can also be used for model selection. Therefore, the mean of the log predictive likelihood is used as an additional forecast metric. For a motivation and detailed description of the predictive likelihood see Geweke & Amisano (2010).

Table 2.1 contains the results for the one quarter and one year ahead forecasting performance. Overall, it turns out that the MDS models forecast quite well. For one quarter ahead inflation the forecasting performance of the Bernoulli model is similar to the forecasting performance of the AR(1) model and the full model containing all predictors. For one year ahead inflation the full model seems to overfit the data, as it forecasts poorly. Variable selection in the Bernoulli model delivers forecasting improvements over the full model including all predictors, but does not improve over the simple AR(1) model. Forecasting improvements over the full model and a simple

Table 2.1: Forecasting performance for one quarter and one year inflation

		$h = 1$			$h = 4$		
Model	Variance	RMSFE	MAFE	PL	RMSFE	MAFE	PL
MDS uniform	constant	0.34	0.22	4.05	1.25	0.99	2.91
MDS uniform	stochastic	0.34	0.22	4.09	1.17	0.92	2.98
MDS sparse	constant	0.33	0.21	4.08	1.19	0.90	2.84
MDS sparse	stochastic	0.33	0.21	4.07	1.10	0.83	3.04
MDS smooth	constant	0.34	0.22	4.09	1.15	0.90	2.87
MDS smooth	stochastic	0.33	0.22	4.19	1.14	0.90	2.93
Bernoulli	constant	0.35	0.24	2.25	1.35	1.08	2.44
Bernoulli	stochastic	0.35	0.24	2.87	1.36	1.08	2.52
AR(1)	constant	0.37	0.24	2.84	1.35	0.95	2.63
AR(1)	stochastic	0.37	0.24	2.91	1.35	0.95	2.66
Full model	constant	0.37	0.24	0.70	1.40	1.14	2.27
Full model	stochastic	0.37	0.24	1.58	1.39	1.13	2.19
LASSO	constant	0.36	0.24	3.13	1.39	1.12	2.55
LASSO	stochastic	0.36	0.24	3.60	1.34	1.10	2.88
TVP-shrink	constant	1.56	1.01	2.14	2.68	1.95	1.68
TVP-shrink	stochastic	0.42	0.29	3.50	1.42	1.09	2.73
DMA (0.95)	stochastic	0.35	0.23	3.91	1.06	0.81	2.88
DMA (0.99)	stochastic	0.35	0.22	4.03	1.18	0.86	2.82
DMS (0.95)	stochastic	0.37	0.25	4.08	1.18	0.92	2.96
DMS (0.99)	stochastic	0.35	0.22	4.05	1.24	0.91	2.80

The table shows the RMSFE and MAFE in percentage points and the mean log predictive likelihood (PL).

AR(1) model can be achieved by considering dynamic variable selection in the form of MDS. The MDS models forecast better than the Bernoulli models, both in terms of point forecasts and in terms of the predictive likelihood as a forecasting metric. The different priors for the transition probabilities deliver a very similar forecasting performance for one quarter and one year ahead inflation. The specification of the variance turns out to be not crucial. An exception is the TVP regression model, which forecasts poorly with a constant variance specification, as the time-varying coefficients falsely fit the time-varying volatility rather than finding a pattern useful for forecasting in this case. Furthermore, the hierarchical shrinkage in TVP and constant coefficient regression produces less precise forecasts than the MDS models. Only the DMA and DMS (which also allows for changing predictors) approach show a similar forecasting performance compared to the MDS models. This finding stresses the importance of allowing for changing predictors over time using the Phillips curve to forecast inflation.

2.3.3 Full Sample Results

The calculation of variable inclusion probabilities is interesting from an economic perspective, but may also provide an explanation why MDS models provide better inflation forecasts than the Bernoulli models. Figures 2.1 and 2.2 display the inclusion probabilities of the MDS model with the uniform, the smooth and the sparse prior for the transition probabilities and the Bernoulli model for the full sample. The inclusion probabilities are shown for the stochastic variance specification. Overall, the Bernoulli approach assigns higher inclusion probabilities to the variables than the MDS models. This may be one reason why the MDS models deliver better forecasts. Another reason may be that some inclusion probabilities show a sizable degree of time variation, for which the Bernoulli approach cannot account. This demonstrates the usefulness of the MDS model over the Bernoulli model. Comparing the three different priors for the MDS models reveal that under the smooth prior the variable inclusion probabilities are less noisy, as a stronger signal is needed to obtain a regime change compared to the uniform prior. In addition, the sparse prior yields more parsimonious models, as a stronger signal in the data is needed for a variable to be included in the model. However, sometimes the signal in the data is strong enough to yield similar inclusion probabilities for the different prior specifications.

In many cases the Bernoulli model and the MDS model under the uniform prior deliver similar results. In some cases the MDS model even assigns a roughly constant inclusion probability to a variable. In other cases the MDS model also assigns a high probability to one variable, but the probability changes over time. For one quarter inflation *INEXP*, *OIL* and *TBILL* turn out to be important in all approaches and for one year inflation *INEXP*, *HSTARTS*, *UNEMP* and *TBILL* turn out to be important. In particular for one year inflation these variables show a sizeable degree of time variation. The inclusion probabilities of *INEXP*, *HSTARTS* and *UNEMP* switch very rapid over time. This shows that the relevance of predictors does not always change gradually, like it is assumed for example in TVP models. From an economic perspective it is particular interesting that the relevance of *UNEMP* changes that rapidly as it has long been assumed that economic policymakers face a trade-off between unemployment and inflation. These results, however, suggest that this relation might not be stable over time.

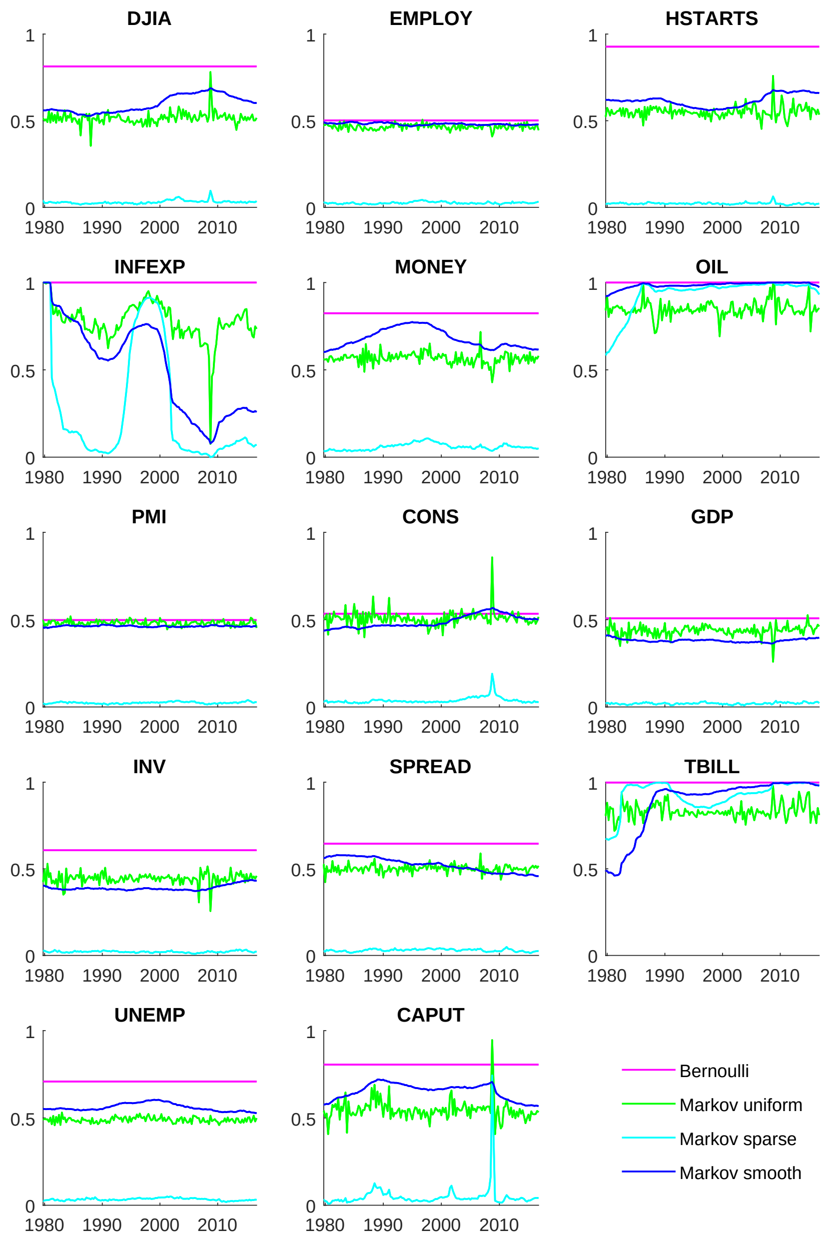


Figure 2.1: Variable inclusion probabilities for one quarter inflation.

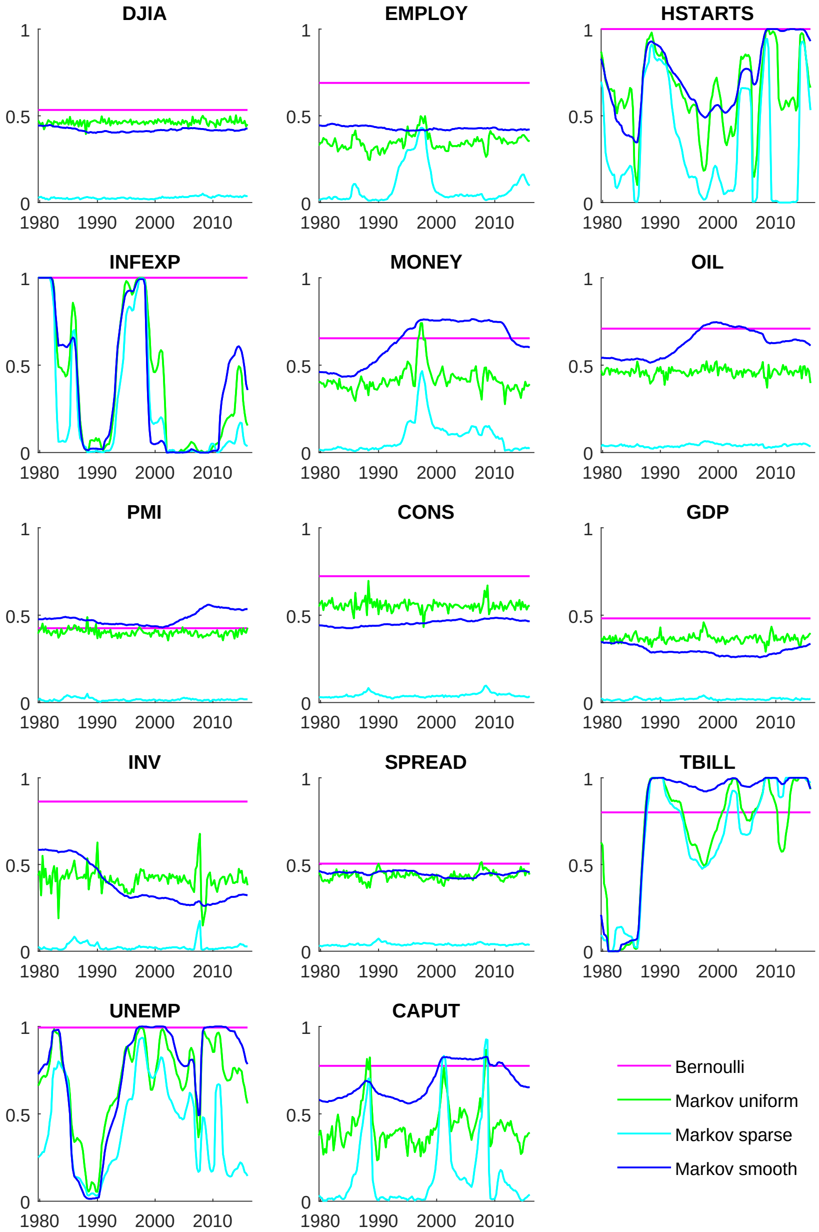


Figure 2.2: Variable inclusion probabilities for one year inflation.

2.4 Conclusion

This study uses the generalized Phillips curve to forecast inflation. While the original Phillips curve is likely to miss some important predictors, a generalized Phillips curve which uses too many predictors may lead to overfitting the data and to imprecise out-of-sample predictions. Thus, this chapter aims to assess which variables are important in determining inflation by using the Bernoulli model. The Bernoulli model, however, is unable to account for model change over time. In order to be able to account for the possibility that the set of predictors changes over time, this chapter introduces the Markov Dimension Switching (MDS) approach. In the MDS approach the set of predictors is allowed to change over time. The empirical application shows that the most important variables in the generalized Phillips curve are inflation expectations, the percentage change of the oil price and the Treasury bill rate for one quarter inflation and the unemployment rate, the Treasury bill rate and the number of newly built houses for one year inflation. Furthermore, for one year inflation the unemployment rate, the Treasury bill rate and the number of newly built houses show a sizeable degree of time variation for which the Bernoulli approach is not able to account, highlighting the importance and benefit of the MDS approach. This is also confirmed in a forecasting exercise, where the MDS model delivers more precise forecasts than the Bernoulli model for one quarter and one year ahead inflation. In addition, the chapter demonstrates that the forecasting performance of the MDS model is competitive in comparison with a range of other plausible alternatives. Taken together, the chapter presents a battery of theoretical and empirical arguments for the potential benefits of the MDS approach.

2.A Gibbs Sampling in Markov Switching Models

This chapter considers Markov switching for each variable. Each Markov switching process S_t can take on the value one or zero and is characterized by a 2×2 transition matrix $\boldsymbol{\mu}$ where $\mu_{kj} = \Pr(S_{t+1} = j|S_t = k)$, $k = 0, 1$ and $j = 0, 1$.[4] In order to draw S_t for $t = 1, \ldots, T$ first the Hamilton filter, proposed by Hamilton (1989), is used followed by the simulation smoother of Chib (1996):

1. Initialize the Hamilton filter using steady state probabilities:
$$\Pr(S_0 = 0) = \frac{1 - \mu_{11}}{2 - \mu_{11} - \mu_{00}},$$
$$\Pr(S_0 = 1) = \frac{1 - \mu_{00}}{2 - \mu_{11} - \mu_{00}}.$$

2. Given $\Pr(S_{t-1} = k|\psi_{t-1})$, where ψ_{t-1} denotes the information set at time point $t-1$, calculate $\Pr(S_t = j|\psi_{t-1})$ as
$$\Pr(S_t = j|\psi_{t-1}) = \sum_{k=0}^{1} \mu_{kj} \Pr(S_{t-1} = k|\psi_{t-1}).$$

3. Given ψ_t update the probabilities as
$$\Pr(S_t = j|\psi_t) = \frac{\mathrm{f}(y_t|S_t = j, \psi_{t-1})\Pr(S_t = j|\psi_{t-1})}{\sum_{j=0}^{1} \mathrm{f}(y_t|S_t = j, \psi_{t-1})\Pr(S_t = j|\psi_{t-1})},$$
where $\mathrm{f}(y_t|S_t = j, \psi_{t-1})$ denotes the likelihood function of the dependent variable.

4. Sample S_T using $\Pr(S_t = T|\psi_T)$.

5. Sample $S_{T-1}, \ldots, S_1$ sequentially using
$$\Pr(S_t = 1|S_{t+1}, \psi_t) = \frac{\Pr(S_{t+1}|S_t = 1)\Pr(S_t = 1|\psi_t)}{\sum_{j=0}^{1} \Pr(S_{t+1}|S_t = j)\Pr(S_t = j|\psi_t)},$$
where $\Pr(S_{t+1}|S_t = j)$ denotes the transition probability and $\Pr(S_t = j|\psi_t)$ is saved from step 3.

[4] For a simplified notation the index i is omitted and the general case of a two state Markov process is considered.

3 Adaptive Learning from Model Space

Dynamic model averaging (DMA) is used extensively for the purpose of economic forecasting. This chapter extends the framework of DMA by introducing adaptive learning from model space. In the conventional DMA framework all models are estimated independently and hence the information of the other models is left unexploited. In order to exploit the information in the estimation of the individual time-varying parameter models, this chapter proposes to not only average over the forecasts but, in addition, to also dynamically average over the time-varying parameters. This is done by approximating the mixture of individual posteriors with a single posterior, which is then used in the upcoming period as the prior for each of the individual models. The relevance of this extension is illustrated in three empirical examples involving forecasting US inflation, US consumption expenditures and forecasting of five major US exchange rate returns. In all applications adaptive learning from model space delivers improvements in out-of-sample forecasting performance.

3.1 Introduction

Forecasting in economics is challenging for three major reasons. First, the existence of many potential predictors can result in a huge number of potential models. While regressions with many predictors may overfit, small models may miss important predictors. This leads to the need for model selection strategies. Second, a useful forecasting model may change over time. For instance, some variables may predict well in recessions while others may predict well in expansions or the set of relevant predictors may change between certain events such as the Great Moderation. This further complicates the statistical problem as a researcher needs to select one model in each period. Third, in case of parameter change the marginal effect of predictors may change over time. However, modeling such change will increase the risk of overfitting the data, resulting in poor out-of-sample predictions. Recently, a growing literature addresses these points by using dynamic model averaging (DMA), proposed by Raftery et al. (2010). Koop & Korobilis (2012) introduce DMA to the economic literature by forecasting inflation. They find a favorable forecasting performance of DMA over simple benchmark regressions and more sophisticated approaches. Studies that use DMA to forecast a variety of different economic time series include:

Buncic & Moretto (2015), Drachal (2016) and Naser (2016) forecasting commodities, Bruyn et al. (2015), Beckmann & Schüssler (2016) and Byrne et al. (2018) forecasting exchange rates, Liu et al. (2015) forecasting stock returns, Gupta et al. (2014) forecasting foreign exchange reserves, Bork & Moller (2015), Risse & Kern (2016) and Wei & Cao (2017) forecasting house price growth, Aye et al. (2015) and Baur et al. (2016) forecasting gold prices, Koop & Korobilis (2011) and Filippo (2015) forecasting inflation and Wang et al. (2016) and Liu et al. (2017) forecasting realized volatility.

While conventional DMA is well established in the economic literature, the aim of this chapter is to extend this framework by introducing adaptive learning from model space (ALM). The conventional DMA approach consists of independently estimating K different time-varying parameters (TVP) models. In order to combine the different models, their individual forecasts are weighted by time-varying inclusion probabilities. The time-varying inclusion probabilities depend on the most recent forecasting performance of each model and allow that the weight placed on each model may change over time. However, as each model is estimated independently the information provided by the other models, the information provided from the time-varying inclusion probabilities is left unexploited in the process. In order to exploit the information in the estimation of the individual TVP models, this chapter proposes to not only average over the forecasts but, in addition, to also dynamically average over the time-varying parameters. This is done by approximating the mixture of individual posteriors with a single posterior in each period. By doing so, the information of all model posteriors is summarized in one single posterior, which is then used in the following period as the prior for each of the individual models. This is attractive because it is often argued that pooling information is optimal relative to pooling forecasts; as the latter introduces an efficiency loss, see Timmermann (2006). For instance it may be the case that at the beginning of the sample most weight is placed on parsimonious models while later in the sample more weight is placed on models with a larger set of predictors. However, these models cannot benefit from this information when estimated independently. In contrast, when averaging over the time-varying parameters some of the variables in the larger models may be shrunk to zero at the beginning of the sample by exploiting the information that they were not relevant at this time.

The relevance of this extension is illustrated in three empirical applications. In the first application, both conventional DMA and the enhanced version ALM are used to forecast US inflation one quarter and one year ahead. Under different settings, ALM compares favorably to conventional DMA. The second application considers forecasting nominal and real US consumption expenditures one quarter and one year ahead. For nominal and real consumption expenditures ALM outperforms conventional DMA. Finally, the third application forecasts five major US end-of-month (log) exchange rate returns one month and one year ahead. It turns out that ALM delivers more precise forecasts than the conventional DMA for all five countries. The finding that ALM yields improvements in out-of-sample forecasting holds in particular for the long horizon in all three applications.

The remainder of this chapter is organized as follows. Section 2 lays out and discusses the econometric framework. Section 3 presents the empirical findings and the last section concludes.

3.2 Econometric Framework

3.2.1 Baseline Dynamic Model Averaging

Consider a set of TVP models M_k, $k = 1, \dots, K$, which can be written as

$$y_t = \boldsymbol{z}_t^{(k)} \boldsymbol{\theta}_t^{(k)} + \epsilon_t^{(k)}, \tag{3.1}$$

$$\boldsymbol{\theta}_t^{(k)} = \boldsymbol{\theta}_{t-1}^{(k)} + \boldsymbol{\eta}_t^{(k)}, \tag{3.2}$$

where $\epsilon_t^{(k)} \sim N(0, H_t^{(k)})$ and $\boldsymbol{\eta}_t^{(k)} \sim N(\mathbf{0}, \boldsymbol{Q}_t^{(k)})$. The predictor vector $\boldsymbol{z}_t^{(k)}$ for each model can be of different dimension and the predictor set of the different models need not overlap. Let $L_t = k$ if the process is modelled by model M_k at time t. Conditioning on $L_t = k$, the state vector $\boldsymbol{\theta}_t^{(k)}$ of each model can be estimated independently using the Kalman filter. Assuming that $\boldsymbol{\theta}_{t-1}^{(k)} | L_{t-1} = k, \boldsymbol{y}^{t-1} \sim N(\hat{\boldsymbol{\theta}}_{t-1|t-1}^{(k)}, \boldsymbol{\Sigma}_{t-1|t-1}^{(k)})$, Kalman filtering proceeds using

$$\boldsymbol{\theta}_t^{(k)} | L_{t-1} = k, \boldsymbol{y}^{t-1} \sim N(\hat{\boldsymbol{\theta}}_{t|t-1}^{(k)}, \boldsymbol{\Sigma}_{t|t-1}^{(k)}), \tag{3.3}$$

where

$$\hat{\boldsymbol{\theta}}_{t|t-1}^{(k)} = \hat{\boldsymbol{\theta}}_{t-1|t-1}^{(k)} \tag{3.4}$$

and

$$\boldsymbol{\Sigma}_{t|t-1}^{(k)} = \boldsymbol{\Sigma}_{t-1|t-1}^{(k)} + \boldsymbol{Q}_t^{(k)}. \tag{3.5}$$

Followed by the updating equations to complete the estimation

$$\boldsymbol{\theta}_t^{(k)}|L_t = k, \boldsymbol{y}^t \sim N(\hat{\boldsymbol{\theta}}_{t|t}^{(k)}, \boldsymbol{\Sigma}_{t|t}^{(k)}), \tag{3.6}$$

where

$$\hat{\boldsymbol{\theta}}_{t|t}^{(k)} = \hat{\boldsymbol{\theta}}_{t|t-1}^{(k)} + \boldsymbol{\Sigma}_{t|t-1}^{(k)} \boldsymbol{z}_t^{\prime(k)} (H_t^{(k)} + \boldsymbol{z}_t^{(k)} \boldsymbol{\Sigma}_{t|t-1}^{(k)} \boldsymbol{z}_t^{\prime(k)})^{-1} (y_t - \boldsymbol{z}_t^{(k)} \hat{\boldsymbol{\theta}}_{t|t-1}^{(k)}) \tag{3.7}$$

and

$$\boldsymbol{\Sigma}_{t|t}^{(k)} = \boldsymbol{\Sigma}_{t|t-1}^{(k)} - \boldsymbol{\Sigma}_{t|t-1}^{(k)} \boldsymbol{z}_t^{\prime(k)} (H_t^{(k)} + \boldsymbol{z}_t^{(k)} \boldsymbol{\Sigma}_{t|t-1}^{(k)} \boldsymbol{z}_t^{\prime(k)})^{-1} \boldsymbol{z}_t^{(k)} \boldsymbol{\Sigma}_{t|t-1}^{(k)}. \tag{3.8}$$

In order to run the Kalman filter, one needs to know the variance $H_t^{(k)}$ of the observation equation and the covariance matrix $\boldsymbol{Q}_t^{(k)}$ of the transition equation. Estimating or simulating $H_t^{(k)}$ and $\boldsymbol{Q}_t^{(k)}$ running Markov-Chain-Monte-Carlo methods for each model would be computationally demanding. Therefore, it is convenient to use approximations that allow to estimate each model with only one iteration of the Kalman filter. The covariance matrix $\boldsymbol{Q}_t^{(k)}$ appears in the Kalman Filter only in equation (3.5). Following Raftery et al. (2010), equation (3.5) is replaced by

$$\boldsymbol{\Sigma}_{t|t-1}^{(k)} = \frac{1}{\lambda} \boldsymbol{\Sigma}_{t-1|t-1}^{(k)} \tag{3.9}$$

or, equivalently, $\boldsymbol{Q}_t^{(k)} = (1 - \lambda^{-1}) \boldsymbol{\Sigma}_{t-1|t-1}^{(k)}$, where λ is called the forgetting factor with $0 < \lambda \leq 1$. Forgetting factor approaches have a long tradition in the state space literature and a detailed motivation is given, e.g., by Fagin (1964) and Jazwinsky (1970). The forgetting factor implies that observations which are lagged by i periods receive the weight λ^i. The idea is similar to applying a rolling window regression with a window size of $\frac{1}{1-\lambda}$. Typically the value for λ is set close to one, in order to favor a gradual evolution of coefficients. Raftery et al. (2010) set $\lambda = 0.99$. For quarterly data, this means that observations from five years ago receive around 80% as much weight as observation of the last period. Setting $\lambda = 0.95$ implies that observations five years ago receive only 35% as much weight as the observation of the last period and would allow for higher degrees of parameter change. This suggests that the range of plausible values should be close to one. In section 2.3 a way to estimate λ over a small grid of values is discussed. With this simplification there is no need to estimate $\boldsymbol{Q}_t^{(k)}$ anymore. To estimate each model with only one Kalman filter iteration requires a method for estimating $H_t^{(k)}$. Following Koop & Korobilis (2012), $H_t^{(k)}$ is estimated by using an Exponentially Weighted Moving Average (EWMA)

$$H_t^{(k)} = \kappa H_{t-1}^{(k)} + (1-\kappa)(y_t - \boldsymbol{z}_t^{(k)} \hat{\boldsymbol{\theta}}_{t|t-1}^{(k)})^2, \tag{3.10}$$

with $0 < \kappa \leq 1$. This estimator is a weighted average of $H_{t-1}^{(k)}$ and the squared residuals at time t, with κ a decay factor, similar to the forgetting factor λ, with effective window size $\frac{\kappa}{2} - 1$. Therefore, the value for κ is also typically set close to one. RiskMetrics (1996) set $\kappa = 0.97$ for monthly data and Koop & Korobilis (2012) set $\kappa = 0.98$ for quarterly data. See RiskMetrics (1996) for general properties of the EWMA estimator. After having replaced $\boldsymbol{Q}_t^{(k)}$ and $H_t^{(k)}$ with equations (3.9) and (3.10), all results are available in closed form and only one iteration of the Kalman filter is required for the estimation of each model.

As a next step, a way to combine the models is needed. Raftery et al. (2010) propose to calculate time-varying model probabilities by using the following model prediction equation with forgetting factor α:

$$\pi_{k|t-1} = \frac{\pi_{k|t-1}^{\alpha}}{\sum_{l=1}^{K} \pi_{l|t-1}^{\alpha}} \tag{3.11}$$

and a model updating equation

$$\pi_{k|t} = \frac{\pi_{k|t-1} p_k(y_t|\boldsymbol{y}^{t-1})}{\sum_{l=1}^{K} \pi_{l|t-1} p_l(y_t|\boldsymbol{y}^{t-1})}, \tag{3.12}$$

where p_k denotes the predictive likelihood of model k. The predictive likelihood is a measure of forecasting performance and is defined as the predictive density evaluated at the actual outcome y_t. The predictive density is given by

$$y_t|L_{t-1} = k, \boldsymbol{y}^{t-1} \sim N(\boldsymbol{z}_t^{(k)} \hat{\boldsymbol{\theta}}_{t|t-1}^{(k)}, H_t^{(k)} + \boldsymbol{z}_t^{(k)} \boldsymbol{\Sigma}_{t|t-1}^{(k)} \boldsymbol{z}_t^{\prime(k)}). \tag{3.13}$$

In order to understand the role of the forgetting factor α, write $\pi_{t|t-1,k}$ as

$$\pi_{k|t-1} \propto \prod_{i=1}^{t-1} [p_k(y_{t-i}|y^{t-i-1})]^{\alpha^i}. \tag{3.14}$$

Thus, the model probabilities change over time according to the forecasting performance (measured by the predictive likelihood) in the recent past of each model. The forgetting factor α discounts the past forecasting performance in the same fashion as the forgetting factor λ and therefore controls the frequency of model change. As a special case, $\alpha = 1$ corresponds to conventional model averaging using the marginal likelihood. Hence, similar considerations as for λ apply and suggest to set α close to one. Section 2.3 discusses a way to estimate α. Now, given the model probabilities we can forecast using DMA via

$$\hat{y}_t^{DMA} = \sum_{l=1}^{K} \pi_{k|t-1} \boldsymbol{z}_t^{(k)} \hat{\boldsymbol{\theta}}_{t|t-1}^{(k)} \tag{3.15}$$

or using dynamic model selection (DMS) as

$$\hat{y}_t^{DMS} = \boldsymbol{z}_t^{(k^*)} \hat{\boldsymbol{\theta}}_{t|t-1}^{(k^*)}, \tag{3.16}$$

where k^* refers to the model with the maximum model probability at time $t-1$. Thus, each model is estimated independently using the Kalman filter and then either the forecast of each individual model is weighted by its probability at period $t-1$ or the forecast of one model is selected with the highest probability at time $t-1$.

3.2.2 Adaptive Learning from Model Space

The aim of this chapter is to go one step further by not only performing DMA over the individual forecast of each model but also doing DMA over the time-varying model parameters $\boldsymbol{\theta}_t^{(k)}$. Using conventional DMA, each model is estimated independently and cannot exploit the information in the other models. By averaging over $\boldsymbol{\theta}_t^{(k)}$ each model is not estimated independently anymore and uses the information of all other $K-1$ models. This is attractive because it may be the case that at the beginning of the sample most weight is placed on parsimonious models while later in the sample more weight is placed on models with a larger set of predictors. However, these models cannot benefit from this information when estimated independently. In contrast, when averaging over $\boldsymbol{\theta}_t^{(k)}$, some of the variables in the larger models may be shrunk to zero at the beginning of the sample by exploiting the information that they were not relevant at this time.

In order to average over $\boldsymbol{\theta}_t^{(k)}$ at each period t, a single Gaussian $q(\boldsymbol{\theta}_t) = N(\boldsymbol{\theta}_t | \overline{\boldsymbol{\theta}}_{t|t}, \overline{\boldsymbol{\Sigma}}_{t|t})$ is used to approximate a mixture of Gaussians $p(\boldsymbol{\theta}_t) = \sum_{k=1}^{K} \pi_{k|t} N(\boldsymbol{\theta}_t | \hat{\boldsymbol{\theta}}_{t|t}^{(k)}, \boldsymbol{\Sigma}_{t|t}^{(k)})$.[1] The Kullback-Leibler divergence is a measure of the dissimilarity between two distributions. Therefore, the two moments $\overline{\boldsymbol{\theta}}_{t|t}$ and $\overline{\boldsymbol{\Sigma}}_{t|t}$ can be determined

[1]Note that the vector $\boldsymbol{z}_t^{(k)}$ may include a different set of variables for each model k. Hence, $\hat{\boldsymbol{\theta}}_{t|t}^{(k)}$ may also correspond to a different set of variables. In order to account for this, zeros can be placed in the corresponding elements of $\boldsymbol{z}_t^{(k)}$ in case certain variables are not included.

by minimizing the Kullback-Leibler divergence between $q(\boldsymbol{\theta}_t)$ and $p(\boldsymbol{\theta}_t)$, denoted by $\mathrm{KL}(q||p)$, with respect to $\overline{\boldsymbol{\theta}}_{t|t}$ and $\overline{\boldsymbol{\Sigma}}_{t|t}$. The minimization problem is given by $q = \arg\min_q \mathrm{KL}(q||p)$ and the solution to this problem is given by

$$\overline{\boldsymbol{\theta}}_{t|t} = \sum_{k=1}^{K} \pi_{k|t} \hat{\boldsymbol{\theta}}_{t|t}^{(k)}, \tag{3.17}$$

$$\overline{\boldsymbol{\Sigma}}_{t|t} = \sum_{k=1}^{K} \pi_{k|t} (\boldsymbol{\Sigma}_{t|t}^{(k)} + (\hat{\boldsymbol{\theta}}_{t|t}^{(k)} - \overline{\boldsymbol{\theta}}_{t|t})(\hat{\boldsymbol{\theta}}_{t|t}^{(k)} - \overline{\boldsymbol{\theta}}_{t|t})'). \tag{3.18}$$

In the graphical model literature, this is called weak marginalization, as it preserves the first two moments, see Lauritzen (1992). The center of the distribution $q(\boldsymbol{\theta}_t)$ is just the weighted average of the mean $\hat{\boldsymbol{\theta}}_{t|t}^{(k)}$ from all individual models. Thus, models with a higher posterior probability receive more weight. The total variance $\boldsymbol{\Sigma}_{t|t}^{(k)}$ arises from two sources of variability. The first source is the weighted average of the covariance-matrix $\boldsymbol{\Sigma}_{t|t}^{(k)}$ of the individual models and reflects the uncertainty about $\boldsymbol{\theta}_t$ which comes from the estimation of the individual models. And the second source is the weighted average of the squared difference between $\hat{\boldsymbol{\theta}}_{t|t}^{(k)}$ and $\overline{\boldsymbol{\theta}}_{t|t}$ and reflects the uncertainty through the heterogeneity between the different models. In order to complete estimation, the posterior $q(\boldsymbol{\theta}_t)$ serves as a prior for each model in the upcoming period and $\hat{\boldsymbol{\theta}}_{t-1|t-1}^{(k)}$ and therefore $\boldsymbol{\Sigma}_{t-1|t-1}^{(k)}$ in equation (3.4) and (3.5) are replaced by $\overline{\boldsymbol{\theta}}_{t-1|t-1}$ and $\overline{\boldsymbol{\Sigma}}_{t-1|t-1}$. Thus, the time-varying parameter vector $\boldsymbol{\theta}_t$ is estimated by exploiting the information of all K models and its elements are shrunken towards the parameters of models that receive a higher weight. For example, they can be shrunk towards zeros, if models that include the corresponding variable receive only little weight or they may be shrunken towards the value they have in models which receive a high weight.

3.2.3 Estimation of Hyperparamter

In order to estimate a model, one has to determine the values of the hyperparameters λ, κ and α. Previous consideration suggests to set them close to one. Furthermore, the two hyperparamters λ and κ are estimated over a grid of values by treating different values as different models, i.e. by setting $\lambda = \lambda^{(k)}$ and $\kappa = \kappa^{(k)}$. Thus, if models discounting past data more strongly yield a better forecasting performance (measured by the predictive likelihood) in the recent past (which is controlled by α), a higher weight is placed on them. However, it is not possible to estimate the forgetting factor α in this fashion. Fortunately, Beckmann & Schüssler (2016) provide a way to sum over a grid of values $\alpha_v \in (\alpha_1, \alpha_2, \dots, \alpha_a)$ by replacing equations (3.11) and (3.12) with

$$\pi_{k|t-1} = \sum_{v=1}^{a} \pi_{k|t-1,\alpha_v} p(\alpha_v | I_{t-1}), \tag{3.19}$$

where $\pi_{k|t-1,\alpha_v} = \frac{\pi_{k|t}^{\alpha_v}}{\sum_{l=1}^{K} \pi_{l|t}^{\alpha_v}}$ and

$$\pi_{k|t} = \sum_{v=1}^{a} \frac{p_k(y_t|\boldsymbol{y}^{t-1})\pi_{k|t-1,\alpha_v}}{\sum_{l=1}^{K} p_l(y_t|\boldsymbol{y}^{t-1})\pi_{k|t-1,\alpha_v}} p(\alpha_v|I_t). \tag{3.20}$$

The posterior at time t of a particular grid point of the forgetting factor α is given by

$$p(\alpha_z|I_t) = \frac{\sum_{k=1}^{K} p_k(y_t|\boldsymbol{y}^{t-1})\pi_{k|t-1,\alpha_z} p(\alpha_z|I_{t-1})}{\sum_{v=1}^{a}\sum_{l=1}^{K} p_l(y_t|\boldsymbol{y}^{t-1})\pi_{l|t-1,\alpha_v} p(\alpha_v|I_{t-1})}. \tag{3.21}$$

3.3 Empirical Applications

3.3.1 Forecasting Inflation

This section considers one quarter and one year ahead forecasts for core inflation as measured by the Personal Consumption Expenditure (PCE) deflator. A standard set of variables is considered as potential predictors, reflecting the major theoretical explanations of inflation as well as variables which have been found to be useful in forecasting inflation in other studies. Potential predictors are the percentage change in the Dow Jones Industrial Average, the percentage change in employment, the log of housing starts, University of Michigan survey of inflation expectations, the percentage change in the money supply (M1), the percentage change of Spot Crude Oil Price (WTI), the change in the Institute of Supply Management index (Manufacturing), the percentage change in real personal consumption expenditures, the percentage change in real GDP, the percentage change in real Gross Private Domestic Investment (Residential), the spread between the ten year and three month Treasury bill, the three month Treasury bill and the unemployment rate. In addition, following Koop & Korobilis (2012), each model contains one intercept and two lags of inflation. All variables are quarterly, seasonally adjusted and are obtained from the FRED database of the Federal Reserve Bank of St. Louis. The data are observed for the period 1978Q2 to 2016Q3 and the period from 1992Q1 to 2016Q3 is used to evaluate the out-of-sample forecast performance.

Table 3.1: Forecasting performance for one quarter and one year inflation

		$h = 1$		$h = 4$	
Model	Hyerparamter	RMSFE	MAFE	RMSFE	MAFE
DMA	grid	0.35	0.21	1.10***	0.84
DMS	grid	0.40	0.24	1.09***	0.86
ALM	grid	0.36	0.23	1.00	0.80
DMA	0.99	0.37	0.22	1.22***	0.92
DMS	0.99	0.36	0.23	1.27***	0.96
ALM	0.99	0.35	0.22	1.05	0.82
DMA	0.95	0.36	0.23	1.08***	0.83
DMS	0.95	0.40	0.25	1.22***	0.97
ALM	0.95	0.37	0.23	1.00	0.80

The table shows the RMSFE and MAFE in percentage points for three different settings of the forgetting factor α (controls the change between models) and λ (controls the change in coefficients). In the first ($\alpha = \lambda = 0.95$), in the second ($\alpha = \lambda = 0.99$) and in the third both are estimated from a small grid $\alpha, \lambda \in (0.95, 0.99)$. The DM test calculates the statistic for the null hypotheses of equal squared forecast errors between conventional DMA/DMS and ALM. Asterisks (*10%, **5%, ***1%) denote the level of significance at which the null hypotheses are rejected.

Based on this set of predictors, the performance of conventional DMA and ALM is compared. Furthermore, the performance for different choices of the forgetting factors is investigated. Koop & Korobilis (2012) set $\kappa = 0.98$ and focus on a modest range for the forgetting factors, i.e. $\alpha, \lambda \in [0.95, 0, 99]$, which they find to deliver a favorable forecasting performance over simple benchmark regressions and more sophisticated approaches. Thus, this set of values is used to forecast inflation. In order to not just rely on one value for the whole period, this section also considers estimating $\alpha \in [0.95, 0.99]$ and $\lambda \in [0.95, 0.99]$ dynamically. This allows the framework to switch between a gradual change in both coefficients and models ($\alpha = \lambda = 0.99$), a more rapid change in both coefficients and models ($\alpha = \lambda = 0.95$) or a mix of the two over time.

Table 3.1 contains the results for the one quarter and one year ahead forecasting performance in terms of the root mean squared forecast error (RMSFE) and in terms of the mean absolute forecast error (MAFE). In addition, it shows the results of the Diebold-Mariano test (DM-test) proposed by Diebold & Mariano (1995) in order to investigate whether the forecasting errors of the conventional DMA approach differ

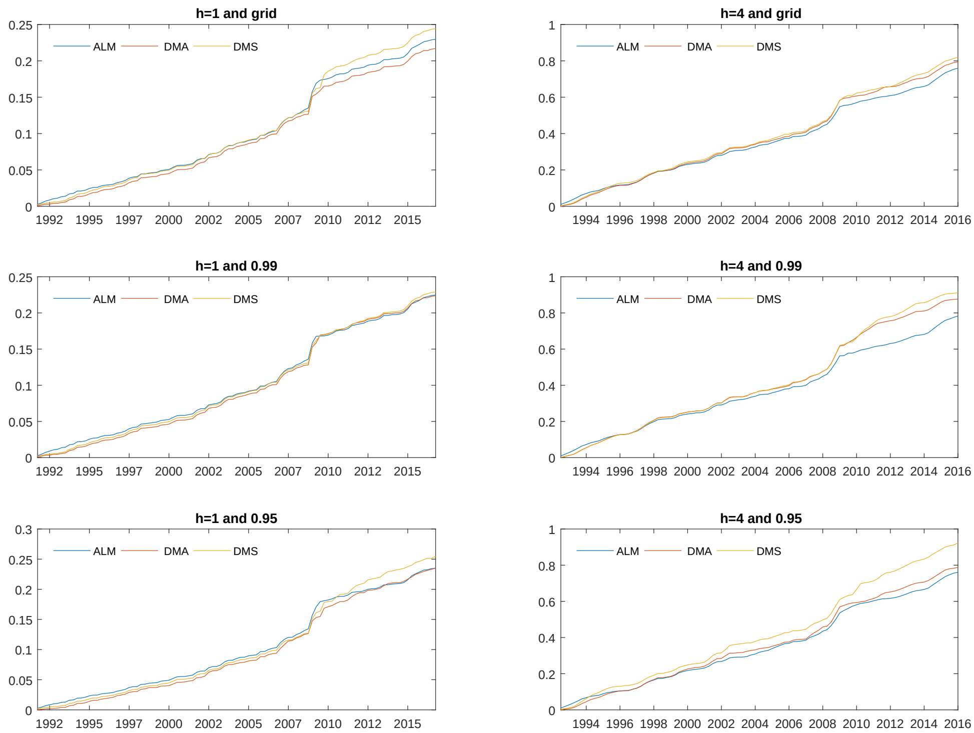


Figure 3.1: Cumulative sum of absolute forecast errors for inflation.

(statistically) significantly from those obtained using the ALM approach. ALM and conventional DMA deliver similar forecasting errors in all settings for one quarter ahead inflation. But for one year ahead ALM delivers significantly smaller forecasting errors. Furthermore, the results show that allowing for a more rapid change between models and in coefficients yields better predictions. However, estimating the forgetting factors seems to be a useful strategy in order to avoid poor forecasts due to a poor selection of values for the forgetting factors. Figure 3.1 compares the cumulative sum of the absolute forecast error over time of ALM and conventional DMA and DMS. This allows to assess the forecasting performance over time. The absolute forecast errors grow roughly linearly over time for both horizons and all three approaches. Only after the financial crisis a big jump can be observed for all setups. Hence, one can conclude that the forecast errors are fairly stable over time. Furthermore, it can be seen that the one year cumulative sum of absolute forecast errors after the financial crisis is lower for ALM compared to the conventional approach.

Table 3.2: Forecasting performance for one quarter and one year consumption expenditures

		$h = 1$		$h = 4$	
Model	Variable	RMSFE	MAFE	RMSFE	MAFE
DMA	nominal	0.59	0.38	2.36***	1.54
DMS	nominal	0.59	0.38	2.75***	1.74
ALM	nominal	0.63	0.38	2.02	1.23
DMA	real	0.42	0.33	1.64***	1.25
DMS	real	0.44	0.34	1.68***	1.29
ALM	real	0.42	0.32	1.42	1.04

The table shows the RMSFE and MAFE in percentage points for nominal and real US consumption expenditures. The DM test calculates the statistic for the null hypotheses of equal squared forecast error between conventional DMA/DMS and ALM. Asterisks (*10%, **5%, ***1%) denote the level of significance at which the null hypotheses are rejected.

3.3.2 Forecasting Consumption Expenditures

This section compares the forecasting performance of ALM and conventional DMA for nominal and real consumption expenditures. As potential predictors, the percentage change in the Dow Jones Industrial Average, the percentage change in employment, the log of housing starts, the percentage change in the money supply (M1), University of Michigan survey of inflation expectations, University of Michigan survey of consumer sentiment, the spread between the ten year and three month Treasury bill, the three month Treasury bill, the unemployment rate, the inflation rate and (real) disposable income are used. In addition, each model contains one intercept and two lags of consumption expenditures and the same grid of forgetting factors is used for the estimation as in the inflation application. All variables are quarterly, seasonally adjusted and are obtained from the FRED database of the Federal Reserve Bank of St. Louis. The data are observed for the period 1978Q2 to 2016Q3 and the period from 1992Q1 to 2016Q3 is used to evaluate the out-of-sample forecast performance.

The results turn out to be similar to the ones obtained for inflation. Table 3.2 shows that ALM and conventional DMA deliver similar forecasting errors for one quarter ahead consumption expenditures and for one year ahead, ALM delivers smaller forecasting errors. Moreover, the DM test reveals that the difference between the squared forecast errors is statistically significant for one year ahead predictions. Figure 3.2 shows a similar pattern for the cumulative sum of the absolute forecast errors for the consumption expenditures compared to these obtained for inflation.

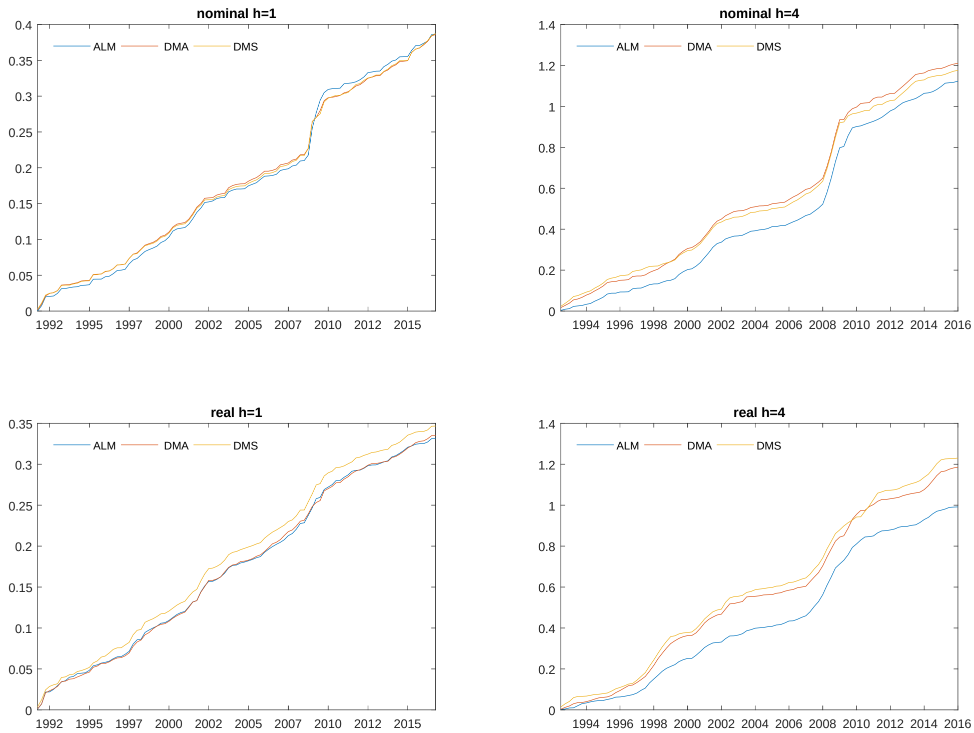


Figure 3.2: Cumulative sum of absolute forecast errors for consumption expenditures.

Again, the absolute forecast errors grow roughly linearly over time for both horizons and all three approaches with the exception of the period after the financial crisis. While the cumulative sum of the absolute forecast errors is similar for the short horizon, for the long horizon the cumulative sum of the absolute forecast errors obtained from the ALM approach are below the ones obtained from conventional DMA for the entire period.

3.3.3 Forecasting Exchange Rates

This section considers one month and one year ahead forecasting of five major US end-of-month (log) exchange rate returns. The five different countries are Canada, Denmark, the United Kingdom, Japan and Sweden. The monthly data range from 1975M2 to 2017M4 and the forecasting results are obtained after a training period of 50 months. Four regressors, based on economic theory, in addition to an intercept are considered as potential predictors. The first regressor is based on the uncovered interest parity condition (UIP) and is defined as

$$UIP_t = i_t - i_t^*, \tag{3.22}$$

where i_t denotes the nominal interest rate and i_t^* is the foreign nominal interest rate (measured by the money market rate and obtained from the International Financial Statistics database). The second regressor is based on the deviation from the purchasing power parity condition (PPP) and is defined as

$$PPP_t = p_t - p_t^* - s_t, \tag{3.23}$$

where p_t denotes the log of the domestic price level (measured as the consumer price index and obtained from the FRED database), p_t^* the log of foreign price level and s_t denotes the log of the nominal exchange rate (measured as end-of-period exchange rates and obtained from the FRED database). The third regressor is based on the asymmetric Taylor (1993) rule (ATR) and is defined as

$$ATR_t = 1.5(\pi_t - \pi_t^*) + 0.1(g_t - g_t^*) + 0.1(s_t + p_t^* - p_t), \tag{3.24}$$

where π_t is the domestic inflation rate, π_t^* is the foreign inflation rate, g_t is the domestic output gap and g_t^* is the foreign output gap. The output gap is measured as the deviation of real output (measured by industrial production and obtained from the FRED database) from an estimate of potential output calculated using the Hodrick & Prescott (1996) filter. The parameter values (1.5, 0.1, 0.1) are a standard choice in the literature, due to Molodtsova & Papell (2009). The last regressor is based on the deviation from the monetary fundamentals (MF) and is defined as

$$MF_t = (m_t - m_t^*) - (prod_t - prod_t^*) - s_t, \tag{3.25}$$

where m_t is the log of the domestic money supply (measured as M1 if available and otherwise as M3 and obtained from the OECD database), m_t^* is the log of the foreign money supply and $prod_t^{(*)}$ is the log of the domestic (foreign) industrial production.

For the estimation of the two forgetting factors α and λ a wider range as before is considered, i.e. $\alpha, \lambda \in (0.80, 0.90, 0.95, 0.99, 1)$, which in one extreme nests the special case of constant parameter models ($\lambda = 1$) and no model change ($\alpha = 1$) and in the other extreme allows for a very rapid change in both coefficients and models ($\alpha = \lambda = 0.80$). The wider grid allows the predictive information in macroeconomic fundamentals for exchange rate returns to change fast over time, as suggested by Beckmann & Schüssler (2016). For the decay factor κ a tight grid around the value 0.97 (which is recommended by RiskMetrics (1996) for monthly data), i.e. $\kappa \in (0.95, 0.96, 0.97, 0.98, 0.99)$, is considered.

Table 3.3: Forecasting performance for one month and one year exchange rate

		$h = 1$		$h = 12$	
Model	Country	RMSFE	MAFE	RMSFE	MAFE
DMA	Canada	2.12*	1.50	11.69***	7.59
DMS	Canada	2.24**	1.57	13.14***	8.26
ALM	Canada	2.04	1.45	7.13	5.37
DMA	Denmark	3.17*	2.40	21.21***	14.81
DMS	Denmark	3.35***	2.51	21.71***	15.65
ALM	Denmark	3.10	2.39	12.39	10.33
DMA	United Kindom	2.99	2.22	15.63***	12.35
DMS	United Kindom	3.00	2.24	16.30***	12.67
ALM	United Kindom	2.97	2.24	11.06	08.62
DMA	Japan	3.42*	2.57	20.93***	14.85
DMS	Japan	3.59	2.70	13.14***	16.11
ALM	Japan	3.28	2.49	12.39	10.03
DMA	Sweden	3.23	2.44	17.90***	13.35
DMS	Sweden	3.43***	2.54	19.98***	14.69
ALM	Sweden	3.21	2.41	13.48	10.82

The table shows the RMSFE and MAFE in percentage points for five major US exchange rates returns. The DM test calculates the statistic for the null hypotheses of equal squared forecast errors between conventional DMA/DMS and ALM. Asterisks (*10%, **5%, ***1%) denote the level of significance at which the null hypotheses are rejected.

Table 3.3 displays the results for the one month and one year ahead forecasting performance. The results show a clear pattern. It turns out that ALM delivers a better forecasting performance for all five countries than conventional DMA and the difference in forecasting performance is statistically significant for one year ahead predictions in all cases and in most cases for one month ahead predictions. Figure 3.3 compares the cumulative sum of the absolute forecast error over time of ALM and conventional DMA and DMS. The cumulative sum of absolute forecast errors for one month grows roughly linear over time. In contrast, the cumulative sum of the absolute forecast error of one year exhibits some jumps over time. This shows that the absolute forecast errors of one year ahead are less stable than for one month ahead. Moreover, while for the short horizon the cumulative sum of absolute forecast errors looks very similar for all three approaches, for the long horizon the cumulative sum of absolute forecast errors is smaller for ALM than for conventional DMA or DMS for the entire sample. Furthermore, ALM exhibits fewer and smaller jumps than conventional DMA or DMS.

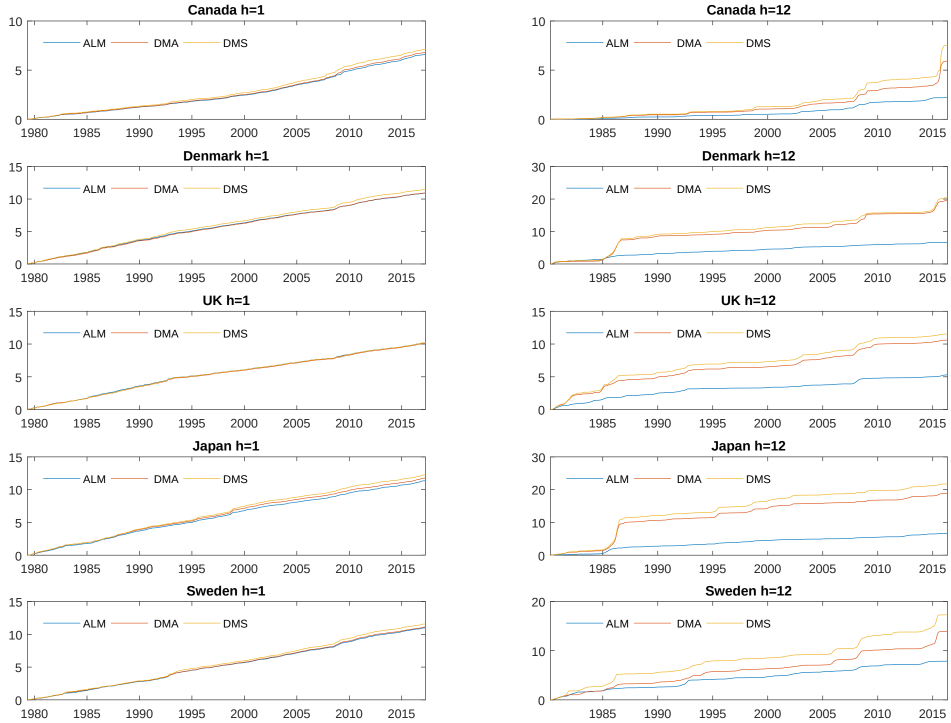


Figure 3.3: Cumulative sum of absolute forecast errors for exchange rates.

3.4 Conclusion

DMA has been used extensively for the purpose of economic forecasting as it dynamically addresses both model and parameter uncertainty. This chapter extends this framework by considering ALM. ALM dynamically averages not only over the forecasts of the individual models but, in addition, dynamically averages over the time-varying parameters. Therefore, it exploits the information of all models in the estimation of the time-varying parameters. This is done by approximating the mixture of individual posteriors with a single posterior, which is then used in the upcoming period as the prior for each of the individual models. The relevance of this extension is illustrated in three empirical applications involving US inflation, US consumption expenditures and five major US exchange rate returns. It turns out that ALM leads to improved out-of-sample predictions in all applications.

4 Forecasting with many Predictors using Bayesian additive Regression Trees

Forecasting with many predictors provides the opportunity to exploit a much richer base of information. However, macroeconomic time series are typically rather short, raising problems for conventional econometric models. This paper explores the use of Bayesian additive regression trees (Bart) from the machine learning literature to forecast macroeconomic time series in a predictor rich environment. The interest lies in forecasting nine key macroeconomic variables of interest for government budget planning, central bank policy making and business decisions. It turns out that Bart is a valuable addition to existing methods for handling high dimensional data sets in a macroeconomic context.

4.1 Introduction

Machine learning methods gain in popularity in many fields. Although, they are still rarely used in macroeconomic applications, but have the potential to revolutionize the way economists do econometrics. Government statistical agencies collect data on a wide range of macroeconomic variables, e.g. measures of output, capacity, employment and unemployment, prices, wages, housing, inventories, orders, stock prices, interest rates, exchange rates and monetary aggregates. Forecasting with many predictors provides the opportunity to exploit a much richer base of information. However, macroeconomic time series are typically rather short. A large number of predictors combined with only a small number of observations raises problems for conventional econometric methods. Intuitively, there is not enough information in the data to estimate large models in an unrestricted fashion. So far Factor model estimated with principal components, see Stock & Watson (2002), and shrinkage methods like the Lasso estimator, see Stock & Watson (2012) and Korobilis (2013*b*), have found to be useful to address these problems. The Factor model saves degree of freedom by summarizing the data through a few factors which are then used as predictors. The Lasso shrinks the coefficients towards zero in a data driven-way in order to avoid overfitting. This paper explores the use of Bayesian additive regression trees (Bart), proposed by Chipman et al. (2010), for forecasting with many predictors in a macroeconomic context.

The Bart model is a sum-of-tree model and is attractive for several reasons. In order to address the problem of a short span of observations relative to the number of potentially relevant explanatory variables the Bart model provides built-in variable selection. Furthermore, most econometric models are linear. In contrast, the Bart model allows for non-linearity and interaction effects between predictors in a natural way. But, it is well known that large tree models tend to overfit (i.e. they fit the noise in the data rather than detecting a pattern which is useful for forecasting). The Bayesian approach provides the opportunity to address this problem by using prior distributions to regularize the fit of each individual tree, so that each tree only explains a small fraction of the variation in the response variable. In addition, the Bayesian approach allows to account for parameter, tree and forecasting uncertainty by providing a predictive distribution for the response variable which reflects the different sources of uncertainty.

The main contribution of this paper is to provide a forecasting comparison between the Bart, the Lasso approach and the Factor approach using a large data set of over hundred quarterly US macroeconomic time series. Interest lies in forecasting nine key macroeconomic variables, which are of interest for government budget planning, central bank policy making and business decisions. The first four measure economic activity and are Real Gross Domestic Product (GDP), Industrial production, Unemployment rate and Real Personal Consumption Expenditure (PCE). The next three measure the price level in the economy and are GDP Deflator, PCE Deflator and the Consumer Price Index (CPI) and the last two are the Effective Federal Funds Rate and the 10-Year Treasury Constant Maturity Rate. It turns out that Bart outperforms Lasso and Factor approaches and therefore is a valuable addition to existing methods for handling high dimensional data sets in a macroeconomic context.

The remainder of the chapter is structured as follows. Section 2 provides an overview of the Bart model, Section 3 contains the empirical results and Section 4 concludes.

4.2 Bayesian Additive Regression Trees

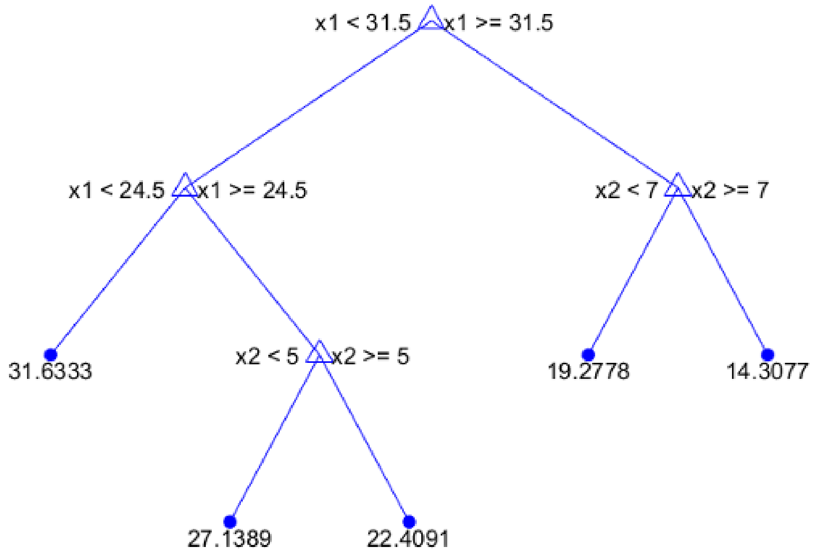


Figure 4.1: Example of a regression tree with two variables.

4.2.1 A sum-of-trees Model

The Bart model proposed by Chipman et al. (2010) provides predictions of the response variable Y by a sum of Bayesian regression trees. Given an $N \times p$ matrix X of predictor variables, let $x_n = [x_{n1}, \ldots, x_{np}]$ denote the nth row (i.e, the nth observation) of X. The Bart model can then be written as

$$Y_n = \sum_{j=1}^{m} g(x_n; T_j, M_j) + \epsilon_n, \quad \epsilon_n \sim N(0, \sigma^2), \tag{4.1}$$

where $g(x_n; T_j, M_j)$ is a regression tree as described in Chipman et al. (1998), m denotes the number of regression trees, T_j denotes a tree consisting of a set of interior node decision rules and a set of terminal nodes, and $M_j = \{\mu_{1j}, \mu_{2j}, \ldots, \mu_{bj}\}$ denotes a set of parameter values associated with each of the b_j terminal nodes of T_j.[1] The decision rules are of the form $\{x_{nk} \leq c\}$ vs $\{x_{nk} > c\}$, where c is a threshold value

[1]Note that the individual regression trees are jointly estimated, see Section 4.2.3.

within the range of values of variable x_{nk} and $k = 1, \ldots, p$. Observations which satisfy the splitting rules are sent to the left-hand daughter node and those which do not are send to the right-hand daughter node. The process of splitting the observations at each interior node iteratively carries on until a terminal node is reached. An illustration of a single regression tree is given in Figure (4.1).

Tree models are attractive relative to standard statistical models such as linear regression models, because the decision tree performs variable selection and allows for non-linear interaction effects. However, decision trees are known to tend to overfit the data. Therefore, the idea of the Bart model is to regularize the fit of each individual tree, so that each tree only explains a small and different fraction of the variation in the response variable. In order to regularize the fit of each single tree the Bart model specifies prior distributions for this purpose.

4.2.2 A Regularization Prior

To form the Bayesian model, prior distributions for the parameters are required. The prior distributions are specified with the aim to facilitate computation, regularize the fit of each individual tree, to be controlled by just a few interpretable hyperparameters and not to be in severe conflict with the data. Note that the prior for the Bart model has three components: the tree structure, the terminal-node parameters given the tree structure and the error variance σ^2. It is useful to assume that the priors for the tree components (T_j, M_j) are independent of each other and of σ^2, such that the prior can be written as

$$p((T1, M1), \ldots, (T_m, M_m), \sigma) = \left[\prod_j p(T_j, Mj)\right] p(\sigma^2) \tag{4.2}$$

$$= \left[\prod_j p(M_j|T_j)p(T_j)\right] p(\sigma^2) \tag{4.3}$$

and

$$p(M_j|T_j) = \prod_i p(\mu_{ij}|T_j), \tag{4.4}$$

where $\mu_{ij} \in M_j$. Now the problem boils down to specifying $p(T_j)$, $p(\mu_{ij}|T_j)$ and $p(\sigma^2)$.

The tree prior $p(T_j)$ affects the tree structure and the decision rules of each tree.

It specifies the probability that a node is nonterminal, the distribution on the splitting variable assignments at each interior node and the distribution on the splitting rule assignment in each interior node. In order to regularize the tree structure of each single tree, the probability that a node at depth $d(= 0, 1, 2, \dots)$ is nonterminal, is assumed to be

$$\alpha(1 + d)^{-\beta}, \quad \alpha \in (0, 1), \beta \in [0, \infty). \tag{4.5}$$

The hyperparameter α controls the overall strength of this prior and the hyperparameter β controls how fast the probability decreases in d that a node is nonterminal. This component of the prior for the tree structure has the ability to enforce shallow tree structures. Thereby, limiting the complexity of a single tree. For the distribution on the splitting variable a uniform prior on the available variables is assumed and for the distribution on the splitting rule assignment a uniform prior on the available splitting values is assumed.[2]

The conjugate normal distribution $N(\mu_\mu, \sigma_\mu^2)$ is used as a prior for $\mu_{ij}|T_j$. Under the sum-of-trees model, and because the priors are i.i.d., the induced prior on $\mathrm{E}(Y_n|x_n)$ is $N(m\mu_\mu, m\sigma_\mu^2)$. The parameters at the terminal node μ_{ij} are shrunk towards zero by setting $\mu_\mu = 0$, limiting the effect of the individual tree components by keeping them small. The hyperparamter σ_μ^2 is chosen such that $y_{min} = m\mu_\mu - k\sqrt{m}\sigma_\mu$ and $y_{max} = m\mu_\mu + k\sqrt{m}\sigma_\mu$ for some k. The larger the value of k, the smaller the value of σ_μ^2, resulting in stronger model regularization. For $k = 2$ a 95% prior probability that $\mathrm{E}(Y_n|x_n) \in (y_{min}, y_{max})$ is obtained. This prior specification implies that as k and/or the number of trees m is increased this prior will become tighter.

The conjugate inverse gamma distribution $\mathrm{InvGamma}(v/2, v\lambda/2)$ is used for $p(\sigma^2)$. The hyperparameter λ is set such that the qth quantile of the prior on σ is located at $\hat{\sigma}$ (the standard deviation of y), i.e., $p(\sigma < \hat{\sigma}) = q$ and the hyperparameter v is set to 3 so that the first two moments exists. Note that the adjustable hyperparameters are now α, β, k, and q. In addition, the number of trees m must be chosen.

4.2.3 Gibbs Sampler

Combining equations (4.1) and (4.3) gives the posterior distribution:

[2] The available splitting values are of the form $(xs_l + xs_{l+1})/2$ where xs contains the unique sorted values from the corresponding predictor.

$$p(\mathcal{T}, M, \sigma^2|X, Y) \propto p(Y|X, \mathcal{T}, M, \sigma^2) \times \left[\prod_j p(M_j|T_j)p(T_j)\right] p(\sigma^2), \tag{4.6}$$

where $p(Y|X, \mathcal{T}, M, \sigma^2)$ is the likelihood for a given sum-of-trees model and $\mathcal{T}$ is the set of all trees, i.e., $\mathcal{T} = \{T_1, \ldots T_m\}$. In order to generate draws from the posterior, a Metropolis-within-Gibbs algorithm is employed. The algorithm involves drawing from the full conditionals $p(T_j, M_j|X, Y, \sigma^2, \mathcal{T}_{-j}, M_{-j})$ and $p(\sigma^2|X, Y, \mathcal{T}, M)$. In order to draw from $p(T_j, M_j|X, Y, \sigma^2, \mathcal{T}_{-j}, M_{-j})$, it is useful to note that $p(T_j, M_j|X, Y, \sigma^2, \mathcal{T}_{-j}, M_{-j})$ depends on $(X, Y, \mathcal{T}_{-j}, M_{-j})$ only through $R_j = Y - \sum_{t \neq j} g(X; T_j, M_j)$, the N-vector of partial residuals based on the fit that excludes tree j. Therefore, sampling of (T_j, M_j) given $(X, Y, \sigma^2, \mathcal{T}_{-j}, M_{-j})$ is equivalent to sampling from $p(T_j, M_j|R_j, \sigma^2)$. Because of the conjugate prior for M_j, it is possible to obtain

$$p(T_j|R_j, \sigma^2) \propto p(T_j) \int p(R_j|M_j, T_j, \sigma^2)p(M_j|T_j, \sigma^2)\mathrm{d}M_j \tag{4.7}$$

in closed form up to a normalizing constant. This allows to draw from the full conditionals $p(T_j, M_j|R_j, \sigma^2)$ in two successive steps. First draw from $p(T_j|R_j, \sigma^2)$ and then draw from $p(M_j|T_j, R_j, \sigma^2)$. In order to draw from $p(T_j|R_j, \sigma^2)$ a Metropolis step is used, which selects from one of four proposal moves: GROW (node birth), PRUNE (node death), CHANGE (changing splitting rules) and SWAP (swapping internal nodes), for details see Chipman et al. (1998). Due to the normal-inverse-gamma conjugacy prior, the posterior of the terminal parameter M_j is normal and the posterior of the variance σ^2 is inverse gamma, from which it is easy to sample. The analytic-expressions for these posteriors can be found in Gelman et al. (2004). To generate one draw from the posterior in (4.6) the sampler proceeds as follows:

$$\begin{aligned}
1 &: T_1|R_1, \sigma^2 \\
2 &: M_1|T_1, R_1, \sigma^2 \\
3 &: T_2|R_2, \sigma^2 \\
4 &: M_2|T_2, R_2, \sigma^2 \\
&\vdots \\
2m-1 &: T_m|R_m, \sigma^2 \\
2m &: M_m|T_m, R_m, \sigma^2 \\
2m+1 &: \sigma^2|T_1, M_1, \ldots, T_m, M_m.
\end{aligned}$$

4.2.4 Cross Validation

In order to estimate the model, one needs to select the number of trees m and the hyperparamters α, β, k, and q. These are chosen out of all combinations from the grid $m \in (200, 300, 400)$, $\alpha \in (0.5, 0.1, 0.05, 0.01)$, $\beta \in (1, 2)$, $k \in (2, 3)$ and $q \in (0.75, 0.9)$[3] by using cross-validation, which computes a quasi-out-of-sample score by estimating the model with a subset of data and validating the omitted data. A training data set Y_N, with N observations, is partitioned into R blocks,

$$Y_N = [\mathcal{Y}_1 \quad \mathcal{Y}_2 \quad \dots \quad \mathcal{Y}_R] \tag{4.8}$$

for

$$\mathcal{Y}_r = [y_{n_r} \quad y_{n_r+1} \quad \dots \quad y_{n_{r+1}-1}], \tag{4.9}$$

and the full set of observations with block $\mathcal{Y}_r$ deleted is denoted as

$$\mathcal{Y}^{(r)} = [\mathcal{Y}_1 \quad \dots \quad \mathcal{Y}_{r-1} \quad \mathcal{Y}_{r+1} \quad \dots \quad \mathcal{Y}_R]. \tag{4.10}$$

Define $\mathcal{X}^{(r)}$ in the same way as a matrix of realization for X_N with block r deleted:

$$\mathcal{X}^{(r)} = [\mathcal{X}_1 \quad \dots \quad \mathcal{X}_{r-1} \quad \mathcal{X}_{r+1} \quad \dots \quad \mathcal{X}_R]. \tag{4.11}$$

This partition allows to judge how well a particular model, that was estimated only on data $\mathcal{Y}^{(r)}$, predicts the values of $\mathcal{Y}_r$. In order to evaluate the out-of-sample predictions, three different forecast metrics are considered. The first two are the mean squared forecast error (MSE) and the mean absolute forecast error (MAE). The MSE is given by

$$MSE = \frac{1}{N} \sum_{r=1}^{R} \sum_{n=n_r}^{n_{r+1}-1} (y_n - \hat{y}_n)^2 \tag{4.12}$$

and the MAE is given by

$$MAE = \frac{1}{N} \sum_{r=1}^{R} \sum_{n=n_r}^{n_{r+1}-1} |y_n - \hat{y}_n|, \tag{4.13}$$

where $\hat{y}_n$ denotes a point forecast of the response variable. These two only evaluate

[3]Note that similar values are employed by Chipman et al. (2010).

the point forecasts and ignore the remaining part of the predictive distribution. In order to address this issue the sum of log predictive likelihoods is used as an additional forecast metric. The predictive likelihood is the predictive density evaluated at the actual outcome

$$PL = \sum_{r=1}^{R} \sum_{n=n_r}^{n_{r+1}-1} \log[p(y_n|\mathcal{Y}^{(r)}, \mathcal{X}^{(r)})] \tag{4.14}$$

and has the advantage of evaluating the forecasting performance of the entire predictive density, see Geweke & Amisano (2010) for a detailed motivation.

4.3 Forecasting with many Predictors

The data set consists of 123 quarterly US macroeconomic time series spanning the period from 1959Q1 to 2017Q2. All series were downloaded from the St. Louis Fed FRED database. A complete description can be found in Table 4.4. The data set covers information about the real economy (output, labor, consumption, orders and inventories), money, prices and financial markets (interest rates, exchange rates, stock market indexes). All series are seasonally adjusted and transformed to be approximately stationary. The transformation codes are listed in Table 4.4. The interest lies in forecasting nine key macroeconomic variables, which are of interest for government budget planning, central bank policy making and business decisions. The first four measure economic activity and are Real GDP (rGDP), Industrial production (IND), Unemployment rate (UNEM) and real PCE (rPCE), the second three measure the price level in the economy and are GDP deflator (GDPdef), PCE deflator (PCEdef) and CPI and the last two are the Effective Federal Funds Rate (FED) and the 10-Year Treasury Constant Maturity Rate (GS10). All variables given in Table 4.4 are employed as predictors and in addition four lags of the dependent variable are used.

Following Stock & Watson (2002) and Korobilis (2013*b*) Factor models and shrinkage methods are considered as main competitors. The Factor model is estimated using a two step approach. The Factors are estimated by principal components. These are used as observations in a regression model. Factor models with one to three Factors are considered in the forecasting comparison. In the class of shrinkage estimators, the hierarchical Lasso prior proposed by Park & Casella (2008) is used. Following Korobilis (2013*b*) the hyperparamters of the hierarchical inverse gamma prior are set either to 0.01 or 0.001, see Park & Casella (2008) or Korobilis (2013*b*) for details.[4] In addition an AR(1) model is used to provide a simple benchmark model

[4]Note that for the first own lag a non-informative prior is employed.

which does not suffer from overparametrization.

The Bart model is employed with four different specifications. The first three are Bart models specified by cross validation. In this specification the first 100 observations are used to train the model and the rest of the observations are used to evaluate the out-of-sample performance. As described in Section 2.4., the 100 observations are divided into $R = 10$ blocks and the three forecasting metrics (MSE, MAE and PL) are used to select the optimal combinations of hyperparamters and number of trees. The fourth specification uses the values $m = 200$, $\alpha = 0.1$, $\beta = 1$, $k = 2$ and $q = 0.9$,[5] which turned out to be a quite effective specification in the forecasting exercise for all nine variables and may be useful as a benchmark specification (BM) in other applications.

The forecasts are compared using the MSE and the MAE as measures of point forecasts and PL as measures for the density forecast performance. Table 4.1 contains the results for the MAE, Table 4.2 for the MSE and Table 4.3 for the PL. The MAE (MSE) of each model is divided by the MAE (MSE) obtained from an AR(1) model. Consequently, values above one mean that the AR(1) model dominates and values below one mean the opposite. The PL of the AR(1) model is subtracted from the PL of each model. This time negative values imply that the AR(1) model dominates and positive values imply the opposite. All three tables contain results for one-quarter and one year-ahead predictions.

Overall the results clearly reveal that the Bart model serves as a valuable addition to existing methods for handling large dimensional data. The Bart model outperforms the Factor and the Lasso model for both point and density forecasts for most variables and forecasting horizons. There is always an Bart specification which performs better than the Factor and Lasso approaches with just a few exceptions. The Factor approach turns out to be the best model in terms of MAE only for one quarter ahead forecasts of rGDP. The Lasso approach turns out to be best model in terms of MAE only for forecasting one year ahead CPI, in terms of MSE it turns out to be the best model for forecasting one year ahead PCEdef, CPI and GS10 and in terms of density forecasts it turns out to be the best model for forecasting one quarter ahead CPI and one year ahead GS10. In all other cases a Bart specification turns out to be better than the Factor and Lasso approaches. Furthermore, the Bart model is also highly competitive compared to a simple AR(1) model. It tends to produce more precise

[5]These are the values used by Chipman et al. (2010), with the exception that the strength of the prior which regularizes the tree structure of each single tree is higher. A high degree of regularization of the tree structure may be important in order to avoid overfitting the data.

point forecasts than an AR(1) model for most variables and forecasting horizons. In terms of MAEs the performance of the Bart model with the BM specification is only worse in forecasting IND one quarter ahead and in forecasting one year ahead it is only worse for rPCE, IND and rGDP. For the MSE, as a measure for point forecasts, the Bart BM model is only worse for one quarter ahead forecasts of GS10 and for one year ahead for rPCE. In terms of density predictions the Bart BM model performs even better compared to an AR(1) model as only the forecasting performance for one year ahead rPCE is worse. For the other specifications of the Bart model the results are similar. Finally, comparing the different Bart specifications none of the four outperforms the others. However, the BM specification turns out to never be the worst specification of the four in terms of point and density forecasts. In this sense it may provide an useful and simple choice in other macroeconomic applications.

4.4 Conclusion

This paper explores the use of Bayesian additive regression trees (Bart) from the machine learning literature for forecasting with many predictors in a macroeconomic context. Using many predictors in a macroeconomic context raises problems to conventional econometric methods. The Bart model is attractive because it has built in variable selection in order to address these problems. Moreover it allows for non-linear and interaction effects in a natural way. In a forecasting comparison for nine key macroeconomic variables, which are of interest for government budget planning, central bank policy making and business decisions, it turns out that the Bart model is a valuable addition to existing methods for handling high dimensional data sets.

4.A Forecasting Results

Table 4.1: Out of sample Results MAE relative to AR(1)

Variable	BartMSE	BartMAE	BartPL	BartBM	Fac1	Fac2	Fac3	Lasso1	Lasso2
					one quarter ahead				
rGDP	0.94	0.97	0.94	0.94	**0.92**	1.03	1.01	0.98	0.98
rPCE	**0.95**	0.99	0.97	0.95	0.98	1.04	1.01	1.06	1.06
IND	**1.01**	1.03	1.08	1.05	1.09	1.15	1.10	1.20	1.20
UNEM	0.93	**0.92**	0.93	0.93	1.04	1.04	0.95	1.10	1.11
GDPdef	1.18	1.02	1.10	**0.99**	2.78	1.94	1.86	2.12	2.10
PCEdef	**0.91**	1.33	0.95	0.91	1.86	1.34	1.38	1.02	1.02
CPI	0.64	0.66	0.64	**0.61**	0.95	0.91	0.85	0.64	0.64
FED	1.33	**0.89**	0.90	1.03	1.07	1.11	1.15	1.28	1.27
GS10	0.99	**0.98**	1.02	0.99	1.02	1.01	1.05	1.02	1.02
					one year ahead				
rGDP	1.05	1.04	**1.02**	1.02	1.03	1.15	1.16	1.08	1.08
rPCE	1.06	1.09	1.06	**1.04**	1.09	1.16	1.19	1.25	1.26
IND	**1.02**	1.03	1.06	1.02	1.03	1.18	1.14	1.12	1.13
UNEM	0.98	0.95	0.95	**0.92**	1.11	1.14	1.00	1.03	1.03
GDPdef	0.90	**0.83**	0.84	0.87	2.90	1.97	2.03	2.22	2.23
PCEdef	0.79	**0.75**	0.83	0.76	1.77	1.33	1.41	0.80	0.81
CPI	0.50	0.48	0.50	0.49	0.92	0.92	0.80	**0.45**	0.46
FED	0.97	0.93	**0.90**	0.95	0.93	0.97	1.11	1.02	1.04
GS10	0.96	0.98	0.95	**0.90**	0.98	1.00	0.99	0.96	0.95

The tables shows the forecasting performance of the Bart model with four different specifications, the Factor model with one to three Factors and the Lasso approach with two different hierarchical priors. The forecasting performance is measured by the mean absolute forecasting error (MAE) and values below indicate that the model outperforms the AR(1) model.

Table 4.2: Out of sample Results MSE relative to AR(1)

Variable	BartMFE	BartMAE	BartPL	BartBM	Fac1	Fac2	Fac3	Lasso1	Lasso2
				one quarter ahead					
rGDP	0.83	0.83	0.93	**0.82**	0.86	1.00	0.92	0.91	0.90
rPCE	0.92	0.91	**0.86**	0.89	1.00	1.05	0.95	1.14	1.14
IND	**0.90**	0.96	1.02	0.98	1.12	1.26	0.96	1.42	1.44
UNEM	0.86	0.86	0.86	**0.85**	1.06	1.09	0.91	1.29	1.30
GDPdef	0.85	0.90	1.02	**0.84**	5.4	3.51	2.93	3.47	3.46
PCEdef	**0.82**	1.35	0.83	0.82	2.14	1.43	1.47	0.88	0.88
CPI	0.54	0.56	**0.53**	0.53	0.91	0.96	0.84	0.54	0.54
FED	0.91	**0.65**	0.68	0.83	1.00	1.13	1.39	1.13	1.11
GS10	1.02	**1.01**	1.05	1.02	1.07	1.07	1.15	1.07	1.06
				one year ahead					
rGDP	0.93	1.07	0.97	**0.90**	1.05	1.22	1.20	1.17	1.16
rPCE	1.16	1.26	1.16	**1.09**	1.27	1.31	1.34	1.49	1.49
IND	0.88	0.96	**0.86**	0.87	1.06	1.29	1.06	1.24	1.25
UNEM	0.97	**0.82**	0.86	0.83	1.09	1.14	0.90	1.12	1.13
GDPdef	0.91	0.78	**0.74**	0.80	6.21	3.88	4.04	4.49	4.44
PCEdef	0.63	0.60	0.68	0.61	2.08	1.41	1.64	**0.53**	0.54
CPI	0.37	0.35	0.36	0.37	0.86	0.98	0.75	0.29	**0.29**
FED	0.86	0.78	**0.70**	0.70	0.85	0.92	1.36	0.91	0.93
GS10	0.98	0.98	0.97	0.98	1.01	1.04	1.12	0.97	**0.96**

The tables shows the forecasting performance of the Bart model with four different specifications, the Factor model with one to three Factors and the Lasso approach with two different hierarchical priors. The forecasting performance is measured by the mean squared forecasting error (MSE) and values below indicate that the model outperforms the AR(1) model.

Table 4.3: Out of sample Results PL - PL of AR(1)

Variable	BartMSE	BartMAF	BartPL	BartBM	Fac1	Fac2	Fac3	Lasso1	Lasso2
				one quarter ahead					
rGDP	15.67	**18.31**	5.73	16.10	5.62	6.49	9.19	11.49	12.24
rPCE	8.49	12.77	**13.04**	11.36	-1.49	2.44	7.45	-1.63	-1.79
IND	**17.57**	15.63	3.69	15.15	-3.62	0.15	10.79	-8.67	-9.89
UNEM	13.46	13.25	13.97	**16.25**	-3.72	2.50	11.93	-11.69	11.83
GDPdef	-9.22	6.92	3.95	**12.22**	-103.43	-70.71	-62.21	-69.11	-69.14
PCEdef	15.57	-1.77	**16.39**	15.61	-47.26	-20.61	-20.78	13.60	13.00
CPI	0.54	40.30	40.44	41.86	5.11	8.09	15.88	41.13	**43.41**
FED	29.79	16.99	15.11	**30.81**	6.52	5.63	4.32	6.05	6.08
GS10	**1.14**	0.57	-1.8	0.58	-3.90	-3.65	-5.99	-1.85	-1.26
				one year ahead					
rGDP	9.01	-0.87	**15.79**	12.26	-3.02	-5.01	-3.39	-1.15	-1.03
rPCE	-6.17	-12.11	-6.32	**-0.55**	-11.46	-12.41	-12.41	-21.80	-22.27
IND	12.13	3.52	**15.01**	13.31	-0.76	-10.34	3.81	-9.74	-9.66
UNEM	1.23	**15.38**	10.75	15.03	-4.62	-3.14	10.77	-3.32	-4.55
GDPdef	29.05	30.42	**30.43**	29.16	-3.02	-5.01	-3.39	-72.14	-71.26
PCEdef	32.04	**33.64**	17.84	33.21	-48.75	-20.53	-28.90	32.76	32.84
CPI	68.23	**72.13**	69.70	70.81	6.02	7.92	19.96	73.71	73.77
FED	9.94	14.46	**21.37**	21.52	9.02	7.42	1.94	8.81	7.69
GS10	1.72	0.46	2.41	3.29	3.67	-1.77	-3.00	3.89	**4.03**

The tables shows the forecasting performance of the Bart model with four different specifications, the Factor model with one to three Factors and the Lasso approach with two different hierarchical priors. The forecasting performance is measured by the sum of log predictive likelihoods (PL) and positive values indicate that the model outperforms the AR(1) model.

4.B Data

Table 4.4: Data

No.	Name	ID	TC
1	Real Gross Domestic Product, 3 Decimal	GDPC96	5
2	Gross Domestic Product: Implicit Price Deflator	GDPDEF	5
3	Real Personal Consumption Expenditures	PCECC96	5
4	Personal Consumption Expenditures: Chain-type Price Index	PCECTPI	5
5	Real Gross Private Domestic Investment, 3 Decimal	GPDIC96	5
6	Real Imports of Goods & Services, 3 Decimal	IMPGSC96	5
7	Real Exports of Goods & Services, 3 Decimal	EXPGSC96	5
8	Real Change in Private Inventories	CBIC96	1
9	Real Final Sales of Domestic Product	FINSLC96	5
10	Gross Saving	GSAVE	5
11	Real Government Consumption Expenditures & Gross Investment	GCEC96	5
12	State & Local Government Current Expenditures	SLEXPND	6
13	State & Local Government Gross Investment	SLINV	6
14	Real Disposable Personal Income	DPIC96	6
15	Personal Income	PINCOME	6
16	Personal Saving	PSAVE	5
17	Private Residential Fixed Investment	PRFI	6
18	Private Nonresidential Fixed Investment	PNFI	6
19	Personal Consumption Expenditures: Durable Goods	PCDG	5
20	Personal Consumption Expenditures: Nondurable Goods	PCND	5
21	Personal Consumption Expenditures: Services	PCESV	5
22	Gross Private Domestic Investment: Chain-type Price Index	GPDICTPI	6
23	Compensation of Employees: Wages & Salary Accruals	WASCUR	6
24	Net Corporate Dividends	DIVIDEND	6
25	Corporate Profits After Tax	CP	6
26	Corporate: Consumption of Fixed Capital	CCFC	6
27	Housing Starts: Total: New Privately Owned Housing Units Started	HOUST	4
28	Privately Owned Housing Starts: 1-Unit Structures	HOUST1F	4
29	Privately Owned Housing Starts: 5-Unit Structures or More	HOUST5F	4
30	Housing Starts in Midwest Census Region	HOUSTMW	4
31	Housing Starts in Northeast Census Region	HOUSTNE	4
32	Housing Starts in South Census Region	HOUSTS	4
33	Housing Starts in West Census Region	HOUSTW	4

Table 4.4: Data(continued)

No.	Name	ID	TC
34	Industrial Production Index	INDPRO	5
35	Industrial Production: Consumer Goods	IPCONGD	5
36	Industrial Production: Durable Consumer Goods	IPDCONGD	5
37	Industrial Production: Nondurable Consumer Goods	IPNCONGD	5
38	Industrial Production: Materials	IPMAT	5
39	Industrial Production: Durable Materials	IPDMAT	5
40	Industrial Production: nondurable Materials	IPNMAT	5
41	Industrial Production: Business Equipment	IPBUSEQ	5
42	Industrial Production: Final Products (Market Group)	IPFINAL	5
43	Capacity Utilization: Manufacturing	CUMFNS	1
44	Civilians Unemployed - Less Than 5 Weeks	UEMPLT5	5
45	Civilians Unemployed for 5-14 Weeks	UEMP5TO14	5
46	Civilians Unemployed for 15-26 Weeks	UEMP15T26	5
47	Civilians Unemployed for 27 Weeks and Over	UEMP27OV	5
48	Civilian Unemployment Rate	UNRATE	2
49	Total Nonfarm Payrolls: All Employees	PAYEMS	5
50	All Employees: Nondurable Goods Manufacturing	NDMANEMP	5
51	All Employees: Durable Goods Manufacturing	DMANEMP	5
52	All Employees: Construction	USCONS	5
53	All Employees: Goods-Producing Industries	USGOOD	5
54	All Employees: Financial Activities	USFIRE	5
55	All Employees: Wholesale Trade	USWTRADE	5
56	All Employees: Trade, Transportation & Utilities	USTPU	5
57	All Employees: Retail Trade	USTRADE	5
58	All Employees: Natural Resources & Mining	USMINE	5
59	All Employees: Professional & Business Services	USPBS	5
60	All Employees: Leisure & Hospitality	USLAH	5
61	All Employees: Information Services	USINFO	5
62	All Employees: Education & Health Services	USEHS	5
63	All Employees: Service-Providing Industries	SRVPRD	5
64	All Employees: Total Private Industries	USPRIV	5
65	All Employees: Government	USGOVT	5
66	Average Hourly Earnings: Manufacturing	AHEMAN	6
67	Average Hourly Earnings: Construction	AHECONS	6
68	Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing	AWHMAN	5
69	Average Weekly Hours: Overtime: Manufacturing	AWOTMAN	5

Table 4.4: Data(continued)

No.	Name	ID	TC
70	Civilian Employment-Population Ratio	EMRATIO	5
71	Civilian Participation Rate	CIVPART	5
72	Business Sector: Output Per Hour of All Persons	OPHPBS	5
73	Nonfarm Business Sector: Unit Labor Cost	ULCNFB	5
74	Commercial and Industrial Loans at All Commercial Banks	BUSLOANS	6
75	Real Estate Loans at All Commercial Banks	REALLN	6
76	Total Consumer Credit Owned and Securitized, Outstanding	TOTALSL	5
77	Total Loans and Leases at Commercial Banks	LOANS	6
78	Bank Prime Loan Rate	MPRIME	2
79	1-Year Treasury Constant Maturity Rate	GS1	2
80	3-Year Treasury Constant Maturity Rate	GS3	2
81	5-Year Treasury Constant Maturity Rate	GS5	2
82	10-Year Treasury Constant Maturity Rate	GS10	2
83	Effective Federal Funds Rate	FEDFUNDS	2
84	3-Month Treasury Bill: Secondary Market Rate	TB3MS	2
85	6-Month Treasury Bill: Secondary Market Rate	TB6MS	2
86	Moody's Seasoned Aaa Corporate Bond Yield	AAA	2
87	Moody's Seasoned Baa Corporate Bond Yield	BAA	2
88	M1 Money Stock	M1SL	6
89	M2 Money Stock	M2SL	6
90	Currency Component of M1	CURRSL	6
91	Demand Deposits at Commercial Banks	DEMDEPSL	6
92	Savings Deposits - Total	SAVINGSL	6
93	Total Checkable Deposits	TCDSL	6
94	Travelers Checks Outstanding	TVCKSSL	6
95	Currency in Circulation	CURRCIR	6
96	MZM Money Stock	MZMSL	6
97	Velocity of M1 Money Stock	M1V	5
98	Velocity of M2 Money Stock	M2V	5
99	Total Nonrevolving Credit Outstanding	NONREVSL	6
100	Total Consumer Credit Outstanding	TOTALSL	6
101	Consumer Price Index for All Urban Consumers: All Items	CPIAUCSL	6
102	Consumer Price Index for All Urban Consumers: Commodities	CUSR0000SAC	6
103	Consumer Price Index for All Urban Consumers: All Items Less Energy	CPILEGSL	6
104	Consumer Price Index for All Urban Consumers: All Items Less Food	CPIULFSL	6

Table 4.4: Data(continued)

No.	**Name**	**ID**	**TC**
105	Consumer Price Index for All Urban Consumers: Energy	CPIENGSL	6
106	Consumer Price Index for All Urban Consumers: Food	CPIUFDSL	6
107	Consumer Price Index for All Urban Consumers: Apparel	CPIAPPSL	6
108	Consumer Price Index for All Urban Consumers: Medical Care	CPIMEDSL	6
109	Consumer Price Index for All Urban Consumers: Transportation	CPITRNSL	6
110	Producer Price Index: All Commodities	PPIACO	6
111	S&P 500 Index	SP500	5
112	Spot Oil Price: West Texas Intermediate	WTISPLC	5
113	US / UK Foreign Exchange Rate	EXUSUK	5
114	Switzerland / US Foreign Exchange Rate	EXSZUS	5
115	Japan / US Foreign Exchange Rate	EXJPUS	5
116	Canada / US Foreign Exchange Rate	EXCAUS	5
117	ISM Manufacturing: PMI Composite Index	PMI	1
118	ISM Manufacturing: New Orders Index	NAPMNOI	1
119	ISM Manufacturing: Inventories Index	NAPMII	1
120	ISM Manufacturing: Employment Index	NAPMEI	1
121	ISM Manufacturing: Prices Index	NAPMPRI	1
122	ISM Manufacturing: Production Index	NAPMPI	1
123	ISM Manufacturing: Supplier Deliveries Index	NAPMSDI	1

This table summarizes information regarding the time series. Transformation code (TC): 1-level; 2-first difference; 3-second difference; 4-log-level; 5-first difference of logarithm; 6-second difference of logarithm. All times series have been downloaded from FRED.

5 On the Time-Varying Effects of Economic Policy Uncertainty on the US Economy

This chapter is joint work with Alexander Schlösser. We study the time-varying impact of Economic Policy Uncertainty (EPU) on the US Economy by using a VAR with time-varying coefficients. The coefficients are allowed to evolve gradually over time which allows us to discover structural changes without imposing them a priori. We find three different regimes which match the three major business cycles of the US economy, namely the Great Inflation, the Great Moderation and the Great Recession. This finding is in contrast to previous literature which typically imposes two regimes a priori. Furthermore, we distinguish the effect of EPU on real economic activity and on financial markets.

5.1 Introduction

In the context of the Great Recession Economic Policy Uncertainty (EPU) has been recognized as a major driver of the business cycle. A high degree of EPU has the potential to dampen economic activity. Baker et al. (2016) pioneered a newspaper based index to measure EPU. They used their EPU index to provide empirical evidence that EPU shocks cause a decline in both, employment and industrial production. Based on this index a large literature established further empirical evidence that EPU has a negative impact on economic activity, e.g. Bloom (2009), Baker et al. (2012), Colombo (2013) or Caldara et al. (2016). We contribute to this literature by investigating whether the effect of EPU on the US economy is time-varying.

To model the possibly time-varying impact of EPU shocks on the economy we use the time-varying parameter VAR (TVP-VAR) of Primiceri (2005). In the TVP-VAR the coefficients are allowed to evolve gradually over time. Thereby it is possible to detect structural changes without imposing them a priori. However, this flexible structure does not come without costs. First, it bears a high risk of overfitting and, second, estimation is only feasible with a small number of variables.

The first problem is typically tackled by imposing tight priors which regularize the amount of time-variation. The strength of these priors depends on a small set of hyperparameters which have to be set by the researcher. The ideal choice is, however, subject to a trade-off. While an overly loose prior may result in overfitting, an overly tight prior may suppress possible time-variation which we want to discover. Most applications of the TVP-VAR use fixed values on an ad hoc basis or the values used by Primiceri (2005). It is, however, unclear whether these values should be employed in other applications. Moreover, previous applications do not take into account that uncertainty about these hyperparameters may influence inference. We therefore estimate these hyperparameters using a fully Bayesian approach proposed by Amir-Ahmadi et al. (2018). We find that estimating the hyperparameters is important since using the benchmark values of Primiceri (2005) would underestimate some amount of time-variation. In order to address the second problem we follow Korobilis (2013*a*) and augment our TVP-VAR with a few factors which capture information from a large data set without introducing a degrees of freedom problem, instead of selecting a few variables from over 100 potential variables. This enables us to investigate simultaneously the impact of EPU on variables which represent real economic activity and on variables which mirror the activity on financial markets. This step turns out to be empirically important since EPU has an impact on a wide range of different variables. The impulse responses of macroeconomic variables share strong similarities over time while those of financial variables differ from the former.

Our main contribution is that we provide empirical evidence of a time-varying impact of EPU on the US economy by calculating time-varying impulse response functions. In principle, time-varying impulse response functions can vary along three dimensions, the initial impact, the overshooting behavior and the persistence of the shock. It turns out that the time-varying impulse responses vary across all three dimensions. During the 1970s, the Great Inflation, the initial impact was relatively high but was followed by overshooting which dampened the net impact of the shock. During the Great Moderation EPU shocks had a smaller impact on the economy. Finally, during the Great Recession the initial impact of EPU shocks again increased and had a persistent effect on the economy, preventing a quick recovery. We therefore find three different regimes which match the three major business cycles of the US economy, namely the Great Inflation, the Great Moderation and the Great Recession.

By modeling the time-varying effects of EPU on the US economy, we contribute to the growing literature focusing on the regime dependence of uncertainty shocks. Alessandri & Mumtaz (2014) use a threshold model and condition on the state of financial markets. Caggiano et al. (2017) and Popp & Zhang (2016) employ a smooth transition model and show that the effect depends on whether the economy is in recession or non-recession.[1] Castelnuovo et al. (2017) use an interacted VAR model and examine whether the effects of uncertainty are greater when the economy is at the zero lower bound. Of course, while these approaches have their individual appeal, we stress that we neither have to define ex ante a certain number of regimes, nor do we have to condition on a threshold variable, such as recession/ non-recession or a certain stance of monetary policy. Instead, we let the data guide us by allowing for time-varying model parameters. By doing so, we find three different regimes in contrast to previous research which typically imposes two regimes a priori. This is exactly the gap we are filling. Benati (2013) heads in a similar direction by using a TVP-VAR. We differ from Benati (2013) by also allowing for time variation in the autoregressive coefficients which turns out to be crucial for our empirical results. Furthermore, we additionally consider the period of the Great Inflation. Summing up, we extend the literature about the non-linear effects of EPU shocks on the US economy in a time-varying parameter environment over the last five decades.

The remainder of the chapter is structured as follows. Section 2 provides a brief overview of the underlying econometric model, Section 3 contains the empirical results and Section 4 concludes.

5.2 Methodology

5.2.1 TVP-FAVAR

In this section we discuss our econometric framework. We start with the TVP-VAR model based on Primiceri (2005). The model can be written in state space form as

$$\boldsymbol{y}_t = \boldsymbol{z}_t'\boldsymbol{\beta}_t + \boldsymbol{\Omega}_t^{1/2}\boldsymbol{\epsilon}_t, \tag{5.1}$$
$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, \tag{5.2}$$
$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} + \boldsymbol{v}_t, \tag{5.3}$$
$$\log\boldsymbol{\sigma}_t = \log\boldsymbol{\sigma}_{t-1} + \boldsymbol{w}_t, \tag{5.4}$$

where $\boldsymbol{z}_t' = \boldsymbol{I}_n \otimes [\boldsymbol{y}_{t-1}', \ldots, \boldsymbol{y}_{t-p}']$, $\boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{I}_n)$, $\boldsymbol{\eta}_t \sim N(\boldsymbol{0}, \boldsymbol{Q})$, $\boldsymbol{v}_t \sim N(\boldsymbol{0}, \boldsymbol{S})$ and $\boldsymbol{w}_t \sim N(\boldsymbol{0}, \boldsymbol{W})$ and the covariance matrix $\boldsymbol{\Omega}_t$ is decomposed as

$$\boldsymbol{\Omega}_t = \boldsymbol{A}_t^{-1}\boldsymbol{\Sigma}_t\boldsymbol{\Sigma}_t'(\boldsymbol{A}_t^{-1})', \tag{5.5}$$

[1] The conclusions drawn from these models might be too general. E.g., the recessions in 1990 and 2001 were relatively mild compared to the recessions in 1981 and 2007 so that the simple classification recession vs non-recession might miss the relevance of the respective depth of the recession.

where $\boldsymbol{\Sigma}_t$ is a diagonal matrix and $\boldsymbol{A}_t$ is a lower triangular matrix with ones on the main diagonal. Let $\boldsymbol{\alpha}_t$ denote the $n(n-1)/2$ vector of below-diagonal elements of $\boldsymbol{A}_t$ and let $\boldsymbol{\sigma}_t$ denote the vector consisting of all n diagonal elements in $\boldsymbol{\Sigma}_t$. Following Primiceri (2005), Cogley & Sargent (2005), Bernanke et al. (2005) and others we use two lags for the estimation.

In this setup the autoregressive coefficients, the covariances and the log standard deviation are allowed to evolve over time according to a random walk process and thereby allows us to detect structural breaks or regime changes. However, in contrast to regime switching models it does not need to impose a fixed number of regime changes prior to estimation as the parameters are allowed to take on a different value in each period. This flexible model structure, however, bears the risk of overfitting. The covariance matrix $\boldsymbol{Q}$ controls how much $\boldsymbol{\beta}_t$ is likely to change from t to $t+1$. Typically, researchers put a tight prior on $\boldsymbol{Q}$ in order to impose gradual changes in the parameters over time. The exact choice, however, is not straightforward. While an overly tight prior on $\boldsymbol{Q}$ may suppress possible time variation, an overly loose prior may result in overfitting the data. We employ similar priors to those used in Primiceri (2005),

$$\boldsymbol{\beta}_0 \sim N(\hat{\boldsymbol{\beta}}_{OLS}, V(\hat{\boldsymbol{\beta}}_{OLS})), \tag{5.6}$$

$$\boldsymbol{\alpha}_0 \sim N(\hat{\boldsymbol{\alpha}}_{OLS}, V(\hat{\boldsymbol{\alpha}}_{OLS})), \tag{5.7}$$

$$log\boldsymbol{\sigma}_0 \sim N(log\hat{\boldsymbol{\sigma}}_{OLS}, \boldsymbol{I}_n), \tag{5.8}$$

$$\boldsymbol{Q} \sim IW(k_{\boldsymbol{Q}} \cdot V(\hat{\boldsymbol{\beta}}_{OLS}), v_1), \tag{5.9}$$

$$\boldsymbol{S} \sim IW(k_{\boldsymbol{S}} \cdot V(\hat{\boldsymbol{\alpha}}_{OLS}), v_2), \tag{5.10}$$

$$\boldsymbol{W} \sim IW(k_{\boldsymbol{W}} \cdot \boldsymbol{I}_n, v_3), \tag{5.11}$$

where OLS denotes the OLS estimator using a training sample of ten years, $k_{\boldsymbol{Q}}, k_{\boldsymbol{S}}$ and $k_{\boldsymbol{W}}$ are hyperparameters set by the researcher and v denotes the degrees of freedom and is set such that the inverse Wishart prior has a finite mean and variance.

The importance of the hyperparameters $k_{\boldsymbol{S}}, k_{\boldsymbol{W}}$ and in particular of $k_{\boldsymbol{Q}}$ in this setup has been highlighted by Primiceri (2005) and Cogley & Sargent (2005). Nevertheless, most applications with this setup use fixed values on an ad hoc basis or the estimated values of Primiceri (2005).[2] It is, however, unclear whether the estimated values of Primiceri (2005) should be employed in other applications. Furthermore, previous applications using this model class do not take into account that uncertainty about the hyperparameters may influence inference. Therefore, we estimate the

[2]Primiceri (2005) estimates the hyperparameters over a small grid by maximizing the marginal likelihood.

hyperparameters $k_{\boldsymbol{Q}}, k_{\boldsymbol{S}}$ and $k_{\boldsymbol{W}}$ jointly with all other model parameters using a fully Bayesian approach as proposed by Amir-Ahmadi et al. (2018). This approach estimates the hyperparameters in a data-based fashion and takes the surrounding uncertainty into account.

The approach of Amir-Ahmadi et al. (2018) exploits the finding that only the prior of $\boldsymbol{X}$, $\boldsymbol{X} \in \{\boldsymbol{Q}, \boldsymbol{S}, \boldsymbol{W}\}$, depends on $k_{\boldsymbol{X}}$, and that, conditional on $\boldsymbol{X}$, all other model densities are independent from $k_{\boldsymbol{X}}$. Thus, the conditional posterior is

$$p(k_{\boldsymbol{X}}|\boldsymbol{X}) \propto p(\boldsymbol{X}|k_{\boldsymbol{X}})p(k_{\boldsymbol{X}}), \tag{5.12}$$

where $p(\boldsymbol{X}|k_{\boldsymbol{X}})$ denotes the prior of $\boldsymbol{X}$ and $p(k_{\boldsymbol{X}})$ the prior of $k_{\boldsymbol{X}}$, and can be obtained by a Metropolis-within-Gibbs step, as all other model densities cancel out in the acceptance probability.[3] We formulate relatively non-informative hierarchical inverse gamma priors for $p(k_{\boldsymbol{X}})$.

The curse of dimensionality typically forces researchers to include only a small number of variables in their VAR models. Primiceri (2005) for example uses the 3-Month Treasury Bill Yield as a measure for monetary policy, the unemployment rate as a measure of economic activity and inflation measured by the growth rate of a chain weighted GDP Price Index. Though, a variety of other measures exists and results may be sensitive to such choices. Furthermore, these variables may not fully represent the economy so that it may be necessary to include further variables, e.g., variables which capture information about the nature of consumers, organizations, business, financial or housing markets in order to be able to model the complex structure of the economy and to overcome the problem of non-fundamentalness.[4] Thus, instead of selecting a few variables from a set of over 100 potential variables, we follow Bernanke et al. (2005) and increase the information set used in a VAR by augmenting it with a few factors which capture the information of a large data set without introducing a degrees of freedom problem. That is, $\boldsymbol{y}_t$ consists of k factors and further variables of interest, in our case the monetary policy rate and EPU. Thereby our results are less sensitive to the concrete choice of variables.

We estimate the factors and model parameters following Stock & Watson (2005) and Korobilis (2013*a*) and use a simple two step approach. In the first step the factors $\boldsymbol{f}_t(k \times 1)$ are estimated using the first k principal components (PC) obtained from the singular value decomposition of the data matrix $\boldsymbol{x}_t$ $(m \times 1)$ with $k \ll m$. The

[3]For more details see Amir-Ahmadi et al. (2018).

[4]This concern typically arises if the econometrician does not use an information set which is identical or at a minimum closely overlapping to that used by policy makers, see Lippi & Reichlin (1994).

data matrix $\boldsymbol{x}_t$ contains our panel of macroeconomic variables. The PC estimates are then treated as observations. In the second step the parameters can be estimated conditional on these observed factors. Each observed variable x_{it}, for $i = 1, \ldots, m$, is linked to the k factors, to the monetary policy rate (R_t) and Economic Policy Uncertainty (epu_t) via the factor regression

$$x_{it} = \boldsymbol{\lambda}_i^f \boldsymbol{f}_t + \lambda_i^R R_t + \lambda_i^{epu} epu_t + \epsilon_{it} \tag{5.13}$$

where $\boldsymbol{\lambda}^f$ is $(1 \times k)$, λ^R, λ^{epu} are scalars and $\epsilon_{it} \sim N(0, \sigma_i^2)$. In order to model the dependence between factors and policy variables, the VAR model (5.1) is augmented with the obtained factors $\boldsymbol{y}_t = [\boldsymbol{f}_t', R_t, epu_t]'$. Following Bernanke et al. (2005) and Korobilis (2013*a*) we estimate the model using three factors.

5.2.2 Identification

We follow Canova & Nicolo (2002) and Uhlig (2005) and identify an EPU shock by placing sign restrictions on the contemporaneous responses of some of the variables in $\boldsymbol{x}_t$. In contrast to a Cholesky based identification scheme we avoid imposing zero restrictions and hence avoid an ordering of the variables which may be difficult to establish in an economically reasonable fashion. For instance, Caggiano et al. (2014) assume uncertainty to be slow-moving while Gilchrist et al. (2014) assume it to be fast-moving.[5] Instead, we impose restrictions in accordance with economic theory. We assume that the uncertainty shock has a negative contemporaneous effect on consumption and investment as well as unemployment as suggested by the *precautionary savings* and *real options* channel developed by Romer (1990) and Bernanke (1983). We leave the other shocks unidentified. Scholl & Uhlig (2008) or Rafiq & Mallick (2008) also identify a single shock. However, a monetary policy or demand shock may be associated with the same sign pattern. To distinguish the EPU shock from these shocks (i.e., shocks with the same sign pattern), we further assume that an EPU shock has the largest contemporaneous impact on EPU itself among all shocks. This approach is similar to maximizing the fraction of the forecast error variance at horizon zero which has been pioneered by Uhlig (2004*b*) and Uhlig (2004*a*) and used by Benati (2013) to identify an EPU-shock.

These sign and magnitude restrictions are implemented using the algorithm of

[5]Furthermore, simulation experiments suggest that identification based on sign restrictions performs well relative to identification methods based on contemporaneous zero restrictions, see Canova & Pina (1998), and that a standard Cholesky assumption can severely distort the impulse response functions, see Carlstrom et al. (2009).

Ramirez et al. (2010). First, we draw an $n \times n$ matrix, $\boldsymbol{J}$, from the N($\mathbf{0}$,$\boldsymbol{I}$) distribution. Second, calculating $\bar{\boldsymbol{Q}}$ from the QR decomposition of $\boldsymbol{J}$ provides a candidate structural impact matrix as $\boldsymbol{A}_{0,t} = \boldsymbol{A}_t^{-1}\boldsymbol{\Sigma}_t\bar{\boldsymbol{Q}}$. The candidate contemporaneous impulse response functions of $x_{1,t}, \ldots, x_{m,t}$ are then given by

$$\boldsymbol{IRF}_{0,t} = \boldsymbol{\Lambda} \times \boldsymbol{A}_{0,t}, \tag{5.14}$$

where $\boldsymbol{\Lambda}$ denotes the $m \times n$ matrix of factor loadings. The candidate matrix $\boldsymbol{A}_{0,t}$ is accepted if it satisfies the specified restrictions. In this setup, the shock is only set identified. Therefore, we follow the suggestion in Fry & Pagan (2011) and collect for each draw from the posterior 100 candidates $\boldsymbol{A}_{0,t}$ which satisfy the restrictions. Out of this set of 'admissible models' we select the one with elements closest to the median across these 100 candidates in order to deal with the multiple models problem.

5.3 Empirical Results

5.3.1 Data

We estimate the model based on a large data set covering 125 time series for the US economy. All variables are seasonally adjusted if necessary and standardized for the estimation. To incorporate the stance of monetary policy and the level of uncertainty we make use of the effective Federal Funds Rate and the historical version of the EPU index. Our quarterly data set starts in 1959:Q1 and ends in 2014:Q3. The names and the transformation codes of all series can be found in Table 5.1.

5.3.2 Impulse Responses

Figures 5.1a to 5.1j display the impulse response functions (IRF) for 12 selected time series of the US economy. Each plot consists of seven subplots. The three dimensional graph on the left displays the change of the response pattern over time and has been generated by allowing for time-variation in $\boldsymbol{\beta}_t$, $\boldsymbol{A}_t$ and $\boldsymbol{\Sigma}_t$. The upper three subfigures on the right display the effect of EPU on the respective variable for three different time periods, while the lower three subfigures display the difference between the three time periods along with 68% and 95% credible regions. A glance at the upper three subfigures of all variables shows that the initial impact credibly differ from zero at the 95% credible region. Since the model has been estimated with standardized data, the IRFs are standardized back such that the magnitude can be interpreted in the unit of measurement with respect to Table 5.1. In a first step, we focus on the sign of the impulse response for the selected variables followed by a second step, in which we look in more detail at the time-varying effect. In a third step we briefly comment on the magnitude and the economic significance of an EPU shock.

On the Sign of the Effect

The responses of the variables are in line with the theoretical sign pattern. The impulse responses of real GDP, consumption, investment and unemployment are depicted in Figure 5.1a to 5.1d. The former three respond negatively and the latter positively to an EPU shock. Therefore the signs are in accordance with the *precautionary savings* and *real options* channel developed by Romer (1990) and Bernanke (1983), respectively.

The IRF of the velocity of M1, housing starts, the S&P500 and the 10-Year Treasury Rate are depicted in Figure 5.1e to 5.1h. The negative response of the velocity of M1 reveals that a shock in EPU slows down economic activity. This does not come as a surprise since a shock depresses consumption, investment and increases unemployment which causes a reduction in short-term transactions. The IRF of housing starts is negative as expected. Both, the responses of the S&P500 as well as of the 10-Year Treasury Rate, are negative. The effect on the latter is in line with Scheffel (2015) who also found a 'flight to safety effect'.

Finally, we turn to Figures 5.1i and 5.1j which depict the IRFs of the ISM Manufacturing Composite Index and the capacity utilization in manufacturing. Both respond negatively, showing the adverse effects of EPU on business conditions in the US. Summing up, the signs are in accordance with economic theory.

On the Magnitude of the Effect

Beside the sign pattern and the shape of the IRFs, the magnitude of an EPU shock on the respective variable is of economic interest. In the following, we will briefly focus on the economic relevance of an EPU shock for most important macroeconomic variables. For the remaining variables the magnitude is economically reasonable as well since it is neither to high nor to low.

The initial impact on real GDP ranges between almost -0.3% for the Great Inflation and -0.2% for the Great Moderation. The dynamic effect (presistence) increases with the onset of the Great Recession supporting the slow recovery hypothesis. The initial impact on investment ranges between -1.2% during the Great Inflation and approximately -1.0% during the Great Recession. If we compare the size of the initial impact on investment with the size of the initial impact on consumption, which ranges approximately between -0.25% during the Great Inflation and approximately -0.15% for the Great Recession, it becomes obvious that consumers are less sensitive to an EPU shock than investors. This finding is in line with Prüser & Schlösser (2017) who found a similar result for European economies. In the case of unemployment, the initial impact varies slightly around an increase of 0.05 percentage points. Therefore, the empirical effect of a shock in EPU on unemployment is economically negligible.

On the Time Variation of the Effect

We now turn our attention to the degree of time variation in the IRFs. In principle time-varying IRFs can vary along three dimensions, the initial impact, the overshooting behavior[6] and the persistence of the shock. Investigating Figure 5.1a to 5.1j reveals substantial time variation in the IRFs of all variables. The profiles of the IRFs vary across all three dimensions, i.e., the initial impact of the shock, the overshooting behavior and the persistence of the shock. First, we focus on the time profile of the response of real GDP, depicted in the left subfigure of Figure 5.1a, followed by a wider macroeconomic perspective.

Focussing on Figure 5.1a, the response of real GDP, reveals that the first dimension, the initial impact, is relatively high during the 1970s, starts to decline in the early 1980s and stays stable until the early 2000s before it finally becomes larger with the onset of the financial crisis. The overshooting behavior, the second dimension, is pronounced during the 1970s, almost disappears from the 1980s onwards until the early 2000s and finally shows up again. Lastly, also the third dimension, the persistence of the shock, changes over time. From the 1970s onwards, the dynamic effects of the shocks are short-lived, became more persistent in the early 2000s and were most persistent with the onset of the financial crisis.

We will now look on the wider macroeconomic impact. Most variables under consideration reflect the macroeconomic conditions of the US economy and, compared to the two financial variables, namely the S&P500 as well as the 10-Year Treasury Rate, share strong similarities across these three dimensions. Those variables are real gross investment, real consumption, the civilian unemployment rate or the ISM Manufacturing PMI Composite Index (see Figure 5.1b, 5.1d, 5.1c, 5.1i). Their profiles differ, beside minor movements, mostly in terms of the effect size. Interestingly, their IRF profiles (across the three dimensions) mimic a close relationship with the three major business cycles which have been identified previously, namely the Great Inflation, the Great Moderation and the Great Recession. During the Great Inflation, the initial impact of an uncertainty shock was relatively high followed by overshooting which

[6] Within this framework, overshooting refers to a non monotonic IRF.

dampens the net effect. With the onset of the Great Moderation, the initial impact became smaller and the overshooting behavior almost disappeared. Finally, since the onset of the Great Recession, the initial impact again increased, the shocks became more persistent and overshooting returned. Summing up, the macroeconomic effects of a shock in EPU seem to depend on the major business cycles of the US economy rather than on business cycles at lower frequencies. This seems plausible since both, the Great Inflation and the Great Recession, were periods with fundamental economic turmoil while the Great Moderation has only been interrupted by two mild recessions in 1990 and 2001.

After having identified that the impact of EPU on a wide range of different variables changes between the Great Inflation, the Great Moderation and the Great Recession we now ask whether the responses differ credibly across these business cycles. Therefore, we choose a reference date for each business cycle and calculate credible regions for the difference between these dates. The reference date for the Great Inflation is 1970 Q2, for the Great Moderation, 1996 Q1 and for the Great Recession, 2009 Q1.[7] The results are depicted in the lower right part of Figure 5.1a to 5.1j. We focus on the responses of Real GDP, depicted in Figure 5.1a. However, the responses of the remaining macroeconomic variables give very similar results. The comparison of the response of the US economy to a shock in EPU during the Great Inflation and the Great Moderation reveals that they do not differ credibly from each other. The pattern changes if we compare the Great Inflation with the Great Recession and the Great Moderation with the Great Recession. For the former, we clearly obtain a difference at the 68% credible region and almost a difference at the 95% credible region. The responses of the latter differ credibly at the 95% credible region. Note that the Great Inflation and the Great Moderation differ credibly from each other if we fix the error covariance matrix at its posterior mean as done in following Section. There we will argue that the overshooting and the persistence of the responses credibly separate these episode from each other.

Impulse Responses with Fixed Error Covariance Matrix

During the previous analysis we allowed $\boldsymbol{\beta}_t, \boldsymbol{A}_t$ and $\boldsymbol{\Sigma}_t$ to vary over time. However, this approach does not allow to disentagle which component of the model causes the changing pattern in the IRFs. The time-variation might be due to either changes in the autoregressive coefficients $\boldsymbol{\beta}_t$, or due to changes in the error covariance matrix $\boldsymbol{\Omega}_t$. Differentiating between the two sources is important since it allows us to figure out whether the economy responds differently to an EPU shock (captured by changes in

[7] The dates are chosen arbitrarily and choosing different dates would give similar results.

$\boldsymbol{\beta}_t$) or if the shock size and the shock structure have changed (captured by changes in $\boldsymbol{\Omega}_t$). In order to differentiate between these two sources, we recalculated the IRFs for real GDP thereby keeping the error covariance matrix $\boldsymbol{\Omega}_t$ fixed at its posterior mean. Figure 5.2 depicts the IRF for this setup. The initial impact is the same over time by construction. But the IRF still varies along the other two dimensions, the overshooting behavior and the persistence of the shock. Overshooting is pronounced during the Great Inflation, does not appear during the Great Moderation and finally returns during the Great Recession. Furthermore, the persistence of the shock increases strongly with the onset of the Great Recession. Interestingly, testing whether the three periods differ from each other suggests credible differences between all three subperiods at the 95% credible region. This finding suggests that substantial structural changes occurred moving from one regime to another and highlights our key finding that we identified three different regimes, namely the Great Inflation, the Great Moderation and the Great Recession. The Great Inflation and the Great Recession differ in how the economy responds to an EPU shock. In the Great Inflation period the responses are characterized by an overshooting behavior and the responses in the Great recession are characterized by a persistent effect of an EPU shock.

5.3.3 On the Relevance of Estimating the Hyperparameters

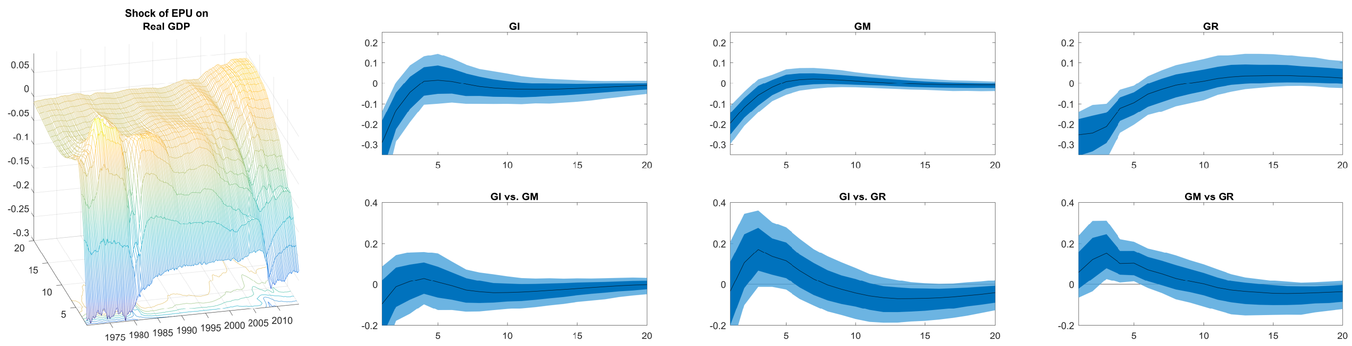
We now turn to the importance of estimating the hyperparameters by using the approach of Amir-Ahmadi et al. (2018). Therefore, we compare our estimated hyperparameters with those usually used in the literature. Furthermore, we compare our IRFs with the ones obtained by using the benchmark values. The posterior distributions of $k_{\boldsymbol{Q}}$, $k_{\boldsymbol{W}}$ and $k_{\boldsymbol{S}}$ are depicted in Figure 5.3. The red line in each histogram corresponds to the typically used benchmark values. The benchmark values for $k_{\boldsymbol{Q}}$ and $k_{\boldsymbol{S}}$ are not even in the domain of the corresponding posterior distribution while the benchmark value of $k_{\boldsymbol{W}}$ is quite close to the mode of the relevant posterior distribution. Thus, if we had used the benchmark values commonly used in the literature, we would have underestimated the amount of time variation in the autoregressive coefficients, $\boldsymbol{\beta}_t$, as well as in the contemporaneous correlation of the error term $\boldsymbol{\alpha}_t$. To highlight the importance of estimating the hyperparameters for our results, Figure 5.4 displays the impulse responses derived by using the commonly used benchmark values. Comparing Figure 5.1a with Figure 5.4 clearly demonstrates the difference. Using the benchmark values suppresses a substantial amount of time variation in the IRF. The resulting IRF differs across the three dimensions described above, namely the initial impact, the overshooting behavior and the persistence profile of the shock. But without estimating the hyperparameters, the correlation with the three major business cycles of the US economy would be less clear.

5.4 Conclusions

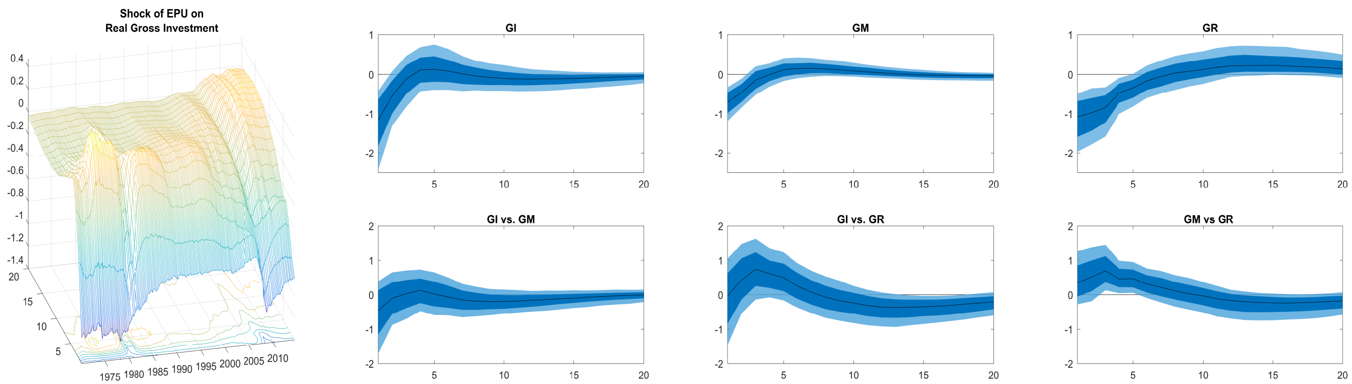
We estimate a time-varying parameter VAR in the spirit of Primiceri (2005) with data-based hyperparameters estimated in a fully Bayesian approach to investigate the time-varying impact of EPU shocks on the US economy. The TVP-VAR coefficients are allowed to evolve gradually over time. Thereby, it is possible to detect structural changes without imposing them a priori. To increase the information set in our model and, at the same time, keep the estimation of the model feasible, we follow Korobilis (2013*a*) and augment our TVP-VAR with a few factors. This enables us to investigate simultaneously the impact of EPU on variables which represent real economic activity and on variables which mimic the activity on financial markets.

Our main results are threefold: First, we find empirical evidence of a time-varying impact of EPU on the US economy. Interestingly, the shape of the IRFs strongly correlates with the three major business cycles of the US economy, namely the Great Inflation, the Great Moderation and the Great Recession and therefore, our econometric approach has discovered three different regimes. This is in contrast to previous research which typically imposes two regimes a priori. The time-varying impulse responses vary across three dimensions, the initial impact, the overshooting behavior and the persistence. During the 1970s, the Great Inflation, the initial impact was relatively high but was followed by overshooting which dampened the net impact of the shock. During the Great Moderation EPU shocks had a smaller impact on the economy. Finally, during the Great Recession the initial impact of EPU shocks again increased and had a more persistent effect on the economy, preventing a quick recovery. Second, we find that estimating the hyperparameters is important since using the benchmark values of Primiceri (2005) would underestimate the amount of time-variation. Third, we find that the responses of macroeconomic variables share strong similarities over time and that the response of financial variables differs from these.

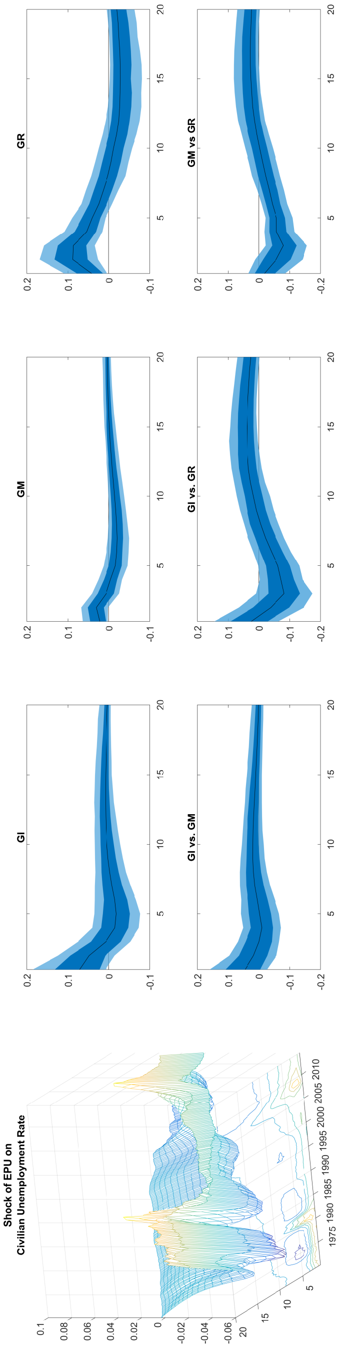
5.A Figures



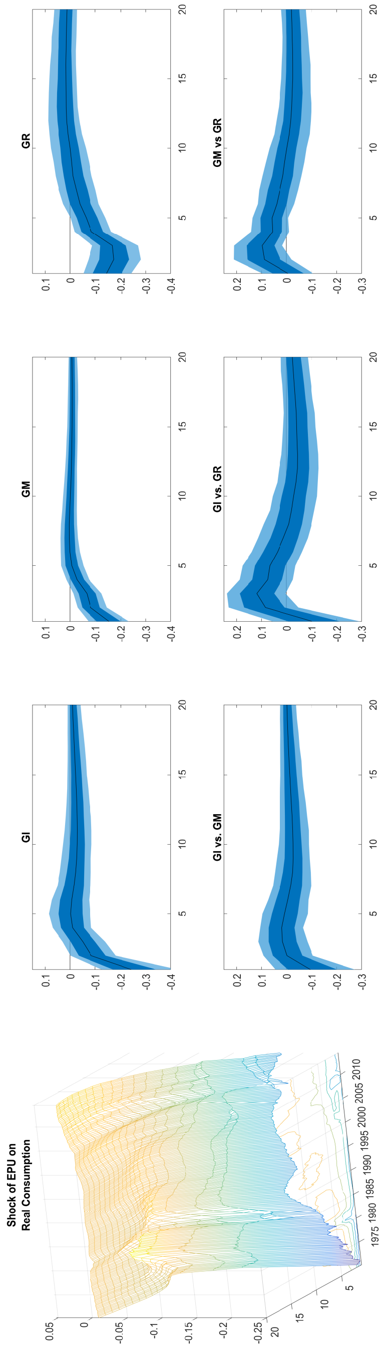
(a) *IRF of GDP to an increase in Economic Policy Uncertainty by one standard deviation.* The three dimensional graph displays the change in the pattern of the response over time. The upper three subfigures display the effect of economic policy uncertainty on the respective variable for three different time periods while the lower three subfigures display the difference between the three time periods along with 68% and 95% credible regions. 'GI' denotes Great Inflation, 'GM' Great Moderation and 'GR' Great Recession. Note that the three dimensional IRFs have been generated by allowing for time-variation in $\boldsymbol{\beta}_t, \boldsymbol{A}_t$ and $\boldsymbol{\Sigma}_t$.



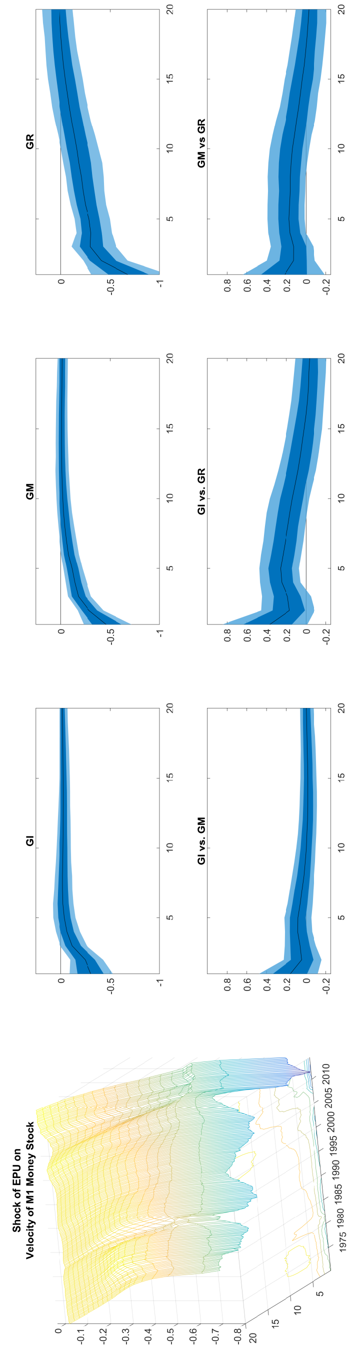
(b) *Impulse Response of Investment to an increase in Economic Policy Uncertainty by one standard deviation.*



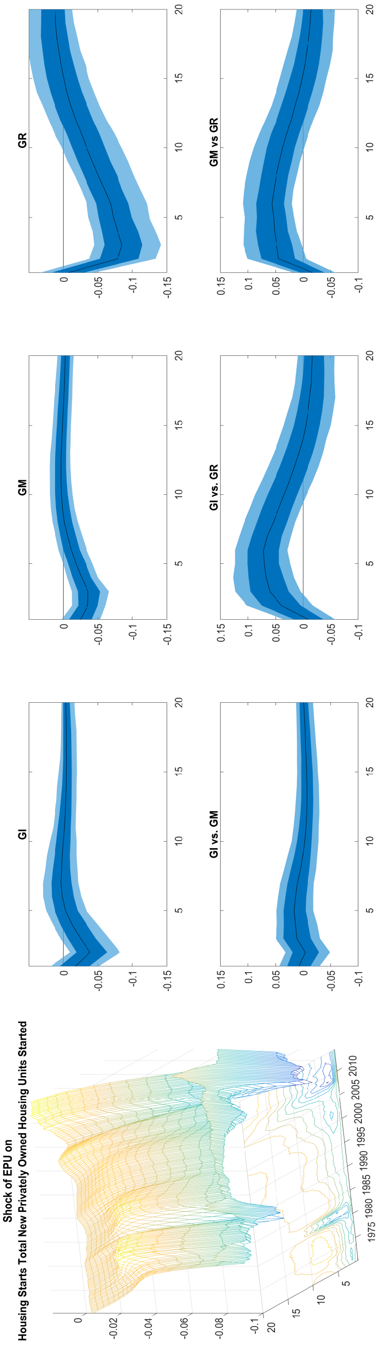
(c) *Impulse Response of Unemployment to an increase in Economic Policy Uncertainty by one standard deviation.*



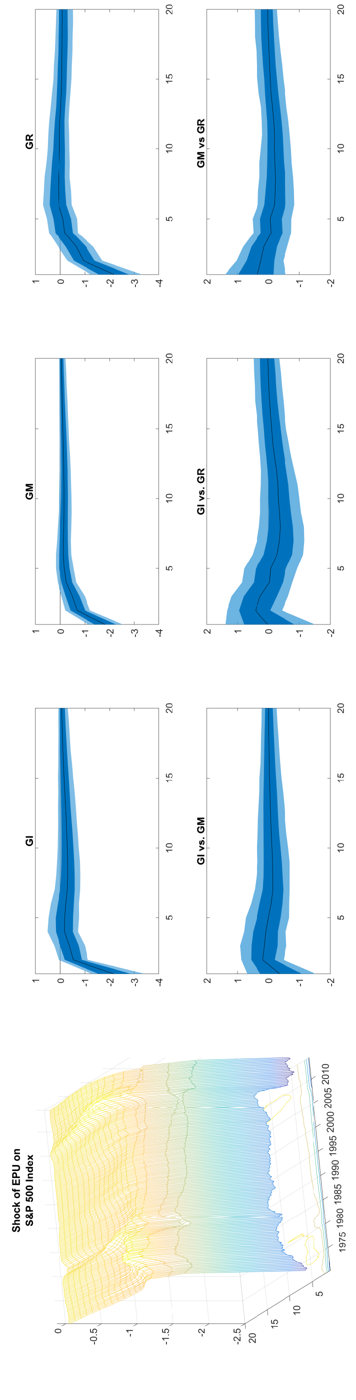
(d) *Impulse Response of Consumption to an increase in Economic Policy Uncertainty by one standard deviation.*



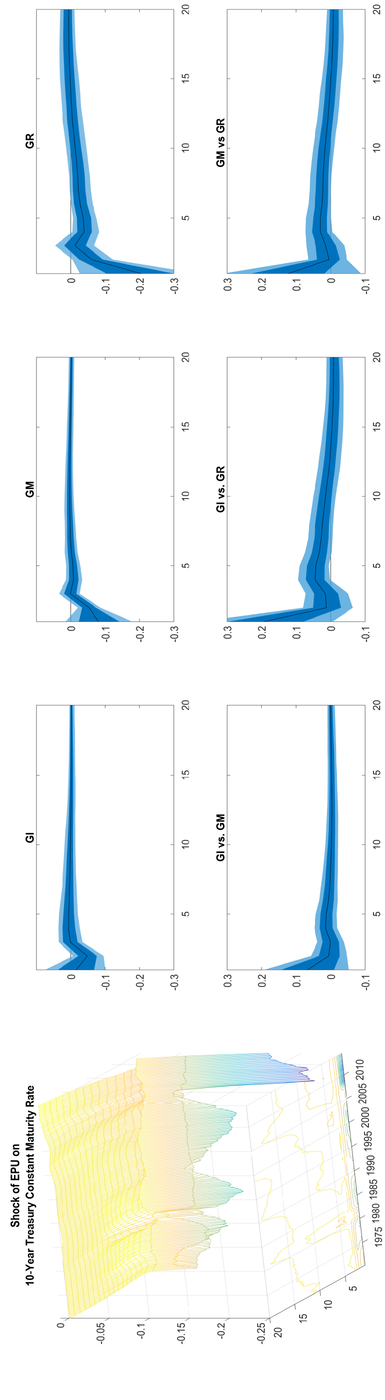
(e) *Impulse Response of Velocity of M1 to an increase in Economic Policy Uncertainty by one standard deviation.*



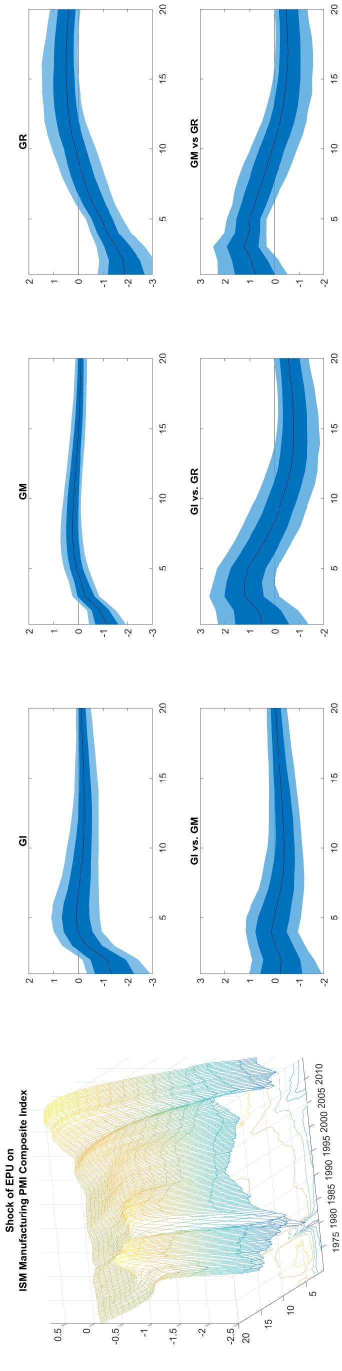
(f) *Impulse Response of Housing Starts to an increase in Economic Policy Uncertainty by one standard deviation.*



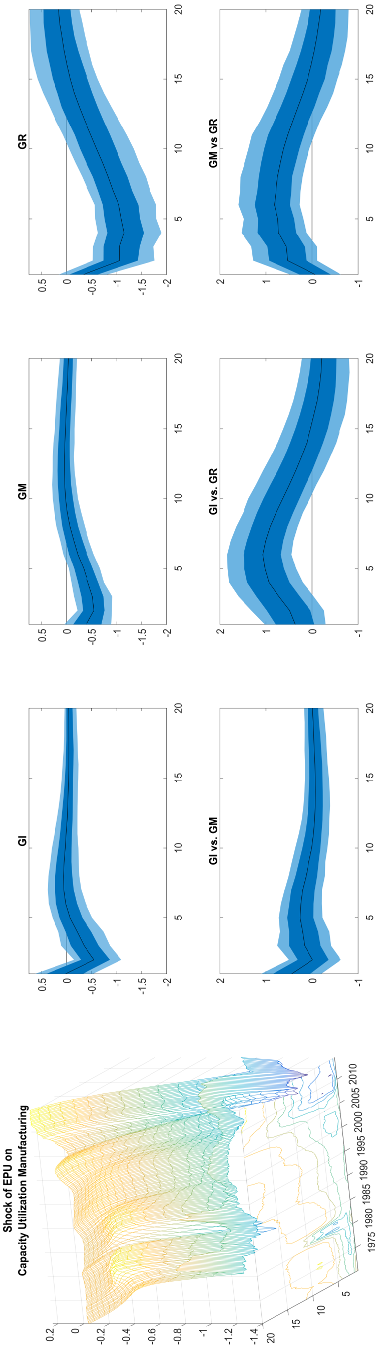
(g) *Impulse Response of S&P500 to an increase in Economic Policy Uncertainty by one standard deviation.*



(h) *Impulse Response of 10-Year Treasury Rate to an increase in Economic Policy Uncertainty by one standard deviation.*



(i) *Impulse Response of ISM Composite Index to an increase in Economic Policy Uncertainty by one standard deviation.*



(j) *Impulse Response of Capacity Utilization to an increase in Economic Policy Uncertainty by one standard deviation.*

Figure 5.1: *Impulse Response Functions.*

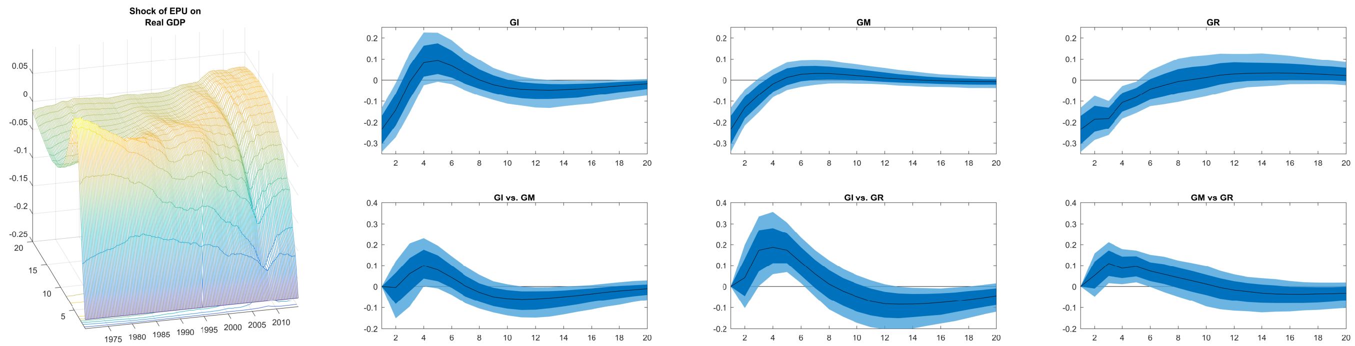
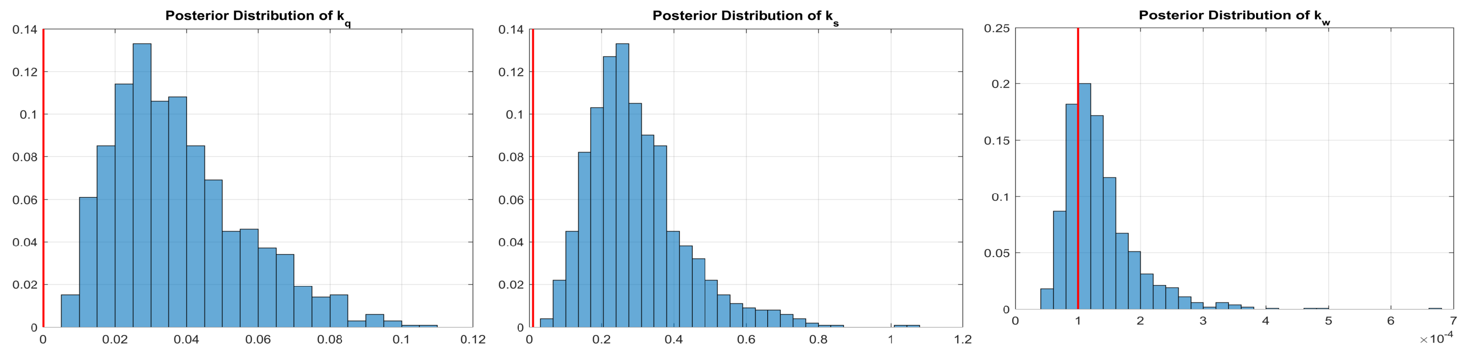


Figure 5.2: *Response of Real GDP to a shock in Economic Policy Uncertainty if the covariance matrix* $\boldsymbol{\Omega_t}$ *is fixed at its posterior mean. 'GI' denotes Great Inflation, 'GM' Great Moderation and 'GR' Great Recession.*



(a) (b) (c)

Figure 5.3: *Posterior distributions of $k_{\boldsymbol{Q}}$, $k_{\boldsymbol{S}}$ and $k_{\boldsymbol{W}}$. The red lines correspond to the benchmark values used in Primiceri (2005).*

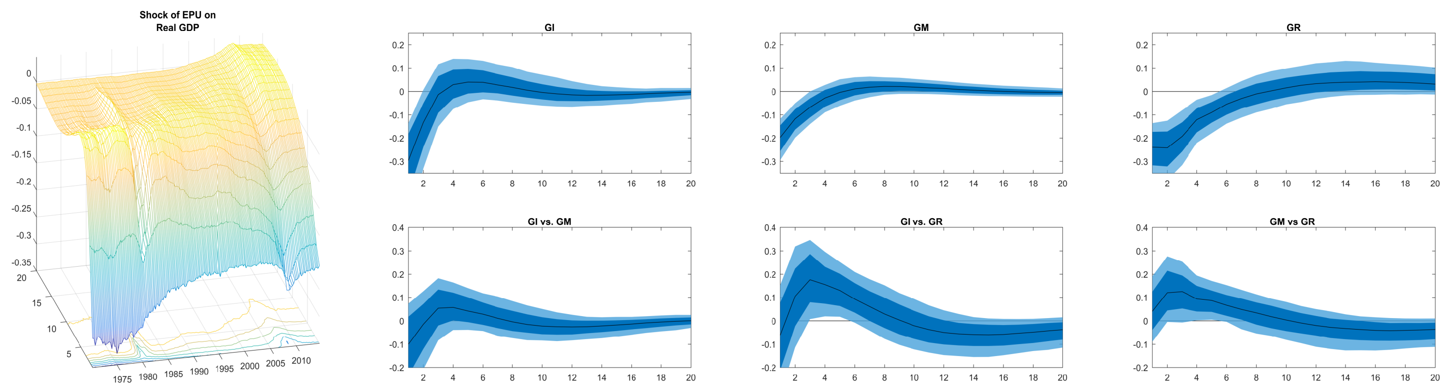


Figure 5.4: *Response of Real GDP to a shock in Economic Policy Uncertainty estimated by using the benchmark values suggested in Primiceri (2005). 'GI' denotes Great Inflation, 'GM' Great Moderation and 'GR' Great Recession*

5.B Data

Table 5.1: Data

No.	Name	ID	TC
1	Real Gross Domestic Product, 3 Decimal	GDPC96	5
2	Gross Domestic Product: Implicit Price Deflator	GDPDEF	5
3	Real Personal Consumption Expenditures	PCECC96	5
4	Personal Consumption Expenditures: Chain-type Price Index	PCECTPI	5
5	Real Gross Private Domestic Investment, 3 Decimal	GPDIC96	5
6	Real Imports of Goods & Services, 3 Decimal	IMPGSC96	5
7	Real Exports of Goods & Services, 3 Decimal	EXPGSC96	5
8	Real Change in Private Inventories	CBIC96	1
9	Real Final Sales of Domestic Product	FINSLC96	5
10	Gross Saving	GSAVE	5
11	Real Government Consumption Expenditures & Gross Investment	GCEC96	5
12	State & Local Government Current Expenditures	SLEXPND	6
13	State & Local Government Gross Investment	SLINV	6
14	Real Disposable Personal Income	DPIC96	6
15	Personal Income	PINCOME	6
16	Personal Saving	PSAVE	5
17	Private Residential Fixed Investment	PRFI	6
18	Private Nonresidential Fixed Investment	PNFI	6
19	Personal Consumption Expenditures: Durable Goods	PCDG	5
20	Personal Consumption Expenditures: Nondurable Goods	PCND	5
21	Personal Consumption Expenditures: Services	PCESV	5
22	Gross Private Domestic Investment: Chain-type Price Index	GPDICTPI	6
23	Compensation of Employees: Wages & Salary Accruals	WASCUR	6
24	Net Corporate Dividends	DIVIDEND	6
25	Corporate Profits After Tax	CP	6
26	Corporate: Consumption of Fixed Capital	CCFC	6
27	Housing Starts: Total: New Privately Owned Housing Units Started	HOUST	4
28	Privately Owned Housing Starts: 1-Unit Structures	HOUST1F	4
29	Privately Owned Housing Starts: 5-Unit Structures or More	HOUST5F	4
30	Housing Starts in Midwest Census Region	HOUSTMW	4
31	Housing Starts in Northeast Census Region	HOUSTNE	4
32	Housing Starts in South Census Region	HOUSTS	4
33	Housing Starts in West Census Region	HOUSTW	4

Table 5.1: Data(continued)

No.	Name	ID	TC
34	Industrial Production Index	INDPRO	5
35	Industrial Production: Consumer Goods	IPCONGD	5
36	Industrial Production: Durable Consumer Goods	IPDCONGD	5
37	Industrial Production: Nondurable Consumer Goods	IPNCONGD	5
38	Industrial Production: Materials	IPMAT	5
39	Industrial Production: Durable Materials	IPDMAT	5
40	Industrial Production: nondurable Materials	IPNMAT	5
41	Industrial Production: Business Equipment	IPBUSEQ	5
42	Industrial Production: Final Products (Market Group)	IPFINAL	5
43	Capacity Utilization: Manufacturing	CUMFNS	1
44	Civilians Unemployed - Less Than 5 Weeks	UEMPLT5	5
45	Civilians Unemployed for 5-14 Weeks	UEMP5TO14	5
46	Civilians Unemployed for 15-26 Weeks	UEMP15T26	5
47	Civilians Unemployed for 27 Weeks and Over	UEMP27OV	5
48	Civilian Unemployment Rate	UNRATE	2
49	Total Nonfarm Payrolls: All Employees	PAYEMS	5
50	All Employees: Nondurable Goods Manufacturing	NDMANEMP	5
51	All Employees: Durable Goods Manufacturing	DMANEMP	5
52	All Employees: Construction	USCONS	5
53	All Employees: Goods-Producing Industries	USGOOD	5
54	All Employees: Financial Activities	USFIRE	5
55	All Employees: Wholesale Trade	USWTRADE	5
56	All Employees: Trade, Transportation & Utilities	USTPU	5
57	All Employees: Retail Trade	USTRADE	5
58	All Employees: Natural Resources & Mining	USMINE	5
59	All Employees: Professional & Business Services	USPBS	5
60	All Employees: Leisure & Hospitality	USLAH	5
61	All Employees: Information Services	USINFO	5
62	All Employees: Education & Health Services	USEHS	5
63	All Employees: Service-Providing Industries	SRVPRD	5
64	All Employees: Total Private Industries	USPRIV	5
65	All Employees: Government	USGOVT	5
66	Average Hourly Earnings: Manufacturing	AHEMAN	6
67	Average Hourly Earnings: Construction	AHECONS	6
68	Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing	AWHMAN	5
69	Average Weekly Hours: Overtime: Manufacturing	AWOTMAN	5

Table 5.1: Data(continued)

No.	Name	ID	TC
70	Civilian Employment-Population Ratio	EMRATIO	5
71	Civilian Participation Rate	CIVPART	5
72	Business Sector: Output Per Hour of All Persons	OPHPBS	5
73	Nonfarm Business Sector: Unit Labor Cost	ULCNFB	5
74	Commercial and Industrial Loans at All Commercial Banks	BUSLOANS	6
75	Real Estate Loans at All Commercial Banks	REALLN	6
76	Total Consumer Credit Owned and Securitized, Outstanding	TOTALSL	5
77	Total Loans and Leases at Commercial Banks	LOANS	6
78	Bank Prime Loan Rate	MPRIME	2
79	1-Year Treasury Constant Maturity Rate	GS1	2
80	3-Year Treasury Constant Maturity Rate	GS3	2
81	5-Year Treasury Constant Maturity Rate	GS5	2
82	10-Year Treasury Constant Maturity Rate	GS10	2
83	Effective Federal Funds Rate	FEDFUNDS	2
84	3-Month Treasury Bill: Secondary Market Rate	TB3MS	2
85	6-Month Treasury Bill: Secondary Market Rate	TB6MS	2
86	Moody's Seasoned Aaa Corporate Bond Yield	AAA	2
87	Moody's Seasoned Baa Corporate Bond Yield	BAA	2
88	M1 Money Stock	M1SL	6
89	M2 Money Stock	M2SL	6
90	Currency Component of M1	CURRSL	6
91	Demand Deposits at Commercial Banks	DEMDEPSL	6
92	Savings Deposits - Total	SAVINGSL	6
93	Total Checkable Deposits	TCDSL	6
94	Travelers Checks Outstanding	TVCKSSL	6
95	Currency in Circulation	CURRCIR	6
96	MZM Money Stock	MZMSL	6
97	Velocity of M1 Money Stock	M1V	5
98	Velocity of M2 Money Stock	M2V	5
99	Total Nonrevolving Credit Outstanding	NONREVSL	6
100	Total Consumer Credit Outstanding	TOTALSL	6
101	Consumer Price Index for All Urban Consumers: All Items	CPIAUCSL	6
102	Consumer Price Index for All Urban Consumers: Commodities	CUSR0000SAC	6
103	Consumer Price Index for All Urban Consumers: All Items Less Energy	CPILEGSL	6
104	Consumer Price Index for All Urban Consumers: All Items Less Food	CPIULFSL	6

Table 5.1: Data(continued)

No.	Name	ID	TC
105	Consumer Price Index for All Urban Consumers: Energy	CPIENGSL	6
106	Consumer Price Index for All Urban Consumers: Food	CPIUFDSL	6
107	Consumer Price Index for All Urban Consumers: Apparel	CPIAPPSL	6
108	Consumer Price Index for All Urban Consumers: Medical Care	CPIMEDSL	6
109	Consumer Price Index for All Urban Consumers: Transportation	CPITRNSL	6
110	Producer Price Index: All Commodities	PPIACO	6
111	S&P 500 Index	SP500	5
112	Spot Oil Price: West Texas Intermediate	WTISPLC	5
113	US / UK Foreign Exchange Rate	EXUSUK	5
114	Switzerland / US Foreign Exchange Rate	EXSZUS	5
115	Japan / US Foreign Exchange Rate	EXJPUS	5
116	Canada / US Foreign Exchange Rate	EXCAUS	5
117	ISM Manufacturing: PMI Composite Index	PMI	1
118	ISM Manufacturing: New Orders Index	NAPMNOI	1
119	ISM Manufacturing: Inventories Index	NAPMII	1
120	ISM Manufacturing: Employment Index	NAPMEI	1
121	ISM Manufacturing: Prices Index	NAPMPRI	1
122	ISM Manufacturing: Production Index	NAPMPI	1
123	ISM Manufacturing: Supplier Deliveries Index	NAPMSDI	1
124	Total Borrowings of Depository Institutions from the Federal Reserve	BORROW	6
125	SPOT MARKET PRICE INDEX:BLS & CRB: ALL COMMODITIES(1967=100)	PSCCOM	5

This table summarizes information regarding the time series. Transformation code (TC): 1-level; 2-first difference; 3-second difference; 4-log-level; 5-first difference of logarithm; 6-second difference of logarithm. All times series have been downloaded from FRED with the exception of series No.125 which has been retrieved from Datastream.

5.C The Gibbs Sampler for the TVP-VAR

Here we briefly describe the Markov Chain Monte Carlo (MCMC) algorithm which allows to sample from the joint posterior distributions of all coefficients. The algorithm is the same as in Del Negro & Primiceri (2015), but adds the Metropolis-within-Gibbs step to sample the hyperparameter ($k_{\boldsymbol{Q}}$, $k_{\boldsymbol{S}}$ and $k_{\boldsymbol{W}}$) as in Amir-Ahmadi et al. (2018). To draw from the joint posterior distributions, we draw from the following conditional posterior distributions:

1. Draw $\boldsymbol{\Sigma_t}$ from its conditional distribution $p(\boldsymbol{\Sigma_t}|\boldsymbol{y^T}, \boldsymbol{\beta^T}, \boldsymbol{\alpha^T}, \boldsymbol{I_n}, \boldsymbol{Q}, \boldsymbol{S}, \boldsymbol{W}, \boldsymbol{s^T}, k_{\boldsymbol{Q}}, k_{\boldsymbol{S}}, k_{\boldsymbol{W}})$, where $\boldsymbol{s^T}$ denotes the indicator vector needed to use the mixtures of normals approach suggested by Kim et al. (1998) to sample $\boldsymbol{\Sigma_t}$.[8]
2. Draw $\boldsymbol{\beta}^T$ from its conditional distribution $p(\boldsymbol{\beta^T}|-)$ by making use of the simulation smoother developed by Carter & Kohn (1994).[9]
3. Draw $\boldsymbol{\alpha}_t$ from its conditional distribution $p(\boldsymbol{\alpha^T}|-)$ by making use of the simulation smoother developed by Carter & Kohn (1994).
4. Draw $\boldsymbol{Q}|-$, $\boldsymbol{S}|-$ and $\boldsymbol{W}|-$ using standard expression from Inverse Wishart, see Primiceri (2005).
5. Draw $k_{\boldsymbol{X}}$, $\boldsymbol{X} \in \{\boldsymbol{Q}, \boldsymbol{W}, \boldsymbol{S}\}$ using the same Gaussian random walk Metropolis-Hastings algorithm with an automatic tuning step as in Amir-Ahmadi et al. (2018):
   a) At each Gibbs iteration i, draw a candidate $k^*_{\boldsymbol{X}}$ from $N(k^{i-1}_{\boldsymbol{X}}, \sigma^2_{k\boldsymbol{X}})$.
   b) Calculate the acceptance probability $\alpha^i_{k_{\boldsymbol{X}}} = \min\left(1, \frac{p(\boldsymbol{X}|k^*_{\boldsymbol{X}})p(k^*_{\boldsymbol{X}})}{p(\boldsymbol{X}|k^{i-1}_{\boldsymbol{X}})p(k^{i-1}_{\boldsymbol{X}})}\right)$.
   c) Accept the candidate draw by setting $k^i_{\boldsymbol{X}} = k^*_{\boldsymbol{X}}$ with probability $\alpha^i_{k_{\boldsymbol{X}}}$. Otherwise set $k^i_{\boldsymbol{X}} = k^{i-1}_{\boldsymbol{X}}$.
   d) Calculate the average acceptance ratio $\overline{\alpha}_{k_{\boldsymbol{X}}}$. Adjust $\sigma_{k_{\boldsymbol{X}}}$ at every qth iteration according to $\sigma^{New}_{k_{\boldsymbol{X}}} = \sigma_{k_{\boldsymbol{X}}} \frac{\overline{\alpha}_{k_{\boldsymbol{X}}}}{\alpha^*}$, with α^* being the target average acceptance ratio. This step is not used after a pre-burn-in phase.

[8] T is a superscript and therefore denotes a sample from the corresponding variable for $t = 1, ..., T$.
[9] The notation $\theta|-$ represents the conditional posterior of θ conditional on the data and draws of all other coefficients.

6. Draw $\boldsymbol{s^T}$, needed to use the mixtures of normals approach, see Kim et al. (1998).

6 The Effects of Economic Policy Uncertainty on European Economies: Evidence from a TVP-FAVAR

This chapter is joint work with Alexander Schlösser. We use a time-varying parameter FAVAR model to investigate the effects of Economic Policy Uncertainty (EPU) on a wide range of macroeconomic variables for eleven European Monetary Union (EMU) countries. First, we are able to distinguish between a group of fragile countries (GIIPS-countries) and a group of stable countries (Northern countries), where the former suffered most due to EPU shocks. Second, we find that EPU shocks are transmitted through various channels, such as the real options-, the financial- and the precautionary savings channel and that private investors and financial market participants react more sensitively than consumers to EPU shocks. Third, we discover that the transmission of EPU shocks is quite stable over time.

6.1 Introduction

Economic policy uncertainty (EPU), as a driving force of the business cycle, has recently gained increasing attention. Events, such as the financial and sovereign debt crisis in Europe, have triggered uncertainty about a possible bailout of Greece, the continuity of the Eurozone and in general the capacity of policy makers to solve the crisis. The International Monetary Fund (2012), Baker et al. (2012) and, more recently, the ECB (2016) argue that increased EPU in Europe has hampered growth and the possibility of a faster recovery in the Eurozone after the crisis. This raises several questions. First, has EPU potentially harmed growth in the Euro area since the introduction of the Euro or has the effect of EPU changed since the financial crisis. Put differently, has the transmission or the size of EPU shocks varied over time? Second, the theory of optimum currency areas suggests that asymmetric shocks undermine the business cycle synchronization. Therefore, it is important to investigate whether all countries respond similarly to an EPU shock. Third, three major transmission channels have been identified in the theoretical literature. The first describes how EPU affects investment, the second investigates the effect on private consumption and the third explains how EPU affects financial variables. This raises the question which channels are most important for each country and which variables are affected most by EPU shocks.

Baker et al. (2016) developed a newspaper-based index to measure EPU.[1] Recently, many empirical studies used this index to quantify the effect of EPU on the economy. Most studies use a constant parameter SVAR with a small set of variables, e.g., Bloom (2009), Baker et al. (2012), Colombo (2013), Caldara et al. (2016) or Aastveit et al. (2017).[2] A small constant parameter SVAR, however, is unable to investigate whether the transmission of EPU shocks changes over time, whether Euro areas' economies respond symmetrically to EPU shocks and which transmission channels are most important. In contrast our TVP-FAVAR model allows us to address these questions.

We contribute to the literature in at least three dimensions. First, previous studies have focused on one or two of the three channels discussed above. The TVP-FAVAR model allows us to investigate simultaneously how EPU shocks are transmitted through all three channels, which channel is relevant for each country, which further variables are affected and which variable is most affected by EPU shocks. Second, some studies have considered the effect of EPU shocks at the European aggregate level of different macroeconomic variables. However, these studies might miss that uncertainty shocks may have a heterogeneous effect across different European economies. Also, some effects may cancel out in the aggregate. For example, an EPU shock may hit fragile countries like the GIIPS-countries much harder compared to stable countries like Germany or France. Hence, we use the TVP-FAVAR model to investigate whether European economies respond differently to uncertainty shocks. Furthermore, by using Bayesian estimation techniques, we are able to calculate *credible* intervals of the difference between the countries' responses, in order to investigate whether these are credibly different from zero. Third, the TVP-part of our model allows us to investigate whether the transmission of uncertainty shocks or the size of shocks have changed over time. These three points allow us to gain a range of new insights about the transmission of EPU shocks.

[1] Other measures of uncertainty exist, e.g. the measure of Jurado et al. (2015). That measure captures macroeconomic uncertainty, bases on non-European data and has no obvious political component. In contrast, the measure of Baker et al. (2016) includes both components, a macroeconomic and a political, as also suggested by our historical decomposition (see section 4.5), and therefore mirrors the type of uncertainty we want to capture. Baker et al. (2016) document robustness and reliability of their index by a comparison of their algorithm based index with an index constructed from human reading of the same articles. Both indices are very similar.

[2] Notable exceptions are Benati (2013), Caggiano et al. (2014) and Popp & Zhang (2016).

In order to use the TVP-FAVAR model, we follow Stock & Watson (2005) and Korobilis (2013*a*) and estimate the unobserved factors using principal components in order to avoid implausible identification restrictions (needed in a MCMC estimation scheme). Conditional on the estimated factors we use the TVP-VAR model of Primiceri (2005) which is well established in the literature to model time variation in the autoregressive coefficients and the covariance matrix of the error terms. Primiceri (2005) specifies a small number of hyperparameters to control the degree of time variation allowed for in the coefficients. Their choice will affect posterior inference and influence the amount of time variation in the coefficients. Several studies use "benchmark" values, which may not be appropriate. Given the importance of these parameters, we estimate them jointly with all other model parameters using a fully Bayesian approach as recently proposed by Amir-Ahmadi et al. (2018). By estimating them, we start with a very flexible model, allowing for time-varying autoregressive coefficients, time-varying variances and time-varying covariances, after which the model endogenously decides which dimension of time variation is supported by the data.

Our analysis differs from the existing SVAR (e.g., Bloom (2009), Colombo (2013), Österholm & Stockhammar (2014), Alam (2015) and others) and FAVAR (e.g., Popp & Zhang (2016) focusing on the US) literatures in several ways. We extend the former because our FAVAR allows us to investigate all three channels for several countries at the same time. Our study is different from both types of empirical studies in that we allow all model parameters to change continuously over time. Our empirical framework permits us to study the effect of uncertainty independently of certain regimes, e.g., a "normal" or a "recessionary" regime like in Popp & Zhang (2016) and Caggiano et al. (2014) for the US economy. This is attractive since the effect of uncertainty on the economy might also depend on the political leadership of the economy or on other economic circumstances, such as plans to leave the EU. Benati (2013) considers a possible time-varying effect of EPU on aggregate Euro area variables by using a TVP-VAR. However, Benati (2013) only allows for time variation in the covariance matrix of the error terms and only considers the effect of EPU shocks at the European aggregate level. Lastly, our empirical approach allows us to investigate whether Euro area countries respond differently to an EPU shock, a question which has not been addressed so far by the empirical literature.

We strengthen the evidence that EPU shocks have hampered growth in Europe, as argued by International Monetary Fund (2012), Baker et al. (2012) and the ECB (2016). In particular, economic growth in the GIIPS countries suffered most from EPU shocks. Furthermore, investors request a higher risk premium for bonds of these countries, while for the northern countries a safe haven effect appears. This stresses that the effect of EPU on European countries is quite asymmetric. We further find that, while EPU shocks are transmitted through all three channels, investors and financial market participants react more sensitively than consumers to EPU. Finally, we discover that the transmission of EPU shocks is quite stable over time and, therefore, that EPU shocks played an important role even before the Euro crisis.

The remainder of the chapter is structured as follows. Section 2 summarizes the theoretical and empirical literature, Section 3 provides a brief overview of the underlying econometric model, Section 4 contains empirical results and Section 5 concludes.

6.2 Transmission Channels

Political uncertainty can affect the economy in several ways. A first channel is the real-options channel considered by Bernanke (1983). The premise is that investment and employment decisions are costly to revert. If decision makers are uncertain about the future of the economy, they might adopt a wait-and-see attitude and postpone investment and hiring. Thus, the option value is high when uncertainty is high and vice versa.[3] Several studies provided evidence supporting this channel, for example Bloom et al. (2007), Carrière-Swallow & Céspedes (2013) and Meinen & Röhe (2017) in terms of investment and Alexopoulos & Cohen (2009), Scheffel (2015) and Netšunajev & Glass (2017) in terms of employment.

A second channel developed by Romer (1990) explains why uncertainty affects consumption. If future income is uncertain, consumption is subject to a similar degree of irreversibility and leads households to postpone their consumption decision until uncertainty has resolved. This channel is sometimes referred to as *precautionary savings* channel. That is, uncertainty can affect the intertemporal consumption decisions made by households. Benito (2006), Carrière-Swallow & Céspedes (2013) and Caldara et al. (2016) provide evidence on this channel.[4]

Consumption and investment constitute approximately 75% of Euro area's Gross Domestic Product (GDP). Obviously, if either one or both are negatively affected by EPU, GDP should decrease as well. This indirect effect has been documented by Donadelli (2015) and Baker et al. (2016). Furthermore, a decrease in GDP can be

[3]The empirical importance of this channel is underpinned by a statement of the FOMC of October 2001: "Several [survey] participants reported that uncertainty about the economic outlook was leading firms to defer spending projects until prospects for economic activity became clearer." A similar statement can be found in the Minutes of the FOMC from December 15-16, 2009.

[4]In order to provide further evidence that households' decisions are affected by EPU we extend the data set by the consumer confidence indicator.

interpreted as a slowdown in aggregate demand, which leads, under the assumption of constant supply, to a reduction in inflation. Colombo (2013) and Belke & Osowski (2018) include inflation in their data set and both find evidence in favor of a negative relationship.

Through a third channel, uncertainty might affect financial markets. Within this third channel we differentiate between the effect of EPU on the cost of finance and the effect on the stock market. The former, known as the risk premium effect, describes that an increase in uncertainty may increase the variance of expected profitability of firms, which increases their perceived riskiness. Subsequently, investors require higher interest rates to be compensated for the higher risk, so that issuance of additional debt becomes more costly and adversely affects investment. Gilchrist et al. (2014) explore this hypothesis in a general equilibrium model and empirical evidence is provided by Nodari (2014) and Waisman et al. (2015). The latter study describes that, due to the effects of EPU on investment and the risk premium, stock prices are affected as well. Financial theory suggests that stock prices are determined by the sum of expected future cash flows, discounted at the appropriate risk-adjusted discount rate. Thus, a decrease in cash flow or an increase in the risk-adjusted discount rate may lower the stock price.[5] Studies like Chang et al. (2015) or Chen et al. (2016) find empirical support for this effect.

The effect of EPU on credit is less well explored in the literature. (Bordo et al. 2016, p. 90) provide a theoretical reasoning and empirical evidence on this transmission channel. They argue that "following the Great Recession, bankers complained that delays implementing financial reform under the Dodd-Frank Act created regulatory policy uncertainty that restrained lending, which, in turn, slowed economic recovery." Using a small VAR model they are able to show that EPU has a significant negative effect on bank loans.

6.3 Methodology

6.3.1 TVP-FAVAR

The VAR model introduced by Sims (1980) has become a popular tool to model dynamic relationships among macroeconomic variables and can be written in reduced form as

$$\boldsymbol{y}_t = \sum_{j=1}^{p} \boldsymbol{B}_j \boldsymbol{y}_{t-j} + \boldsymbol{u}_t, \tag{6.1}$$

[5] Expected future cash flow also depends on consumption which itself is adversely affected by EPU.

for $t = 1, \ldots, T$, where T denotes the total number of periods, $\boldsymbol{y}_t$ is a $n \times 1$ vector of endogenous variables, $\boldsymbol{B}_j$ are $n \times n$ coefficient matrices and $\boldsymbol{u}_t \sim N(\boldsymbol{0}, \boldsymbol{\Omega})$ are reduced form errors, with $\boldsymbol{\Omega}$ a $n \times n$ covariance matrix. Because of the curse of dimensionality VAR models typically only include a small number of variables. However, more variables may be necessary to avoid an omitted variable bias and to model complex relations between macroeconomic variables. Bernanke et al. (2005) propose to increase the information set used in a VAR by augmenting it with a few factors which capture information of a large data set without introducing a degrees of freedom problem. That is, $\boldsymbol{y}_t$ consists of k factors and further variables of interest, in our case economic policy uncertainty and the monetary policy rate. The use of the FAVAR model allows us to investigate through which channels an EPU shock is transmitted and how it affects European economies, which would not be possible within a standard small scale VAR.

Estimating the latent factors and model parameters jointly in one step, by making use of MCMC methods, allows for full treatment of uncertainty surrounding the latent factors and model parameters. Nevertheless, identification restrictions are needed in this approach, which lead to flat (unidentified) impulse response functions, useless for economic interpretation.[6] Thus, we follow Stock & Watson (2005) and Korobilis (2013*a*) and apply a simpler two step approach. In the first step, the factors $\boldsymbol{f}_t (k \times 1)$ are estimated using the first principal components (PC) obtained from the singular value decomposition of the data matrix $\boldsymbol{x}_t$ $(m \times 1)$ with $k \ll m$. The data matrix $\boldsymbol{x}_t$ contains our panel of macroeconomic variables. The PC estimates are then treated as observations, have an economic meaning and approximate the true factors in case of constant loadings. In the second step the parameters can be estimated conditional on these observed factors. Each of the observed variables x_{it} for $i = 1, \ldots, m$ is linked to the k factors, to economic policy uncertainty (epu_t) and the monetary policy rate (R_t) via the factor regression

$$x_{it} = \boldsymbol{\lambda}_i^f \boldsymbol{f}_t + \lambda_i^R R_t + \lambda_i^{epu} epu_t + \epsilon_{it} \tag{6.2}$$

where $\boldsymbol{\lambda}^f$ is $(1 \times k)$, λ^R, λ^{epu} are scalars and $\epsilon_{it} \sim N(0, \sigma_i^2)$. In order to model the dependence between factors and policy variables, the VAR model (6.1) is augmented with the obtained factors $\boldsymbol{y}_t = [\boldsymbol{f}_t', R_t, epu_t]'$.

So far, we have assumed time invariant parameters and a time invariant covariance matrix of the error terms. This a priori assumption is possibly too restrictive, given that events such as the introduction of the euro or the financial crisis might have changed the transmission or the average size of shocks over time. To model time

[6] For a discussion of these aspects see Korobilis (2013*a*).

variation in the parameter matrix $\boldsymbol{B}_j$ and covariance matrix $\boldsymbol{\Omega}$, we use the TVP-VAR model of Primiceri (2005). This allows us to investigate whether the transmission of an economic policy uncertainty shock, the size or the correlation of the shocks have changed over time. By assuming that coefficients evolve as multivariate random walks the TVP-FAVAR can be written in state space form as

$$\boldsymbol{y}_t = \boldsymbol{z}_t'\boldsymbol{\beta}_t + \boldsymbol{u}_t, \tag{6.3}$$
$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, \tag{6.4}$$

where $\boldsymbol{z}_t' = \boldsymbol{I}_n \otimes [\boldsymbol{y}_{t-1}', \ldots, \boldsymbol{y}_{t-p}']$, $\boldsymbol{\beta}_t = \text{vec}([\boldsymbol{B}_{1,t}, \ldots, \boldsymbol{B}_{p,t}]')$ and $\boldsymbol{\eta}_t \sim N(\boldsymbol{0}, \boldsymbol{Q})$ is a state disturbance term with system covariance matrix $\boldsymbol{Q}$ of dimension $l \times l$, with $l = n(np+1)$. The covariance matrix $\boldsymbol{Q}$ is important in determining the amount of time-variation in the regression's coefficient. Setting this matrix to zero would lead to constant coefficients over time, nesting a constant coefficient VAR as a special case. An increased variability of the coefficients however bears the risk of overfitting the data. This suggests to impose some structure on $\boldsymbol{Q}$, as discussed in the next section. The covariance matrix $\boldsymbol{\Omega}_t$ is decomposed as

$$\boldsymbol{\Omega}_t = \boldsymbol{A}_t^{-1}\boldsymbol{\Sigma}_t\boldsymbol{\Sigma}_t'(\boldsymbol{A}_t^{-1})', \tag{6.5}$$

where $\boldsymbol{\Sigma}_t$ is a diagonal matrix and $\boldsymbol{A}_t$ is a lower triangular matrix with ones on the main diagonal. Let $\boldsymbol{a}_t$ denote the $n(n-1)/2$ vector of below-diagonal elements of $\boldsymbol{A}_t$ and let $\boldsymbol{\sigma}_t$ denote the vector consisting of all n diagonal elements in $\boldsymbol{\Sigma}_t$. Then the complete model can be written in state space form as

$$\boldsymbol{y}_t = \boldsymbol{z}_t'\boldsymbol{\beta}_t + \boldsymbol{A}_t^{-1}\boldsymbol{\Sigma}_t\boldsymbol{\epsilon}_t, \tag{6.6}$$
$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, \tag{6.7}$$
$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} + \boldsymbol{v}_t, \tag{6.8}$$
$$\log\boldsymbol{\sigma}_t = \log\boldsymbol{\sigma}_{t-1} + \boldsymbol{w}_t, \tag{6.9}$$

where $\boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{I}_n)$, $\boldsymbol{v}_t \sim N(\boldsymbol{0}, \boldsymbol{S})$ and $\boldsymbol{w}_t \sim N(\boldsymbol{0}, \boldsymbol{W})$.[7] The priors are similar to those used in Primiceri (2005),

$$\boldsymbol{\beta}_0 \sim N(\hat{\boldsymbol{\beta}}_{OLS}, V(\hat{\boldsymbol{\beta}}_{OLS})), \tag{6.10}$$
$$\boldsymbol{\alpha}_0 \sim N(\hat{\boldsymbol{\alpha}}_{OLS}, V(\hat{\boldsymbol{\alpha}}_{OLS})), \tag{6.11}$$
$$log\boldsymbol{\sigma}_0 \sim N(log\hat{\boldsymbol{\sigma}}_{OLS}, \boldsymbol{I}_n), \tag{6.12}$$
$$\boldsymbol{Q} \sim IW(k_Q \cdot V(\hat{\boldsymbol{\beta}}_{OLS}), v_1), \tag{6.13}$$

[7] We adopted the additional assumption of $\boldsymbol{S}$ being block diagonal. This assumption is not crucial but simplifies inference and increases the efficiency of the algorithm. See Primiceri (2005) for details.

$$\boldsymbol{S} \sim IW(k_S \cdot V(\hat{\boldsymbol{\alpha}}_{OLS}), v_2), \tag{6.14}$$
$$\boldsymbol{W} \sim IW(k_W \cdot \boldsymbol{I}_n, v_3), \tag{6.15}$$

where OLS denotes the OLS estimator using the full sample[8], k_Q, k_S and k_W are hyperparameters set by the researcher and v denotes the degrees of freedom and is set such that the inverse Wishart prior has a finite mean and variance.

6.3.2 Estimation of Hyperparameters

The choice of hyperparameters k_S, k_W, and in particular of k_Q, is crucial, as it determines the amount of time variation in the autoregressive coefficients (see Primiceri, 2005). Researchers typically impose tight priors on $\boldsymbol{Q}$, thus controlling the amount of time variation, in order to avoid overfitting the data.[9] Unfortunately, the choice of the hyperparameters will affect posterior inference and influence the amount of time variation in the coefficients, a fact which is largely ignored in applications. Only a few studies, such as Primiceri (2005), select the hyperparameters over a small grid by maximizing the marginal likelihood. Inference is then conditioned on the selected hyperparameters. However, given the importance of the choice of these parameters, it may be desirable to sample the hyperparameters k_Q, k_S, k_W in a data-based fashion and take the uncertainty surrounding k_Q, k_S, k_W into account. Hence, we sample/estimate the hyperparameters k_Q, k_S, k_W jointly with all other model parameters using a fully Bayesian approach as proposed by Amir-Ahmadi et al. (2018). That has the additional advantage that one does not need to specify a grid. For example, the values in the grid specified in Primiceri (2005) for k_Q are not in the domain of the posterior we found for k_Q. The approach of Amir-Ahmadi et al. (2018) is based on the observation that only the prior of $\boldsymbol{X}$, $\boldsymbol{X} \in \{\boldsymbol{Q}, \boldsymbol{S}, \boldsymbol{W}\}$, depends on $k_{\boldsymbol{X}}$, and that, conditional on $\boldsymbol{X}$, all other model densities are independent of $k_{\boldsymbol{X}}$. Thus, the conditional posterior

$$p(k_{\boldsymbol{X}}|\boldsymbol{X}) \propto p(\boldsymbol{X}|k_{\boldsymbol{X}})p(k_{\boldsymbol{X}}), \tag{6.16}$$

where $p(\boldsymbol{X}|k_{\boldsymbol{X}})$ denotes the prior of $\boldsymbol{X}$ and $p(k_{\boldsymbol{X}})$ the prior of $k_{\boldsymbol{X}}$ can be obtained by a Metropolis-within-Gibbs step, as all other model densities cancel out in the acceptance probability.[10] We formulate relatively non-informative hierarchical inverse gamma priors for $p(k_{\boldsymbol{X}})$.

[8]The same prior specification can be found in Gambetti & Musso (2017).
[9]See, for example, the discussion in Primiceri (2005) and Cogley & Sargent (2005).
[10]For more details see Amir-Ahmadi et al. (2018).

6.3.3 Identification

Crucially, the FAVAR allows to investigate the effect of a shock in a policy variable on a wide range of macroeconomic variables. To obtain the vector moving average (VMA) representation we rewrite equations (6.1) and (6.2) as

$$\boldsymbol{x}_t = \Lambda \boldsymbol{y}_t + \Phi(L)\boldsymbol{x}_t + \boldsymbol{\epsilon}_t^{\boldsymbol{x}} \tag{6.17}$$

$$\boldsymbol{y}_t = \Psi_t(L)\boldsymbol{y}_t + \boldsymbol{A}_t^{-1}\boldsymbol{\Sigma}_t\boldsymbol{\epsilon}_t^{\boldsymbol{y}} \tag{6.18}$$

where Λ is $m \times n$ and $\Phi(L)$ as well as $\Psi_t(L)$ are lag polynomials. Inserting (6.18) in (6.17), we obtain the VMA representation

$$\boldsymbol{x}_t = \tilde{\Phi}(L)\boldsymbol{\Lambda}\tilde{\Psi}_t(L)\boldsymbol{A}_t^{-1}\boldsymbol{\Sigma}_t\boldsymbol{\epsilon}_t^{\boldsymbol{y}} + \tilde{\Phi}(L)\boldsymbol{\epsilon}_t^{\boldsymbol{x}} \tag{6.19}$$

with $\tilde{\Phi}(L) = (I - \Phi(L))^{-1}$ and $\tilde{\Psi}_t(L) = (I - \Psi_t(L))^{-1}$. Our preferred identification strategy is based on the assumption that EPU influences the real economy with a delay. This allow us to stay agnostic about the sign of the IRFs. The Cholesky decomposition imposes the identifying assumption that the latent factors do not respond to a EPU shock within the same period. Fortunately, we do not need to impose this assumption on all the variables in $\boldsymbol{x_t}$. Instead, in accordance with Bernanke et al. (2005), we categorize the variables in $\boldsymbol{x_t}$ to be "fast-moving" and "slow-moving".[11] Fast-moving variables are assumed to respond contemporaneously to an unanticipated change in EPU, while slow-moving variables do not. We employ the same assumption for the monetary policy shock. Table A.1 of the supplementary appendix contains details on the classification of the variables.[12] As a robustness check, we calculate IRFs using an alternative identification strategy. In order to identify the uncertainty shock, without using a recursive structure, we employ sign restrictions.[13] We assume that the uncertainty shock has a negative contemporaneous effect on consumption and investment as predicted by the precautionary savings and real option channel. We left the other shocks unidentified. However, a monetary or demand shock may be associated with the same sign pattern. To distinguish the uncertainty shock from these shocks (i.e. shocks with the same sign pattern), we further assume that uncertainty shock has a larger contemporaneous impact on uncertainty. These sign and magnitude restrictions are implemented using the algorithm of Ramirez et al. (2010).[14] The

[11] Popp & Zhang (2016) use the same strategy to identify uncertainty shocks in a FAVAR model.

[12] We also estimate the model with two different orderings. First, $\boldsymbol{y}_t = [\boldsymbol{f}_t', epu_t, R_t]'$. Second, $\boldsymbol{y}_t = [epu_t, \boldsymbol{f}_t', R_t]'$. Our main findings are qualitatively similar for both orderings.

[13] For the implementation of sign restriction in a FAVAR model, see Ellis et al. (2014).

[14] In this setup, the shock is only set identified. Therefore, we follow a suggestion in Fry & Pagan (2011) and keep for each draw from the posterior 100 candidates which satisfy the assumption and out of this set of "admissible models" we select the one with elements closest to the median across these 100 candidates.

IRFs for this identification strategy are qualitatively similar and are available up on request.

6.4 Empirical Results

6.4.1 Data

We estimate the model with data from eleven EMU countries which can be splitted in two groups. The first group consists of Greece, Italy, Ireland, Portugal and Spain. Those countries have been in focus in the run up of and during the sovereign debt crisis. We refer to these as the "GIIPS countries". The second group consists of the remaining countries France, Germany, Finland, Austria, Netherlands and Belgium, called "northern countries". For each country we consider nine macroeconomic variables consisting of gross domestic product, investment, consumption, the GDP deflator, the unemployment rate, credit to the non-financial privat sector, 10 year government bond yields, a stock market index and consumer confidence.[15] All variables are seasonally adjusted if necessary and standardized for the estimation of the PC. The set of further variables consists of EPU and the EURIBOR to approximate the ECB interest rate on main refinancing operations. We use quarterly data ranging from 1997:Q1 until 2016:Q1 restricted by data availability. The data sources and variable transformations can be found in the Table A.1 of the supplementary appendix.

6.4.2 Model Estimation

We estimate the VAR model using two lags.[16] Furthermore, in our preferred specification, we include three factors in our model. In addition, we consider models with up to six factors. As highlighted by Stock & Watson (1998), while the space spanned by the factors is still estimated consistently when the number of factors is overestimated though efficiency is reduced, an underestimation of the number of factors results in an inconsistent model as potentially important dynamics will not be captured by the factors. (Bernanke et al. 2005, p. 406) argue that "if the additional information was irrelevant then adding one factor to the VAR would render the estimation less precise, but the estimate should remain unbiased. We would thus not expect the estimated response to change considerably." This is exactly what we find in our estimation. Increasing the number of factors gives qualitatively similar results, but the IRFs are

[15]These 99 macroeconomic variables form the data matrix $\boldsymbol{x}_t$ from which we extract the factors.

[16]We estimate the autocorrelation of the error term for each equation and find no evidence for serious autocorrelation.

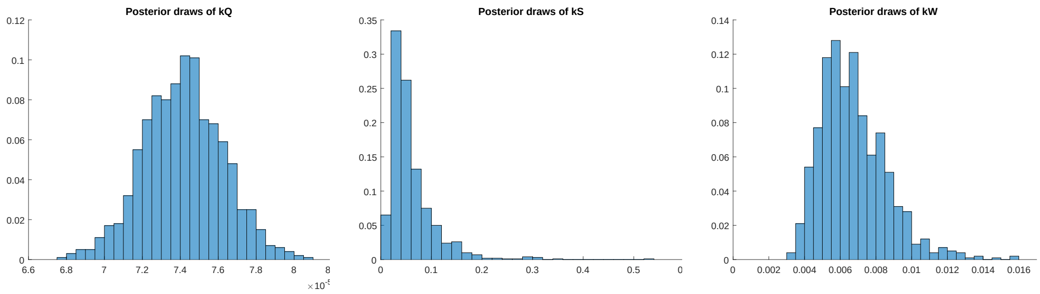


Figure 6.1: Posterior distributions of k_Q, k_S and k_W. The hyperparameter k_Q controls the amount of time variation in the autoregressive coefficients, k_S the amount of time variation in the contemporaneous covariance, and k_W the amount of time variation in the stochastic volatility.

becoming more volatile and less smooth, suggesting that more factors overfit the data.

We initiate the Gibbs sampler with a preliminary (pre) burn-in phase of 100,000 draws in order to adjust the variance of the proposal density of $k_{\boldsymbol{X}}$ in the Metropolis-within-Gibbs step.[17] The proposal variance is adjusted to achieve a target acceptance rate of 50%. After the pre burn-in, we use a burn-in of 50,000 draws with fixed proposal variance followed by 100,000 draws to approximate the posterior, where we retained every 100th draw to deal with autocorrelation in the chain. This leaves us with 1000 draws for inference. In order to judge the mixing properties of the algorithm, the autocorrelation functions of all parameters are investigated. Therefore, we calculate the inefficiency factor (IF) for all estimated coefficients, see Figure C.1 of the supplementary appendix. According to Primiceri (2005), IFs ≤ 20 are regarded as satisfactory. All IFs are below 4, indicating a well mixing chain. Stability of the model has been verified by calculating the modulus of all eigenvalues for each draw.

6.4.3 Posterior Distributions of Hyperparameters

Prior to discussing the economic effects of EPU, we focus on the posterior distributions of k_Q, k_S and k_W and the amount of time variation in the corresponding set of coefficients. Figure 6.1 shows the posterior distributions of all three parameters. The posterior either mimics a normal or an inverse gamma distribution. The domains of the posterior distributions differ strongly. While the posterior median of $k_Q = 7.4061\text{e-}05$ is very small, indicating that the autoregressive parameters do

[17] A detailed explanation of the Metropolis-within-Gibbs step as well as the whole algorithm can be found in Section B of the supplementary appendix.

not change much over time, the posterior medians of $k_S = 0.0925$ and $k_W = 0.0076$ indicate moderate time variation in the contemporaneous correlation and the stochastic volatility. Indeed, Figure 6.2 reveals that a small posterior of k_Q leads to fairly constant coefficients over time. While there is little time variation in the volatility of EPU, the volatility of R shows a sizeable amount of time variation.[18] Comparing our estimated hyperparameters with those used by Primiceri (2005) who sets $k_Q = 0.01^2$, $k_S = 0.1^2$ and $k_W = 0.01^2$, and which are often taken as given in other studies, highlights the importance of sampling them in a data-based fashion. Those values would have yielded more time variation in the autoregressive coefficients and less time variation in the contemporaneous correlation as well as in the stochastic volatility.[19] The importance of sampling the hyperparameters, instead of using the same values as Primiceri (2005), is also documented by Amir-Ahmadi et al. (2018).

It should be stressed that we start with a very flexible model, allowing for time-varying autoregressive coefficients, time-varying variances and time-varying covariances, after which the model endogenously decides which aspect is supported by the data. Thereby, our econometric design can discover which part of the model is time varying. Simply starting with a constant coefficient model implies the risk to work with misspecified regression equations and would not been able to investigate if the transmission of EPU shocks changes over time. In our case, the estimation process reduces the amount of time variation in the autoregressive coefficients, basically switching off this part of the model. On the other hand it allows for time varying volatility and covariances. This result is in line with empirical findings for US data. Sims & Zha (2006), for example, find that a VAR with constant coefficients and a time varying covariance matrix delivers the best fit for US data.

6.4.4 Transmission of EPU Shocks

Figure 6.3 and Figures 6.6 to 6.13 show the impulse response functions for all countries and all variables. Each Figure consists of twelve Subfigures. The first eleven Subfigures display the country specific response of the respective variable to a shock in EPU over the whole sample period and the 12th panel contains Bayes p-values and indicates how

[18]The decline of the median volatility of EPU since 2008 may at first come as a surprise. However, this is due to the fact that a large amount of variation in EPU is explained by other shocks in the model, in particular the increase of EPU since 2008. We will discuss this issue further in the historical decomposition of EPU in Section 6.4.6.

[19]Estimation results with the hyperparameter values choosen by Primiceri (2005) are available upon request. The major difference appears in the stochastic volatility of R which becomes time invariant. This suggests that inappropriate benchmark values erroneously suppress or increase time variation of the model parameters.

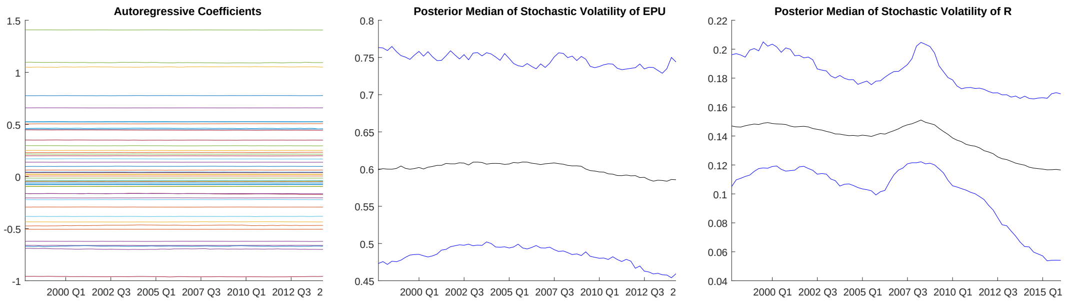


Figure 6.2: Posterior medians of the autoregressive coefficients, the stochastic volatility of EPU and the stochastic volatility of R along with 95% credible region.

credibly the response differs from zero.[20] We reduce for convenience the dimensionality of this plot by averaging over time. This is plausible since there is no credible time variation in the IRFs to an EPU shock. The IRFs are standardized such that the effect can be interpreted in the initial unit of measurement.

Figure 6.3 contains the IRFs of GDP growth. We observe a negative effect in all countries, in line with the earlier theoretical considerations. For example in Spain, an increase in the logarithm of EPU by one standard deviation decreases GDP growth in period one after the shock by 0.1%. The size as well as the development of the IRFs across all countries seem reasonable, i.e., the effect is neither too strong nor too weak and dies off after approximately ten quarters in Greece and Spain and after approximately three quarters in all other countries. We observe that the fragile countries are on average hit more strongly by an EPU shock compared to the stable countries. The distinction between the two subgroups will become more apparent in the discussion of all other IRFs.

Turning our attention to investment, depicted in Figure 6.6, a negative effect in all countries is observed, providing evidence in favor of the real options channel and highlights that business decisions are adversely affected. Comparing the effect sizes of GDP and investment reveals that investment is hit more strongly. This holds especially for the fragile countries. The IRFs of the stable and the fragile countries exhibit substantial heterogeneity. For example, Greeks' investment decreases by almost 0.4% in the period after the shock while in Germany investment decreases by only 0.1%. Furthermore, the effect dies off earlier on average in the stable countries than in the fragile countries.

The response of consumption to a shock in EPU, as summarized in Figure 6.7, is negative and therefore provides evidence on the precautionary savings channel. Thus,

[20] The Bayes p-value is calculated as the frequentists' p-value but has a Bayesian interpretation. For this reason we use the expression "credible" instead of "significant".

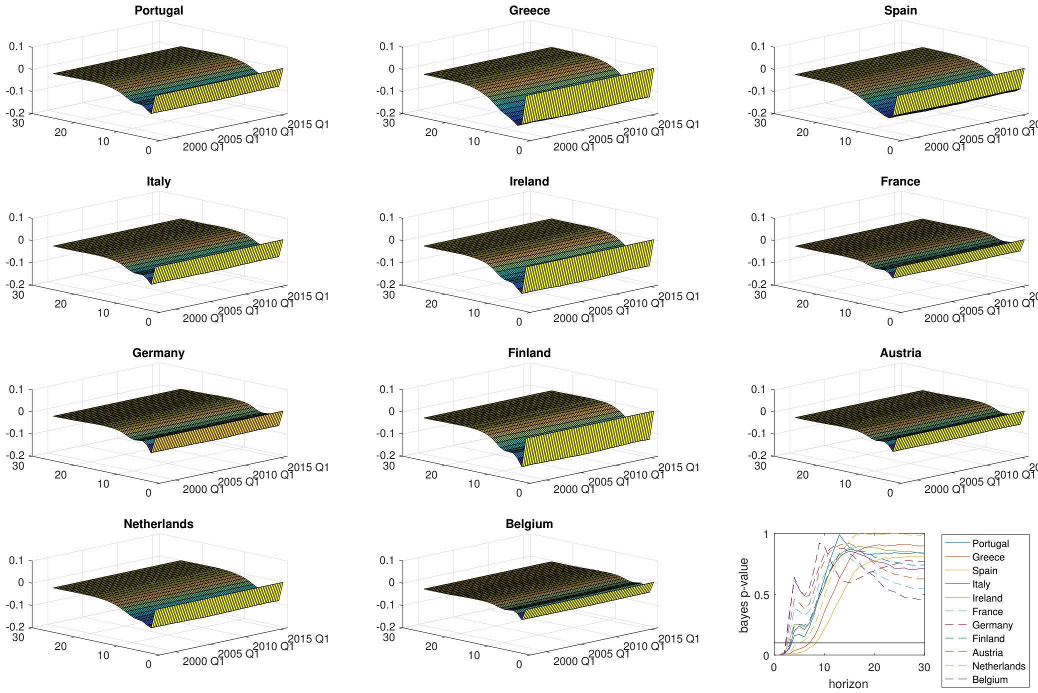


Figure 6.3: *Response of GDP growth to one standard deviation shock in EPU.* The first eleven plots display the country specific response while the twelfth plot provides information about the Bayes p-value. The Bayes p-value is defined as one minus the largest coverage region of the credible bands which does not include the value zero.

not only business decisions are affected by EPU but also the consumption decision of households. Again the response of the GIIPS countries is stronger. A comparison of the effect sizes of investment and consumption shows that in many countries, e.g., Greece or the Netherlands, investment is hit harder than consumption. This suggests that investors react more sensitively than consumers.

Inflation is negatively affected as expected. Figure 6.8 again reveals heterogeneity. There are three exceptions with a positive response, namely Germany, Finland and Austria. For Finland and Austria, the effect is not credibly different from zero while the effect of Germany is quite persistent. However, the effect size is very small in all countries suggesting that the effect of EPU on inflation is negligible.

The IRFs of unemployment, depicted in Figure 6.9, are positive for all countries, confirming that a shock in EPU postpones hiring decisions as suggested by the “wait-and-see” attitude. Similar to the previous variables, the country set can be grouped into fragile and stable countries. Even though the effect is credible, its impact is extremely small.

Figure 6.10 visualizes the responses of credit. The effect is quite strong, especially in the fragile countries. For example, in Greece the credit volume decreases in period 2 after the shock by 0.4%. The effect is smaller in the stable countries but not negligible.[21]

The IRFs of the 10 year government bond, depicted in Figure 6.11, provide evidence on first part of the financial channel and reveal a very interesting and intuitive pattern. While we observe an increase of the interest rate up to 1 percentage point (Greece) for the fragile countries, the stable countries experience a decrease in interest rates. That is, investors request higher risk premia for the fragile countries, while for the stable countries, a safe haven effect appears. This result indicates that caution is needed if one argues that uncertainty in general increases risk premia because this effect might be country group specific and is in line with the finding of Boumparis et al. (2017) who find that EPU affects creditworthiness of euro zone countries.

The effect of EPU on stock markets, the second part of the financial channel, is given in Figure 6.12. The IRFs are negative and quite large, with stock market declines up to -5%. The effect is short-lived and homogeneous. This indicates that financial market participants do not differentiate between the current state of the economy in different countries and that the degree of financial integration between the EMU countries is high.

Finally, Figure 6.13 provides an overview about the effect of EPU on consumer confidence. The pattern of the IRFs is very similar to the one of the stock market, i.e., the effect is negative, as expected, short-lived and homogeneous, but weaker. This suggests that consumers are less sensitive to uncertainty compared to stock market traders.

6.4.5 Credibility of Cross-Country Heterogeneity

The discussion of the IRFs reveals cross-country heterogeneity, i.e., the GIIPS countries respond more strongly to a shock in EPU than the northern countries. To judge whether the response of the GIIPS countries differs credibly from the response of the northern countries, we calculate the posterior distribution of the difference between the responses. The posterior distribution is calculated by subtracting the IRF of a benchmark country from the IRF of the GIIPS countries for each draw from the

[21] Whether the effect is due to financial reforms as raised by Bordo et al. (2016) or a consequence of the decrease in investment and consumption needs to be investigated in future research and is beyond the scope of this chapter.

posterior of the IRFs. Subsequently, two countries differ from each other if the resulting distribution differs credibly from zero. We choose Germany to be the benchmark country.

Figure 6.4 displays the difference of the IRFs between Greece and Germany. The IRFs differ for seven out of nine variables, namely for gross domestic (GDP) product, the GDP deflator (GDPD), credit to the non-financial private sector (CR), the 10 year government bond yield (LTI), the unemployment rate (U), consumption (C) and investment (I) at the 95% and 68% credible region. The only exceptions are the stock market index (SP) and the consumer confidence indicator (CCI). Both responses do not credibly differ from zero. A similar pattern holds for the remaining countries depicted in Section D of the supplementary appendix. The IRFs differ in at least five variables for each country in the case of the 95% credible region and in at least seven variables in the case of the 68% credible region. These results provide credible evidence for cross-country heterogeneity, i.e. the GIIPS countries respond more strongly to a shock in EPU than Germany.

6.4.6 Historical Decomposition of EPU

To assess if the newspaper-based index developed by Baker et al. (2016) measures political and economic uncertainty in our modelling context, the type of uncertainty we want to model, we perform a historical shock decomposition. The historical decomposition reveals the cumulative contribution of each structural shock to the evolution of EPU.

Figure 6.5 shows the time series and the historical decomposition. Our decomposition reveals that own shocks had the largest impact on EPU. The largest shocks occur between 2001:Q3 and 2003:Q1 and are represented by the first group of blue bars in Figure 6.5. The first blue bar (right panel) of this group is due to 9/11 which was an exogenous event from a European perspective. Furthermore, 2002 was characterized by substantial uncertainty regarding growth perspectives of the global economy. At the end of 2002, the upcoming Iraq war led to higher policy uncertainty. Summing up, the exogeneity of EPU between 2001:Q3 and 2003:Q1 is reasonable. At the beginning of the sample, monetary policy played an important role in reducing policy uncertainty (represented by the light blue bars), i.e., the decline of the ECB interest rate on the main refinancing operations, starting in 2001 and ending in 2005 during the stagnation period in Europe, reduced EPU in Europe.[22] This negative stimulus continued during the period of the rise in ECB interest rate,

[22]This period was characterized by sluggish growth of the global economy in 2003 and passed over into the Iraq war, which further affected the global economy. With the onset of the Iraq war, the ECB reduced its main interest rate by 0.5% because of expected adverse effects.

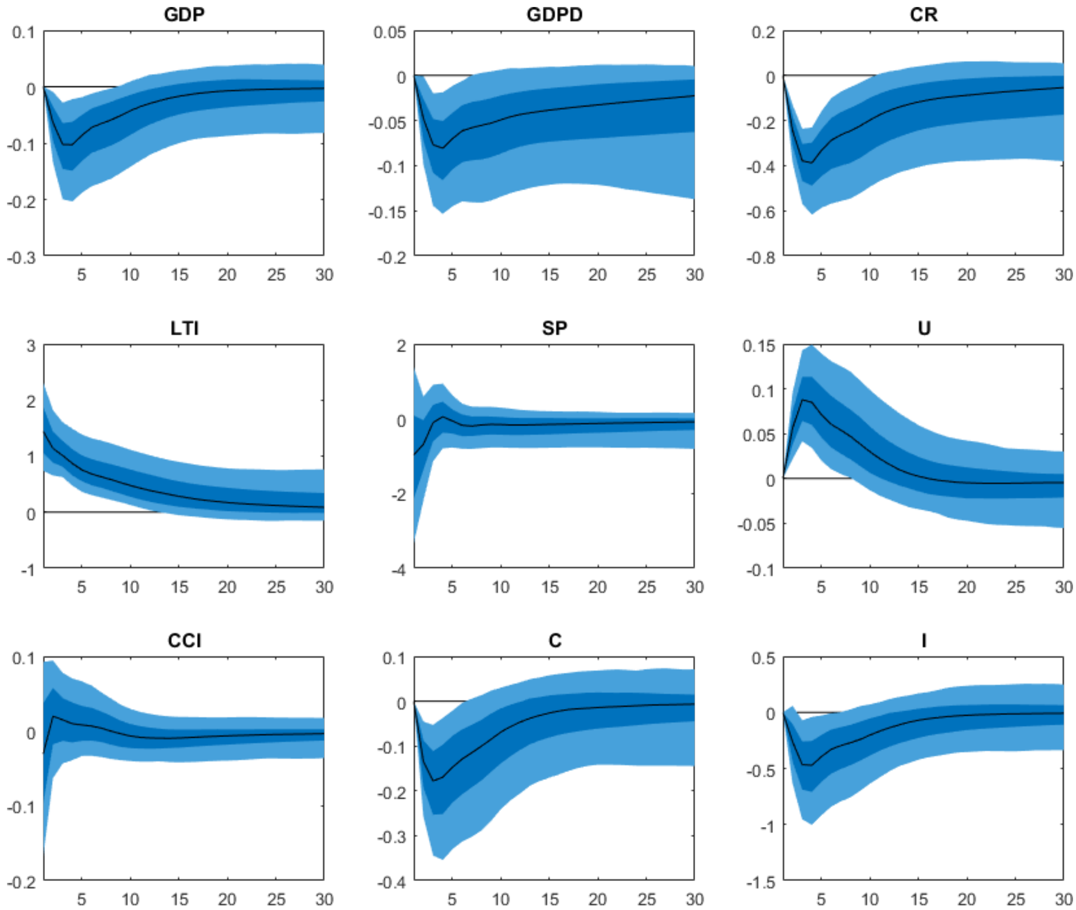


Figure 6.4: *Greece: Credibility of cross-country heterogeneity.* The Figure displays the median difference between the IRF for all variables compared to the benchmark country (Germany) along with 68% and 95% credible regions. The difference between the IRFs are interpreted in the unit of measurement.

starting in 2005:Q3. Between 2005 and 2010 own shocks had a negative impact on EPU. In that period European countries as well as the global economy returned to a growth path and EPU was historically low (see right panel of Figure 6.5). At the same time the first factor (the yellow bars in Figure 6.5), representing one of the driving forces of the economy, started to have a positive impact on EPU. This reveals that the increase of EPU, especially since the outbreak of the financial crisis, was caused by the driving forces of the economy. Plausible as the European economy was hit by the financial crisis followed by the sovereign debt crisis. Therefore, the

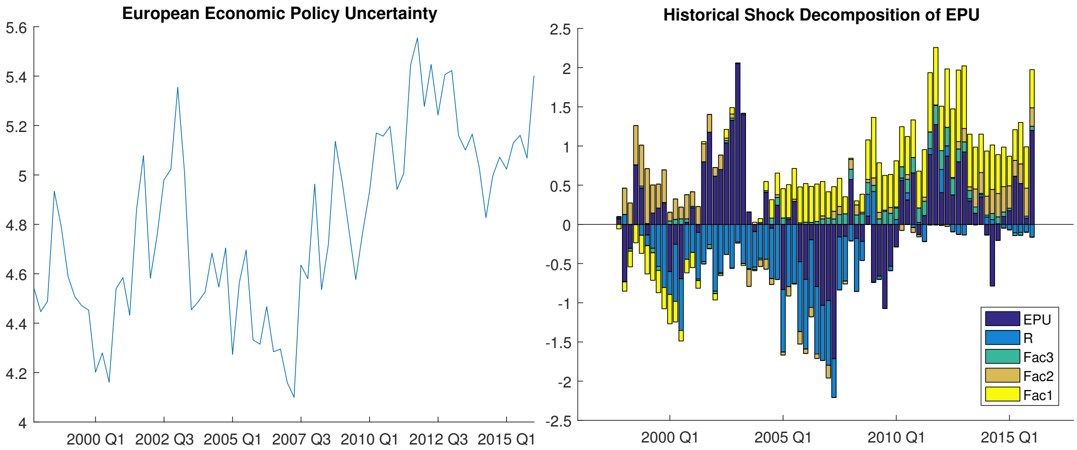


Figure 6.5: The *left* panel displays log EPU from 1997:Q1 until 2016:Q1. The *right* panel depicts the historical decomposition of EPU into the structural shocks. The historical decomposition reveals the cumulative effect of each structural shock on the evolution of the time series. The dark blue bars represent the structural shock of EPU, the light blue bars the shock of R and the remaining colors the shock of the corresponding factor.

recent high level of EPU is at least partly endogenous. This finding is in line with Benati (2013), who also focused on the endogeneity of EPU for Europe in aggregate as well as the UK, USA and Canada. He used Granger causality tests to examine the endogeneity and was able to reject the null of no Granger causality. Nevertheless, there are also some clearly exogenous shifts in EPU between 2011 and 2014. Those shifts are due to events such as the referendum in Greece in October 2011 or the debt cut for Greece in March 2012. The assignment of certain political events to large cumulative contributions of the structural shocks in EPU provides a justification for using the newspaper index by Baker et al. (2016) to measure political and economic uncertainty. And even more important, the historical decomposition shows that the index reflects political as well as economic uncertainty and therefore measures the type of uncertainty we want to model.

6.5 Conclusions

This study estimates a TVP-FAVAR model following Stock & Watson (2005) and Korobilis (2013*a*) in order to avoid implausible identification restrictions (needed in a MCMC estimation scheme) and estimates the unobserved factors using principal components. Conditional on the estimated factors we use the TVP-VAR model of Primiceri (2005) to model time variation in autoregressive coefficients and the covariance matrix. The majority of previous studies uses potentially inappropriate benchmark values for the hyperparameters, which control the amount of time variation in the coefficients and the covariance matrix. We instead estimate the hyperparameters jointly with all other model parameters using a fully Bayesian approach as proposed by Amir-Ahmadi et al. (2018). We find that the hyperparameters shrink the amount of time variation in the autoregressive coefficients, but allow for time varying volatility and covariances. This finding demonstrates the importance and benefit of estimating the hyperparameters using a fully Bayesian approach.

In order to investigate the effect of EPU on the European economies it is desirable to first consider the three theoretical channels (i.e. the real options, the precautionary savings and the financial channel) through which EPU shocks are potentially transmitted, second to allow for heterogeneous effects between different countries and third to allow for time variation in the transmission of EPU shocks. Our TVP-FAVAR allows us to address all three points. We cater the first two points by investigating the impact of EPU shocks on 100 variables, i.e., nine macroeconomic variables consisting of the gross domestic product, investment, consumption, the GDP deflator, the unemployment rate, credit to the non-financial private sector, 10 year government bond yields, a stock market index, and consumer confidence for eleven EMU countries, namely Greece, Italy, Ireland, Portugal, Spain, France, Germany, Finland, Austria, Netherlands and Belgium. The third point is considered by using the TVP-VAR model of Primiceri (2005).

We discover that EPU shocks are transmitted through all three channels and hit fragile countries (GIIPS-countries) credibly harder than more stable countries (northern countries). Furthermore, investors and financial market participants react more sensitively than consumers to uncertainty, since investment and stock prices are affected by EPU shocks more strongly than consumption and consumer confidence. While most IRFs differ only in magnitude and not in sign the response of the long term interest rates to EPU shocks has a different sign across countries. For the GIIPS countries we observe an increase of the interest rate up to 1 percentage point (Greece) and for the northern countries we observe a decrease in interest rates. That is, investors request a higher risk premium for the fragile countries, while for the stable countries a safe haven effect appears. This stresses that the effect of EPU on European countries is quite asymmetric. Lastly we find that the transmission of EPU shocks is quite stable over time.

Our findings demonstrate the economy wide effects of EPU and that it indeed might be one of the sources for the slow recovery of the European economy and in particular of the GIIPS-countries. This stresses the importance for the EU to decrease EPU in the future. During the Eurozone crisis, policy makers' response in putting together a workable plan were rather slow and the major institutions did not always agree on how to solve the crisis leading to increased EPU in the Euro area. Therefore, it is important for the EU to agree on quick and transparent ways to reform struggling countries, how to handle the high level of government debt and how to deal with single market leavers. This is especially important for a faster recovery of the GIIPS-countries.

6.A IRFs to EPU Shocks

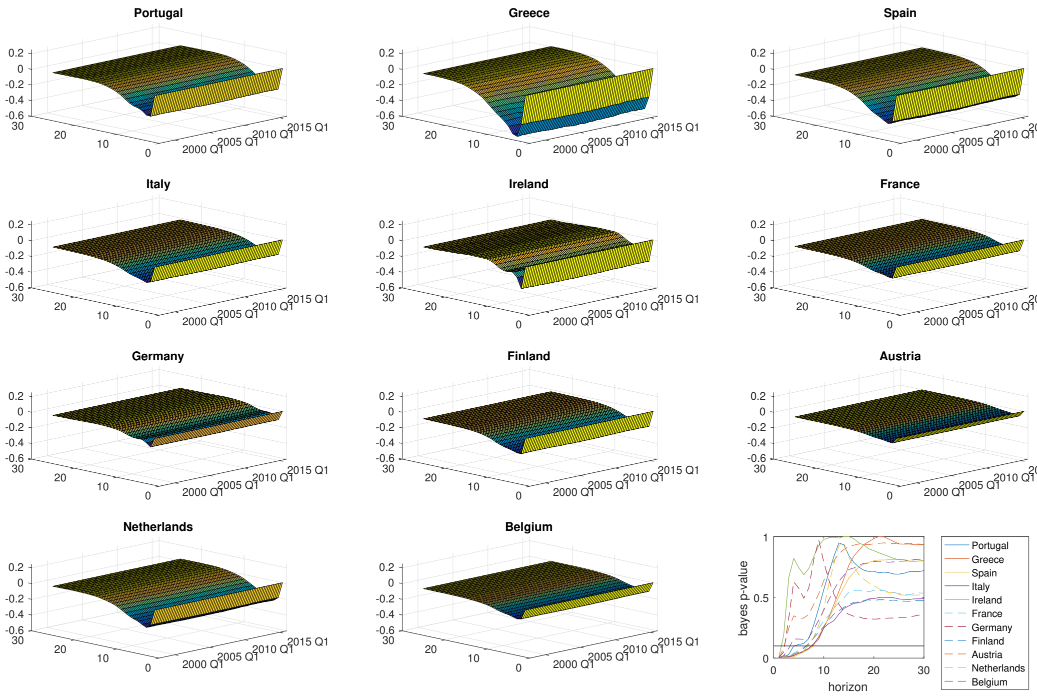


Figure 6.6: *Response of investment growth to one standard deviation shock in EPU.* For details see Figure 6.3.

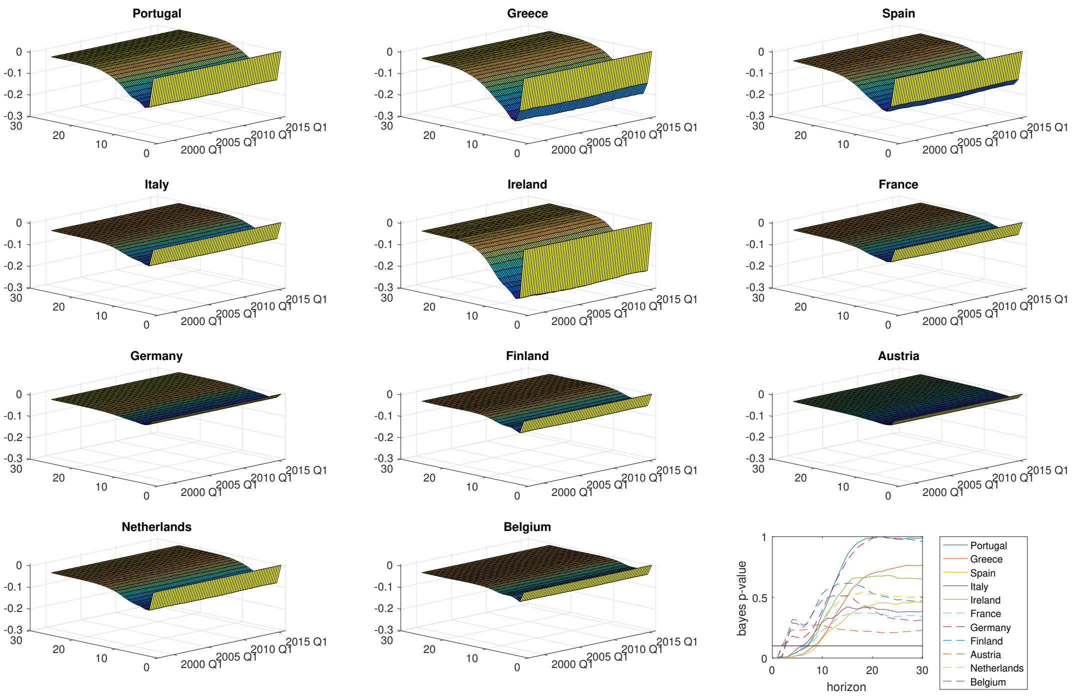


Figure 6.7: *Response of consumption growth to one standard deviation shock in EPU.* For details see Figure 6.3.

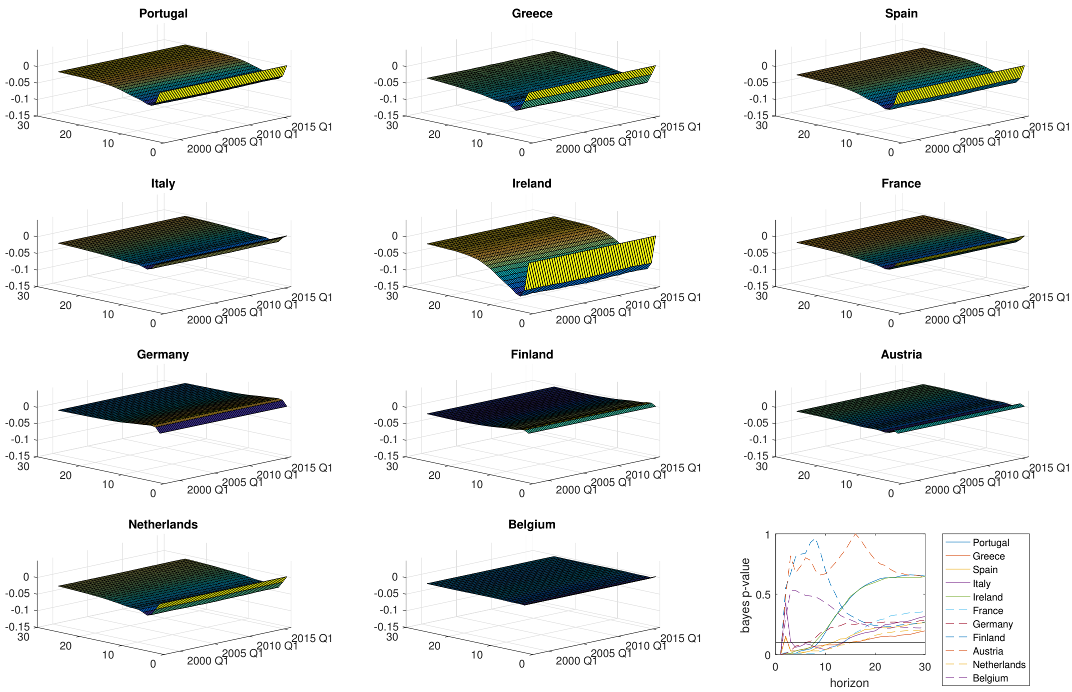


Figure 6.8: *Response of inflation to one standard deviation shock in EPU*. For details see Figure 6.3.

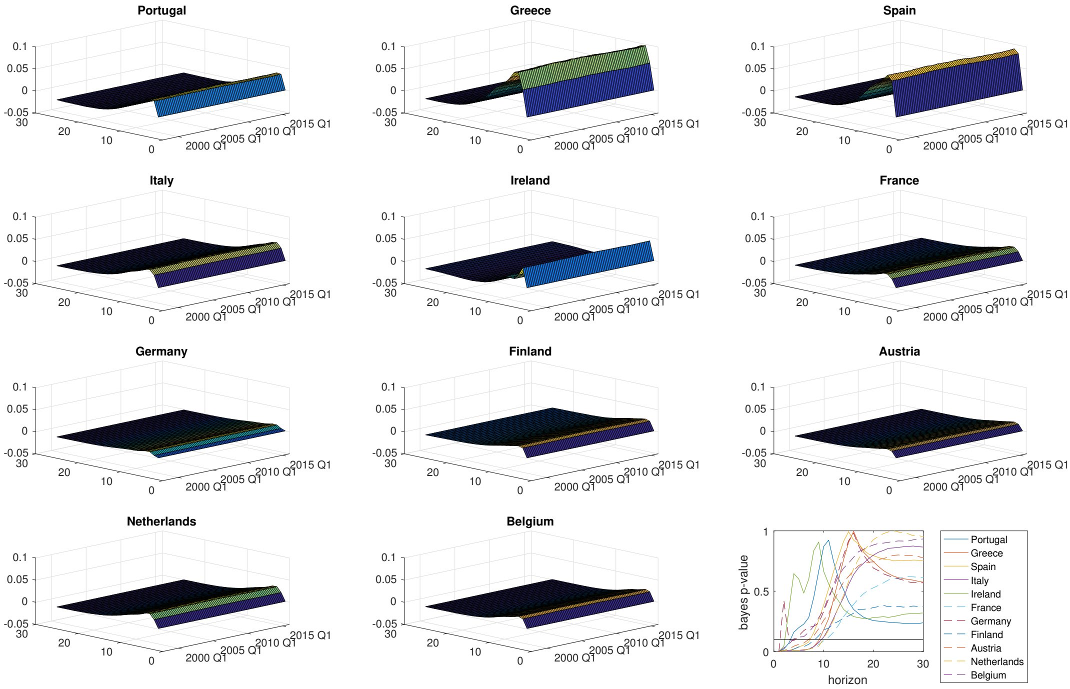


Figure 6.9: *Response of the change of the unemployment rate to one standard deviation shock in EPU.* For details see Figure 6.3.

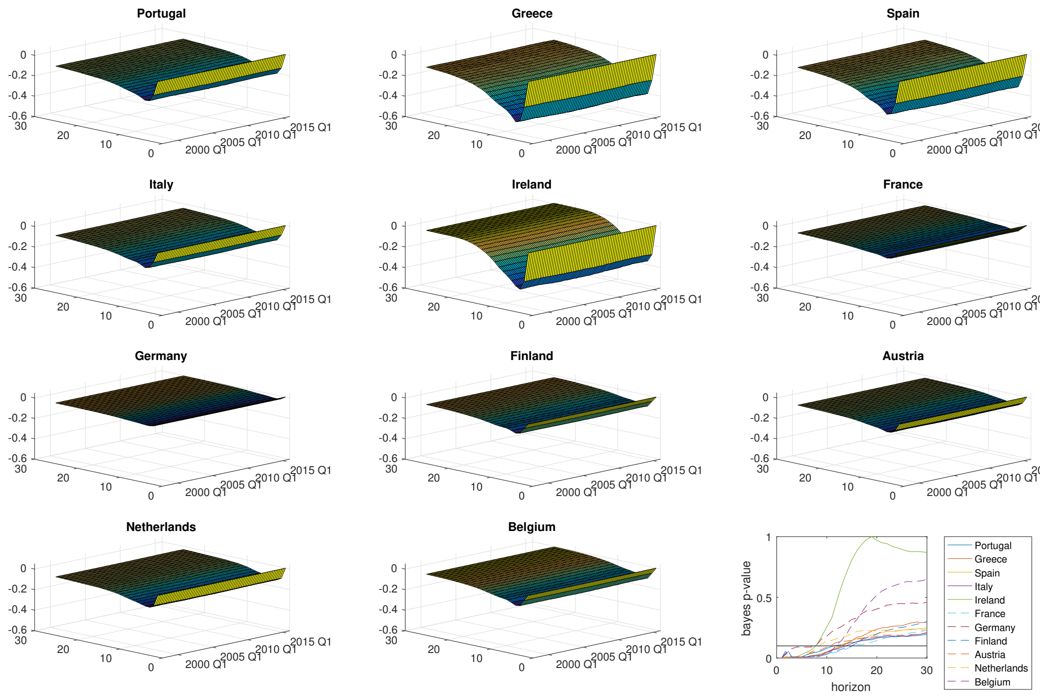


Figure 6.10: *Response of credit growth to one standard deviation shock in EPU*. For details see Figure 6.3.

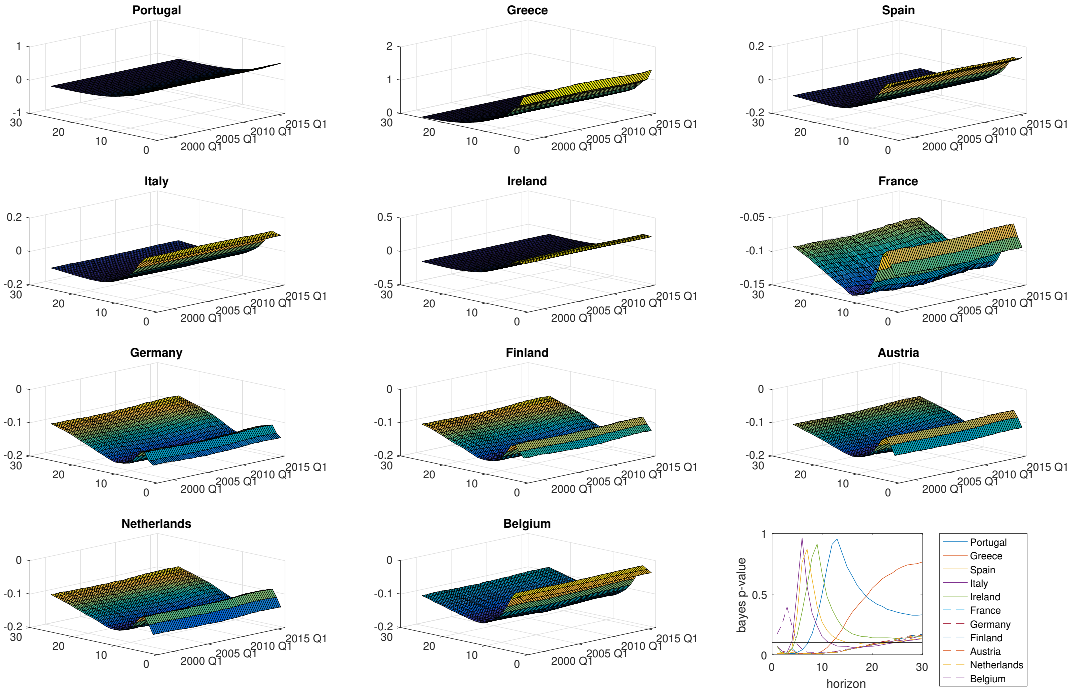


Figure 6.11: *Response of LTI to one standard deviation shock in EPU*. For details see Figure 6.3.

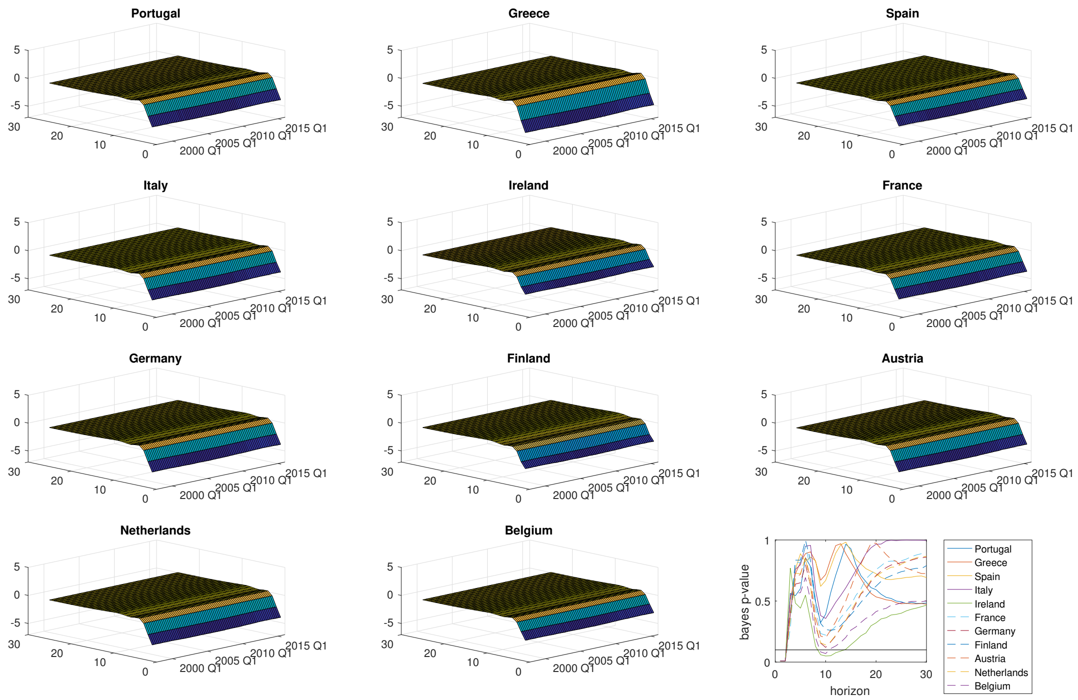


Figure 6.12: *Response of stock market return to one standard deviation shock in EPU.* For details see Figure 6.3.

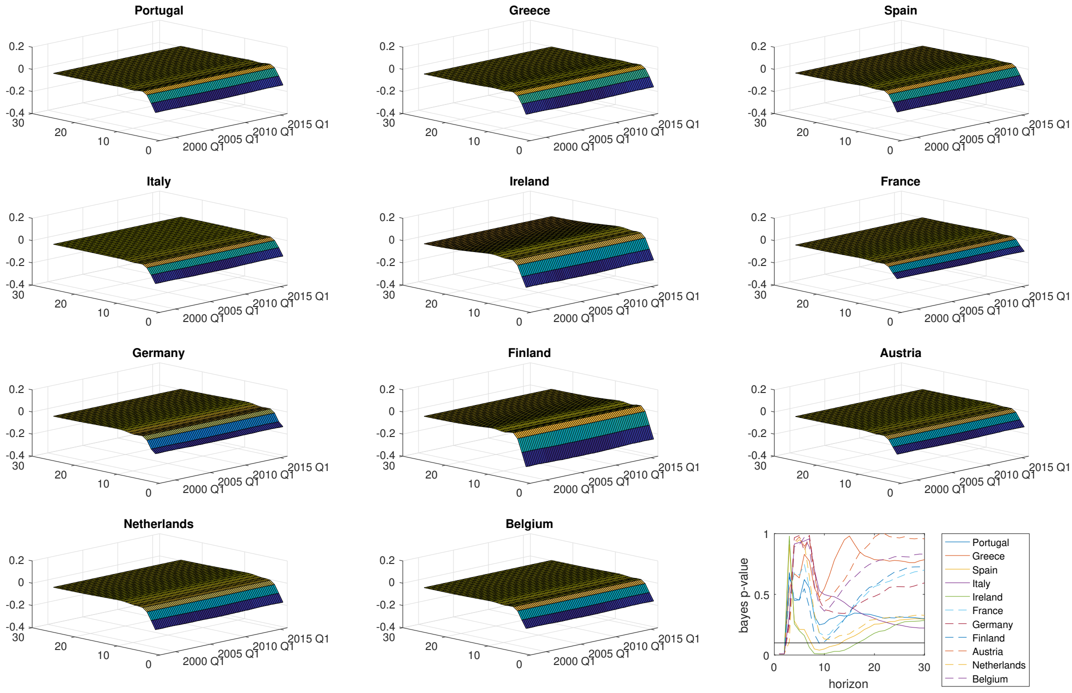


Figure 6.13: *Response of consumer confidence percentage change to an one standard deviation shock in EPU*. For details see Figure 6.3.

7 House Prices and Interest Rates - Bayesian Evidence from Germany

This chapter is joint work with Christoph Hanck and 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 & 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.

7.1 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 autoregessive 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 & 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 & Tallman, 1999).[1] We are the first, to the 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.

[1]A third strategy would be to control for overfitting by choosing the amount of shrinkage that yields a desired in-sample fit, which seems to be promising for large BVARs. See e.g. Bańbura et al. (2010).

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. As variable selection in previous VAR studies is to some extent arbitrary and leads to different results, we extend the existing literature by calculating posterior variable inclusion probabilities using a new Bayesian approach of Ding & Karlsson (2014). We find that it delivers interesting results, shedding light on which variables besides the interest rate may be important for explaining German house prices. In our structural analysis, we employ two different identification strategies: a recursive identification scheme and sign restrictions. Both strategies deliver similar results. Finally, 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 identification of the system.

The remainder of this chapter 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.

7.2 Methodology

7.2.1 Bayesian Vector Autoregression

It has recently become popular to use prior information in VARs to overcome the problem of overfitting (e.g., Koop & Korobilis, 2010). In comparison to theory-based Dynamic Stochastic General Equilibrium models, 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

$$\begin{aligned} \boldsymbol{y}_t' &= \sum_{i=1}^{p} \boldsymbol{y}_{t-i}' \boldsymbol{A}_i + \boldsymbol{x}_t' \boldsymbol{C} + \boldsymbol{u}_t' \\ &= \boldsymbol{z}_t' \boldsymbol{\Gamma} + \boldsymbol{u}_t', \end{aligned} \tag{7.1}$$

with $\boldsymbol{y}_t$ a $m \times 1$ vector containing the variables at time t, $\boldsymbol{A}_i$ is a $m \times m$ matrix of parameters, $\boldsymbol{x}_t$ a vector of d deterministic variables at time t, $\boldsymbol{C}$ is a $d \times m$ parameter matrix, $\boldsymbol{z}'_t = (\boldsymbol{y}'_{t-1}, ..., \boldsymbol{y}'_{t-p}, \boldsymbol{x}'_t)$ a $1 \times k$ vector with $k = mp + d$, $\boldsymbol{\Gamma} = (\boldsymbol{A}'_1, ..., \boldsymbol{A}'_p, \boldsymbol{C}')'$ a $k \times m$ matrix and $\boldsymbol{u}_t \sim N(\boldsymbol{0}, \boldsymbol{\Psi})$ 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 (7.1) 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)

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\Gamma} + \boldsymbol{U}, \tag{7.2}$$

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

$$f(\boldsymbol{Y}|\boldsymbol{\Gamma}, \boldsymbol{\Psi}) = 2\pi^{-mT/2}|\boldsymbol{\Psi}|^{-T/2} \exp\left\{-\frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Psi}^{-1}(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\Gamma})'(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\Gamma})\right]\right\}. \tag{7.3}$$

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

The Normal-Wishart Prior

The Normal-Wishart prior is a natural conjugate prior for normal multivariate regressions, allowing for computationally convenient system estimation. Let us rewrite the likelihood function in the form of a Normal-Wishart distribution. Define the OLS estimate $\widehat{\boldsymbol{\Gamma}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}$ and $\boldsymbol{S} = (\boldsymbol{Y} - \boldsymbol{Z}\widehat{\boldsymbol{\Gamma}})'(\boldsymbol{Y} - \boldsymbol{Z}\widehat{\boldsymbol{\Gamma}})$. Adding and subtracting $\boldsymbol{Z}\widehat{\boldsymbol{\Gamma}}$ in the exponential in (7.3) yields

$$
\begin{aligned}
f(\boldsymbol{Y}|\boldsymbol{\Gamma},\boldsymbol{\Psi}) \propto |\boldsymbol{\Psi}|^{-T/2} \exp\left\{-\frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Psi}^{-1}\boldsymbol{S}\right]\right\} \\
\times \exp\left\{-\frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Psi}^{-1}(\boldsymbol{\Gamma}-\widehat{\boldsymbol{\Gamma}})'\boldsymbol{Z}'\boldsymbol{Z}(\boldsymbol{\Gamma}-\widehat{\boldsymbol{\Gamma}})\right]\right\}.
\end{aligned}
\tag{7.4}
$$

The prior for $\boldsymbol{\Gamma}$ and $\boldsymbol{\Psi}$ is specified as a conditional matricvariate normal distribution[2] (MN), where $\boldsymbol{X} \sim MN(\boldsymbol{M},\boldsymbol{Q},\boldsymbol{P})$ iff $\operatorname{vec}(\boldsymbol{X}) \sim N(\operatorname{vec}(\boldsymbol{M}),\boldsymbol{Q}\otimes\boldsymbol{P})$ and an inverse Wishart distribution $iW(\boldsymbol{\Sigma},v)$, where $\boldsymbol{\Sigma}$ is the scale matrix and v denotes the degrees of freedom,

$$
\boldsymbol{\Gamma}|\boldsymbol{\Psi} \sim MN(\underline{\boldsymbol{\Gamma}},\boldsymbol{\Psi},\underline{\boldsymbol{\Omega}}), \tag{7.5}
$$

$$
\boldsymbol{\Psi} \sim iW(\underline{\boldsymbol{S}},\underline{v}). \tag{7.6}
$$

The posterior using the Normal-Wishart prior becomes

$$
\begin{aligned}
f(\boldsymbol{\Gamma},\boldsymbol{\Psi}|\boldsymbol{Y}) \propto |\boldsymbol{\Psi}|^{(\overline{v}+m+k+1)/2} \\
\times \exp\left\{-\frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Psi}^{-1}(\boldsymbol{\Gamma}-\overline{\boldsymbol{\Gamma}})'\overline{\boldsymbol{\Omega}}^{-1}(\boldsymbol{\Gamma}-\overline{\boldsymbol{\Gamma}})\right]\right\} \\
\times \exp\left\{-\frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Psi}^{-1}\overline{\boldsymbol{S}}\right]\right\},
\end{aligned}
\tag{7.7}
$$

where $\overline{\boldsymbol{\Omega}} = (\underline{\boldsymbol{\Omega}}^{-1} + \boldsymbol{Z}'\boldsymbol{Z})^{-1}$, $\overline{\boldsymbol{\Gamma}} = \overline{\boldsymbol{\Omega}}(\underline{\boldsymbol{\Omega}}^{-1}\underline{\boldsymbol{\Gamma}} + \boldsymbol{Z}'\boldsymbol{Z}\widehat{\boldsymbol{\Gamma}})$, $\overline{v} = T + \underline{v}$ and $\overline{\boldsymbol{S}} = \underline{\boldsymbol{S}} + \boldsymbol{S} + (\underline{\boldsymbol{\Gamma}} - \widehat{\boldsymbol{\Gamma}})'(\underline{\boldsymbol{\Omega}} + (\boldsymbol{Z}'\boldsymbol{Z})^{-1})^{-1}(\underline{\boldsymbol{\Gamma}} - \widehat{\boldsymbol{\Gamma}})$. By definition of the Normal-Wishart distribution the conditional and marginal posterior distributions are

$$
\boldsymbol{\Gamma}|\boldsymbol{Y},\boldsymbol{\Psi} \sim MN(\overline{\boldsymbol{\Gamma}},\boldsymbol{\Psi},\overline{\boldsymbol{\Omega}}) \tag{7.8}
$$

and

$$
\boldsymbol{\Psi}|\boldsymbol{Y} \sim iW(\overline{\boldsymbol{S}},\overline{v}). \tag{7.9}
$$

This shows that the mean of the conditional posterior distribution $\overline{\boldsymbol{\Gamma}}$ is a weighted average of $\underline{\boldsymbol{\Gamma}}$ and $\widehat{\boldsymbol{\Gamma}}$. The weights are given by the inverse of the variance-covariance

[2] An excellent summary of matricvariate distributions can be found in the appendix of Karlsson (2013).

matrix of the prior, $\underline{\mathbf{\Omega}}^{-1}$, and by $\boldsymbol{Z}'\boldsymbol{Z}$. Overall, smaller values of $\underline{\mathbf{\Omega}}$ imply that the prior receives a higher weight relative to $\widehat{\mathbf{\Gamma}}$. Hence, if the posterior mean is not to be strongly influenced by the prior, one assigns a high variance to the prior.

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

$$\underline{\mathbf{\Gamma}}_{ij} = \begin{cases} 1, & \text{first own lag, } i = j \\ 0, & i \neq j \end{cases} . \tag{7.10}$$

The specification of the prior variances for the Normal-Wishart prior is a bit more tricky. Integrating $\mathbf{\Psi}$ out of $MN(\underline{\mathbf{\Gamma}}, \mathbf{\Psi}, \underline{\mathbf{\Omega}})$ shows that the marginal prior distribution of $\mathbf{\Gamma}$ is matricvariate t with $km \times km$ variance-covariance matrix $V(\gamma) = \frac{1}{\underline{v}-m-1}\underline{\boldsymbol{S}} \otimes \underline{\mathbf{\Omega}}$, where $\gamma = \text{vec}(\mathbf{\Gamma})$. 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 standard BVAR prior of Sims & Zha (1998). With $V(\gamma_j) = \frac{\tilde{s}_j}{\underline{v}-m-1}\underline{\mathbf{\Omega}}$, where γ_j, with $j = 1, ..., m$, is a $k \times 1$ vector such that $\gamma = (\gamma_1', ..., \gamma_m')'$, the diagonal elements of $\underline{\mathbf{\Omega}}$ are set to

$$\omega_{ii} = \begin{cases} (\lambda_0\lambda_1)^2/(l^{\lambda_3}s_r)^2, & \text{for lag } l \text{ of variable } r,\ i = (l-1)m + r \\ (\lambda_0\lambda_4)^2, & \text{for the constant, } i = mp + 1 \\ (\lambda_0\lambda_5)^2, & \text{for the deterministic variables, } i = mp + 2, ..., k \end{cases} \tag{7.11}$$

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

$$\underline{\boldsymbol{S}} = (\underline{v} - m - 1)\,\text{diag}\left(\left(\frac{s_1}{\lambda_0}\right)^2, ..., \left(\frac{s_m}{\lambda_0}\right)^2\right). \tag{7.12}$$

[3]Note 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 & Karlsson (1997), the normal-Wishart prior.

To ensure that the prior variance exists $\underline{v}$ is set to $m + 2$ (Kadiyala & 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 & Nicolo 2002).

In addition to the Minnesota prior, the Bayesian literature uses other priors to introduce beliefs about unit roots and 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 & 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., $\sum_i \boldsymbol{A}_i = \boldsymbol{I}_m$, or similarly the belief that the recent average of lagged values of a variable $\overline{y}_{0i} = \frac{1}{p}\sum_{n=1-p}^{0} y_{n,i}$ 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 $\boldsymbol{Y}$ and $\boldsymbol{Z}$. Let $y_{i,j}$ with $i, j = 1, ..., m$ represent the elements of the first m rows of $\boldsymbol{Y}$ and $z_{i,s}$ with $i = 1, ..., m$; $s = 1, ..., k$ represent the elements of the first m rows of $\boldsymbol{Z}$. Then

$$y_{i,j} = \begin{cases} \mu_5 \overline{y}_{0i}, & i = j \\ 0, & \text{otherwise} \end{cases} \tag{7.13}$$

and

$$z_{i,s} = \begin{cases} \mu_5 \overline{y}_{0i}, & i = j,\ s \leq mp \\ 0, & \text{otherwise} \end{cases}. \tag{7.14}$$

When $\mu_5 \to \infty$ 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 $i = 1, ..., m$ represent the elements of row $m+1$ of $\boldsymbol{Y}$ and $z_{s,m+1}$ with $s = 1, ..., k$ represent the elements of row $m+1$ of $\boldsymbol{Z}$. The additional row can then be written as

$$y_{i,m+1} = \mu_6 \overline{y}_{0i} \tag{7.15}$$

and

$$z_{s,m+1} = \begin{cases} \mu_6 \overline{y}_{0i}, & s \leq mp \\ \mu_6, & s = mp+1 \\ 0, & \text{otherwise} \end{cases}, \tag{7.16}$$

with

$$i = \begin{cases} s, & s \leq m \\ s-m, & m < s \leq 2m \\ \vdots & \\ s-m(p-1), & m(p-1) < s \leq mp \end{cases}. \tag{7.17}$$

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 & Zha 1998).

7.2.2 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 & Tallman (1999). Examples for this approach include Summers (2001), Brandt & Freeman (2002), Wright (2009) and Litterman (1986*b*). 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

$$\ln m(\boldsymbol{Y}|\Xi, M_i) = \sum_{t=1}^{T} \ln m(\boldsymbol{y}_t|\{\boldsymbol{y}_s\}_{s=0}^{t-1}, \Xi, M_i), \tag{7.18}$$

with $m(\boldsymbol{Y}|\Xi, M_i) = \int f(\boldsymbol{Y}|\boldsymbol{\theta}_i, \Xi, M_i)\pi(\boldsymbol{\theta}_i)\,\mathrm{d}\,\boldsymbol{\theta}_i$ 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 $\boldsymbol{y_t}$ 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 M competing models, the posterior probabilities for each model are

$$\pi(M_i|\Xi, \boldsymbol{Y}) = \frac{m(\boldsymbol{Y}|\Xi, M_i)\pi(M_i)}{\sum_{i=1}^{M} m(\boldsymbol{Y}|\Xi, M_i)\pi(M_i)}. \tag{7.19}$$

The posterior odds for two competing models M_1 and M_2 are

$$\frac{\pi(M_1|\Xi, \boldsymbol{Y})}{\pi(M_2|\Xi, \boldsymbol{Y})} = \frac{m(\boldsymbol{Y}|\Xi, M_1)}{m(\boldsymbol{Y}|\Xi, M_2)} \times \frac{\pi(M_1)}{\pi(M_2)}. \tag{7.20}$$

The model choice can be based on the posterior odds, but it is more common to use the Bayes factors $BF_{1,2} = m(\boldsymbol{Y}|\Xi, M_1)/m(\boldsymbol{Y}|\Xi, M_2)$ 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 $Mt(\boldsymbol{M}, \boldsymbol{P}, \boldsymbol{Q}, v)$, where $\boldsymbol{M}$ is the mean matrix, $\boldsymbol{Q}$ and $\boldsymbol{P}$ are positive definite symmetric scale matrices and v is the degree of freedom,

$$\boldsymbol{Y} \sim Mt(\boldsymbol{Z}\underline{\boldsymbol{\Gamma}}, (\boldsymbol{I}_T + \boldsymbol{Z}\underline{\boldsymbol{\Omega}}\boldsymbol{Z}')^{-1}, \underline{\boldsymbol{S}}, \underline{v}). \tag{7.21}$$

[4] This is seen critically by pure Bayesians, as Carlin & 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).

As shown by Giannone et al. (2015), (7.21) 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 & Freeman (2009) and Deryugiana & 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 & 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 & 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. $\boldsymbol{P} = (\boldsymbol{I}_q, \boldsymbol{0}_{q\times(m-q)})'$ be a $m \times q$ selection matrix such that $\boldsymbol{Y}_1 = \boldsymbol{Y}\boldsymbol{P}$, the matrix of the variables of interest. Then, for $\underline{\boldsymbol{\Gamma}}_1 = \underline{\boldsymbol{\Gamma}}\boldsymbol{P}$ and $\underline{\boldsymbol{S}}_1 = \boldsymbol{P}'\underline{\boldsymbol{S}}\boldsymbol{P}$ the marginalized marginal likelihood is

$$\boldsymbol{Y}_1 \sim Mt(\boldsymbol{Z}\underline{\boldsymbol{\Gamma}}_1, (\boldsymbol{I}_T + \boldsymbol{Z}\underline{\boldsymbol{\Omega}}\boldsymbol{Z}')^{-1}, \underline{\boldsymbol{S}}_1, \underline{v} - m + q). \tag{7.22}$$

7.2.3 Conditional Forecasts

Conditional forecasts are a useful tool for policy analysis. Conveniently, such forecasts are invariant to the identification of the system, 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 & Zha (1999) provide a framework for calculating the conditional forecast distribution using a Gibbs sampling algorithm. To illustrate the idea, write (7.1) at time T (with

[5] Somewhat simplified relative to Giannone et al. (2015), we do not simulate the posterior mode of the distribution of the hyperparameters, but numerically maximize (7.21). This substantially reduces the computational burden, see footnote 7.

a constant $\boldsymbol{c}$ as the only deterministic variable for simplicity) as

$$\boldsymbol{y}_T' = \sum_{i=1}^{p} \boldsymbol{y}_{T-i}'\boldsymbol{A}_i + \boldsymbol{c} + \boldsymbol{\epsilon}_T'\boldsymbol{A}_0^{-1}, \tag{7.23}$$

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

$$\boldsymbol{y}_{T+h}' = \boldsymbol{c}\boldsymbol{K}_{h-1} + \sum_{i=1}^{p} \boldsymbol{y}_{T+1-i}'\boldsymbol{N}_i(h) + \sum_{j=1}^{h} \boldsymbol{\epsilon}_{T+j}'\boldsymbol{M}_{h-j}, \tag{7.24}$$

where

$$\boldsymbol{K}_0 = \boldsymbol{I}, \tag{7.25}$$

$$\boldsymbol{K}_l = \boldsymbol{I} + \sum_{j=1}^{l} \boldsymbol{K}_{l-j}\boldsymbol{A}_j, \quad l = 1, 2, ...; \tag{7.26}$$

$$\boldsymbol{N}_i(1) = \boldsymbol{A}_i, \quad i = 1, ..., p; \tag{7.27}$$

$$\boldsymbol{N}_i(h) = \sum_{j=1}^{h-1} \boldsymbol{N}_i(h-j)\boldsymbol{A}_j + \boldsymbol{A}_{h+i-1}, \quad i = 1, 2, ..., p, \quad h = 2, 3, ..; \tag{7.28}$$

$$\boldsymbol{M}_0 = \boldsymbol{A}_0^{-1}, \tag{7.29}$$

$$\boldsymbol{M}_l = \sum_{j=1}^{l} \boldsymbol{M}_{l-j}\boldsymbol{A}_j, \quad l = 1, 2, ...; \tag{7.30}$$

with the convention that $\boldsymbol{A}_j = \boldsymbol{0}$ for $j > p$. The future values $\boldsymbol{y}_{T+h}'$ in (7.24) depend on a systematic part and structural shocks. By writing (7.24) as

$$\boldsymbol{y}_{T+h}' - \boldsymbol{c}\boldsymbol{K}_{h-1} - \sum_{i=1}^{p} \boldsymbol{y}_{T+1-i}'\boldsymbol{N}_i(h) = \sum_{j=1}^{h} \boldsymbol{\epsilon}_{T+j}'\boldsymbol{M}_{h-j}, \tag{7.31}$$

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

$$\boldsymbol{r} = \boldsymbol{R}\boldsymbol{\epsilon}, \tag{7.32}$$

where $\boldsymbol{r}$ 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 $\boldsymbol{R}$ is a $q \times g$ matrix containing the impulse response of the constrained variables to the structural shock $\boldsymbol{\epsilon}_t$ at horizon $1, 2, ..., h$, 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 (7.32) 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

$$\widehat{\boldsymbol{\epsilon}} = \boldsymbol{R}'(\boldsymbol{R}\boldsymbol{R}')^{-1}\boldsymbol{r}. \tag{7.33}$$

The conditional forecasts are then calculated by inserting $\widehat{\boldsymbol{\epsilon}}$ into (7.24). The Gibbs sampling algorithm of Waggoner & 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.

7.3 Simulation Study

7.3.1 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)

$$\boldsymbol{y}_t' = \boldsymbol{y}_{t-1}' \begin{pmatrix} 1 & 0.015 & -0.006 \\ 0.005 & 1 & 0.036 \\ -0.044 & -0.05 & 1 \end{pmatrix} + \boldsymbol{y}_{t-2}' \begin{pmatrix} -0.04 & -0.005 & -0.0261 \\ -0.035 & -0.061 & -0.011 \\ 0.028 & -0.0074 & 0.005 \end{pmatrix} + \boldsymbol{u}_t' \tag{7.34}$$

and

$$
\begin{aligned}
\boldsymbol{y}_t' = \boldsymbol{y}_{t-1}' &\begin{pmatrix} 1 & 0.015 & -0.006 \\ 0.005 & 1 & 0.036 \\ -0.044 & -0.05 & 1 \end{pmatrix} + \boldsymbol{y}_{t-2}' \begin{pmatrix} -0.04 & -0.005 & -0.0261 \\ -0.035 & -0.061 & -0.011 \\ 0.028 & -0.0074 & 0.005 \end{pmatrix} \\
+ \boldsymbol{y}_{t-3}' &\begin{pmatrix} 0.001 & 0.062 & -0.004 \\ 0.01 & -0.029 & -0.004 \\ -0.01 & -0.003 & 0.03 \end{pmatrix} + \boldsymbol{y}_{t-4}' \begin{pmatrix} -0.001 & -0.062 & 0.0274 \\ 0.009 & 0.006 & 0.0311 \\ 0.46 & 0.022 & -0.01 \end{pmatrix} \\
+ \boldsymbol{y}_{t-5}' &\begin{pmatrix} 0.001 & -0.0002 & 0.0004 \\ 0.002 & -0.0002 & -0.00031 \\ 0 & 0.0005 & -0.0001 \end{pmatrix} + \boldsymbol{y}_{t-6}' \begin{pmatrix} -0.0009 & 0.0005 & -0.0004 \\ -0.0004 & 0.0001 & 0.00021 \\ 0 & -0.0001 & 0.0003 \end{pmatrix} \\
+ \boldsymbol{u}_t', &
\end{aligned}
\tag{7.35}
$$

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

Figures 7.1 and 7.2 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 for example the case for German house prices (see Figures 7.7 or 7.10). 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 more volatile than inflation.

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 functions[6] 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

[6] The impulse response functions from the BVAR are estimated by taking the median of 1000 impulse response functions drawn. The Cholesky decomposition is used, being one of the approaches used in the empirical application below.

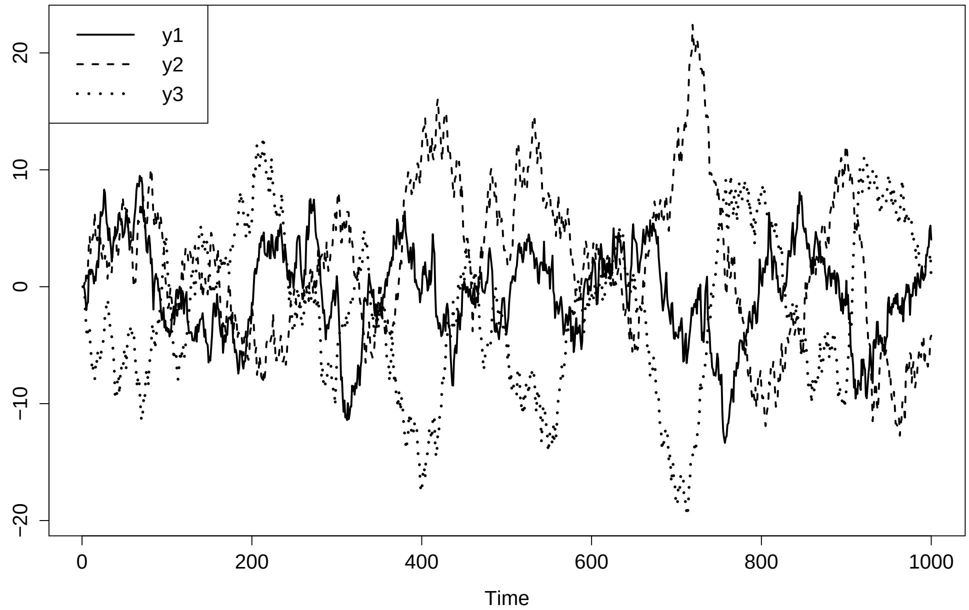


Figure 7.1: DGPa

$$\overline{AMSE} = \frac{1}{m}\sum_{i=1}^{m}\frac{MSE_i}{MSE_i(RW)}, \tag{7.36}$$

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

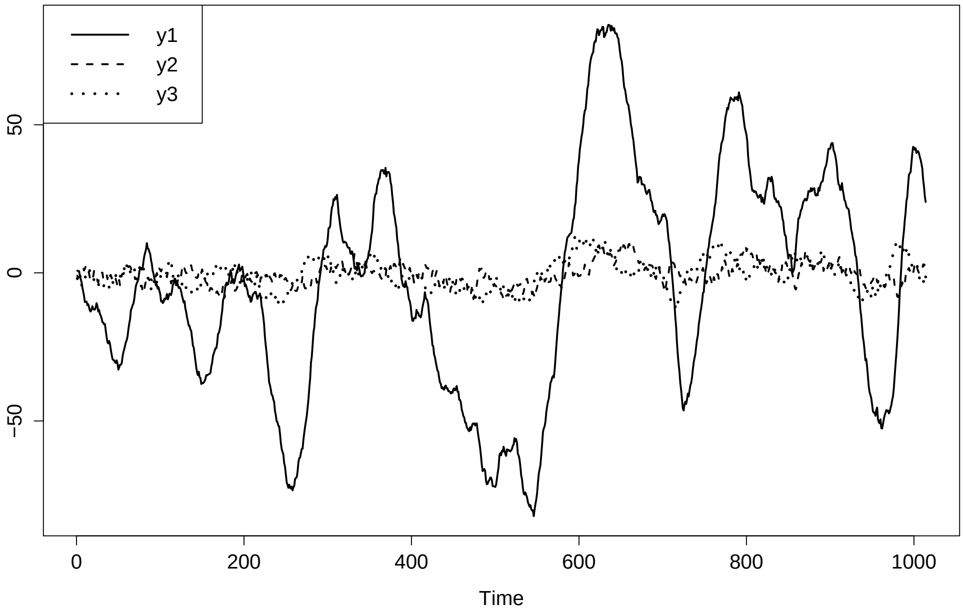


Figure 7.2: DGPb

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 7.1 provides a summary of the different setups.[7] 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

[7] A single setup with 1000 replications is already fairly computationally demanding. For the simulations, we used the HPC cluster Peregrine at the Rijksuniversiteit Groningen.

Table 7.1: 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.

μ_6, as a researcher only observing a small stretch of the data might do the same. We do not need to choose values for λ_5, as no deterministic variables are included in the model here.

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 setups 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 setups 3 and 4.

7.3.2 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 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

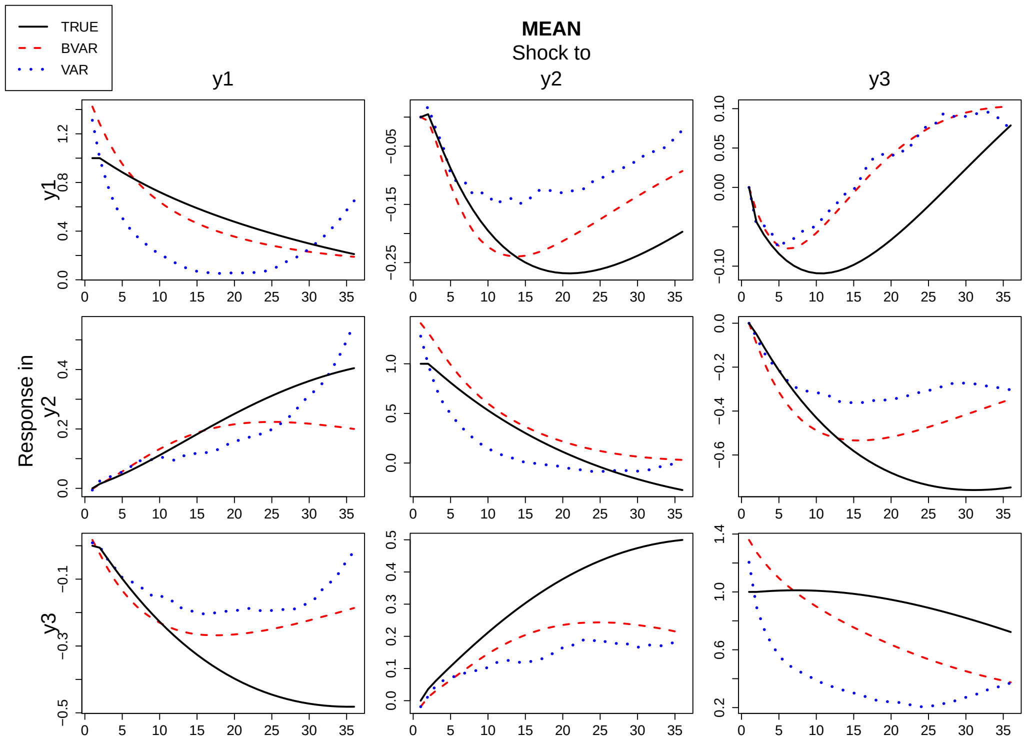


Figure 7.3: Comparison between the mean of the estimated impulse responses of the BVAR and VAR with the true impulse responses for DGPa and the fourth setup. The choice of hyperparameters is based on the marginal likelihood.

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 7.3 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 7.4 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

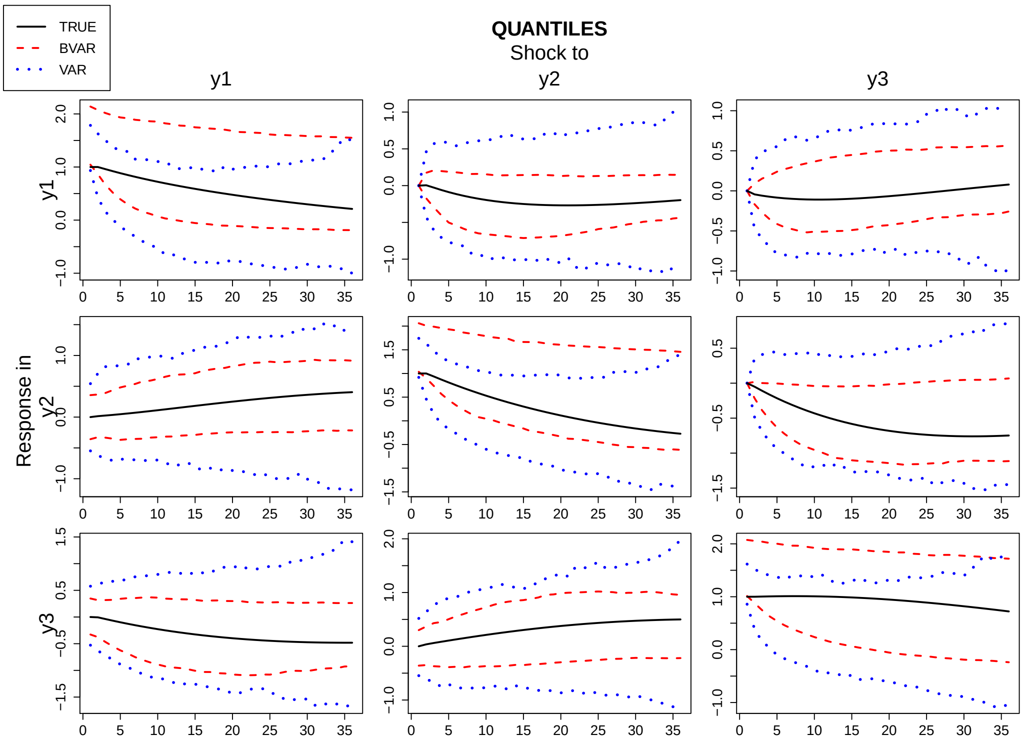


Figure 7.4: Comparison between the 0.05 and 0.95 quantiles of the estimated impulse responses of the BVAR and VAR with the true impulse responses for DGPa and the fourth setup. The choice of hyperparameters is based on the marginal likelihood. Note that the true impulse responses here are the same as in Figure 7.3, but the scale of the y-axis is different.

than the VAR against overspecifying the lag length or/and adding irrelevant variables.

To investigate the performance of the four different approaches we compare the MSE ratios. Figure 7.5 and Figures 7.13-7.16 report the MSE ratio for each approach and all simulations of DGPa for each impulse response function. The results for DGPb 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

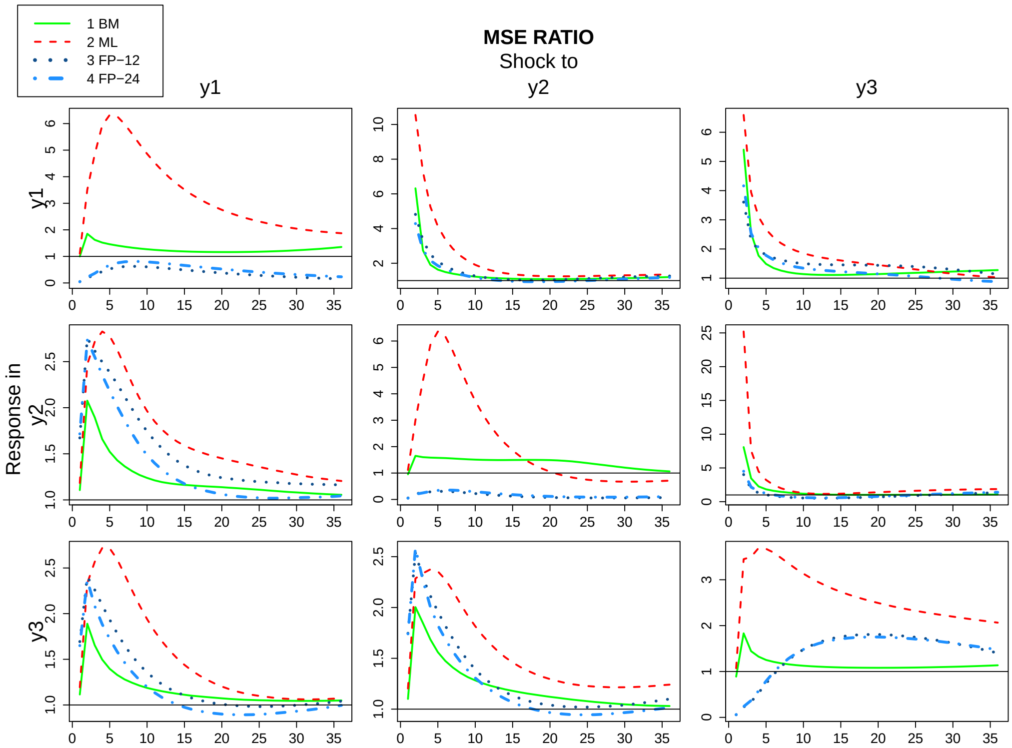


Figure 7.5: Comparison between the four different approaches for DGPa and setup 1. The hyperparameter choice in the first approach is based on the benchmark (BM), the second is based on the marginal likelihood (ML), the third is based on the forecast performance (FP) of 1-to-12 period ahead forecasts 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.

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 result is consistent with Ding & 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

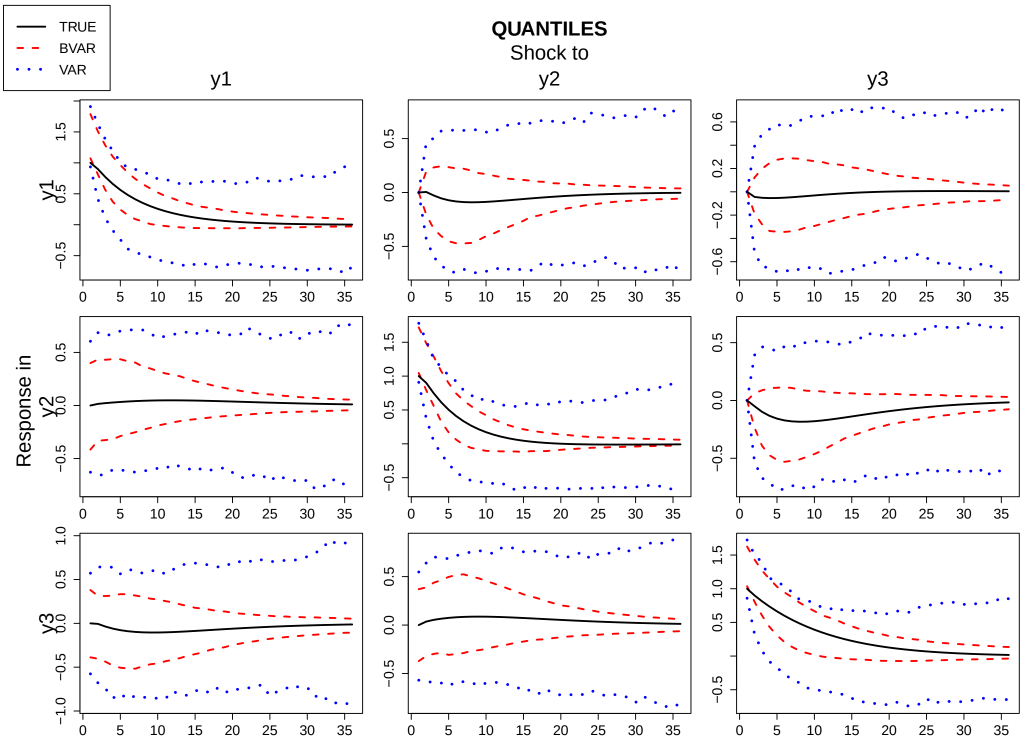


Figure 7.6: Comparison between the 0.05 and 0.95 quantiles of the estimated impulse responses of the BVAR and VAR with the true impulse responses for DGPa where first own lags are set to 0.9 instead of 1 and the fourth setup. The choice of hyperparameters is based on the marginal likelihood.

sample and the predictive likelihood only uses a smaller fraction of the data.

These 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 may be 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 simulation are consistent with the previous results. As an example, Figure 7.6 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.

7.4 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 performs variable selection via the marginalized marginal likelihood. Based on the selected variables, Section 4.3 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 are sufficient to stop the increase of house prices.

7.4.1 Theoretical Considerations about the Data

We use the real estate price index 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. This price index has also be found useful in an de Meulen et al. (2014). The index is measured with monthly frequency starting in 2007 and is based on offer prices placed on ImmobilienScout24, for more information, see (Bauer et al. 2013). Figure 7.7 shows the time series of the log house prices index.

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 7.10. 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 in the direct effect of monetary policy on house prices.

[8] This 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 term interests besides house prices.

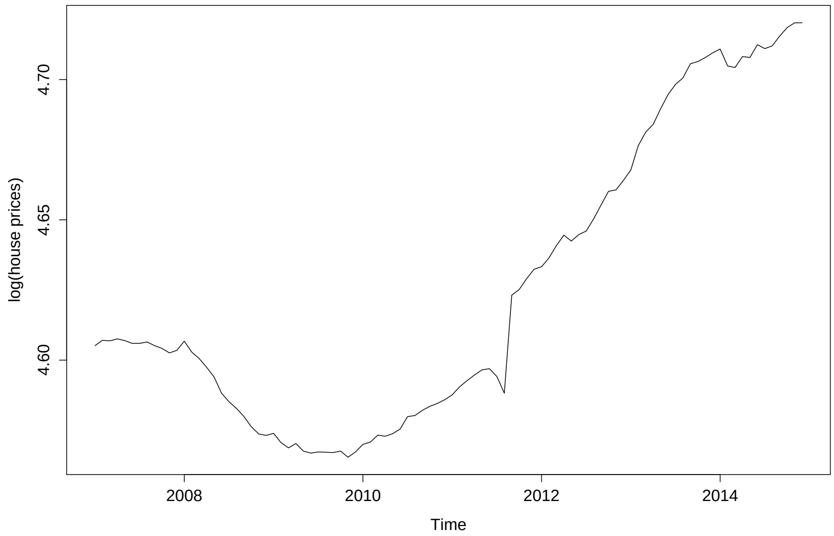


Figure 7.7: Log-house price indices

Several studies analyze the link between central banks' monetary policy decisions and house prices. Examples include Jarociński & Smets (2008), Gopolicy & Hofmann (2008), Musso et al. (2011), Setzer & Greiber (2007) and Demary (2009). All find evidence that monetary policy influences house prices. Gross domestic product (GDP), the consumer price index (CPI) or both are the only variables that all these studies use as control variables in their VAR models. A recent study for Switzerland by Berlemann & Freese (2013) include stock prices in their VAR model. 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 & Zan (2002) and McQuinn & O'Reilly (2007) find evidence of a long-run cointegration relationship between house prices, income and interest rates.

Thus, such previous studies do not seem to agree as to which variables belong in a VAR to model house prices. We aim to contribute to answering this question by calculating posterior variable inclusion probabilities. Furthermore, in contrast to previous studies we focus our attention on Germany and investigate the time period after the financial crisis. We therefore provide evidence for the recent turbulent period

Table 7.2: Variable selection

Variables	Inclusion Prob.	Best Model	Model size	Size Prob.
Production	0.08		2	0.79
DAX	0.01		3	0.20
Income	0.09		4	0.01
CPI	0.09		5	0.00
Eonia	0.15			
Rent	0.05			
r	0.64	X		
Employment	0.10			

The table shows the posterior variable inclusion probabilities for each variable based on formula (7.19), where the marginal likelihood is replaced by the marginalized marginal likelihood (7.22) for house prices. Each model receives equal prior weights. Posterior variable inclusion probabilities are then simply the sum of all posterior probabilities of all models which include the concerned variable. Size Prob. is calculated in the same fashion for each model size, where the model size gives the number of variables included in the BVAR. The best model corresponds to the variable combination which yields the highest marginalized marginal likelihood for house prices.

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. By focusing on this specific period, we contribute to finding an explanation for this recent house price boom in Germany.

We further consider rents, also taken from ImmobilenScout24. 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

Table 7.3: Sign restrictions

Response in	Demand shock	Supply shock	r shock
CPI	+	-	-
Prod	+	+	-
r			+
House_P	+		-

Restrictions are imposed for the period where the shock hits the system. An increase is denoted by + in the respective variable and - represents a decrease.

variables are expressed in logs.

7.4.2 Empirical Variable Selection

We compute posterior variable inclusion probabilities for all potentially useful variables considered in Section 4.1. In total we consider all possible combinations of our eight variables, for VAR models with up to four variables in addition to house prices. The posterior probabilities are calculated based on (7.19), where the marginal likelihood is replaced by the marginalized marginal likelihood for house prices and each model receives equal prior weights. Posterior variable inclusion probabilities are then simply the sum of all models which include the concerned variable. The hyperparameters for each model 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. The hyperparameters and lag lengths are chosen out of all combinations of $\tilde{p} \in \{1, 2, \ldots, 12\}$, $\lambda_0 \in \{0.1, 0.5, 0.9\}$, $\lambda_1 \in \{0.1, 0.5, 0.9\}$, $\lambda_3 \in \{0.1, 1, 2, 4\}$, $\lambda_4 \in \{1, 2, 3\}$, $\mu_5 \in \{0, 0.1, 1\}$ 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 each model. Moreover, one dummy variable for the month 2011:09 is added to account for a potential outlier in the time series of house prices, see Figure 7.7. However, the dummy not have an important influence on the 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 & Freeman 2006).

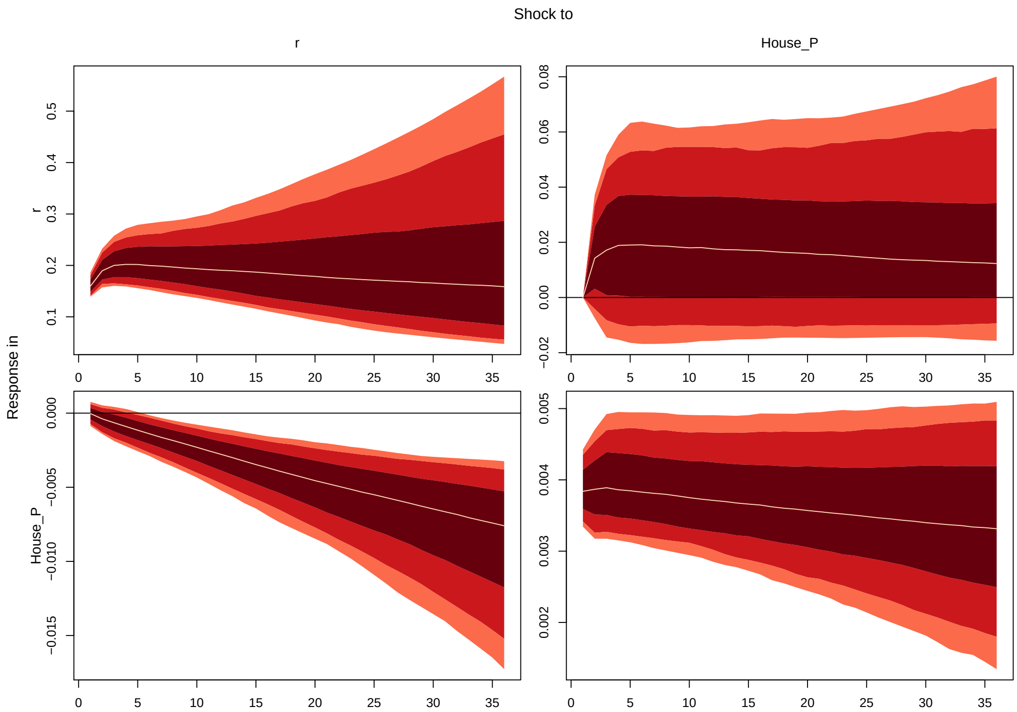


Figure 7.8: Median impulse response functions with 68%, 90% and 95% error bands, with r is ordered first and house prices second.

Table 7.2 shows the posterior variable inclusion probabilities. The short term and long term interest rates are the only variables with a probability above 10%. In line with our arguments in Section 4.1 the long term interest rates receives a much higher probability than the Eonia. Furthermore, parsimonious models receive the highest probability and the model which yields the highest marginalized marginal likelihood for house prices just includes the long term interest rate. This provides evidence that the interest rate is more important for modeling house prices compared to the other macroeconomic variables and that a larger information set does not help to explain house prices.

Nevertheless, to account for the uncertainty of these results and to be able to compare them with previous literature we additionally consider a model where *CPI* and *Prod* is added[9], also bearing in mind that the simulation results have shown that

[9]Note that under the condition that r is included in the BVAR adding *Prod* yields the highest marginalized marginal likelihood and conditioning on r and *Prod* adding *CPI* to the BVAR yields

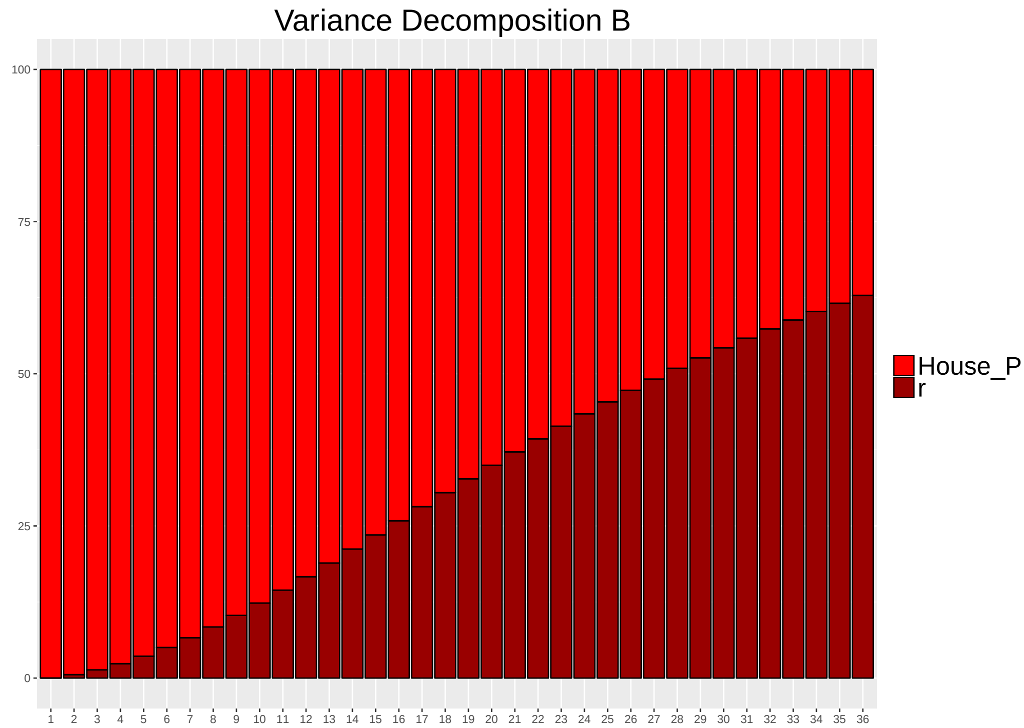


Figure 7.9: The figure shows the forecast error variance decomposition in percent for house prices with, r is ordered first and house prices second.

the BVAR is robust to adding irrelevant variables. The inclusion of these variables will provide further evidence that highlights/shows the importance of interest rates relative to these two variables for the development of house prices. Therefore this confirms the results of the variable selection exercise, by showing that this two variables are not useful to explain house prices. Furthermore, this allows us to use conventional sign restrictions as an additional identification strategy.

7.4.3 Empirical Results

Innovation Accounting

First, we discuss the results of the structural analysis of the two-variable BVAR preferred by the variable selection exercise. As expected, Figure 7.8 shows that a positive shock to long term interest rates has a negative effect on house prices. In

the highest marginalized marginal likelihood.

order to quantify the relative importance of the different shocks, Figure 7.9 reports the proportion that each shock contributes to the forecast error variance of house prices over a 36 month horizon, based on the median impulse response functions. 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.

We now turn to the discussion of the structural analysis of the BVAR model including house prices, *CPI*, interest rates and production. We use sign restrictions as our preferred identification strategy for this model, which avoids an economically challenging ordering of the variables of interest. We use the algorithm by Ramirez et al. (2010) to find an $\boldsymbol{A}_0$ which satisfies the imposed sign restrictions. Note that $\boldsymbol{A}_0$ is not unique, as one could find many $\boldsymbol{A}_0$ matrices that satisfy the sign restrictions but have elements of different magnitude. We hence generate 1000 $\boldsymbol{A}_0$ matrices that satisfy the sign restrictions for each Gibbs draw and then retain the $\boldsymbol{A}_0$ matrix that is closest to the median of these 1000 matrices (Fry & Pagan 2011). We impose standard sign restrictions for the demand, supply and interest rate shock, see Table 7.3. The house price shock is left unidentified in order to capture the effects of omitted variables and other shocks conceptually not belonging to any of the three categories identified. Figure 7.17 shows the impulse response functions and Figure 7.18 the variance decomposition. The results show that the negative impulse response in house prices at all horizons to an interest rate shock is robust to the inclusion of further control variables.

In order to show that our results are also robust to the identification strategy, we also estimate impulse responses by employing a recursive identification of the system. 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. Figure 7.20 shows the impulse response functions, Figure 7.21 shows the variance decomposition and Table 7.4 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. Taken together, the estimated impulse responses are robust to different identification assumptions and to adding additional variables.

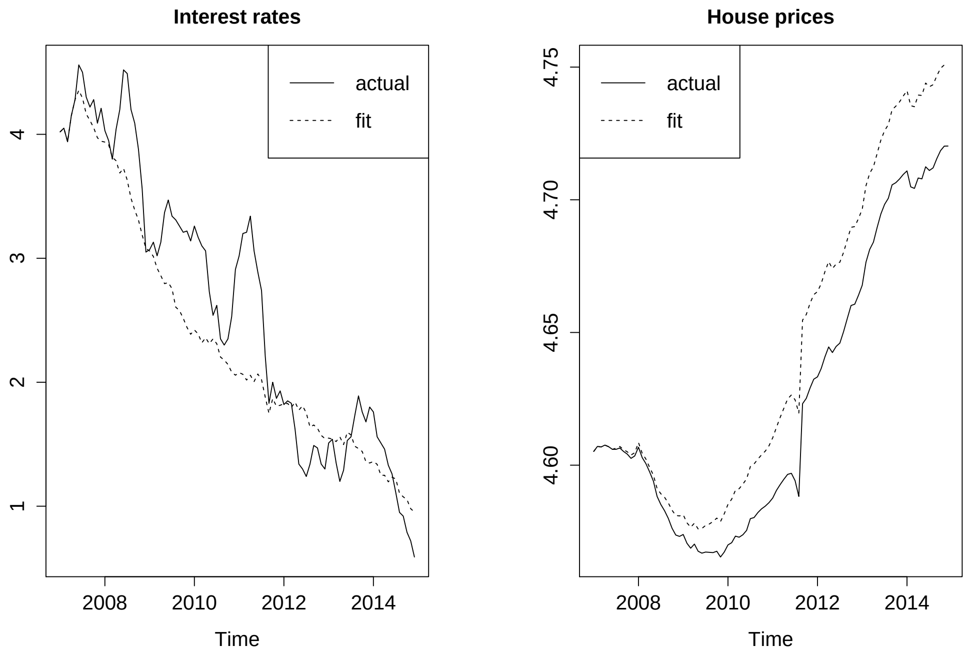


Figure 7.10: The figure shows the actual and the fitted path of the interest rate and house prices in Germany, by setting all interest rates shocks to zero.

Counterfactual Analysis

To gain further insights into the development of house prices, we perform an in-sample counterfactual simulation conditional on the first p observations, using our four-variable BVAR model. We shall turn to the out-of-sample counterfactual analysis as described in Section 2.3 afterwards. First, we simulate the reduced form VAR by setting all coefficients to their mean posterior 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 7.10 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 mostly had a positive effect on the interest rate. However, this cannot entirely explain the decrease of the house prices from 2008 to 2010. Figure 7.11 shows the likely development of house prices to Germany if no

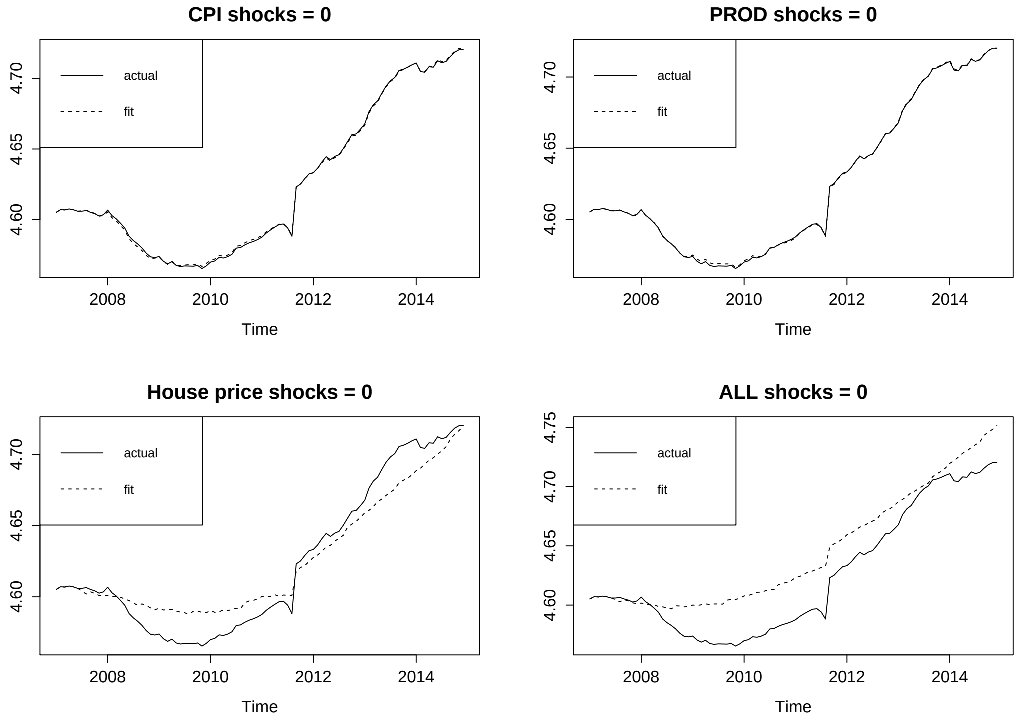


Figure 7.11: The figure shows the actual and the fitted path of house prices in Germany, by setting either all individual shocks (residuals) of the corresponding variable to zero or all shocks in the VAR system.

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. Overall, the results confirm the result of our variable selection exercise.

We now analyze how house prices might develop if the interest rate returned to its initial value of 4%, for example 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

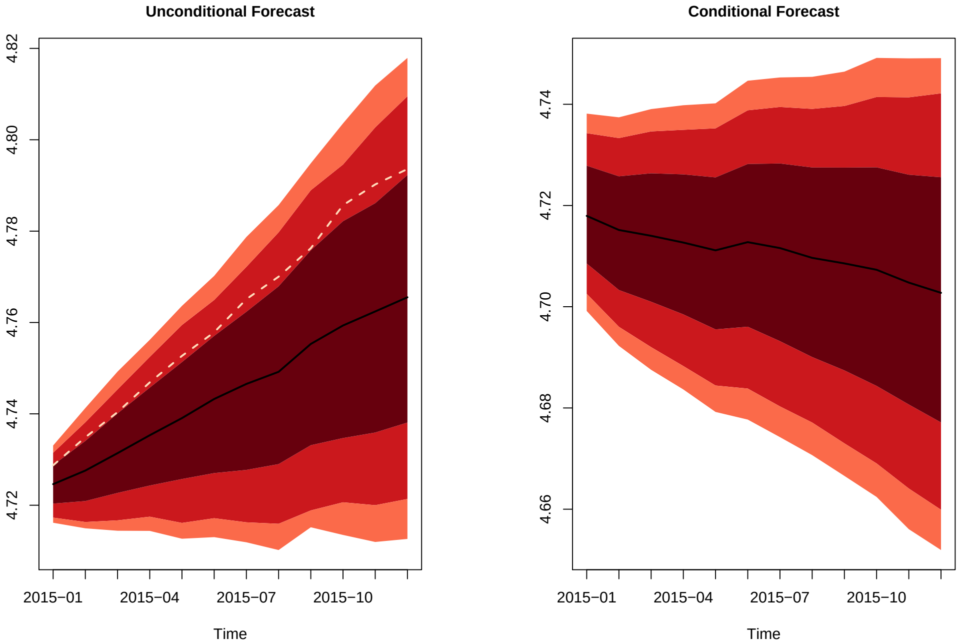


Figure 7.12: Conditional and unconditional forecasts with 68%, 90% and 95% error bands for house prices in Germany. The conditional forecast is based on the assumption that interest rates are at 4% over the forecast horizon and the dotted line represents the actual values. Furthermore interest rates are included in the model.

with the results of an unconditional forecast. The unconditional forecast accounts for both parameter and forecast uncertainty. Figure 7.12 shows the result for the two variable BVAR and 7.19 shows the result for the four variable BVAR. In both models 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.[10]

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

7.5 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. builting 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.

7.A Figures

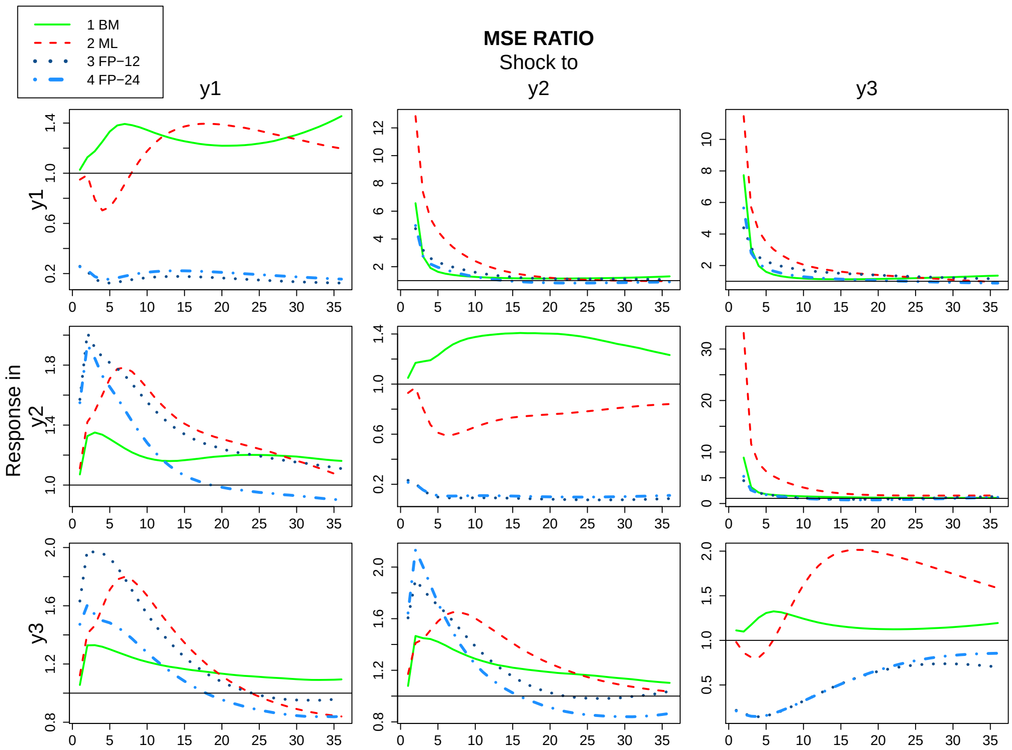


Figure 7.13: Comparison between the four different approaches for DGPa and setup 2. For further explanation see Figure 7.5.

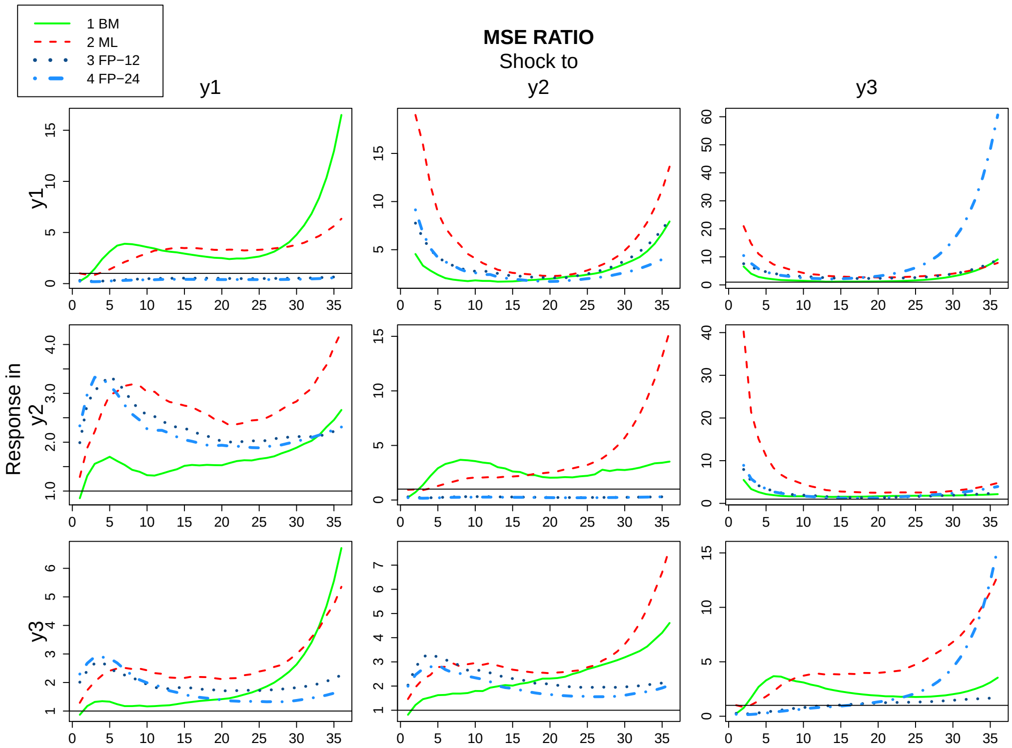


Figure 7.14: Comparison between the four different approaches for DGPa and setup 3. For further explanation see Figure 7.5.

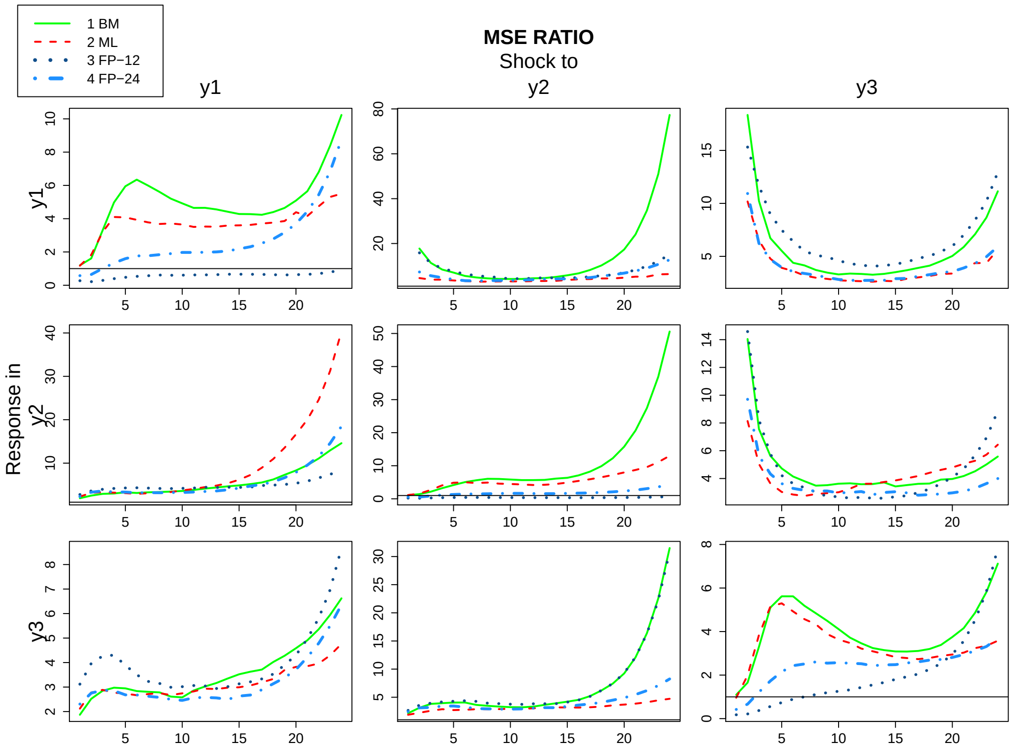


Figure 7.15: Comparison between the four different approaches for DGPa and setup 4. For further explanation see Figure 7.5.

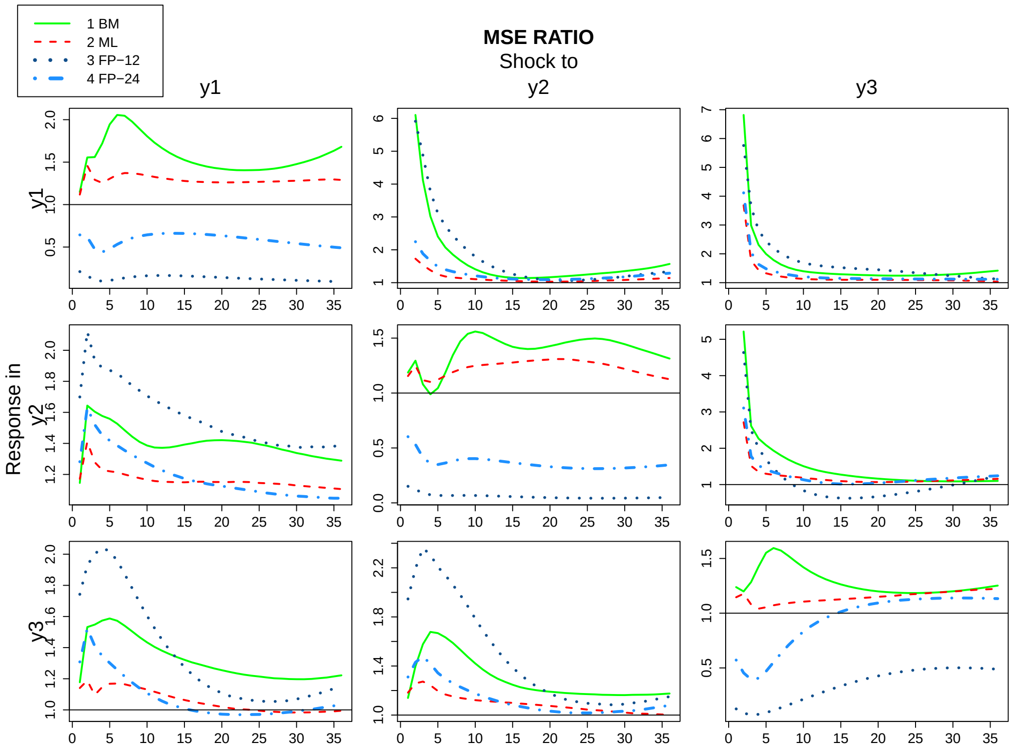


Figure 7.16: Comparison between the four different approaches for DGPa and setup 5. For further explanation see Figure 7.5.

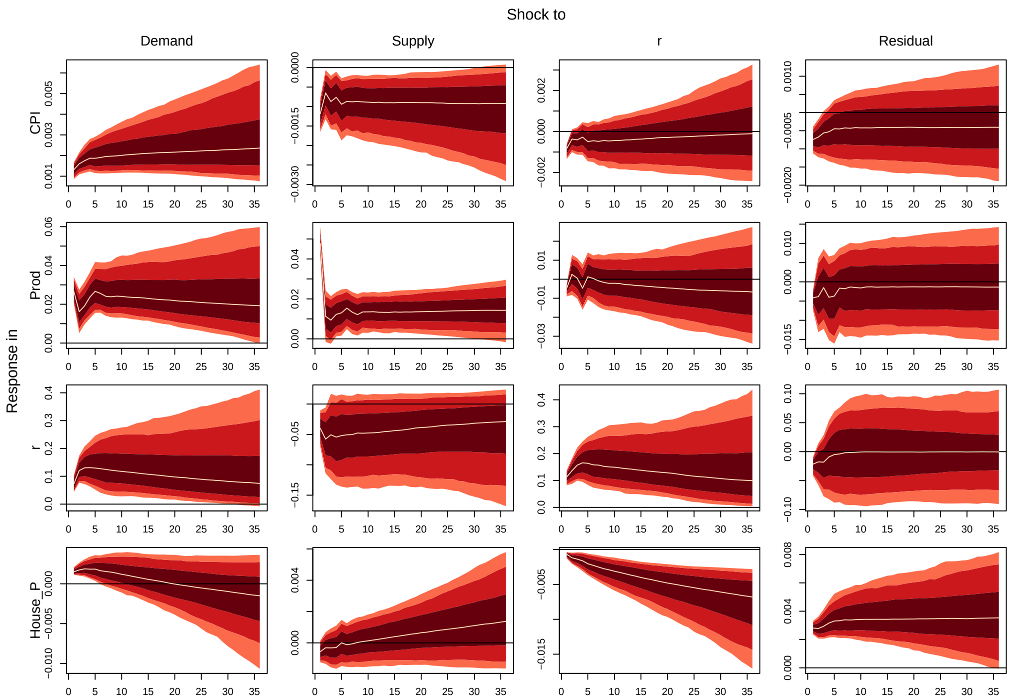


Figure 7.17: Median impulse response functions with 68%, 90% and 95% error bands, identified through sign restrictions.

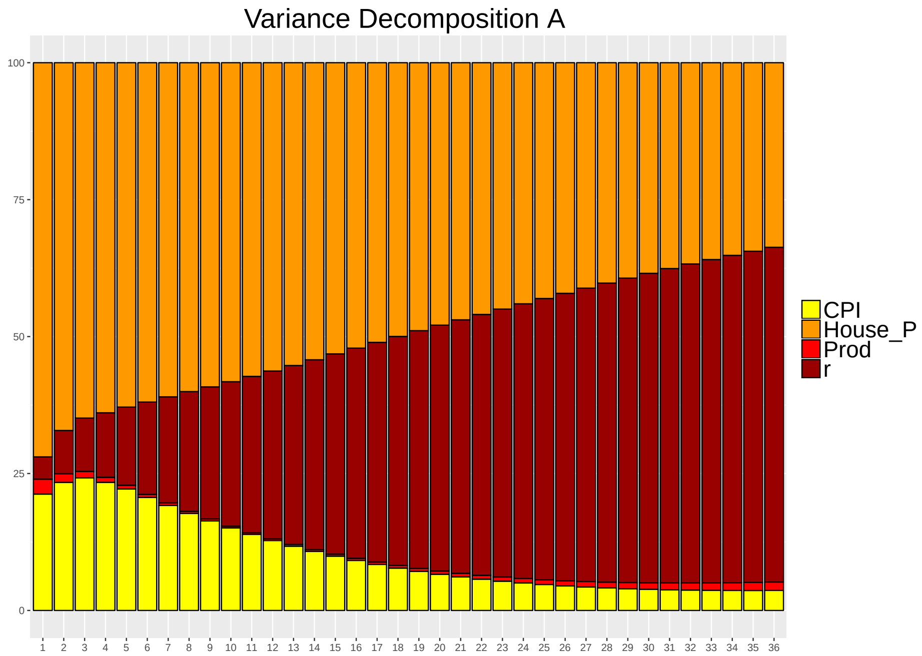


Figure 7.18: The figure shows the forecast error variance decomposition in percent for house prices, identified through sign restrictions.

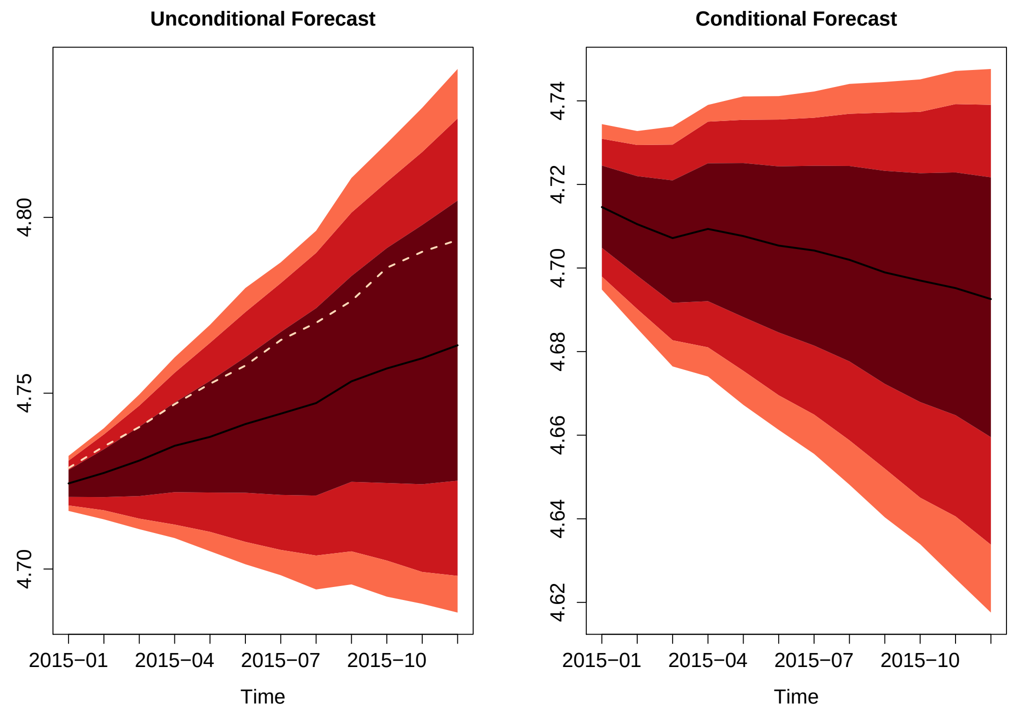


Figure 7.19: Conditional and unconditional forecasts with 68%, 90% and 95% error bands for house prices in Germany. The conditional forecast is based on the assumption that interest rates are at 4% over the forecast horizon and the dotted line represents the actual values. Furthermore, the variables production and *CPI* are included in the model.

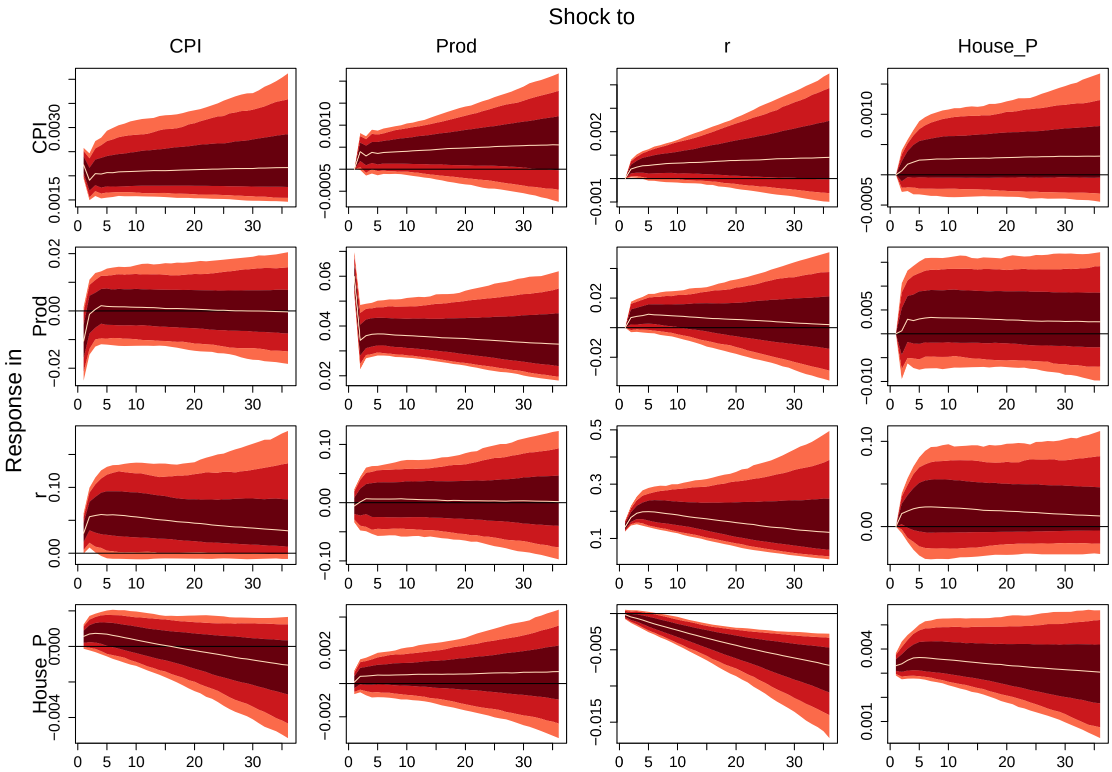


Figure 7.20: Median impulse response functions with 68%, 90% and 95% error bands, identified through cholesky decomposition.

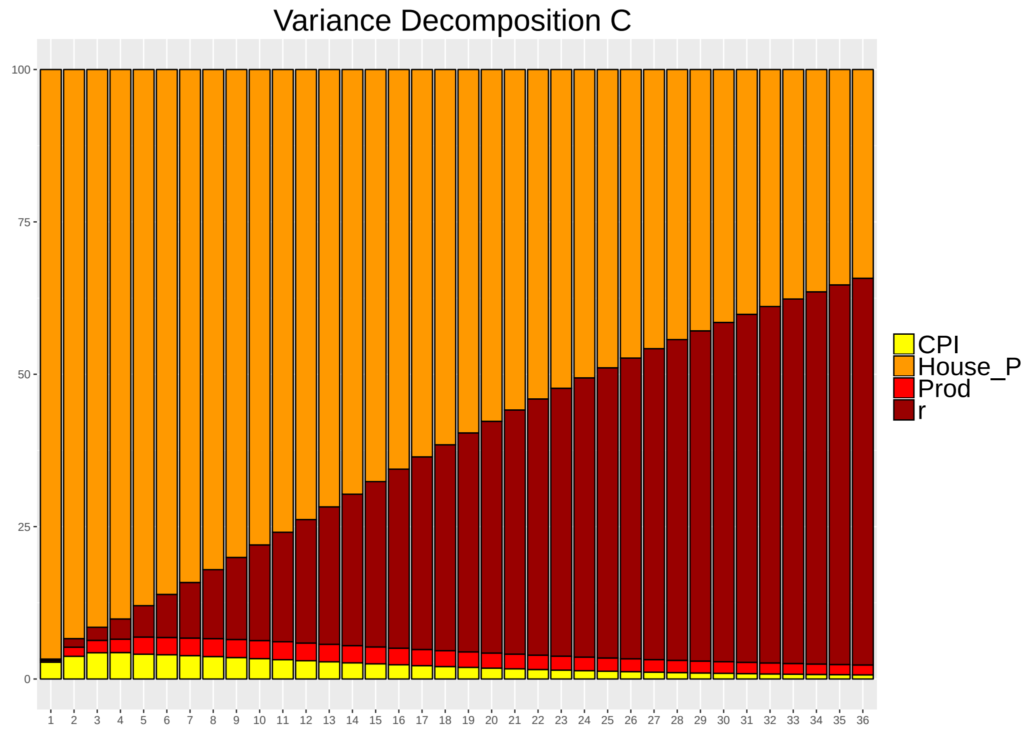


Figure 7.21: The figure shows the forecast error variance decomposition in percent for house prices, identified through cholesky decomposition.

7.B Tables

Table 7.4: 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

8 Conclusion

This thesis explores several Bayesian methods for empirically modeling macroeconomic and financial time series. A common problem in all empirical studies is that the number of observations is small relative to the number of potential variables. This raises problems for conventional econometric methods. Intuitively, there is not enough information in the data to estimate large models in an unrestricted fashion. Therefore, one major contribution of this thesis is to explore econometric methods which are useful in such an environment.

One approach to address this issue is to use Factor methods. The Factor approach saves degrees of freedom by summarizing the data through few factors which are then used as variables. This thesis uses this approach to investigate the effect on economic policy uncertainty on many macroeconomic as well as financial variables. It turns out that economic policy uncertainty affects both macroeconomic and financial variables but in a different way. Furthermore, shrinkage methods (like the Lasso or Minnesota prior) turned out to be useful to overcome the problem of overfitting in rich parametrized models, e.g. the VAR model, with only a small number of observations. The use of prior information provides a natural way to shrink the coefficients towards zero. Empirical or hierarchical Bayes methods provides a strategy to do this in a data driven way. This allows to use a richly parametrized VAR model to investigate which factors have driven the recent house price boom in Germany. In turns out that the major source of this boom was the decrease of long-term interest rates. Moreover, Bayesian additive regression trees are attractive to address the problem of a short span of observations relative to the number of potentially relevant explanatory variables by providing built-in variable selection. In addition, they allow for non-linear interaction effects between the predictor in a natural way. In an extensive forecasting comparison using a large data set of US macroeconomic time series the Bayesian additive regression trees outperform the Factor and Lasso model for nine key macroeconomic variables.

Finally, Bayesian model selection and averaging is considered in an environment with many predictors. These approaches are useful in order to find the relevant variables to address the question of which variables are important to explain for example inflation or exchange rates movements. However, the relevance of the predictors may change over time. In this case, only asking whether a variable is important or not is not addressing the right question. A researcher may not be interested in assessing whether a variable is important, but rather when it is. In order to address this issue this thesis introduced two novel modelling approaches: Markov Dimension Switching and Additive learning from model space. Both approaches turned out to be useful in forecasting macroeconomic variables. Given the promising results of this thesis, it may be beneficial for future empirical macroeconomic research to use methods which allow to include a large set of variables in order to derive new insights or more precise forecasts of future outcomes.

Bibliography

Aastveit, K., Natvik, G. & Sola, S. (2017), 'Economic uncertainty and the influence of monetary policy', *Journal of International Money and Finance* **76**, 50–67.

Alam, M. R. (2015), 'Economic policy uncertainty in the US: Does it matter for Canada?', *Economics Bulletin* **35**(4), 2725–2732.

Alessandri, P. & Mumtaz, H. (2014), 'Financial regimes and uncertainty shocks', *Queen Mary University of London, School of Economics and Finance, Working Papers* **857**.

Alexopoulos, M. & Cohen, J. (2009), 'Uncertain times, uncertain measures', *University of Toronto, Working Papers* **352**.

Amir-Ahmadi, P., Matthes, C. & Wang, M.-C. (2018), 'Choosing prior hyperparameters: With applications to time-varying parameter models', *Journal of Business & Economic Statistics* .

an de Meulen, P., Micheli, M. & Schmidt, T. (2014), 'Forecasting real estate prices in Germany: The role of consumer confidence', *Journal of Property Research* **31**(3), 244–263.

Ang, A., Bekaert, G. & Wei, M. (2007), 'Do macro variables, asset markets, or surveys forecast inflation better?', *Journal of Monetary Economics* **54**(4), 1163–1212.

Atkeson, A. & Ohanian, L. (2001), 'Are Phillips curves useful for forecasting inflation?', *Federals Reserve Bank of Minneapolis* **25**(1).

Aye, G., Gupta, R., Hammoudeh, S. & Joong, W. (2015), 'Forecasting the price of gold using dynamic model averaging', *International Review of Financial Analysis* **41**, 257–266.

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.

Baker, S. R., Bloom, N. & Davis, S. J. (2012), 'Has economic policy uncertainty hampered the recovery?', *Chicago Booth Research Paper* **12**.

Baker, S. R., Bloom, N. & Davis, S. J. (2016), 'Measuring economic policy uncertainty', *Quarterly Journal of Economics* **131**(4), 1593–1636.

Bańbura, M., Goammpme, D. & 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. & 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.

Baur, D., Beckmann, J. & Czudaj, R. (2016), 'A melting pot - Gold price forecasts under model and parameter uncertainty', *International Review of Financial Analysis* **48**, 282–291.

Bauwens, L. & Korobilis, D. (2013), *Bayesian Methods*, Handbook of Research Methods and Applications in Empirical Macroeconomics, chapter 16, pp. 363–380.

Beckmann, J. & Schüssler, R. (2016), 'Forecasting exchange rates under parameter and model uncertainty', *International Money and Finance* **60**, 267–288.

Belke, A. & Osowski, T. (2018), 'International effects of euro area versus US policy uncertainty: A FAVAR approach', *Economic Inquiry* .

Belmonte, M., Koop, G. & Korobilis, D. (2014), 'Hierachical shrinkage in time-varying parameter models', *Journal of Forecasting* **33**(1), 80–94.

Benati, L. (2013), 'Economic policy uncertainty and the great recession', *Journal of Applied Econometrics* . Revise and Resubmit.

Benito, A. (2006), 'Does job insecurity affect household consumption?', *Oxford Economic Papers* **58**(1), 157–181.

Berger, J. O. (1985), *Statistical decision theory and Bayesian analysis*, Springer series in statistics, Springer, New York, Berlin, Heidelberg.

Berlemann, M. & Freese, J. (2013), 'Monetary policy and real estate prices: A disaggregated analysis for Switzerland', *International Economics and Economic Policy* **10**(4), 469–490.

Bernanke, B. S. (1983), 'Irreversibility, uncertainty and cyclical investment', *Quarterly Journal of Economics* **98**(1), 85–106.

Bernanke, B. S., Boivin, J. & Eliasz, P. (2005), 'Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach', *Quarterly Journal of Economics* **120**(1), 387–422.

Bharat, B. & 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.

Bloom, N. (2009), 'The impact of uncertainty shocks', *Econometrica* **77**(3), 623–685.

Bloom, N., Bond, S. & Van Reenen, J. (2007), 'Uncertainty and investment dynamics', *Review of Economic Studies* **74**(2), 391–415.

Bordo, M. D., Duca, J. V. & Koch, C. (2016), 'Economic policy uncertainty and the credit channel: Aggregate and bank level US evidence over several decades', *Journal of Financial Stability* **26**, 90–106.

Bork, L. & Moller, S. (2015), 'Forecasting house prices in the 50 states using dynamic model averaging and dynamic model selection', *International Journal of Forecasting* **31**(1), 63–78.

Boumparis, P., Milas, C. & Panagiotidis, T. (2017), 'Economic policy uncertainty and sovereign credit rating decisions: Panel quantile evidence for the eurozone', *Journal of International Money and Finance* **79**, 39–71.

Brandt, T. P. & Freeman, J. R. (2002), 'Moving mountains: Bayesian forecasting as policy evaluation', *Meeting of the Midwest Political Science Association, Chicago, Il* .

Brandt, T. P. & Freeman, J. R. (2006), 'Advances in Bayesian time series modeling and the study of politics: Theory testing, forecasting and policy analyis', *Political Analysis* **14**(1), 1–36.

Brandt, T. P. & Freeman, J. R. (2009), 'Modeling macro-political dynamics', *Political Analysis* **17**(2), 113–142.

Bruyn, R., Gupta, R. & Eyden, R. (2015), 'Can we beat the randomwalk model for the south african rand-US dollar and south african rand-UK pound exchange rates? evidence from dynamic model averaging?', *Emerging Markets Finance and Trade* **51**(3), 502–524.

Buncic, D. & Moretto, C. (2015), 'Forecasting copper prices with dynamic averaging and selection models', *The North American Journal of Economics and Finance* **33**, 1–38.

Byrne, J., Korobilis, D. & Ribeiro, P. (2018), 'On the sources of uncertainty in exchange rate predictability', *International Economic Review* **59**(1), 1468–2354.

Caggiano, G., Castelnuovo, E. & Groshenny, N. (2014), ‘Uncertainty shocks and unemployment dynamics in US recessions’, *Journal of Monetary Economics* **67**, 78–92.

Caggiano, G., Castelnuovo, E. & Nodari, G. (2017), ‘Uncertainty and monetary policy in good and bad times’, *Bank of Finland Research and Discussion Paper* (8).

Caldara, D., Fuentes-Albero, C., Gilchrist, S. & Zakrajsek, E. (2016), ‘The macroeconomic impact of financial and uncertainty shocks’, *European Economic Review* **88**, 185–207.

Canova, F. & Nicolo, G. (2002), ‘Monetary disturbances matter for business fluctuations in the G-7’, *Journal of Monetary Economics* **49**(6), 1131–59.

Canova, F. & Pina, J. (1998), ‘Monetary policy misspecification in VAR models’, *CEPR Working Paper* (2333).

Carlin, B. P. & Louis, T. A. (2000), *Bayes and empirical Bayes methods for data analysis*, 2nd ed edn, FL: Chapmann & Hall/CRC.

Carlstrom, C., Fuerst, T. & Paustian, M. (2009), ‘Monetary policy shocks, choleski identification and DNK models’, *Journal of Monetary Economics* **56**(7), 1014–1021.

Carrière-Swallow, Y. & Céspedes, L. F. (2013), ‘The impact of uncertainty shocks in emerging economies’, *Journal of International Economics* **90**(2), 316–325.

Carriero, A., Clark, T. & Marcellio, M. (2015), ‘Bayesian VARs: Specification choice and forecast accuracy’, *Journal of Applied Econometrics* **30**(1), 46–73.

Carter, C. K. & Kohn, R. (1994), ‘On gibbs sampling for state space models’, *Biometrika* **81**(3), 541–553.

Castelnuovo, E., Caggiano, G. & Pellegrino, G. (2017), ‘Estimating the real effects of uncertainty shocks at the zero lower bound’, *European Economic Review* **100**, 257–272.

Chan, J. C., Koop, G., Leon-Gonzalez, R. & Strachan, R. W. (2012), ‘Time varying dimension models’, *Journal of Business & Economic Statistics* **30**(3), 358–367.

Chang, T., Chen, W. Y., Gupta, R. & Nguyen, D. K. (2015), ‘Are stock prices related to the political uncertainty index in OECD countries? Evidence from the bootstrap panel causality test’, *Economic Systems* **39**(2), 288–300.

Chen, J., Jiang, F. & Tong, G. (2016), ‘Economic policy uncertainty in China and stock market expected returns’. Available at: http://dx.doi.org/10.2139/ssrn.2808862.

Chib, S. (1996), 'Calculating posterior distributions and modal estimates in Markov mixture models', *Journal of Econometrics* **75**(1), 79–97.

Chipman, H., George, E. & McCulloch, R. (1998), 'Bayesian CART model search', *Journal of the American Statistical Association* **93**(443), 935–948.

Chipman, H., George, E. & McCulloch, R. (2010), 'BART: Bayesian additive regression trees', *The Annals of Applied Statistics* **4**(1), 266–298.

Christiano, L., Eichenbaum, M. & Evans, C. L. (1999), 'Monetary policy shocks: What have we learned and to what end?', *Handbook of macroeconomics* **1**, 65–148.

Cogley, T. & Sargent, T. J. (2005), 'Drifts and volatilities: Monetary policies and outcomes in the post WWII US', *Review of Economic Dynamics* **8**(2), 262–302.

Colombo, V. (2013), 'Economic policy uncertainty in the US: Does it matter for the euro area?', *Economics Letters* **121**(1), 39–42.

Del Negro, M. & Primiceri, G. (2015), 'Time-varying structural vector autoregressions and monetary policy: A corrigendum', *Review of Economic Studies* **82**(4), 1342–1345.

Demary, M. (2009), 'The link between output, inflation, monetary policy and housing price dynamics', *MPRA Working Paper* .

Deryugiana, E. B. & Ponomarenko, A. (2013), 'A large Bayesian vector autoregression model for russia', *BOFIT Discussion Paper* .

Diebold, F. X. & Mariano, R. S. (1995), 'Comparing predictive accuracy', *Journal of Business & Economic Statistics* **13**(3), 253–263.

Ding, S. & Karlsson, S. (2014), 'Model averaging and variable selection in VAR models', *manuscript* .

Doan, T., Littermann, R. B. & Sims, C. (1984), 'Forecasting and conditional projection using realistic prior distributions', *Econometric Reviews* **3**(1), 1–100.

Donadelli, M. (2015), 'Google search-based metrics, policy-related uncertainty and macroeconomic conditions', *Applied Economics Letters* **22**(10), 801–807.

Drachal, K. (2016), 'Forecasting spot oil price in a dynamic model averaging framework. Have the determinants changed over time?', *Energy Economics* **60**, 35–46.

ECB (2016), The impact of uncertainty on activity in the euro area, Technical report, Available at: http://crisisobs.gr/en/2016/12/the-impact-of-uncertainty-on-activity-in-the-euro-area/.

Ellis, C., Mumtaz, H. & Zabcyk, P. (2014), 'What lies beneath? A time-varying FAVAR model for the UK transmission mechanism', *The Economic Journal* **124**(576), 668–699.

Fagin, S. (1964), 'Recursive linear regression theory, optimal filter theory, and error analyses of optimal systems', *IEEE International Convention Record Part 1* pp. 216–40.

Filippo, G. D. (2015), 'Dynamic model averaging and cpi inflation forecasts: A comparison between the Euro area and the United States', *Journl of Forecasting* **34**(8), 619–648.

Fry, R. & Pagan, A. (2011), 'Sign restrictions in structural vector autoregressions: A critical review', *Journal of Economic Literature* **49**(4), 938–60.

Gambetti, L. & Musso, A. (2017), 'Loan supply shocks and the business cycle', *Journal of Applied Econometrics* **32**(4), 764–782.

Gelman, A., Charlin, S. & Rubin, D. (2004), *Bayesian Data Analysis*, 2 edn, Chapman & Hall / CRC.

Gerlach, R., Carter, C. & Kohn, R. (2000), 'Efficient Bayesian inference in dynamic mixture models', *Journal of American Statisitcal Association* **95**(451), 819–828.

Geweke, J. & Amisano, G. (2010), 'Hierachical Markov normal mixture models with application to financal asset returns', *Journal of Applied Econometrics* **26**(1), 1–29.

Geweke, J., Koop, G. & Van Dijk, H. (2011), *The oxford handbook of Bayesian econometrics*, Oxford University Press.

Giannone, D., Lenza, M. & Primiceri, G. E. (2015), 'Prior selection for vector autoregressions', *Review of Economics and Statistics* **97**(2), 436–451.

Gilchrist, S., Sim, J. & Zakrajšek, E. (2014), 'Uncertainty, financial frictions, and investment dynamics', *NBER Working Papers* **20038**.

Gopolicy, C. & Hofmann, B. (2008), 'House prices, money, credit and the macroeconomy', *Oxford Reviews of Economic Policy* **24**(1), 180–205.

Groen, J., Paap, R. & Ravazzolo, F. (2013), 'Real-time inflation forecasting in a changing world', *Journal of Business & Economic Statistics* **31**(1), 29–44.

Gupta, R., Hammoudeh, S., Kim, W. & Simo-Kengne, B. (2014), 'Forecasting China's foreign exchange reserves using dynamic model averaging: The roles of macroeconomic fundamentals, financial stress and economic uncertainty', *The North American Journal of Economics and Finance* **28**, 170–189.

Hamilton, J. (1989), 'A new approach to the economic analysis of nonstationary time series and the business cycle', *Econometrica* **57**(2), 357–384.

Hodrick, E. & Prescott, E. (1996), 'Postwar US business sycles: An empirical investigation', *Journal of Money, Credit and Banking* pp. 1–6.

International Monetary Fund (2012), 'World economic outlook: Coping with high debt and sluggish growth'.

James, W. & Stein, C. (1961), 'Estimation with quardratic loss', *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* **1**, 361–379.

Jarociński, M. & Smets, F. (2008), 'House prices and the stance of monetary policy', *ECB Working Paper* .

Jazwinsky, A. (1970), *Stochastic Processes and Filtering Theory*, New York: Academic Press.

Jorda, O., Schularich, M. & Taylor, A. M. (2015), 'Interest rates and house prices: Pill or poison?', *FRBSF Economic Letter* . Federal Reserve Bank of San Francisco.

Jurado, K., Ludvigson, S. & Ng, S. (2015), 'Measuring uncertainty', *Amercian Economic Review* **105**(3), 1177–1216.

Kadiyala, K. R. & 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.

Kim, S., Shephard, N. & Chib, S. (1998), 'Stochastic volatility: Likelihood inference and comparison with ARCH models', *The Review of Economic Studies* **65**(3), 361–393.

Koop, G. & Korobilis, D. (2010), 'Bayesian multivariate time series methods for empirical macroeconomics', *Foundations and Trends in Econometrics* .

Koop, G. & Korobilis, D. (2011), 'UK macroeconomic forecasting with many predictors: Which models forecast best and when do they do so?', *Economic Modelling* **28**(5), 2307–2318.

Koop, G. & Korobilis, D. (2012), 'Forecasting inflation using dynamic model averaging', *International Economic Review* **53**(3), 867–886.

Korobilis, D. (2013*a*), 'Assessing the transmission of monetary policy using time-varying parameter dynamic factor models', *Oxford Bulletin of Economics and Statistics* **75**(2), 157–179.

Korobilis, D. (2013*b*), 'Hierachical shrinkage prior for dynamic regressions with many predictors', *International Journal of Forecasting* **29**(1), 43–59.

Korobilis, D. (2013*c*), 'VAR forecasting using Bayesian variable selection', *Journal of Applied Econometrics* **28**(2), 204–230.

Lauritzen, S. (1992), 'Propagation of probabilities, means and variances in mixed graphical association models', *Journal of the American Statistical Association* **87**(420), 1098–1108.

Lippi, M. & Reichlin, L. (1994), 'VAR analysis, nonfundamental representations, blaschke matrices', *Journal of Econometrics* **63**(1), 307–325.

Litterman, R. (1986*a*), 'Forecasting with Bayesian vector autoregressions five years of experence', *Journal of Business & Economic Statistics* **4**(1), 25–38.

Litterman, R. B. (1979), 'Techniques of forecasting using vector autoregressions', *Working Paper, Federal Reserve Bank of Minneapolis* .

Litterman, R. B. (1986*b*), 'Forecasting with Bayesian vector autoregressions - Five years of experience', *Journal of Business & Economic Statistics* **4**(1), 25–38.

Liu, J., Wei, Y., Ma, F. & Wahab, M. (2017), 'Forecasting the realized range-based volatility using dynamic model averaging approach', *Economic Modelling* **61**, 12–26.

Liu, L., Ma, F. & Wang, Y. (2015), 'Forecasting excess stock returns with crude oil market data', *Energy Economics* **48**, 316–324.

McQuinn, K. & O'Reilly, G. (2007), 'A model of cross-country house prices', *Research Technical Paper* .

Meinen, P. & Röhe, O. (2017), 'On measuring uncertainty and its impact on investment: Cross-country evidence from the euro area', *European Economic Review* **92**, 161–179.

Molodtsova, T. & Papell, D. (2009), 'Out-of-sample exchange rate predicitability with Taylor rule fundamentals', *Journal of International Economics* **77**(2), 167–180.

Musso, A., Neri, S. & 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.

Naser, H. (2016), 'Estimating and forecasting the real prices of crude oil: A data rich model using dynamic model averaging approach', *Energy Economics* **56**, 75–87.

Netšunajev, A. & Glass, K. (2017), 'Uncertainty and employment dynamics in the euro area and the US', *Journal of Macroeconomics* **51**, 48–62.

Nodari, G. (2014), 'Financial regulation policy uncertainty and credit spreads in the US', *Journal of Macroeconomics* **41**, 122–132.

Österholm, P. & Stockhammar, P. (2014), 'The euro crisis and swedish GDP growth - A study of spillovers', *Applied Economics Letters* **21**(16), 1105–1110.

Park, T. & Casella, G. (2008), 'The Bayesian Lasso', *Journal of American Statistical Association* **103**(482), 681–686.

Popp, A. & Zhang, F. (2016), 'The macroeconomic effects of uncertainty shocks: The role of the financial channel', *Journal of Economic Dynamics and Control* **69**, 319–349.

Primiceri, G. E. (2005), 'Time varying structural vector autoregressions', *Review of Economic Studies* **72**(3), 821–852.

Prüser, J. & Schlösser, A. (2017), 'The effects of economic policy uncertainty on european economies: Evidence from a TVP-FAVAR', *Ruhr Economic Papers* (708).

Rafiq, M. & Mallick, S. (2008), 'The effect of monetary policy on output in emu3: A sign restriction approach', *Journal of Macroeconomics* **30**, 1756–1791.

Raftery, A., Kárný, M. & Ettler, P. (2010), 'Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill', *Technometrics* **52**(1), 52–66.

Ramirez, J., D., W. & Zha, T. (2010), 'Structural vectorautoregressions: Theory of identification and algorithms for inference', *The review of economic studies* **77**(2), 665–696.

RiskMetrics (1996), Technical document-fourth edition, Technical report, Available at: http://www.riskmetrics.com/system/files/private/td4e.pdf.

Risse, M. & Kern, M. (2016), 'Forecasting house-price growth in the euro area with dynamic model averaging', *The North American Journal of Economics and Finance* **38**, 70–85.

Robertson, J. & Tallman, E. (1999), 'Vector autoregressions: Forecasting and reality', *Federal Reserve Bank of Atlanta Economic Review* **84**(1), 4–18.

Romer, C. D. (1990), 'The great crash and the onset of the great depression', *The Quarterly Journal of Economics* **105**(3), 597–624.

Scheffel, E. M. (2015), 'Accounting for the political uncertainty factor', *Journal of Applied Econometrics* **31**, 1048–1064.

Scholl, A. & Uhlig, H. (2008), 'New evidence on the puzzles: Results from agnostic identification on monetary policy and exchange rates', *Journal of International Economics* **76**(1), 1–13.

Setzer, R. & 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* pp. 179–212.

Sims, C. A. & Zha, T. (1998), 'Bayesian methods for dynamic multivariate models', *International Economic Review* **4**(39), 949–68.

Sims, C. A. & Zha, T. (2006), 'Were there regime switches in US monetary policy?', *The American Economic Review* **96**(1), 54–81.

Stock, J. H. & Watson, M. (1998), 'Diffusion indexes', *NBER Working Papers* **6702**.

Stock, J. H. & Watson, M. W. (2005), 'Implications of dynamic factor models for VAR analysis', *NBER Working Papers* **11467**.

Stock, J. & Watson, M. (1999), 'Forecasting inflation', *Journal of Monetary Economics* **44**(2), 293–335.

Stock, J. & Watson, M. (2002), 'Forecasting using principal components from a large number of predictors', *Journal of the American Statistical Association* **97**(460), 1167–1179.

Stock, J. & Watson, M. (2007), 'Why has US inflation become harder to forecast?', *Journal of Money, Credit and Banking* **39**(7), 3–33.

Stock, J. & Watson, M. (2010), 'Model inflation after the crisis', *National Bureau of Economic Research Working Paper* (16488).

Stock, J. & Watson, M. (2012), 'Generalized shrinkage methods for forecasting using many predictors', *Journal of Business and Economic Statistic* **30**(3), 481–493.

Summers, P. (2001), 'Forecasting Australia's economic performance during the Asian crisis', *International Journal of forecasting* **17**(3), 499–515.

Taylor, J. (1993), 'Discretion versus policy rules in practice', *Carnegie-Rochester Conference Series on Public Policy* **39**, 195–214.

Theil, H. & Goldberger, A. S. (1960), 'On pure and mixed statistical estimation in economics', *International Economic Review* **2**(1), 65–78.

Timmermann, A. (2006), Forecast combinations, *in* G. Elliot, C. Granger & A. Timmermann, eds, 'Handbook of Economic Forecasting', Vol. 1, North-Holland, Amsterdam.

Uhlig, H. (2004*a*), 'Do technology shocks lead to a fall in total hours worked?', *Journal of the European Economic Association* **2**(2-3), 361–371.

Uhlig, H. (2004*b*), 'What moves GDP', *Econometric Society 2004 North American Winter Meetings* .

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.

Waggoner, D. F. & Zha, T. (1999), 'Conditional forecasts in dynamic multivariate models', *Review of Economics and Statistics* **81**(4), 639–651.

Waisman, M., Ye, P. & Zhu, Y. (2015), 'The effect of political uncertainty on the cost of corporate debt', *Journal of Financial Stability* **16**, 106–117.

Wang, Y., Ma, F., Wei, Y. & Wu, C. (2016), 'Forecasting realized volatility in a changing world: A dynamic model averaging approach', *Journal of Banking and Finance* **64**, 136–149.

Wei, Y. & Cao, Y. (2017), 'Forecasting house prices using dynamic model averaging approach: Evidence from china', *Economic Modelling* **61**, 147–155.

Wright, J. H. (2009), 'Forecasting US inflation by Bayesian model averaging', *Journal of Forcasting* **28**(4), 131–144.

Zellner, A. (1971), *An introduction to Bayesian inference in econometrics*, John Wiley & Sons.

Eidesstattliche Erklärung

Hiermit versichere ich diese Arbeit selbstständig verfasst zu haben und dabei keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Zitate wurden an den entsprechenden Stellen in der Arbeit kenntlich gemacht.

Essen, 26.09.2018

Jan Prüser

Erklärung zur Koautorenschaft

Das fünfte Kapitel “On the Time-Varying Effects of Economic Policy Uncertainty on the US Economy” ist in Zusammenarbeit mit Alexander Schlösser entstanden. Wir erklären hiermit, dass Jan Prüser an sämtlichen Teilen der Arbeit in etwa proportional beteiligt war.

Essen, 26.09.2018

Jan Prüser Alexander Schlösser

Das sechste Kapitel “The Effects of Economic Policy Uncertainty on European Economies: Evidence from a TVP-FAVAR” ist in Zusammenarbeit mit Alexander Schlösser entstanden. Wir erklären hiermit, dass Jan Prüser an sämtlichen Teilen der Arbeit in etwa proportional beteiligt war.

Essen, 26.09.2018

Jan Prüser Alexander Schlösser

Das siebte Kapitel “House Prices and Interest Rates-Bayesian Evidence from Germany” wurde in einer früheren Version unter dem Titel “The time series properties of German real estate prices: a Bayesian analysis” als Masterarbeit am Prüfungsamt der Universität Essen eingereicht. Seitdem habe ich in Zusammenarbeit mit Christoph Hanck weiter an dem Kapitel gearbeitet und es erheblich erweitert. Wir erklären hiermit, dass Jan Prüser an sämtlichen Teilen der Arbeit überproportional beteiligt war.

Essen, 26.09.2018

Jan Prüser Christoph Hanck