UNIVERSITÄT BERN

Faculty of Economics and Social Sciences

Department of Economics

Priors from DSGE Models for Dynamic Factor Analysis

Gregor Bäurle

08-03

August 2008

DISCUSSION PAPERS

Schanzeneckstrasse 1
Postfach 8573
CH-3001 Bern, Switzerland
http://www.vwi.unibe.ch

Priors from DSGE Models for Dynamic Factor Analysis*

Gregor Bäurle[†]

August 12, 2008

Abstract

We propose a method to incorporate information from Dynamic Stochastic General Equilibrium (DSGE) models into Dynamic Factor Analysis. The method combines a procedure previously applied for Bayesian Vector Autoregressions and a Gibbs Sampling approach for Dynamic Factor Models. The factors in the model are rotated such that they can be interpreted as variables from a DSGE model. In contrast to standard Dynamic Factor Analysis, a direct economic interpretation of the factors is given. We evaluate the forecast performance of the model with respect to the amount of information from the DSGE model included in the estimation. We conclude that using prior information from a standard New Keynesian DSGE model improves the forecast performance. We also analyze the impact of identified monetary shocks on both the factors and selected series. The interpretation of the factors as variables from the DSGE model allows us to use an identification scheme which is directly linked to the DSGE model. The responses of the factors in our application resemble responses found using VARs. However, there are deviations from standard results when looking at the responses of specific series to common shocks.

Keywords: Dynamic Factor Model, DSGE Model, Bayesian Analysis, Forecasting, Transmission of Shocks.

JEL-Classification: C11, C15, C22, C53, E37, E47

*I thank Fabio Canova, Domenico Giannone, Marc Giannoni, Klaus Neusser, Serena Ng, Frank Schorfheide and Mark Watson for helpful discussions and valuable comments. I also thank seminar participants at the Institute for Advanced Studies Vienna, the University of Pennsylvania, participants at the Spring Meeting of Young Economists 2007 in Hamburg and the 62nd European Meeting of the Econometric Society in Budapest. Part of this work was written while visiting the University of Pennsylvania, USA. I am grateful for financial support from the Swiss National Science Foundation. The usual disclaimer applies.

†Economics Department, University of Bern, Schanzeneckstrasse 1, CH-3001 Bern, Switzerland. Tel.: +41 31 631 80 75. Email: gregor.baeurle@vwi.unibe.ch.

Contents

1 Introduction

Dynamic Factor Models are becoming increasingly popular in empirical macroeconomics due to their ability to cope with a large number of data series. The idea is to gather the informational content of a high dimensional data vector in a small dimensional vector of common factors. Each series is decomposed into a sum of a (linear) combination of these common factors and an idiosyncratic term. Assumptions on the correlations between the idiosyncratic terms - in the simplest case there assumed to be orthogonal - make the identification feasible. Compared to a small dimensional Vector Autoregression (VAR), the analysis is more robust with respect to the disturbing influence of idiosyncratic components of the series (such as measurement errors).

From an economist's point of view however, the interpretation of the results and in particular the factors is difficult: Their relationship to economically interpretable concepts is not immediate. This often leads to a purely statistical a analysis. 'Storytelling' in an economically sensible way, which is essential for policy makers, is not immediately possible in the standard setting. Another problem is that even though the parameter space is reduced to some degree compared to a VAR, still many parameters have to be estimated. Generally, the problem with a large parameter space is reflected in the fact that scarcely parameterized models very often have a better forecasting performance than models with a large number of parameters.

In this paper, we propose a Bayesian method which provides a solution to both problems mentioned above: We first rotate the factors such that the relationships between factors and data series leads to a natural economic interpretation. The fact that the factors are only identified up to an invertible rotation is thereby exploited: We use an informative prior on the factor loadings in order to rotate the factors. Even an almost flat prior rotates the factors: As the likelihood is flat, the 'curvature' induced by the prior causes a rotation. By increasing the tightness it is possible to implement believes about the relationship between specific series and factors. Given the economic interpretation induced by the rotation, we then combine prior information from a small scale Dynamic Stochastic General Equilibrium (DSGE) model with information contained in the data in order to estimate process. DSGE models provide a complete description of the dynamics of economic concepts, parameterized only by a small set of deep structural parameters. Thereby, depending on the weight of the prior, the parameter space is shrinked towards a parsimonious representation of the data.

To our knowledge, there is no contribution in the literature which builds prior knowl-

edge from DSGE models into Dynamic Factor Analysis. By contrast, there are sophisticated methods designed for VARs. The availability of VAR techniques motivates the idea to use a Gibbs Sampler for the estimation of a Dynamic Factor Model: Given an initial set of parameters, we draw from the distribution of the unobserved factors. Given the factors, standard regression and, most importantly, VAR techniques can be applied to draw from the distribution of parameters. This draw can again be used to simulate a new set of factors. For a sufficient number of iterations over these two steps, the draws converge to the joint distribution of parameters and states (see Geweke (2005) for a description of Gibbs sampling methods in general and Kim and Nelson (1999) for their application to state space models). In this way, the procedure allows to incorporate Bayesian VAR methods into the estimation of the factor dynamics. The method used in this paper was developed by DelNegro and Schorheide (2004). Intuitively, a sample of artificial data ('dummy observations') is simulated with the DSGE model. This sample is added to actual data and the VAR is estimated over this augmented data set. The size of the dummy observation sample relative to the actual sample gives the weight of the DSGE model restrictions in the estimation. For comparison purposes, the same idea can be used to implement the so-called Minnesota Prior. Instead of a DSGE model, some statistical model (e.g. independent Random Walks for each variable) delivers the set of dummy observations (see Sims (2005) for a general discussion of dummy observation priors).
Factor models are useful to study the transmission of structural shocks to economic variables. Forni, Lippi, and Reichlin (2003) and Giannone and Reichlin (2006) argue that they are more suitable than VARs, as the large information set potentially helps to overcome non-fundamentalness problems. In previous studies, identification has been achieved using merely ad-hoc contemporaneous and long-run restrictions. A main advantage of our setting is, that it is possible to use an identification scheme which is directly linked to the DSGE model. The method relies on the fact that in the DSGE model, the shocks are exactly identified. It builds on the strategy proposed by DelNegro and Schorheide (2004) in the context of VARs. The validity of the method hinges on the assumption that all the factors in our model can indeed be directly related to variables in the DSGE model. We therefore compare the outcome to an agnostic identification strategy relying on sign restrictions. This idea goes back to Faust (1998) and has been elaborated by Uhlig (2005) and Canova (2002) in the context of structural VARs. The latter identification strategy does not necessarily rely on the interpretation of all the factors as variables of the DSGE model. The first strategy - even though widely used in the context of VARs - is novel in the factor model literature. The second scheme has already been recognized as potential

strategy in Stock and Watson (2005). However, pointing to computational problems, they do not apply the method. Interpreting some factors as economic variables, the major computational problems can be solved in our model.

The closest precursor to this paper is Boivin and Giannoni (2006). They estimate a DSGE model with a large data set, interpreting variables in the DSGE model as factors and their observed data as their (imperfect) measures.[1] Our model continuously bridges the gap between a non-structural factor model and the model of Boivin and Giannoni (2006) in the following sense: In the extreme case of degenerate priors on some of the factor loadings and by strictly imposing the restriction of the DSGE model one estimates the DSGE model akin to Boivin and Giannoni (2006).[2] By relaxing restrictions implied by the DSGE model and making the priors for the factor loadings less informative, it is possible to move towards a non-structural factor model.

Our approach is also related to the analysis in Giannone, Reichlin, and Sala (2006). They show that the state variables of a DSGE model can be interpreted as common factors driving the observed variables. However, the focus is slightly different: Giannone, Reichlin, and Sala (2006) model the dynamics of one observed series per variable in the DSGE model in which the number of variables can be larger than the number of shocks. In contrast, we assume that we have the same number of shocks as variables in the DSGE model. Instead, we interpret the variables in the DSGE model as common factors driving a large number of observed variables.

As an application, we estimate the model on quarterly US data from 1985 to 2007. For the DSGE model prior, we use the standard version of the New-Keynesian model as proposed by DelNegro and Schorheide (2004). This model relates output, inflation and interest rates. We therefore select variables from the data set which are supposed to be directly related to these concepts.

A first result is that observed interest rates and observed prices are well described by one corresponding factor even for a very loose prior on the factor loadings. For

[1]The idea to introduce measurement errors into the empirical analysis of DSGE models by means of a factor structure goes back to Sargent (1989) and Altug (1989). It has also been studied in other papers, e.g. Watson (1993) and Ireland (2004).

[2]When the size of the dummy observation sample is infinite, the resulting estimator does not allow for misspecification. Nevertheless, the estimator differs from the one in Boivin and Giannoni (2006) in the sense that we minimize the discrepancy between the unrestricted Maximum Likelihood estimator of a finite order VAR of the factors and the respective values implied by the model parameter (see DelNegro and Schorheide (2004)). If the DSGE model has a VAR representation of the order considered, the estimation is equivalent, but most models used in modern macroeconomics do not have a finite order VAR representation. Another difference is that we do not apply the method suggested by Jacquier, Johannes, and Polson (2004) to reduce the influence of the prior distribution. Boivin and Giannoni (2006) use this method to check the robustness of their estimates with respect to the choice of priors.

the observed output series, the result is not as clear cut: Industrial production is well described by the 'output factor' while GDP loads also on the inflation factor. Posterior marginal data densities suggest that a moderate tightness of the prior for both the factor loadings and the DSGE prior are optimal. Also, we find that including information from the DSGE model improves the forecast performance for output series compared to a simple the non-structural Minnesota prior. For prices and interest rates, the results is ambiguous. Compared to a simple univariate autoregressive model, the performance of estimates using the DSGE prior is markedly better for most of the series. For large weights of the prior distribution the performance gets worse. This also indicates that the DSGE model is too restrictive in some dimensions.

We then analyze the impact of an identified monetary shock on the factors and the observed series. We find that the response of the factors are largely in line with the predictions of the New Keynesian DSGE model, also for small weights of the DSGE prior: A contractionary monetary shock decreases inflation, decreases interest rates (which is assumed in sign restriction identification) and has a negative impact on output growth. There are also some differences, however: The impact on the interest rates is more persistent in the factor model compared to what theory predicts. And, although the distribution of the long-run impact on output is centered around zero, the dispersion is rather wide. Hence, long-run neutrality of monetary shocks cannot be convincingly verified. The analyis of the responses of the observed series reflects the findings on the structure in the estimated observation equation: The impact of the shock on observed prices and interest rates are close the their corresponding factors'. The same is true for Industrial Production. However, the sign of the reaction of GDP is ambiguous.

The paper is structured as follows: Section 2 describes the empirical model and its identification. Section 3 sets out the example for DSGE model from which the prior distribution is inferred. Section 4 contains the estimation method. In section 5, the empirical application of the method is given. Section 6 concludes.

2 Empirical Model and Identification

It is assumed that the data evolves according to the following state space system:

Observation equation:

$$X_t = \Lambda F_t + v_t \tag{1}$$

State equation:

$$\Phi(L) F_t = e_t \tag{2}$$

X_t is a potentially high dimensional vector of N data series observed over T time periods. F_t is a vector of unobserved dynamic factors with a small dimension M.[3] Each variables in X_t loads on at least one factor. Λ is the $N \times M$ matrix of factor loadings. The factors F_t are related to its lagged values by $\Phi(L) = I - \Phi_1 L - \ldots - \Phi_p L^p$.

The error processes are assumed to be Gaussian White noise:

$$\begin{pmatrix} u_t \\ e_t \end{pmatrix} \sim \text{iidN}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} R & 0 \\ 0 & \Sigma \end{bmatrix} \right)$$

where

$$u_t = v_t - \Psi v_{t-1}$$

and R and Ψ are diagonal.[4] Ultimately, we will relate the residuals to structural shocks $\varepsilon_t = H_{VAR}^{-1} e_t$ with $cov(e_t) = I_M$ to analyze the response of the factors and the observed series to these shocks. We assume that H_{VAR} is invertible, hence that there are as many shocks as factors. The identification of H_{VAR} is described in Section 5.6.

A difference between this setting and standard factor models (e.g. Stock and Watson (2002b)), is that the loading matrix Λ is rotated in order to give an economic interpretation to the factors[5]. A structure that can easily be interpreted would be one in which a particular set of variables does only load on one factor. For instance, if different measures of output, e.g. industrial production for different sectors and measures for GDP, load exclusively on one particular factor, this factor can be interpreted as factor 'output'. Ideally, the data series are linked to the factors as follows:

[3]The interpretation of the factors as economic concepts from a DSGE model sidesteps an issue in factor analysis, namely how to identify the number of factors M in the model. In our setting, the number of factors is just the number of concepts the macroeconomist has in mind when she builds a model.

[4]Doz, Giannone, and Reichlin (2006) examine the properties of a quasi-likelihood estimator under omitted serial and cross-sectional correlation. They conclude that the effect is negligible if the size of the cross-section is large and the omitted cross-sectional correlation is limited (approximate factor structure).

[5]The standard procedure is to use an arbitrary statistical normalization.

$$\Lambda_{obj} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

.

Normalizing the non-zero elements to one scales the factors such that the variance of the factor is not changed by premultiplying the factor loadings with Λ_{obj}[6]. One possibility to achieve such a structure is to directly impose the restrictions (see Boivin and Giannoni (2006)). The disadvantage of this approach is that not all of the zero restrictions are necessary to identify the model. Hence, imposing the structure may be too restrictive. A different possibility is to exploit the fact that the factors are only identified up to an invertible rotation. To see this, plug an invertible matrix Q into the system:

$$\begin{aligned} X_t &= \Lambda Q Q^{-1} F_t + v_t \\ Q^{-1}\Phi(L) Q Q^{-1} F_t &= Q^{-1} e_t \end{aligned}$$

Define $\tilde{F}_t = Q^{-1}F_t$, $\tilde{e}_t = Q^{-1}e_t$, $\tilde{\Lambda} = \Lambda Q$ and $\tilde{\Phi}(L) = Q^{-1}\Phi(L)Q$ which yields the following system

$$\begin{aligned} X_t &= \tilde{\Lambda}\tilde{F}_t + v_t \\ \tilde{\Phi}(L)\tilde{F}_t &= \tilde{e}_t \end{aligned}$$

[6]Some normalizations are needed to exactly identify the factors. This is also the case when strictly imposing the restrictions on Λ, see Boivin and Giannoni (2006). An alternative to Λ_{obj} implemented here would be to assume that only one series in each group is related to the factor with loading one. This treats the series asymmetrically, but possibly improves the fit of the model.

The latter system is observationally equivalent to the former. The fact that we can rotate the factors with any invertible transformation Q can be used to make the factors interpretable without strictly imposing the zero restrictions: Starting from an arbitrarily (just) identified model, we can rotate the factors such that $\tilde{\Lambda} = \Lambda Q$ comes as close as possible to the desired factor structure[7]. In our Bayesian setting, the natural way to rotate the factors is to use an informative prior distribution for Λ with mean Λ_{obj}. This 'identifies' the factors in the sense that it puts curvature into the posterior density function for regions in which the likelihood function is flat. It is clear however, that imposing an informative prior for Λ is restrictive to some degree, depending on the tightness of the prior. The exact specification of the prior distribution of parameters in the observation equation is described in section 4. The prior distribution for the parameters in the state equation (2) is based on prior information from economic theory. In the next section, we give an intuition for this approach and describe the economic model we will use in our application.

[7]The idea of rotation has a long tradition in Factor Analysis, see Lawley and Maxwell (1971), but has up to our knowledge not been applied in a dynamic setting.

3 DSGE Model Prior

Presumably among the most popular models in contemporaneous monetary macroeconomics is the standard 'New Keynesian' model. It describes the joint dynamics of output, inflation and the interest rate based on optimizing behavior of a representative consumer and firms which are restricted by some constraints on adjusting prices. The model provides a complete description of comovement between output, inflation and interest rates. When it comes to the empirical implementation, we typically have several data series at our disposal. For example, inflation can be measured by the GDP deflator, consumer prices or producer prices. Often, it is not desirable to chose only one series out of these as immediate measure. We rather want to explain generic comovements of economic variables instead of modeling dynamics of selected data series. It is therefore natural to interpret the factors as variables macroeconomists are interested in and the specific data series as indicators related to that variable. Putting it that way, knowledge from macroeconomic theory can be useful for the estimation of relationships between the factors in our empirical model. The idea to include information from economic theory into the estimation can be implemented using the approach by DelNegro and Schorheide (2004). Their method is developed for the estimation of Bayesian VARs. Inspecting the state equation (2) we see that it has exactly the form of a VAR. By using a Gibbs sampler with data augmentation (in which the factors are interpreted as unknown parameters) it is therefore straight-forward to imbed their method into our framework.

Building in prior information from economic theory into the estimation of the state equation works as follows: Currently, most macroeconomic models fit into the 'Dynamic Stochastic General Equilibrium' (DSGE) framework. They provide a complete description of the dynamic process of a set of macroeconomic variables S_t. Establishing a relationship between S_t and the factors F_t, these dynamics directly translate into restrictions on the matrices Φ and Ω. The information on the restrictions possibly helps getting more precise estimates of the parameters. However, DSGE models are often highly stylized (in the sense that they are based on strong assumptions which simplify the analysis). It is therefore preferable to use some information on the restrictions, but not to strictly impose them. Technically, this can be implemented with 'Dummy' observations: We use observed data augmented with an artificial sample generated with the DSGE model, the 'Dummy' observations, to estimate the factor model. The size of the artificial data relative to the actual data (which is unobserved F_t in our case) gives the weight of the restrictions in the estimation. If the artificial sample is very small, we basically estimate an unrestricted version. If the sample

is very large compared to the actual sample, we only use the actual data to update estimation of the deep parameters of the DSGE models but then directly take the estimates of Φ and Σ implied by the DSGE model.
The likelihood function of a standard VAR depends only on the first and the second moments of the data. It follows that we only need to infer these moments from the DSGE model to augment the actual data set. We illustrate how this can be achieved with the following version of the standard New-Keynesian model (see Del-Negro and Schorheide (2004) and Lubik and Schorfheide (2004) for a derivation). The log-linearized equations are

$$y_t = \mathbb{E}_t y_{t+1} - \frac{1}{\tau}(r_t - \mathbb{E}_t \pi_{t+1}) + (1-\rho_g) g_t + \frac{\rho}{\tau} z_t$$

$$\pi_t = \beta \mathbb{E}_t \pi_{t+1} + \kappa (y_t - g_t)$$

$$r_t = \rho r_{t-1} + (1-\rho)[\psi_1 \pi_t + \psi_2 y_t] + \varepsilon_t^r$$

All the variables in the model are in written in deviations from steady state. The first equation is a standard Euler equation, linking output y_t to the expected real interest rate $r_t - \mathbb{E}_t \pi_{t+1}$, expected output $\mathbb{E}_t y_{t+1}$ and exogenous technology z_t. The Philipps curve can be derived by assuming Calvo Price setting, perfectly competitive labor markets and a linear production function. It relates current inflation π_t to expected inflation $\mathbb{E}_t \pi_{t+1}$, the output gap y_t and an exogenous demand shifter g_t. The third equation is a Taylor rule which attempts to describe the behavior of the Central Bank. The nominal interest rate r_t depends on the lagged nominal interest rate and the reaction of the Central Bank to current inflation, the output gap and a monetary shock ε_t^r. The exogenous components g_t and z_t evolve according to

$$z_t = \rho_z z_{t-1} + \varepsilon_t^z$$

$$g_t = \rho_g g_{t-1} + \varepsilon_t^g$$

The shocks ε_t^z , ε_t^g and the monetary policy shock ε_t^r are assumed to be uncorrelated with each other and across time.
Assuming rational expectations, there are several algorithms to solve the system. We use Sims' method (see Sims (2002)) and therefore define

$$S_t = (y_t, \pi_t, r_t, \mathbb{E}_t y_{t+1}, \mathbb{E}_t \pi_{t+1}, g_t, z_t)'$$

$$\varepsilon_t = (\varepsilon_t^z, \varepsilon_t^g, \varepsilon_t^r)'$$

$$\eta_t = (\eta_{y,t} = y_t - \mathbb{E}_{t-1} y_t, \eta_{\pi,t} = \pi_t - \mathbb{E}_{t-1}\pi_t)'$$

$$\theta = (\psi_1, \psi_2, \rho_R, \beta, \kappa, \tau, \rho_g, \rho_z, \sigma_R, \sigma_g, \sigma_z)'$$

add the equations

$$y_t = \mathbb{E}_{t-1} y_t + \eta_{y,t}$$

$$\pi_t = \mathbb{E}_{t-1} \pi_t + \eta_{\pi,t}$$

and write the system as

$$\Gamma_0(\theta) S_{t+1} = \Gamma_1(\theta) S_t + \Psi(\theta)\varepsilon_t + \Pi(\theta)\eta_t$$

The complete matrices $\Gamma_0(\theta)$, $\Gamma_1(\theta)$, $\Psi(\theta)$ and $\Pi(\theta)$ are given in the appendix[8].
If there is a unique stationary rational expectations solution, it can be casted into the following form:

$$S_t = G(\theta) S_{t-1} + H(\theta)\varepsilon_t$$

The matrices $G(\theta)$ are $H(\theta)$ are complicated non-linear functions of θ. For detailed information regarding the algorithm that maps $\Gamma_0(\theta)$, $\Gamma_1(\theta)$, $\Psi(\theta)$ and $\Pi(\theta)$ into $G(\theta)$ and $H(\theta)$, we refer to Sims (2002). For an extension to indeterminate systems see Lubik and Schorfheide (2004). The central assumption is that the factors represent economic variables which are contained in the DSGE model. We can therefore define the following selection equation relating the factors to the DSGE model variables.

$$F_t^* = Z S_t$$

We denote the DSGE model implied factors by F_t^* as opposed to the factors in the empirical model F_t. Note that even though the S_t is an autoregressive process of order one, this property does not translate into the implied process for F_t^*. Generally, F_t^* has a VAR representation of infinite order:

[8] The state vector contains also the expectations of future variables known at time t. This increases the dimension of the state vector which makes computation more time consuming. On the other hand, it also allows to directly introduce data that measures these expectations in the estimation which might be an interesting extension for future work.

$$F_t^* = \sum_{j=1}^{\infty} \Phi(\theta)_j F_{t-j}^* + e_t^*$$

with $e_t \sim \text{iidN}(0, \Sigma^*(\theta))$. In the empirical model, we will approximate the system by including only a finite number p of lagged factors:

$$F_t^* \approx \sum_{j=1}^{p} \Phi(\theta)_j^* F_{t-j}^* + e_t^*$$

Define

$$F_P^* = \begin{pmatrix} F_p^{*'} & F_{p-1}^{*'} & \dots & F_1^{*'} \\ F_{p+1}^{*'} & F_p^{*'} & \dots & F_2^{*'} \\ \vdots & \vdots & \ddots & \vdots \\ F_{T-1}^{*'} & F_{T-2}^{*'} & \dots & F_{T-p}^{*'} \end{pmatrix}$$

and

$$F^* = \begin{pmatrix} F_{p+1}^{*'} \\ F_{p+2}^{*'} \\ \vdots \\ F_T^{*'} \end{pmatrix}$$

Then define $\Gamma_{FF}^*(\theta) = \mathbb{E}(F^{*'} F^*)$, $\Gamma_{FF_P}^*(\theta) = \mathbb{E}(F_P^{*'} F^*)$ and $\Gamma_{F_P F_P}^*(\theta) = \mathbb{E}(F_P^{*'} F_P^*)$. These moments can easily be calculated given the solution to the DSGE model (see Appendix of DelNegro and Schorheide (2004)). For a given θ, the implied coefficient matrices $\Phi^*(\theta)$ and $\Sigma^*(\theta)$ - the maximum likelihood estimates of Φ and Σ for a truncated VAR on an infinitely large sample of artificial observations - are given by the Yule-Walker equations:

$$\Phi^*(\theta) = \Gamma_{F_P F_P}^{*-1}(\theta) \Gamma_{FF_P}^{*'}(\theta)$$

$$\Sigma^*(\theta) = \Gamma_{FF}^*(\theta) - \Gamma_{FF_P}^*(\theta) \Gamma_{F_P F_P}^{*-1}(\theta) \Gamma_{FF_P}^{*'}(\theta)$$

So the autocovariances up to the order p contain all the relevant information on the VAR parameters implied by the DSGE model. In the estimation, we use these moments to shrink the parameter space of the coefficients in the state equation towards the dynamics implied by the DSGE model. The concrete implementation of this idea is described in the next section.

4 Estimation method

Following Kim and Nelson (1999) and Boivin and Giannoni (2006) we use a Gibbs sampler to estimate the model. In general, Gibbs sampling works as follows. Partition the set of parameters Θ in K subsets, $\Theta = (\Theta_1, \Theta_2, \ldots, \Theta_K)$ and define $\Theta_{-k} = \{\Theta_1, \ldots, \Theta_{k-1}, \Theta_{k+1}, \ldots, \Theta_K\}$. Now suppose it is not possible to draw directly from the distribution of Θ, $p(\Theta)$. But the conditional distributions $p(\Theta_k \mid \Theta_{-k})$ are standard. Starting at an initial value Θ^0_{-k} repeat the following steps for $j = 1, \ldots, J$

(i) Set $\Theta^j = \Theta^{j-1}$

(ii) For each k, draw from $p(\Theta^j_k \mid \Theta^j_{-k})$. Replace the k-th element in Θ^j by the drawn value

(iii) Increase j by one and go back to step (i)

This yields a Markov chain in the parameters Θ:

$$p(\Theta_j | \Theta_{j-1}, ..., \Theta_1) = p(\Theta_j | \Theta_{j-1})$$

Under certain regularity conditions satisfied here[9], the stationary distribution of this Markov chain is $p(\Theta)$. Discarding some initial draws to ensure that the effect of the initial Θ^0_{-k} becomes negligible, which amounts to assuming that the Markov chain has converged to its stationary distribution, we draw from the joint distribution of the parameter vector given the data. Building on Kim and Nelson (1999), Boivin and Giannoni (2006) recognize that given the states F_t, standard methods could be used to draw from the distribution of the parameters of the model and that given the model parameters $(\Lambda, R, \Phi, \Sigma)$, standard methods could be used to sample from the distribution of the states. In our case, $\Theta_1 = (F_1, \ldots, F_T)$, $\Theta_2 = (\Lambda, R)$, $\Theta_3 = \Psi$and $\Theta_4 = (\Phi, \Sigma, \theta)$. Note that the set of parameters is augmented with the vector θ which denotes the parameter of the DSGE model used as prior. We therefore sample iteratively from the following conditional distributions:

$$p(\mathcal{F}^j \mid \Phi^{j-1}, \Sigma^{j-1}, \theta^{j-1}, \Lambda^{j-1}, R^{j-1}, \Psi^j, X) = p(\mathcal{F}^j \mid \Phi^{j-1}, \Sigma^{j-1}, \Lambda^{j-1}, R^{j-1}, \Psi^j, X)$$

$$p(\Lambda^j, R^j \mid \mathcal{F}^j, \Psi^{j-1}, \Phi^{j-1}, \Sigma^{j-1}, \theta^{j-1}, X) = p(\Lambda^j, R^j \mid \mathcal{F}^j, \Psi^{j-1}, X)$$

[9]See Geweke (2005) for details on conditions that ensure convergence.

$$p(\Psi^j \mid \mathcal{F}^j, \Lambda^j, R^j, \Phi^{j-1}, \Sigma^{j-1}, \theta^{j-1}, X) = p(\Psi^j \mid \mathcal{F}^j, \Lambda^j, R^j, X)$$

$$p(\Phi^j, \Sigma^j, \theta^j \mid \mathcal{F}^j, \Lambda^j, R^j, \Psi, X) = p(\Phi^j, \Sigma^j, \theta^j \mid \mathcal{F}^j)$$

where $\Phi^j = (\Phi_1^j, \dots, \Phi_p^j)$, $\mathcal{F}^j = \{F_1^j, \dots, F_T^j\}$ and $X = \{X_1, \dots, X_T\}$. The steps are now described in turn. We drop the index j for notational convenience, but keep in mind that the steps constitute only one iteration of the Gibbs sampler.

4.1 Step 1: Drawing from $\mathbf{p}(\mathcal{F} \mid \boldsymbol{\Phi}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \mathbf{R}, \boldsymbol{\Psi}, \mathbf{X})$

The algorithm described in Kim and Nelson (1999) is used to draw from the joint distribution of the states. The derivation assumes that the state space system (1) and (2) is Markovian of order one. If $p > 1$ the system has to be rewritten into a Markov system by redefining the state vector, see Appendix D[10]. A standard Kalman filter can be used to calculate the distribution of F_t given $X_1, \dots, X_t$ and the model parameters. Define $\mathcal{I}_t = \{X_1, \dots, X_t, \Phi, \Sigma, \theta, \Lambda, R\}$. We want to sample from $p(F_1, \dots, F_T \mid X_1, \dots, X_t, \Phi, \Sigma, \theta, \Lambda, R) = p(F_1, \dots, F_T \mid \mathcal{I}_T)$. Factorize the joint density into a product of conditional densities:

$$p(F_1, \dots, F_T \mid \mathcal{I}_T) \quad = \quad p(F_T \mid \mathcal{I}_T) p(F_{T-1} \mid F_T, \mathcal{I}_T) \dots p(F_1 \mid F_2, \dots, F_T, \mathcal{I}_T)$$

The Markov property of the system implies that $p(F_t \mid F_{t+1}, \dots, F_T, \mathcal{I}_T) = p(F_t \mid F_{t+1}, \mathcal{I}_T)$ and $p(F_t \mid F_{t+1}, \mathcal{I}_T) = p(F_t \mid F_{t+1}, \mathcal{I}_t)$. We therefore write

$$p(F_1, \dots, F_T \mid \mathcal{I}_T) \quad = \quad p(F_T \mid I_T) p(F_{T-1} \mid F_T, I_{T-1}) \dots p(F_1 \mid F_2, \mathcal{I}_1) \qquad (3)$$

The joint distribution of F_t and F_{t+1} given I_t is

$$\begin{pmatrix} F_t \\ F_{t+1} \end{pmatrix} \mid \mathcal{I}_t \sim \mathrm{N}\left(\begin{bmatrix} F_{t|t} \\ \Phi F_{t|t} \end{bmatrix}, \begin{bmatrix} P_{t|t} & P_{t|t}\Phi' \\ \Phi P_{t|t} & \Phi P_{t|t} \Phi' + \Sigma \end{bmatrix} \right)$$

where $F_{t|t} = E(F_t \mid \mathcal{I}_t)$ and $P_{t|t} = \mathbb{V}(F_t \mid \mathcal{I}_t)$ are outputs of the Kalman filter. So the distribution of S_t given S_{t+1} and X_t can be found with the standard formula for multivariate normal distributions:

[10] Also the fact that there is autocorrelation in the observation equation complicates the problem from a numerical point of view. The Kalman filter has to adapted as described in Appendix D.

$$\mathbb{E}(F_t \mid F_{t+1}, \mathcal{I}_t) = F_{t|t} + P_{t|t}\Phi'(GP_{t|t}G' + \Sigma)^{-1}(F_{t+1} - \Phi F_{t|t})$$

$$\mathbb{V}(F_t \mid F_{t+1}, \mathcal{I}_t) = P_{t|t} + P_{t|t}\Phi'(\Phi P_{t|t}\Phi' + \Sigma)^{-1}\Phi P_{t|t}$$

So $F_t \mid F_{t+1}, \mathcal{I}_t$ is normally distributed with expected value and variance that can easily be calculated with the output of the Kalman filter. The last step of the Kalman filter gives us the mean and the variance of $F_T \mid \mathcal{I}_T$. We draw from this distribution. Given this draw, we iteratively draw from $p(F_t \mid X, F_{t+1})$ where F_{t+1} is the value drawn from $p(F_{t+1} \mid \mathcal{I}_{t+1}, F_{t+2})$. According to equation (3), this gives us a draw from the joint distribution of the factors given the parameters of the model and the data.

4.2 Step 2: Drawing from $\mathbf{p}(\mathbf{\Lambda}, \mathbf{R} \mid \mathcal{F}, \mathbf{\Psi}, \mathbf{X})$ and $\mathbf{p}(\mathbf{\Psi} \mid \mathcal{F}, \mathbf{\Lambda}, \mathbf{R}, \mathbf{X})$

Given the states, standard methods can be used to draw from this distribution, see Chib (1993), Bauwens, Lubrano, and Richard (1999) and Boivin and Giannoni (2006). To draw from $p(\Lambda, R \mid \mathcal{F}, \Psi, \Phi, \Sigma, \theta, X)$, we first filter the data and the states $\tilde{X}_t = X_t - \Psi X_{t-1}$ and $\tilde{F}_t = F_t - \Psi F_{t-1}$ such that

$$\tilde{X}_t = \Lambda \tilde{F}_t + u_t$$

Conditional on Ψ and using the assumption that R is diagonal, standard multivariate regression methods can be used to draw from the distribution of Λ and R. We follow Boivin and Giannoni (2006) by using the conjugate prior described in Bauwens, Lubrano, and Richard (1999), p.58. The prior distribution $p(R_n, \Lambda_n \mid \Psi_n)$, where n denotes the respective row in the observation equation, is of the normal-inverted gamma-2 form (as defined in the appendix of Bauwens, Lubrano, and Richard (1999)):

$$R_n \sim \mathrm{iG}_2(3, 0.001)$$

$$\Lambda_n \sim \mathrm{N}(\Lambda_0, R_n M_0^{-1})$$

M_0 is a matrix of parameter that influences the tightness of the priors in the observation equation. The larger the elements of M_0 are, the closer we relate the observed series to the factors a priori. It follows that the posterior distribution is

$$R_n | X, F \sim \mathrm{iG}(\bar{R}_n, T + 0.001)$$

$$\Lambda_n | X, F \sim \mathrm{N}(\bar{\Lambda}_n, R_n \bar{M}_n^{-1})$$

where

$$\bar{\Lambda}_n = \bar{M}_n^{-1}(M_0 \Lambda_0 + \tilde{F}'\tilde{X})$$
$$\bar{R}_n = 3 + u'u + (\Lambda - \Lambda_0)'(M_0^{-1} +' (\tilde{F}'\tilde{F})^{-1})^{-1}(\Lambda - \Lambda_0)$$
$$\bar{M}_n = M_0 + \tilde{F}'\tilde{F}$$

Given the draws from this distribution, we can calculate a draw from the distribution of v_t. Hence, to draw from $p(\Psi \mid \mathcal{F}, \Lambda, R, \Phi, \Sigma, \theta, X)$, standard results for autoregressive processes can be used: Assuming a standard normal prior for Ψ_n we obtain

$$\Psi_n | X, F, \Lambda, R \sim N(\bar{\Psi}_n, \bar{N}_k^{-1})$$

where

$$\bar{\Psi}_n = \bar{N}_n^{-1} R_k^{-1} v_n' v_n \hat{\Psi}_n$$
$$\bar{N}_n = 1 + R_k^{-1} v_n' v_n$$

and $\hat{\Psi}_n$ is the OLS estimate of $v_{nt} = X_{nt} - \Lambda F_t$ on its lagged value (see Chib (1993)).

4.3 Step 3: Drawing from $\mathbf{p}(\mathbf{\Phi}, \mathbf{\Sigma}, \theta \mid \mathcal{F})$

In this step, we invoke the method of DelNegro and Schorheide (2004). We give here a merely intuitive description of their main results. For detailed information we refer to the original paper.
The joint posterior distribution is factorized as follows

$$p(\Phi, \Sigma, \theta \mid \mathcal{F}, \Lambda, R, \Psi, X) = p(\Phi, \Sigma \mid \mathcal{F}, \theta, \Lambda, R, \Psi, X) p(\theta \mid \mathcal{F}, \Lambda, R, \Psi, X)$$

The prior $p(\Phi, \Sigma, \theta)$ is specified hierarchically:

$$p(\Phi, \Sigma, \theta) = p(\Phi, \Sigma \mid \theta) p(\theta)$$

This allows to first draw from the posterior distribution of θ, and the draw from the posterior distribution of Φ and Σ given the draw of θ. The two steps are now described in turn. In what follows we use the following definitions: We parameterize the size of the artificial data T^* relative to the actual sample size actual sample size: $T^* = \lambda T$. The maximum-likelihood estimates of Φ and Σ based on artificial sample and actual sample are denoted by

$$\tilde{\Phi}(\theta) = (\lambda\Gamma^*_{F_pF_p} + F_P'F_P)^{-1}(\lambda\Gamma^*_{F_PF} + F_P'F)$$

$$\tilde{\Sigma}(\theta) = \frac{1}{(\lambda+1)T}[(\lambda\Gamma^*_{FF}+F'F)-(\lambda\Gamma^*_{FF_P}+F'F_P)(\lambda\Gamma^*_{F_PF_P}+F_P'F_P)^{-1}(\lambda\Gamma^*_{F_PF}+F_P'F)]$$

where the definitions of the sample moments $\Gamma_{FF} = F'F$ and $\Gamma_{F_PF_P} = F_P'F_P$ are analogous to their equivalents implied by the DSGE model. That is

$$F_P = \begin{pmatrix} F_p' & F_{p-1}' & \dots & F_1' \\ F_{p+1}' & F_p' & \dots & F_2' \\ \vdots & \vdots & \ddots & \vdots \\ F_{T-1}' & F_{T-2}' & \dots & F_{T-p}' \end{pmatrix}$$

and

$$F = \begin{pmatrix} F_{p+1}' \\ F_{p+2}' \\ \vdots \\ F_T' \end{pmatrix}$$

Step 3.1: Drawing from $\mathbf{p}(\theta \mid \mathcal{F})$ The distribution depends on prior knowledge on specific parameters in the model. Usually, there is no way to obtain a standard posterior distribution for θ. A standard way to draw from a non-standard distribution is a Random Walk Metropolis-Hasting (MH) Algorithm. Given a draw of θ^{j-1} from the previous step, a candidate θ^* is drawn from a proposal distribution:

$$\theta^* = \theta^{j-1} + \epsilon_i$$

Then, the following ratio is calculated:

$$r = \frac{p(F_j \mid \theta^*)p(\theta^*)}{p(F_j \mid \theta^{j-1})p(\theta^{j-1})}$$

We set $\theta^j = \theta^*$ with probability r. If the proposal is rejected, we set $\theta^j = \theta^{j-1}$. The intuition is that we draw from a candidate from an arbitrary distribution and reweigh the draws such that we draw from the desired distribution. For a exposition of MH algorithms and MH within Gibbs algorithms see Geweke (2005). Prerequisite is that the Likelihood can be evaluated for a given θ. The relevant Likelihood is

$$p(F \mid \theta) \propto \frac{|\Gamma^*_{F_pF_p}(\theta) + \Gamma_{F_pF_p}|^{-\frac{M}{2}} |(\lambda+1)T\tilde{\Sigma}(\theta)|^{-\frac{(\lambda+1)T-k}{2}}}{|\Gamma^*_{F_pF_p}(\theta)|^{-\frac{M}{2}} |\lambda T \Sigma^*(\theta)|^{-\frac{\lambda T-k}{2}}}$$

Step 3.2: Drawing from $\mathbf{p}(\boldsymbol{\Phi}, \boldsymbol{\Sigma} \mid \theta, \mathcal{F})$ The prior distribution of Φ and Σ given θ is of the Inverted-Wishart-Normal form:[11]

$$\Sigma|\theta \sim IW(\Sigma^*(\theta), T^* - Np - 1)$$

$$\Theta|\Sigma, \theta \sim N(\Phi^*(\theta), \Sigma \otimes \Gamma_{F_PF_P}(\theta)^{-1})$$

Note that the distributions are centered at the MLE of Φ and Σ on the artificial sample. It follows that

$$\Sigma \mid \theta \sim IW\left(\tilde{\Sigma}(\theta), (1+\lambda)T - Np - 1\right) \tag{4}$$

$$\Phi \mid \Sigma, \theta \sim N\left(\tilde{\Phi}(\theta), \Sigma \otimes (\Gamma^*_{F_PF_P}(\theta) + \Gamma_{F_PF_P})^{-1}\right) \tag{5}$$

The posterior distribution is of the same form as the prior, but it is centered at the MLE on both actual and artificial data. To get an intuition for the result, it is illustrative to decompose the posterior distribution into the likelihood function and the prior distribution:

$$p(\Phi, \Sigma, \theta \mid \mathcal{F}) = \frac{p(\mathcal{F} \mid \Phi, \Sigma, \theta)p(\Phi, \Sigma, \theta)}{p(\mathcal{F})} \propto p(\mathcal{F} \mid \Phi, \Sigma, \theta)p(\Phi, \Sigma, \theta)$$

The likelihood is

$$p(F_t \mid \Phi, \Sigma, \theta) \propto |\Sigma|^{-T/2} \exp\left(-\frac{1}{2}tr(\Sigma^{-1}(F'F - \Phi'F_P'F - F'F_P\Phi + \Phi'F_P'F_P\Phi))\right)$$

Replacing the sample moments as defined above yields

$$p(F_t \mid \Phi, \Sigma, \theta) = |\Sigma|^{-T/2} \exp\left(-\frac{1}{2}tr(\Sigma^{-1}(\Gamma_{FF} - \Phi'\Gamma_{FF_P} - \Gamma'_{FF_P}\Phi + \Phi'\Gamma_{F_PF_P}\Phi))\right)$$

[11] This is a slight abuse of notation: We should vectorize the matrices Θ and Σ.

The density function of the Inverted-Wishart-Normal distribution is

$$
\begin{aligned}
p(\Phi, \Sigma \mid \theta) &= c(\theta)^{-1}|\Sigma|^{-\frac{\lambda T+n+1}{2}} \exp\left(-\frac{1}{2}tr(\Sigma^{-1}(\Gamma^*(\theta) - \Phi'\Gamma^*(\theta) - \Gamma^*(\theta)'\Phi + \Phi'\Gamma^*(\theta)'\Phi))\right) \\
&= p(\Phi, \Sigma \mid \Gamma(\theta))
\end{aligned}
$$

Note that Bayes' Theorem gives

$$
p(\Phi, \Sigma \mid \Gamma(\theta)) \quad \propto \quad p(\Gamma(\theta) \mid \Phi, \Sigma)p(\Phi, \Sigma)
$$

Comparing this to the likelihood $p(F \mid \Phi, \Sigma, \theta)$ we see that the prior distribution of Σ and Φ given θ can be interpreted as augmenting the data set with dummy observations F^* by multiplying the likelihood of 'dummy observations'

$$
p(F^* \mid \theta) \propto |\Sigma|^{-\frac{\lambda T+n+1}{2}} \exp\left(-\frac{1}{2}tr(\Sigma^{-1}(\Gamma^*(\theta) - \Phi'\Gamma^*(\theta) - \Gamma^*(\theta)'\Phi + \Phi'\Gamma^*(\theta)'\Phi))\right)
$$

with an (improper) prior

$$
p(\Phi, \Sigma) = \Sigma^{-\frac{n+1}{2}}
$$

The 'sample size' of the artificial sample is λT, therefore λ is a parameter which reflects the 'tightness' of the DSGE model prior,. The larger λ, the larger the sample compared to the actual sample. If λ is large, the estimates of Φ and Σ will concentrate on the restrictions implied by the DSGE model. Tedious manipulations of

$$
p(\Phi, \Sigma \mid \theta, \mathcal{F}) \propto p(F \mid \Phi, \Sigma)p(\Phi, \Sigma \mid \theta)
$$

show that Φ and Σ given θ and $\mathcal{F}$ are of the Inverted Wishart-Normal form stated above.

To summarize, Step 3 of the Gibbs sampler works as follows: First draw θ^*, accept or reject these draw according to the rule described above. Given the resulting θ_j, use the distributions (4) and (5) to draw Φ_j and Σ_j, respectively. Given this draws, we start a new iteration with step 1.

5 Empirical Application

This section describes the results when we apply the prior from the New Keynesian model to a Dynamic Factor Model. The tightness M_0 of the prior for Λ determines how close the data series are connected a priori to the DSGE model concepts. The value of λ determines the weight of the DSGE model in the estimation. As it is not clear a priori what values should be chosen, we estimate the model over a grid of values for M_0 and λ. We provide some evidence on a optimal weight based on the forecast performance. We also discuss the selection of the optimal weight λ and M_0 based on measures of in-sample fit. In particular, we decompose the variance of the data into the fraction explained by the common factors and the variance of the idiosyncratic component. Additionally, we provide the posterior marginal data density as a selection criterion. Section 5.5 evaluates the estimates of the DSGE model parameters.

We proceed as follows: In section 5.1 we describe the data. Section 5.2 addresses some issues concerning the concrete implementation of the MCMC algorithm. The choice of the prior distribution of the DSGE model parameters is discussed. Section 5.3 discusses the forecast performance. In Section 5.4 we provide the discussion of the optimal weights based on measures of in-sample fit. Section 5.6 discusses how identified monetary shocks influence the common factors and the observed series[12].

5.1 Data

We use quarterly data from 1985 to 2007:3. We do not use data from periods earlier than 1985 because there is evidence for structural break at around 1984 (see e.g. Stock and Watson (2002a)). The data is taken from Federal Reserve Bank of St. Louis Data base and from the Bureau of Labor Statistics. We select data corresponding to the variables contained in the DSGE model: Output, Prices and Interest Rates. The output series include data on real personal income, consumption expenditures, domestic product, industrial production and capacity utilization. Prices indicators are deflators of GDP and consumption expenditures, and consumer prices indexes for several subgroups of goods. Interest rates include bonds with different ratings, Treasury bonds and the FED funds rate. If there was only monthly data available, we took averages to obtain a quarterly series. A complete list with detailed information is given in Table 2.

[12]The calculations are done with our own MATLAB routines. To solve the DSGE model, we adapted the MATLAB code written by Christopher Sims. Gauss routines written by Frank Schorfheide were used to test our code.

A central issue is how the economic concepts contained in the factors relate to the variables in the DSGE model. We adapt the approach taken by DelNegro and Schorheide (2004). They use output growth, inflation and annualized interest rates in levels for the estimation. The following 'observation equation' - which does not correspond to the observation equation (1 - is therefore specified ('obs' refers to the observed series):

$$\Delta y_{t,obs} = \ln \gamma + \Delta y_t + z_t$$

$$\pi_{t,obs} = \ln \pi^* + \pi_t^*$$

$$r_{t,obs} = 4(\ln r^* + \ln \pi^* + r_t)$$

Hence, we also take the growth rate of the price series to measure inflation, the growth rate of the output series, and the interest rate series in levels for our estimation. We adapt these equations by replacing Δy_t, π_t and r_t by their corresponding factors. We do not use annualized interest rates as this introduces undesirable heteroscedasticity in the data and adjust the observation equation accordingly. The series are demeaned, which implies that we omit the constants contained in the equations above.

One further issue is, that in particular in classical analysis of factor models, there is a large and still developing literature of statistical tests to determine the number of factors. We do not attempt to do a methodically sound analysis of our data set in that respect: In our factor model, the number of factor is determined by the number of shocks in the DSGE model. However, the eigenvalues of the covariance matrix of the data provides some indication that three factors are not at odds with the data. We have three eigenvalues which are distinctively different from zero: 78.5, 5.9 and 3.6. The next smaller values are 0.9, 0.8 and 0.5 which are much closer to zero. This is indicative because the number of factors corresponds to the number eigenvalues which go to infinity with increasing cross-sectional dimension.

5.2 Implementation

The prior distribution for θ is taken from DelNegro and Schorheide (2004). Parameters are assumed to be independently distributed according to Table 1. We do not attempt to estimate the steady state values for the interest rate and therefore calibrate $\frac{r^*}{\gamma} = \beta = 0.99$.

We assume the same prior for the coefficients in the observation equation for each series (see section 4). However, we standardize the variance of the series to the stan-

Table 1: Prior Distribution

Parameter	Distribution	Mean	Std.Deviation
ψ_1	gamma	1.5	0.5
ψ_2	gamma	0.125	0.1
ρ_r	beta	0.5	0.2
κ	gamma	0.3	0.15
τ	gamma	2	0.5
ρ_g	beta	0.8	0.1
ρ_z	beta	0.3	0.1
σ_R	inverse gamma-1	0.251	0.139
σ_g	inverse gamma-1	0.630	0.323
σ_z	inverse gamma-1	0.875	0.430

[1] The inverse gamma-1 density is parametrized as in DelNegro and Schorheide (2004): $p(\sigma|\nu, s) \propto \sigma^{-\nu-1}e^{-\nu s^2/2\sigma^2}$ where $\nu = 4$ and s equals 0.2, 0.5 and 0.7, respectively.

[2] Following DelNegro and Schorheide (2004), we truncate the prior density such that the parameter space is restricted to the determinacy region (corresponding to approximately 98.5 % of the prior mass as defined above).

dard deviation of one particular series in the sample: In particular, we standardize all 'output series' to have the same standard deviation as GDP. For the 'price series' we use the GDP deflator and for the 'interest rate series' we use the FED Funds Rate as normalizing series. This makes the estimation more robust to the influence of data series with large variance.

The variance of the innovations for the proposal draws in the MH-algorithm is chosen in order to get an acceptance rate between 0.2 and 0.3. We iterate 200'000 times over Step 1 to 3 described before. To mitigate the effect of the initial values, we discard the first 20% of the draws. For computational reasons we evaluate only every 16st draw, such that we have 10'000 draws to calculate the distribution of the parameters. Convergence is checked by using different initial values and graphically verifying that the recursive means remain stable after the discarded draws.[13]

The number of lags p in the state equation is 4. In the benchmark model, we replace the DSGE prior with a Minnesota prior. The Minnesota prior is implemented with dummy observations as described in the appendix of Lubik and Schorfheide (2005). The lag length of the AR model is also chosen to be 4.

[13]The only coefficient where the recursive mean are only stable after approximately 100'000 is τ.

5.3 Forecast Performance

We build 'rolling forecasts' for the last eight years of the sample. This yields 32 one-period forecasts, 31 two-periods forecasts, etc for each variable. Due to the heavy computational burden we did not reestimate the model for each sample. We evaluate the forecast performance up to a horizon of two years for a grid of values for λ and M_0^{-1}. For each λ and M_0^{-1} forecast horizon h we calculate the covariance matrix of the errors as

$$\Sigma_{forecast}(\lambda, M_0^{-1}, h) = \frac{1}{32-h} \sum_{t=2000:4}^{2007:3} (X_{t+h} - \mathbb{P}_t X_{t+h})(X_{t+h} - \mathbb{P}_t X_{t+h})'$$

where X_t are observed series. Note that if X_t contains only one series, the square root of $\Sigma_{forecast}(\lambda, M_0^{-1}, h)$ corresponds to the root mean squared error (RMSE). Following DelNegro, Schorheide, Smets, and Wouters (2007), we calculate a multivariate statistic for the forecast performance as the inverse of this matrix, divided by 2 to convert from variance to standard error and by the number of variables to obtain an average figure. The percentage improvement in the multivariate statistic across different models is computed by taking the difference multiplied by 100.
We use the mean of the posterior distribution for forecasting. We applied the following algorithm: In each iteration of the MCMC algorithm, forecast the future states given the draws of the states. Then use the draw of Λ to calculate the forecasts for each variable. This results in one forecast for each variable at any given forecast horizon for each draw. So the whole distribution of these linear forecasts can be evaluated. Under a quadratic loss function, the mean of this distribution is the optimal forecast. Hence

$$\mathbb{P}_t X_{t+h} = \frac{1}{J} \sum_{j=1}^{J} \Lambda_j \mathbb{P}_t S_{t+h,j} + \Psi_j^h (X_t - \Lambda_j S_{t,j})$$

$$\mathbb{P}_t S_{t+h,j} = \Phi_j^h S_{t,j}$$

Note that the forecast with the mean of the estimates is not equal to the mean of the distribution of forecasts. An extension would be the standard practice to use a two step procedure for forecasting with Dynamic Factor Models: In a first step, the factors (sometimes called 'diffusion indices') are estimated. In the second step, the variables of interest are regressed on the factors and on their own lags. The resulting equation is used for forecasts. This procedure potentially improves the forecasts for all weights of the prior. But as we are mainly interested in relative performance across different priors, we do not follow this approach.

RMSE for selected series Table 3 shows the RMSE for selected series for different forecast horizons and λ's. For most of the series, the factor model outperforms the AR(4) forecast substantially. Generally, the forecast error is minimized for moderately positive values of λ. The optimal λ depends on the series and also on the forecast horizon. For some series, the gain is quite big with respect to the estimation with zero weight.

Multivariate statistic In Table 4 to 7, the multivariate statistics for different groups of series are given. The multivariate statistics confirm the result that the factor model shows a superior performance compared to the univariate AR(4) forecasts. Again, the optimal value for λ is positive, but small. Increasing the weight to large values results in a worse performance. For the tightness of the prior in the observation equation, the results are ambiguous: For output series, a tighter link increases the gain, for prices there is no effect and for interest right a looser prior is to be favoured. However, the differences are small. Comparing the forecast performance to the factor model with a Minnesota prior, one can see that the Minnesota prior also performs well. However, there is still a gain from using the DSGE prior for the output series. Only for the one period forecast of prices, the Minnesota prior performs better.
To summarize, including prior information improves the forecast performance considerably for most of the variables at all forecast horizons compared to the simple AR forecast. Note that there is also a gain over a horizon of two years. However, the longer the horizon, the less reliable are the figures as the number of periods used for the evaluation decreases. The optimal weight is clearly positive, but small: The values hoover around a value of 1. Also the Minnesota prior performs considerably better than the AR forecasts. Therefore, the gain of the DSGE prior compared to this model is small.

5.4 Selection of weights of prior

The previous analysis focused on the out-of-sample forecasting performance of the model for different prior weights. In this section, we provide information on in-sample fit. First, we informally use the fraction of the data which is explained by the common factors. Then, a formal assessment based on the posterior probabilities is provided.

5.4.1 Variation explained by common factors

The ratio of standard deviation of ΛF_t, the common component of the observed series and the standard deviation of the actual data is informally used as a measure for the explanatory power of the DSGE model for the specific series. Table 11 reveals that the fraction of the variance that is explained by the factors is quite large. The fraction is only slightly influenced by the weight of the DSGE prior. That is, the DSGE prior is not restricting the dynamics of the factors in way that influences the fit of the model. This indicates that the distribution of the factors are mainly determined by the information in the observation equation. The law of motion of the factors has only small effects on the fit.

Comparing the ratios in Table 11 for different values of M_0, we see that for some variables, increasing the weight decreases the 'fit', while it increases for other series. Strikingly, the estimation with a low weight of the DSGE prior and a very loose prior on the observed coefficients detoriate the fit of the interest rate series, without increasing the fit of other series. This indicates that some 'curvature' in the prior density is needed to identify the factors. It is optimal to have a moderately tight prior on the factor loadings in that respect.

5.4.2 Posterior probabilities

In this section, a formal assessment of the in-sample fit is provided: We index each model by its values for the weight of the DSGE prior λ and the the tightness of the prior distribution of the factor loadings M_0, and denote the respective models by $\mathcal{M}_{\lambda,M_0}$. We then calculate the posterior probabilities of each model:

$$p(\mathcal{M}_{\lambda,M_0}|X) = \frac{p(\mathcal{M}_{\lambda,M_0})p(X|\mathcal{M}_{\lambda,M_0})}{p(X)}$$

To compare the different models, we put equal prior weight for the each model

$$p(\mathcal{M}_{\lambda,M_0}) = \frac{1}{\#(models)}$$

Hence, in relative terms, only the posterior marginal data density is used as a measure of fit:

$$p(X|\mathcal{M}_{\lambda,M_0}) = \frac{p(X|\Theta,\mathcal{M}_{\lambda,M_0})p(\Theta|\mathcal{M}_{\lambda,M_0})}{p(\Theta|X,\mathcal{M}_{\lambda,M_0})}$$

The selection criteria favors models with a high likelihood, but imposes a penalty on too loose priors. It works in the same way as the usual information criteria used in standard time series analysis to select the lag length in autoregressions. Indeed, the Schwarz or Bayesian information criterion is derived on a approximation of the posterior data density[14].

The density $p(X|\mathcal{M}_{\lambda,M_0})$ cannot be calculated analytically. However, the harmonic mean estimator of Geweke (1999) can readily be applied. This estimator is based on the following identity:

$$\frac{1}{p(X)} = \int \frac{f(\Theta)}{p(X|\Theta)p(\Theta)} p(\Theta|X) d\Theta$$

where $\int f(\Theta)d\Theta = 1$. This expression is estimated with

$$\frac{1}{p(X)} = \frac{1}{J} \sum_{j=1}^{J} \frac{f(\Theta_j)}{p(X|\Theta_j)p(\Theta_j)}$$

In principle, any function $f(\Theta)$ which integrates to one can be used. A standard choice is

$$f(\Theta) = q^{-1}(2\pi)^{-d/2}|V_\Theta|^{-\frac{1}{2}} e^{-\frac{1}{2}(\Theta-\bar{\Theta})V_\Theta^{-1}(\Theta-\bar{\Theta})} \times I\left[(\Theta-\bar{\Theta})V_\Theta^{-1}(\Theta-\bar{\Theta}) < F^{-1}_{\chi^2_d(q)}\right]$$

$\bar{\Theta}$ refers to the posterior mean and V_Θ is the posterior variance of the draws. The parameter q is deliberately chosen to dampen the effect of extreme draws out of the posterior density. One word of caution maybe necessary at this point: In theory the value of q has no influence on the estimated value of the marginal data density. In practice, the estimation depends to some extend on the value of q due to the finite number of draws. We therefore calculate the $p(X)$ with $q = (0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95)$. We interpret the discrepency between the values of $p(X)$ for the different q's as numerical inaccuracy.

In Table 12 we see the results for different weights of the DSGE prior λ and different values of M_0^{-1}, the tightness of the prior on the factor loadings. It can be seen that there are considerable differences for different values of q, which makes it difficult to order all the models. A robust result, with respect to the choice of q is that for

[14]With a quadratic loss function, we would use the posterior probabilities to calculate a weighted average of statistics of interests implied by the different models. However, it turns out that the model with highest posterior receives a weight of almost one, while the other models receive no weight. Hence, to consider the model with the highest posterior data density does not lead to different conclusion, unless an extreme position on the prior model probabilities $p(\mathcal{M}_{\lambda,M_0})$ is taken.

large values of λ, the fit detoriates. It is maximized for small values, in most cases for $\lambda = 0.25$ or $\lambda = 0.5$. The optimal λ tends to be smaller for small values of M_0. This is intuitive: The more plausible the interpretation of the factors as the selected economic variables is, the better the DSGE model describes the factors. It is hard to tell what the optimal M_0 is. All the differences are well in the range of the numerical inaccuracy. A tendence maybe that for a very tight prior, that is large values of M_0, the fit is slightly worse. Recalling that the criterion is generally penalizing too loose priors, we interpret this as evidence that we should not strictly impose the restrictions on the factor loadings matrix. However, inducing some information on what the series actually measure does not harm the fit and may help to get more precise estimates.

5.5 Parameter estimates

Before studying effects of shocks on factors and observed variables, we summarize findings on the posterior distribution of the estimated parameters.

Parameters of DSGE model The posterior mean of θ is shown in Table 15. While the width of the intervals decreases with increasing prior weight, the location of the intervals of most of the coefficients is only slightly influenced. An exception may be the reaction of the central bank to an increase in inflation. This coefficient increases with increasing weight. The intertemporal elasticity of substitution τ^{-1} is slightly higher for moderate values of λ compared to little prior weight or very large prior weight.

Relationship of factors and observed series We inspect the properties of the model with respect to nine variables, three for each group of variables: Personal Income, total Industrial Production and GDP (as measures for output), the GDP Deflator, the implicit Deflator on Personal Consumption and the Consumer Price Index for all goods (as measures for inflation) and the Federal Funds Rate, 3-Month Treasury US Bills and Moody's Aaa Corporate Bond Yields (as measures for the interest rate). Additionally, we compare the posterior distribution of the DSGE model parameter vector θ over the grid of λ.

The estimates vary only slightly by changing the weight of the prior. Table 13 shows the estimated factor loadings for $\lambda = 1$. It is indeed the case, that the series load more on the factor which is economically related. Especially the interest rate and price series are explained almost entirely by only one factor. Moreover, the loading is centered around one. This is also the case for Industrial Output for moderately tight priors on the factor loadings. For the other output series the result is not as clear cut: GDP loads positively on the output factor, as expected. However, the loading on the price factor is distinctively negative. Real Personal Income is negatively influence by the price factor. Increasing the tightness off the prior on the factor loadings shifts the coefficients towards the mean: For $M_0 = 16$, the loadings are very close to the prior mean.

5.6 Transmission of Shocks

We investigate two questions concerning the transmission of monetary shocks. First, we analyze how these shocks influence the common factors. Having identified the factors as economic variables, we are able to compare the responses to typical results in the literature. Second, the model allows to investigate how structural shocks influence observed series. The response is obviously driven by the response of the common factors. However, deviations of the posterior distribution of the factor loadings from the prior distribution induces more complicated dynamics. We use two different identification schemes. The first, agnostic identification scheme relies on the sign restrictions. The second follows DelNegro and Schorheide (2004) in that we rotate the Cholesky decomposition of the covariance matrix of the residuals e_t with corresponding matrix implied by the DSGE model. In the next section, we describe the two methods. We then document the results for different weights of the DSGE prior. For the tightness of the prior for the factor loadings, we choose $M_0 = \frac{1}{2}$ according to the consideration in the previous section.

5.6.1 Identification of Shocks

Recall that the residuals in the state equation relate to structural shocks as

$$e_t = H_{VAR}\varepsilon_t$$

with $\mathbb{E}(\varepsilon_t\varepsilon_t') = \mathcal{I}_M$. The goal is to estimate the reaction of the factors F_t to shocks ε_t:

$$H_{VAR} = \frac{\partial F_{t+h}}{\partial \varepsilon_t'}$$

Given the responses on impact $\frac{\partial F_t}{\partial \varepsilon_t'}$, one can use $\Phi(L)$ to calculate the responses for $h > 0$. Once this reaction is determined, one can calculate the response of X_t as follows:

$$\frac{\partial X_{t+h}}{\partial \varepsilon_t'} = \Lambda \frac{\partial F_{t+h}}{\partial \varepsilon_t'}$$

The problem of identification arises because H_{VAR} can not be uniquely determined using only information from the reduced form estimation of the factor model. H_{VAR} is only restricted by its relationship to the covariance matrix of the reduced form residuals:

$$\Sigma = H_{VAR}\mathbb{E}(\varepsilon_t \varepsilon_t')H_{VAR}' = H_{VAR}H_{VAR}'$$

It is always possible to plug an orthonormal matrix Ω into the equation above:

$$\Sigma = H_{VAR}\Omega\Omega' H_{VAR}'$$

and define $\tilde{H} = H_{VAR}\Omega$. This matrix also satisfies the restrictions implied by the reduced form estimation. However, it implies potentially very different reactions of F_t to the shocks. Hence, given an arbitrary $\tilde{H}$, there have to be further restrictions on Ω in order to determine the responses H_{VAR}. The two approaches we use differ in the way Ω is chosen.

DSGE Model Rotation DelNegro and Schorheide (2004) propose an approach which relies on the fact that in the DSGE model, the shock are exactly identified. That is, the matrix

$$\frac{\partial \tilde{F}_t}{\partial \varepsilon_t'} = \frac{\partial Z S_t}{\partial \varepsilon_t'} = H(\theta)$$

is uniquely determined. Recall that $H(\theta)$ can be calculated using standard methods to solve linear(ized) DSGE models. Furthermore, there is a unique decomposition of this matrix into the product of a triangular matrix $H_{tr,DSGE}(\theta)$ and an orthonormal matrix $\Omega(\theta)$:

$$H(\theta) = H_{tr,DSGE}(\theta)\Omega(\theta)$$

The idea is to set $\tilde{H}$ to $H_{tr,VAR}$, the Cholesky decomposition of Σ, and then to use $\Omega(\theta)$ as a rotation:

$$H_{VAR}(\theta) = H_{tr,VAR}\Omega(\theta)$$

On impact, the response differ to the extend that $H_{tr,DSGE}(\theta)$ and $H_{tr,VAR}$ differ. That is, if the covariance matrix of residuals is similar to its counterpart in the DSGE model, then the responses on impact will be close. For horizons bigger than zero, there is the influence of Φ which allows for further deviations of the factor model responses to the DSGE model implications.

Sign Restrictions The idea of the second approach is to be 'agnostic': One tries to find restrictions on the sign of the response which are consistent with commonly accepted theories. Depending on the nature of this restrictions, it is possible to reduce the range of possible rotations Ω. This idea goes back to Faust (1998) and has been elaborated by Uhlig (2005) and Canova (2002). We implement the 'pure' sign restriction approach, as opposed to the 'penalty-function approach'. Hence, we do use an additional to select the 'best' of all impulse response vectors. All impulse responses satisfying the sign restrictions are considered to be equally likely. In the 'pure' sign restriction approach, one estimates the impulse-responses and the reduced form coefficients jointly. The impulse responses are parameterized as follows: $H_{sign} = H_{chol}\Omega_{\alpha}$, where Ω_{α} is the orthonormal rotation matrix with one column given by a vector of unit length α. H_{chol} is the Cholesky decomposition of the covariance matrix $\Sigma = H_{chol}H'_{chol}$. Uhlig (2005) shows that the set of impulse response functions can be characterized by a suitable choice of α. The prior for the coefficients in the state equation is formulated as

$$p(\Phi, \Sigma, \alpha) \propto p(\Phi, \Sigma)I(\alpha)$$

where $I(\alpha)$ is one if the sign restrictions are satisfied and zero otherwise. This can be implemented by repeating the following steps for each iteration of the MCMC algorithm:

1. Calculate the Cholesky decomposition of the $\Sigma_j = A_j A'_j$
2. Draw a 3×1-vector $\tilde{\alpha}_j$ with independently and standard normally distributed elements and normalize the length to one: $\alpha_j = \frac{\tilde{\alpha}_j}{||\tilde{\alpha}_j||}$
3. Calculate $a_j = A_j\alpha_j$ and check if implied responses satisfy the sign restriction.
4. If it satisfies the restrictions keep all the parameters drawn in iteration j. If it does not satisfy the restrictions, discard all of the draws.

Note that the posterior distribution of the reduced form coefficients is different from the pure reduced form estimation. Draws for which it is more likely that the sign restrictions are satisfied receive more weight.

In our setting, it is possible to restrict the response of the factors or the response of a set of observed series (or both). We restrict the response of the interest rate factor and the inflation factor, and not the observed series directly. This results in a more robust identification with respect to unimportant idiosyncratic noise. Moreover, it can also be justified by the fact that reaction of observed interest rates and observed inflation are very well described by the reaction of the corresponding common factors: The factor loadings are close to the block-diagonal structure. To the extend that this is true, our approach is meaningful, even if one is not convinced that the factor indeed correspond to the variable in the DSGE model: One can interpret the rotation as a technical device to approximately implement sign restrictions on the response of a whole set of observed variables. To identify the shocks, we impose that a contractionary monetary policy shock does not lead to

- an increase in the price factor
- an increase in the interest rate factor

The restrictions are required to hold for the first 5 period, following Uhlig (2005).

5.6.2 Response of Common Factors

The different identification approaches do not lead to conflicting results: The 80% HPD overlap to a large extend (see Figures 3 to 6). The intervals of the response on impact implied by the DSGE model rotation are more narrow than the intervals implied by the sign restriction. This is especially true for the reaction of the output factor. In that sense, including more information leads to more precise results.
Generally, we find that the response of the factors resembles those from standard VAR analysis: A contractionary monetary shock decreases interest rates (which is assumed in sign restriction identification), decreases inflation and has a negative impact on output growth. The responses on impact in the factor model are close to the responses in the DSGE model. In some dimensions however, there are pronounced differences between the implications of the DSGE model and the factor model: One striking observation is that the reaction of interest rates to a monetary shock is far more persistent in the factor model than in the DSGE model. This is even true for a fairly large weight of the DSGE prior. Figure 7 reveals a further major difference regarding the long-run response of the level of output: Whereas the DSGE model predicts long run neutrality of money shocks, the corresponding response in the factor model does not allow to make precise statements on this. The 80% HPD intervals are quite large. Hence, there seems to be seems to be considerable uncertainty on this effect.

5.6.3 Response of Observed Series

The responses of prices and interest rates are very well described by the responses of the common factors, see Figure 8 and 9. This is directly implied by the fact that the posterior distribution of the factor loading matrix is close to its prior mean. The same is true for Industrial Production. This implies that Industrial production reacts in line with the standard view: The response of output to a contractionary monetary policy shock is negative. For other output series, there are considerable deviations (Figure 10): Personal Income shows no reaction and Consumption Expenditures even reacts slightly positive to a contractionary monetary policy shock. Importantly, we confirm Uhlig (2005)'s result that the reaction of GDP to a monetary policy shocks is ambiguous. We want to stress, that this result is also achieved with an agnostic identification approach, not relying on the interpretation of the factors. However, the DSGE prior contributes to tighten the dispersion of the posterior distribution.

6 Conclusion

Dynamic Factor Models are powerful tools to handle large data sets. So far, the theory and also the applications focused mainly on finding statistically meaningful representations of the data. Inspired by the work of Boivin and Giannoni (2006) and DelNegro and Schorheide (2004), we proposed a method to relate the statistical model to economic theory, without fully imposing the theoretical restrictions. Our model continuously bridges the gap between a non-structural factor model and their model in the following sense: In the extreme case of degenerate priors on some of the factor loadings and by strictly imposing the restriction of the DSGE model one estimates the VAR approximation to DSGE model where data is measured with error. By relaxing restrictions implied by the DSGE model and making the priors for the factor loadings less informative, it is possible to move towards a non-structural factor model.

We illustrated our method with an application on US data. A very simple New Keynesian model proved to be useful to reduce the forecast errors for all horizons. Intermediate values for the prior weight perform markedely better than simple AR(4) forecasts. Taking the restrictions from such a simple DSGE model too literally worsens the forecast performance for most of the series. Compared to a Minnesota prior, the magnitude of the gain shrinks considerably.

To evaluate the in-sample fit, we calculate the posterior marginal data density and informally use the fraction of the variation in the observed series explained by the common factors to analyze selected observed series. The fraction of variation explained by the common factors does not deteriorate when the prior information from the DSGE model is included for most of the series. This indicates that the restrictions implied by the DSGE model are supported by the data, once the idiosyncratic components are filtered out. This finding is not supported by the measures of the posterior marginal data densities, however: According to this criterion, a very loose prior is optimal. For the tightness of the prior in the observation equation, the results are ambiguous due to the numerical inaccuracy of the data density estimate.

Analyzing the estimated observation equation reveals that inflation and interest rate series are closely related to their corresponding factors even for a very loose prior on the factor loadings. Observed output series other than industrial production load exclusively on the output factor only if the tightness is increased to more restrictive values.

We finally evaluated the response of the common factor as well as the observed variables to an identified monetary policy shock. Identification is achieved using an agnostic sign restriction approach and a method which more close relates to the DSGE model implied restrictions. Both approaches lead to similar conclusions: The common factors' reaction to a monetary policy shock resembles the predictions from standard theory. However, various measures of output show different reactions to a monetary shock. Industrial production reacts negatively to a contractionary shock, while the response of GDP close to zero. This result is consistent with finding of Uhlig (2005) who finds ambiguous effects of a monetary shock on GDP. We add the evidence that this is not true for all measures of output. Industrial production shows a reaction which is consistent with the standard result that output reacts negatively to a contractionary monetary policy shock.

References

ALTUG, S. (1989): "Time-to-Build and Aggregate Fluctuations: Some New Evidence," *International Economic Review*, 30, 889–920.

ANDERSON, B. D. O., AND J. B. MOORE (1979): *Optimal Filtering*. Prentice Hall.

BAUWENS, L., M. LUBRANO, AND J.-F. RICHARD (1999): *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press.

BOIVIN, J., AND M. GIANNONI (2006): "DSGE Models in a Data-Rich Environment," *NBER Working Paper*, (12772).

CANOVA, F. (2002): "Validating Monetary DSGE models through VARs," *CEPR Working Paper*, (3442).

CHIB, S. (1993): "Bayes Regression with Autocorrelated Errors," *Journal of Econometrics*, pp. 275–294.

DELNEGRO, M., AND F. SCHORHEIDE (2004): "Priors from General Equilibrium Models for VARs," *International Economic Review*, 45(2), 643–673.

DELNEGRO, M., F. SCHORHEIDE, F. SMETS, AND R. WOUTERS (2007): "On the Fit and Forecasting Performance of New Keynesian Models," *Journal of Business and Economic Statistics*, 25(2), 123–162.

DOZ, C., D. GIANNONE, AND L. REICHLIN (2006): "A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models," *European Central Bank Working Paper Series*, (674).

FAUST, J. (1998): "The robubstness of identified VAR conclusions about money," *Carnegie Rochester Conference Series in Public Policy*, 49, 207–244.

FORNI, M., M. LIPPI, AND L. REICHLIN (2003): "Opening the Black Box: Structural Factor Models versus Structural VARs," .

GEWEKE, J. (1999): "Using simulation methods for Bayesian econometric models: inference, development and communication," 18, 1–126, Econometric Reviews.

——— (2005): *Contemporary Bayesian Econometrics and Statistics*. John Wiley & Sons, Inc.

GIANNONE, D., AND L. REICHLIN (2006): "Does information help recovering structural shocks from past observations?," *Journal of the European Economic Association*, 4, 455–465.

GIANNONE, D., L. REICHLIN, AND L. SALA (2006): "VARs, common factors and the empirical validation of equilibrium business cycle models," *Journal of Econometrics*, 132, 257–279.

HAMILTON, J. D. (1994): *Time Series Analysis*. Princeton University Press.

IRELAND, P. N. (2004): "A method for taking models to the data," *Journal of Economic Dynamics and Control*, 28, 1205–1226.

JACQUIER, E., M. JOHANNES, AND N. POLSON (2004): "MCMC Maximum Likelihood for Latent State Models," Manuscript, Columbia University.

KIM, C.-J., AND C. R. NELSON (1999): *State-Space Models with Regime Switching*. The MIT Press.

LAWLEY, D. N., AND A. E. MAXWELL (1971): *Factor Analysis as a Statistical Method*. London, Butterworths, 2 edn.

LUBIK, T., AND F. SCHORFHEIDE (2004): "Testing for Indeterminacy: An Application to U.S. Monetary Policy," *American Economic Review*, 94, 190–217.

——— (2005): "A Bayesian Look at New Open Economy Macroeconomics," *NBER Macroeconomics Annual*.

SARGENT, T. J. (1989): "Two models of measurements and the investment accelerator," *Journal of Political Economy*, 97, 251–87.

SIMS, C. A. (2002): "Solving Linear Rational Expectations Models," *Computational Economics*, 20, 1–20.

——— (2005): "Dummy Observation Priors Revisited," .

STOCK, J. H., AND M. W. WATSON (2002a): "Has the Business Cycle Changed and Why," *NBER Macroeconomics Annual*, pp. 159–218.

——— (2002b): "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics*, 20.

——— (2005): "Implications of Dynamic Factor Models for VAR Analysis," .

Uhlig, H. (2005): "What are the effects of monetary policy on output? Results from an agnostic identification procedure," *Journal of Monetary Economics*, 52.

Watson, M. W. (1993): "Measures of Fit for Calibrated Models," *Journal of Political Economy*, 101.

A Tables

Table 2: Data: From St. Louis Fed Economic Data FRED. Some CPI series directly from BLS (indicated). Download: March 7, 2008. First Panel: Output Series, Second Panel: Inflation Series. Third Panel: Interest Rate Series

Number	Code	Description
1	**DPIC96**	Current Real Disposable Personal Income Bil. of Chained 2000 $ Q SAAR 2008-02-28
2	FINSLC96	Current Real Final Sales of Domestic Product, 3 Decimal Bil. of Chained 2000 $ Q SAAR 2008-02-28
3	**GDPC96**	Current Real Gross Domestic Product, 3 Decimal Bil. of Chained 2000 $ Q SAAR 2008-02-28
4	GNPC96	Current Real Gross National Product Bil. of Chained 2000 $ Q SAAR 2007-12-20
5	NICUR	Current National Income Bil. of $ Q SAAR 2008-02-28
6	PCDGCC96	Current Real Personal Consumption Expenditures: Durable Goods Bil. of Chained 2000 $ Q SAAR 2008-02-28
7	PCECC96	Current Real Personal Consumption Expenditures Bil. of Chained 2000 $ Q SAAR 2008-02-28
8	PCESVC96	Current Real Personal Consumption Expenditures: Services Bil. of Chained 2000 $ Q SAAR 2008-02-28
9	PCNDGC96	Current Real Personal Consumption Expenditures: Nondurable Goods Bil. of Chained 2000 $ Q SAAR 2008-02-28
10	**INDPRO**	Current Industrial Production Index Index 2002=100 M SA 2008-02-15
11	IPBUSEQ	Current Industrial Production: Business Equipment Index 2002=100 M SA 2008-02-15
12	IPCONGD	Current Industrial Production: Consumer Goods Index 2002=100 M SA 2008-02-15
13	IPDCONGD	Current Industrial Production: Durable Consumer Goods Index 2002=100 M SA 2008-02-15
14	IPDMAN	Current Industrial Production: Durable Manufacturing (NAICS) Index 2002=100 M SA 2008-02-15
15	IPDMAT	Current Industrial Production: Durable Materials Index 2002=100 M SA 2008-02-15
16	IPFINAL	Current Industrial Production: Final Products (Market Group) Index 2002=100 M SA 2008-02-15
17	IPMAN	Current Industrial Production: Manufacturing (NAICS) Index 2002=100 M SA 2008-02-15
18	IPMAT	Current Industrial Production: Materials Index 2002=100 M SA 2008-02-15
19	IPMINE	Current Industrial Production: Mining Index 2002=100 M SA 2008-02-15
20	IPNCONGD	Current Industrial Production: Nondurable Consumer Goods Index 2002=100 M SA 2008-02-15
21	IPNMAN	Current Industrial Production: Nondurable Manufacturing (NAICS) Index 2002=100 M SA 2008-02-15
22	IPNMAT	Current Industrial Production: Nondurable Materials Index 2002=100 M SA 2008-02-15
23	IPUTIL	Current Industrial Production: Electric and Gas Utilities Index 2002=100 M SA 2008-02-15
24	MCUMFN	Current Capacity Utilization: Manufacturing (NAICS) % of Capacity M SA 2008-02-15
25	TCU	Current Capacity Utilization: Total Industry % of Capacity M SA 2008-02-15
26	GDPCTPI	Current Gross Domestic Product: Chain-type Price Index Index 2000=100, Q SA 2008-02-28
27	**GDPDEF**	Current Gross Domestic Product: Implicit Price Deflator Index 2000=100 Q SA 2008-02-28
28	GNPCTPI	Current Gross National Product: Chain-type Price Index Index 2000=100 Q SA 2007-12-20
29	GNPDEF	Current Gross National Product: Implicit Price Deflator Index 2000=100 Q SA 2007-12-20
30	GPDICTPI	Current Gross Private Domestic Investment: Chain-type Price Index Index 2000=100 Q SA 2008-02-28
31	**JCXFE**	Current Personal Consumption Expenditures: Chain-type Price Index Index 2000=100 Q SA 2008-02-28
32	PCECTPI	Current Personal Consumption Expenditures: Chain-type Price Index Less Food and Energy, Index 2000=100 Q SA 2008-02-28
33	**CPIAUCSL**	Current Consumer Price Index For All Urban Consumers: All Items Index 1982-84=100 M SA 2008-02-20
34	CUSR0000SAA	Seasonally Adjusted, Area:+U.S. city average Item:+Apparel Base Period:++1982-84=100 (BLS)
35	CUSR0000SAF	Seasonally Adjusted, Area:+U.S. city average Item:+Food and beverages Base Period:++1982-84=100 (BLS)
36	CUSR0000SAG	Seasonally Adjusted, Area:+U.S. city average Item:+Other goods and services Base Period:++1982-84=100 (BLS)
37	CUSR0000SAH	Seasonally Adjusted, Area:++U.S. city average Item:+Housing Base Period:++1982-84=100 (BLS)
38	CUSR0000SAM	Seasonally Adjusted, Area:++U.S. city average Item:+Medical care Base Period:++1982-84=100 (BLS)
39	CUSR0000SAT	Seasonally Adjusted, Area:++U.S. city average Item:+Transportation Base Period:++1982-84=100 (BLS)
40	CPIENGSL	Consumer Price Index for All Urban Consumers: Energy Index 1982-84=100 M SA 2008-02-20
41	CPILEGSL	Consumer Price Index for All Urban Consumers: All Items Less Energy Index 1982-84=100 M SA 2008-02-20
42	CPIUFDSL	Consumer Price Index for All Urban Consumers: Food Index 1982-84=100 M SA 2008-02-20
43	CPIULFSL	Consumer Price Index for All Urban Consumers: All Items Less Food Index 1982-84=100 M SA 2008-02-20
44	**AAA**	Current Moody's Seasoned Aaa Corporate Bond Yield % M NA 2008-03-04
45	BAA	Current Moody's Seasoned Baa Corporate Bond Yield % M NA 2008-03-04
46	CD1M	CD1M Current 1-Month Certificate of Deposit: Secondary Market Rate % M NA 2008-03-04
47	CD3M	3-Month Certificate of Deposit: Secondary Market Rate % M NA 2008-03-04
48	CD6M	CD6M Current 6-Month Certificate of Deposit: Secondary Market Rate % M NA 2008-03-04
49	**FEDFUNDS**	Current Effective Federal Funds Rate % M NA 2008-03-04
50	GS1	Current 1-Year Treasury Constant Maturity Rate % M NA 2008-03-04
51	GS10	Current 10-Year Treasury Constant Maturity Rate % M NA 2008-03-04
52	GS2	Current 2-Year Treasury Constant Maturity Rate % M NA 2008-03-04
53	GS3	Current 3-Year Treasury Constant Maturity Rate % M NA 2008-03-04
54	GS3M	Current 3-Month Treasury Constant Maturity Rate % M NA 2008-03-04
55	GS5	Current 5-Year Treasury Constant Maturity Rate % M NA 2008-03-04
56	GS6M	Current 6-Month Treasury Constant Maturity Rate % M NA 2008-03-04
57	GS7	Current 7-Year Treasury Constant Maturity Rate % M NA 2008-03-04
58	M2OWN	Current M2 Own Rate % M NA 2008-03-07
59	MORTG	Current 30-Year Conventional Mortgage Rate % M NA 2008-03-04
60	MPRIME	Current Bank Prime Loan Rate % M NA 2008-03-04
61	MZMOWN	Current MZM Own Rate % M NA 2008-03-07
62	**TB3MS**	Current 3-Month Treasury Bill: Secondary Market Rate % M NA 2008-03-04
63	TB6MS	Current 6-Month Treasury Bill: Secondary Market Rate % M NA 2008-03-04

Table 3: Percentage decrease of the RMSE relative to AR(4) forecast with $M_0 = 0.5$

Observed Series	h/λ	0.25	0.5	1	2	5	100
GDP	1	19.36	21.08	**23.27**	22.18	18.54	14.24
	4	2.89	3.30	**6.01**	4.75	2.94	1.41
	8	4.95	4.36	**4.99**	3.75	2.76	2.40
	12	4.01	3.54	**4.22**	3.30	3.10	3.44
Industrial Production	1	30.91	**32.91**	32.25	30.13	21.66	13.19
	4	0.43	3.03	7.30	**8.33**	7.79	6.90
	8	3.78	3.24	4.20	5.11	5.89	**6.69**
	12	1.95	1.64	2.86	4.40	6.25	**7.50**
Personal Income	1	**64.50**	62.97	63.52	60.98	60.69	60.97
	4	**-4.07**	-4.62	-5.22	-5.33	-5.28	-5.74
	8	0.64	0.52	0.67	0.74	**0.98**	0.96
	12	**3.36**	3.17	3.35	2.82	2.40	2.29
GDP Defl	1	-1.32	-1.25	-1.46	-0.99	**-0.71**	-0.87
	4	-0.78	-0.35	-0.27	0.90	**1.54**	1.25
	8	-2.50	-2.28	-1.70	-0.76	**-0.17**	-0.89
	12	0.88	0.92	1.27	**1.34**	1.11	0.16
Pers Cons Price Defl	1	-0.05	-0.06	0.19	0.43	**0.69**	0.65
	4	-0.60	-0.53	0.54	1.61	2.82	**2.97**
	8	-0.83	-0.71	0.61	1.54	**2.54**	2.40
	12	0.38	0.41	1.53	1.93	**2.50**	2.20
CPI	1	**2.68**	2.50	2.34	2.30	2.16	1.89
	4	-0.72	**-0.59**	-0.74	-0.96	-1.75	-2.50
	8	-0.14	-0.21	**-0.09**	-0.31	-0.94	-2.09
	12	**1.33**	1.17	1.10	0.41	-0.66	-1.93
Fed Funds Rate	1	**14.93**	13.94	12.23	11.07	11.12	11.73
	4	**1.65**	1.11	-1.04	-4.55	-5.59	-5.06
	8	**12.97**	11.88	10.60	7.32	6.67	6.24
	12	**23.16**	22.33	22.46	20.71	20.72	20.01
3-M T-Bill	1	**14.72**	13.85	12.47	11.54	11.75	12.33
	4	**2.77**	2.30	-0.13	-3.53	-4.40	-3.78
	8	**17.63**	16.66	14.70	11.17	10.34	9.91
	12	**32.60**	31.81	31.16	29.10	28.87	28.11
Moody's Aaa Bonds	1	5.87	6.17	**6.71**	6.06	5.71	5.41
	4	-0.93	-0.31	**2.20**	1.66	1.24	0.19
	8	-0.14	0.23	**1.59**	1.04	0.78	0.27
	12	1.74	2.23	**3.62**	3.34	3.25	3.00

Table 4: Multivariate statistic relative to AR(4) forecast: GDP, IP total, Personal Income GDP Defl, Pers Cons Price Defl, CPI, Fed Funds Rate 3-M T-Bill, Moody's Aaa Bond

h/λ	0.25	0.5	1	2	5	100
$M_0^{-1} = \frac{1}{2}$						
1	**9.33**	9.12	8.43	7.57	7.15	6.80
2	3.26	**3.39**	3.11	2.52	2.29	2.14
3	4.37	**4.52**	4.41	3.98	3.76	3.67
4	4.55	**4.60**	4.40	3.83	3.51	3.35
8	**5.99**	5.84	5.72	5.44	5.28	5.17
12	**6.60**	6.49	6.47	6.33	6.05	5.95
$M_0^{-1} = 1$						
1	**9.22**	9.04	8.48	7.62	7.09	6.95
2	2.76	**2.94**	2.90	2.46	2.09	1.94
3	3.78	4.00	**4.09**	3.84	3.59	3.47
4	4.05	**4.15**	4.13	3.74	3.36	3.23
8	**5.74**	5.61	5.49	5.28	5.12	4.97
12	**6.44**	6.36	6.31	6.20	6.06	5.88
$M_0^{-1} = 2$						
1	9.03	**9.03**	8.57	7.82	6.87	6.61
2	2.29	2.70	**2.74**	2.32	1.84	1.70
3	3.29	3.70	**3.84**	3.51	3.33	3.25
4	3.61	3.88	**3.90**	3.46	3.17	3.01
8	**5.53**	5.46	5.31	4.90	4.94	4.82
12	**6.30**	6.27	6.20	6.00	5.96	5.77

Table 5: Multivariate statistic relative to AR(4) forecast: Output Series

h/λ	0.25	0.5	1	2	5	100
$M_0^{-1} = \frac{1}{2}$						
1	5.86	5.87	5.92	**5.96**	5.88	6.06
2	4.84	4.95	4.98	**5.00**	4.92	5.00
3	5.05	5.08	5.14	5.16	5.12	**5.22**
4	3.65	3.64	**3.67**	3.64	3.62	3.63
8	2.61	2.64	2.62	2.63	2.63	**2.65**
12	1.95	1.99	1.96	2.01	1.99	**2.02**
$M_0^{-1} = 1$						
1	6.02	6.02	5.99	**6.08**	6.00	5.92
2	5.02	5.05	5.11	**5.13**	4.99	4.98
3	5.23	5.27	5.30	**5.35**	5.21	5.12
4	3.76	3.79	3.78	**3.81**	3.71	3.65
8	2.71	2.65	2.68	**2.73**	2.70	2.65
12	1.88	1.93	1.90	2.03	1.99	**2.05**
$M_0^{-1} = 2$						
1	6.18	6.20	**6.27**	6.26	6.24	6.20
2	5.16	5.24	5.26	**5.27**	5.22	5.14
3	5.34	5.41	5.45	**5.47**	5.35	5.28
4	3.86	3.88	3.92	**3.93**	3.88	3.79
8	2.74	2.76	2.74	**2.76**	2.72	2.70
12	1.78	1.86	1.91	1.98	**2.02**	1.99

Table 6: Multivariate statistic relative to AR(4) forecast: Prices

h/λ	0.25	0.5	1	2	5	100
$M_0^{-1} = \frac{1}{2}$						
1	1.04	1.14	**1.35**	1.17	1.22	1.21
2	1.59	1.59	1.76	1.70	**1.80**	1.79
3	0.99	1.05	1.20	1.20	**1.25**	1.26
4	0.89	0.91	1.03	1.05	1.09	**1.10**
8	0.50	0.49	0.56	0.61	0.61	**0.65**
12	0.50	0.49	0.57	0.61	0.62	**0.65**
$M_0^{-1} = 1$						
1	1.13	1.13	1.21	**1.26**	1.23	1.08
2	1.56	1.58	1.66	**1.78**	1.77	1.60
3	1.06	1.08	1.13	**1.25**	1.24	1.09
4	0.95	0.95	1.00	**1.09**	1.08	0.93
8	0.51	0.54	0.56	0.62	**0.63**	0.51
12	0.49	0.55	0.55	0.61	**0.62**	0.50
$M_0^{-1} = 2$						
1	1.12	1.26	1.30	**1.35**	1.27	1.22
2	1.60	1.66	1.75	**1.79**	1.75	1.74
3	1.06	1.14	1.22	**1.26**	1.26	1.23
4	0.94	1.02	1.08	**1.08**	1.08	1.04
8	0.54	0.56	0.57	0.62	**0.64**	0.59
12	0.52	0.53	0.54	0.60	**0.64**	0.60

Table 7: Multivariate statistic relative to AR(4) forecast: Interest Rates

h/λ	0.25	0.5	1	2	5	100
$M_0^{-1} = \frac{1}{2}$						
1	2.12	**2.15**	2.09	2.06	1.97	1.94
2	**2.89**	2.85	2.78	2.78	2.70	2.66
3	**3.09**	3.05	2.99	3.02	2.89	2.91
4	3.25	**3.28**	3.18	3.25	3.09	3.08
8	2.92	**2.94**	2.90	2.85	2.72	2.66
12	2.42	**2.43**	2.36	2.22	2.13	2.04
$M_0^{-1} = 1$						
1	2.07	**2.13**	2.11	2.07	1.99	1.95
2	2.80	**2.86**	2.81	2.80	2.70	2.71
3	2.95	**3.01**	3.00	3.01	2.91	2.90
4	3.16	3.19	**3.21**	3.21	3.14	3.11
8	2.86	2.89	2.91	**2.92**	2.82	2.73
12	2.42	**2.44**	2.43	2.34	2.25	2.03
$M_0^{-1} = 2$						
1	2.00	2.00	**2.10**	2.09	1.99	1.95
2	2.69	2.66	2.77	**2.83**	2.73	2.71
3	2.85	2.86	2.92	**3.00**	2.94	2.89
4	3.10	3.11	3.14	**3.24**	3.14	3.09
8	2.89	2.73	2.90	**2.95**	2.79	2.78
12	2.50	2.26	**2.51**	2.49	2.19	2.27

Table 8: Multivariate statistic relative Minnesota prior forecast: $M_0^{-1} = 1$ and $\lambda = 1$

h	Output	Prices	Interest Rates
1	0.56	-0.23	0.01
2	0.30	-0.10	-0.06
3	0.44	-0.08	-0.07
4	0.02	-0.01	-0.07
8	0.16	0.00	-0.01
12	0.45	0.17	0.25

Table 9: Fraction of Variance explained by Factors for $\lambda = 0.25$ (note that values above one are possible because the estimation method does not require the variance of the components to sum up to one

M_0	0.25	0.5	1	2	4	16
GDP	[0.72 0.89]	[0.58 0.79]	[0.58 0.79]	[0.58 0.79]	[0.61 0.80]	[0.66 0.81]
IP total	[0.89 1.03]	[0.87 1.02]	[0.87 1.01]	[0.86 1.00]	[0.85 0.98]	[0.78 0.89]
Personal Income	[0.25 0.46]	[0.21 0.41]	[0.20 0.40]	[0.22 0.41]	[0.25 0.45]	[0.41 0.60]
GDP Deflator	[0.69 0.95]	[0.64 0.91]	[0.65 0.90]	[0.65 0.87]	[0.63 0.83]	[0.61 0.77]
Pers Cons Defl	[0.24 0.62]	[0.36 0.92]	[0.45 0.93]	[0.52 0.93]	[0.57 0.91]	[0.60 0.81]
CPI	[0.65 0.92]	[0.66 0.92]	[0.65 0.90]	[0.64 0.87]	[0.62 0.83]	[0.62 0.78]
Federal Funds Rate	[0.03 0.13]	[0.75 1.09]	[0.79 1.09]	[0.82 1.07]	[0.84 1.05]	[0.89 1.02]
3-month T-Bill	[0.04 0.13]	[0.77 1.10]	[0.81 1.10]	[0.82 1.08]	[0.85 1.06]	[0.90 1.03]
Moody's Aaa Bond	[0.05 0.16]	[0.37 0.83]	[0.52 0.92]	[0.65 0.97]	[0.74 0.99]	[0.86 1.00]

Table 10: Fraction of Variance explained by Factors for $\lambda = 1$ (note that values above one are possible because the estimation method does not require the variance of the components to sum up to one

M_0	0.25	0.5	1	2	4	16
GDP	[0.59 0.80]	[0.59 0.80]	[0.59 0.79]	[0.59 0.79]	[0.61 0.80]	[0.66 0.82]
IP total	[0.87 1.02]	[0.87 1.01]	[0.87 1.01]	[0.86 1.00]	[0.84 0.98]	[0.78 0.89]
Personal Income	[0.22 0.41]	[0.21 0.40]	[0.21 0.40]	[0.21 0.41]	[0.25 0.45]	[0.42 0.61]
GDP Defl	[0.62 0.92]	[0.64 0.91]	[0.65 0.90]	[0.65 0.88]	[0.63 0.83]	[0.62 0.77]
Pers Cons Defl	[0.31 0.91]	[0.37 0.92]	[0.47 0.95]	[0.54 0.94]	[0.59 0.93]	[0.62 0.83]
CPI	[0.65 0.92]	[0.65 0.92]	[0.65 0.90]	[0.64 0.87]	[0.63 0.84]	[0.63 0.79]
Federal Funds Rate	[0.71 1.09]	[0.73 1.08]	[0.75 1.07]	[0.78 1.05]	[0.83 1.04]	[0.88 1.01]
3-month T-Bill	[0.73 1.10]	[0.77 1.10]	[0.78 1.09]	[0.80 1.06]	[0.83 1.04]	[0.88 1.02]
Moody's Aaa Bond	[0.31 0.80]	[0.40 0.84]	[0.52 0.91]	[0.65 0.96]	[0.73 0.98]	[0.86 1.00]

Table 11: Fraction of Variance explained by Factors for $\lambda = 100$ (note that values above one are possible because the estimation method does not require the variance of the components to sum up to one

M_0	0.25	0.5	1	2	4	16
GDP	[0.59 0.80]	[0.58 0.79]	[0.58 0.79]	[0.59 0.80]	[0.61 0.81]	[0.67 0.82]
IP total	[0.84 1.01]	[0.84 1.00]	[0.85 1.00]	[0.85 1.00]	[0.84 0.98]	[0.78 0.89]
Personal Income	[0.21 0.41]	[0.21 0.40]	[0.21 0.41]	[0.22 0.42]	[0.25 0.45]	[0.42 0.60]
GDP Defl	[0.64 0.93]	[0.65 0.93]	[0.67 0.92]	[0.67 0.90]	[0.65 0.86]	[0.63 0.79]
Pers Cons Defl	[0.29 0.87]	[0.38 0.92]	[0.48 0.97]	[0.56 0.97]	[0.61 0.95]	[0.63 0.84]
CPI	[0.66 0.93]	[0.66 0.92]	[0.65 0.91]	[0.65 0.89]	[0.65 0.86]	[0.63 0.80]
Federal Funds Rate	[0.74 1.10]	[0.75 1.09]	[0.74 1.05]	[0.75 1.02]	[0.79 1.01]	[0.85 0.98]
3-month T-Bill	[0.75 1.10]	[0.78 1.10]	[0.78 1.07]	[0.78 1.03]	[0.80 1.01]	[0.85 0.98]
Moody's Aaa Bond	[0.33 0.81]	[0.42 0.85]	[0.50 0.87]	[0.61 0.91]	[0.69 0.93]	[0.82 0.96]

Table 12: Log-Posterior Marginal Data Density ($\times 10^{-3}$)

		λ					
M	q	0.250	0.500	1.000	2.000	5.00	100.0
0.0625	0.05	$\times$	-3.781	-4.100	-5.844	-11.77	-188.5
	0.1	$\times$	-3.780	-4.100	-5.843	-11.77	-188.5
	0.25	$\times$	-3.779	-4.099	-5.842	-11.77	-188.5
	0.5	$\times$	-3.778	-5.028	-5.873	-11.77	-188.5
	0.75	$\times$	-3.773	-5.023	-5.983	-11.90	-188.5
	0.9	$\times$	-3.778	-5.027	-5.872	-11.88	-188.5
	0.95	$\times$	-3.778	-5.027	-5.872	-11.88	-188.5
0.25	0.05	-3.675	-3.812	-4.557	-6.212	-11.16	-187.7
	0.1	-3.685	-3.811	-4.556	-6.211	-11.55	-187.7
	0.25	-3.684	-3.887	-4.556	-6.210	-11.55	-187.7
	0.5	-3.725	-3.886	-4.555	-6.233	-11.55	-187.7
	0.75	-3.720	-3.948	-4.854	-6.228	-11.58	-187.9
	0.9	-3.724	-3.952	-4.575	-6.233	-11.58	-187.7
	0.95	-3.724	-3.952	-4.575	-6.233	-11.58	-187.8
0.5	0.05	-3.763	-3.650	-4.540	-5.720	-11.00	-187.5
	0.1	-3.762	-3.649	-4.539	-6.061	-11.02	-187.7
	0.25	-3.851	-3.881	-4.539	-6.121	-11.38	-187.7
	0.5	-4.056	-3.880	-4.538	-6.125	-11.38	-187.7
	0.75	-4.387	-3.989	-4.533	-6.120	-11.55	-246.6
	0.9	-4.391	-3.882	-4.537	-6.125	-11.38	-187.7
	0.95	-4.391	-3.994	-4.537	-6.125	-11.38	-187.7
1	0.05	-3.392	-3.490	-4.320	-6.149	-10.85	-187.6
	0.1	-3.409	-3.494	-4.665	-6.148	-11.21	-187.6
	0.25	-3.787	-4.260	-4.665	-6.147	-11.59	-187.6
	0.5	-3.798	-4.504	-4.664	-6.165	-11.67	-187.8
	0.75	-4.719	-4.499	-4.883	-6.336	-11.67	-188.1
	0.9	-4.724	-4.504	-4.887	-6.340	-11.67	-188.1
	0.95	-4.724	-4.504	-4.887	-6.340	-11.67	-188.1
2	0.05	-3.624	-3.630	-4.507	-5.883	-10.90	-187.8
	0.1	-3.720	-3.629	-4.506	-5.883	-10.96	-187.8
	0.25	-4.134	-3.794	-4.505	-6.563	-11.11	-187.8
	0.5	-4.133	-3.964	-4.505	-6.562	-11.20	-187.9
	0.75	-4.128	-3.959	-4.500	-6.557	-11.33	-187.9
	0.9	-4.133	-3.963	-4.504	-6.561	-11.33	-187.9
	0.95	-4.133	-3.963	-4.504	-6.561	-11.33	-187.9
4	0.05	-3.656	-3.708	-4.395	-5.842	-11.35	-187.3
	0.1	-4.355	-3.707	-4.394	-5.842	-11.35	-187.5
	0.25	-4.354	-4.198	-4.393	-6.062	-11.35	-187.5
	0.5	-4.353	-4.197	-4.393	-6.061	-11.35	-187.6
	0.75	-4.348	-4.192	-4.651	-6.375	-11.36	-187.6
	0.9	-4.352	-4.197	-4.563	-6.379	-11.35	-187.6
	0.95	-4.352	-4.197	-4.655	-6.379	-11.35	-187.6
16	0.05	-4.254	-3.984	-4.470	-5.831	-10.77	-187.7
	0.1	-4.253	-3.984	-4.469	-5.830	-10.98	-187.7
	0.25	-4.252	-3.983	-4.469	-5.894	-11.26	-187.7
	0.5	-4.252	-4.083	-4.468	-5.893	-11.26	-187.7
	0.75	-4.247	-4.078	-4.939	-6.035	-11.48	-187.9
	0.9	-4.251	-4.082	-4.944	-6.039	-11.48	-187.9
	0.95	-4.251	-4.082	-4.944	-6.039	-11.48	-187.9

Table 13: 80% HPD invervals of factor loadings for $\lambda = 1$ and different values of M_0

$M_0 = \frac{1}{16}$	Output	Inflation	Interest Rate
GDP	**[0.28 0.75]**	[-2.86 -1.42]	[-0.03 0.67]
IP total	**[0.64 1.19]**	[-1.18 -0.04]	[-0.88 -0.26]
Personal Income	**[-0.01 0.28]**	[-2.11 -0.77]	[0.06 0.65]
GDP deflator	[-0.09 0.14]	**[1.27 2.28]**	[-0.85 -0.41]
Personal Consumption Price Deflator	[-0.09 0.06]	**[0.01 0.93]**	[-0.30 0.07]
CPI	[0.03 0.26]	**[1.36 2.27]**	[-0.93 -0.53]
Federal Funds Rate	[-0.08 0.06]	[-0.69 0.18]	**[0.01 0.40]**
3-month T-Bill	[-0.06 0.09]	[-0.71 0.16]	**[-0.01 0.39]**
Moody's Aaa Bond	[0.00 0.14]	[-0.48 0.32]	**[-0.15 0.20]**

$M_0 = \frac{1}{2}$	Output	Inflation	Interest Rate
GDP	**[0.63 0.97]**	[-1.14 -0.37]	[0.04 0.41]
IP total	**[0.98 1.32]**	[-0.30 0.22]	-0.16 0.18[]
Personal Income	**[0.14 0.43]**	[-1.00 -0.23]	[0.01 0.33]
GDP deflator	[-0.21 0.03]	**[1.00 1.50]**	[-0.36 -0.06]
Personal Consumption Price Deflator	[-0.16 0.05]	**[0.37 0.91]**	[-0.13 0.23]
CPI	[-0.06 0.18]	**[0.95 1.43]**	[-0.31 -0.02]
Federal Funds Rate	[-0.10 0.07]	[-0.26 0.27]	**[0.69 1.10]**
3-month T-Bill	[-0.07 0.11]	[-0.16 0.32]	**[0.72 1.12]**
Moody's Aaa Bond	[-0.00 0.17]	[-0.17 0.30]	**[0.35 0.83]**

$M_0 = 1$	Output	Inflation	Interest Rate
GDP	**[0.64 0.94]**	[-0.81 -0.15]	[0.01 0.30]
IP total	**[1.00 1.26]**	[-0.21 0.21]	[-0.09 0.17]
Personal Income	**[0.17 0.44]**	[-0.80 -0.13]	[-0.03 0.23]
GDP deflator	[-0.19 0.02]	**[0.98 1.38]**	[-0.24 -0.01]
Personal Consumption Price Deflator	[-0.16 0.04]	**[0.51 0.96]**	[-0.07 0.24]
CPI	[-0.05 0.15]	**[0.92 1.32]**	[-0.19 0.03]
Federal Funds Rate	[-0.10 0.06]	[-0.15 0.24]	**[0.77 1.12]**
3-month T-Bill	[-0.07 0.10]	[-0.13 0.25]	**[0.79 1.13]**
Moody's Aaa Bond	[0.00 0.17]	[-0.15 0.25]	**[0.54 0.94]**

$M_0 = 16$	Output	Inflation	Interest Rate
GDP	**[0.83 1.03]**	[-0.16 0.08]	[-0.06 0.11]
IP total	**[1.00 1.12]**	[-0.07 0.08]	[-0.05 0.08]
Personal Income	**[0.51 0.75]**	[-0.20 0.11]	[-0.09 0.10]
GDP deflator	[-0.09 0.03]	**[0.96 1.10]**	[-0.08 0.03]
Personal Consumption Price Deflator	[-0.08 0.04]	**[0.91 1.05]**	[-0.04 0.09]
CPI	[-0.03 0.08]	**[0.94 1.09]**	[-0.06 0.05]
Federal Funds Rate	[-0.06 0.04]	[-0.06 0.07]	**[0.94 1.06]**
3-month T-Bill	[-0.06 0.05]	[-0.06 0.07]	**[0.94 1.06]**
Moody's Aaa Bond	[-0.03 0.08]	[-0.06 0.08]	**[0.90 1.04]**

Table 14: 80% HPD interval of DSGE model parameter θ for different weights λ and $M_0 = 1$

λ	0.25	0.5	1	2	5	100
ψ_1	[1.09 1.67]	[1.10 1.70]	[1.16 1.70]	[1.14 1.69]	[1.68 2.12]	[1.73 2.09]
ψ_2	[0.13 0.32]	[0.15 0.35]	[0.15 0.35]	[0.17 0.36]	[0.08 0.23]	[0.10 0.23]
ρ_r	[0.51 0.78]	[0.55 0.76]	[0.61 0.79]	[0.67 0.81]	[0.70 0.81]	[0.70 0.80]
κ	[0.14 0.54]	[0.16 0.54]	[0.16 0.54]	[0.11 0.34]	[0.05 0.17]	[0.03 0.12]
τ^{-1}	[1.49 2.55]	[1.41 2.47]	[1.75 2.65]	[2.20 3.49]	[1.85 3.24]	[1.84 2.41]
ρ_g	[0.85 0.96]	[0.88 0.97]	[0.90 0.97]	[0.93 0.98]	[0.94 0.99]	[0.95 0.99]
ρ_z	[0.20 0.70]	[0.23 0.70]	[0.36 0.79]	[0.47 0.82]	[0.93 0.98]	[0.95 0.99]
σ_R	[0.04 0.08]	[0.05 0.08]	[0.05 0.09]	[0.06 0.10]	[0.04 0.07]	[0.04 0.06]
σ_g	[0.10 0.21]	[0.09 0.20]	[0.10 0.21]	[0.09 0.21]	[0.27 0.41]	[0.33 0.47]
σ_z	[0.13 0.26]	[0.14 0.27]	[0.14 0.27]	[0.15 0.27]	[0.09 0.14]	[0.10 0.14]

Table 15: 80% HPD interval of DSGE model parameter θ for different weights λ and $M_0 = \frac{1}{2}$

λ	0.25	0.5	1	2	5	100
ψ_1	[1.07 1.56]	[1.03 1.52]	[1.20 1.92]	[1.55 2.09]	[1.58 1.92]	[1.67 2.00]
ψ_2	[0.12 0.31]	[0.14 0.33]	[0.16 0.36]	[0.09 0.27]	[0.11 0.25]	[0.11 0.25]
ρ_r	[0.46 0.74]	[0.52 0.77]	[0.61 0.78]	[0.64 0.78]	[0.64 0.77]	[0.66 0.78]
κ	[0.16 0.56]	[0.13 0.48]	[0.14 0.41]	[0.06 0.24]	[0.04 0.15]	[0.03 0.13]
τ^{-1}	[1.31 2.55]	[1.53 2.65]	[1.97 3.08]	[1.62 2.71]	[1.47 2.34]	[2.45 3.23]
ρ_g	[0.84 0.96]	[0.88 0.97]	[0.90 0.98]	[0.92 0.98]	[0.93 0.98]	[0.94 0.98]
ρ_z	[0.16 0.62]	[0.18 0.70]	[0.41 0.88]	[0.88 0.98]	[0.93 0.97]	[0.94 0.98]
σ_R	[0.04 0.08]	[0.05 0.09]	[0.05 0.09]	[0.04 0.07]	[0.03 0.06]	[0.04 0.06]
σ_g	[0.10 0.21]	[0.09 0.20]	[0.09 0.23]	[0.22 0.36]	[0.27 0.42]	[0.30 0.45]
σ_z	[0.13 0.25]	[0.13 0.26]	[0.12 0.25]	[0.09 0.15]	[0.09 0.14]	[0.11 0.16]

B Figures

Figure 1: Upper panel: Posterior mean of Φ for increasing values of the prior weight λ. Lower panel: Width of posterior 80% HPD interval for increasing values of the prior weight λ.

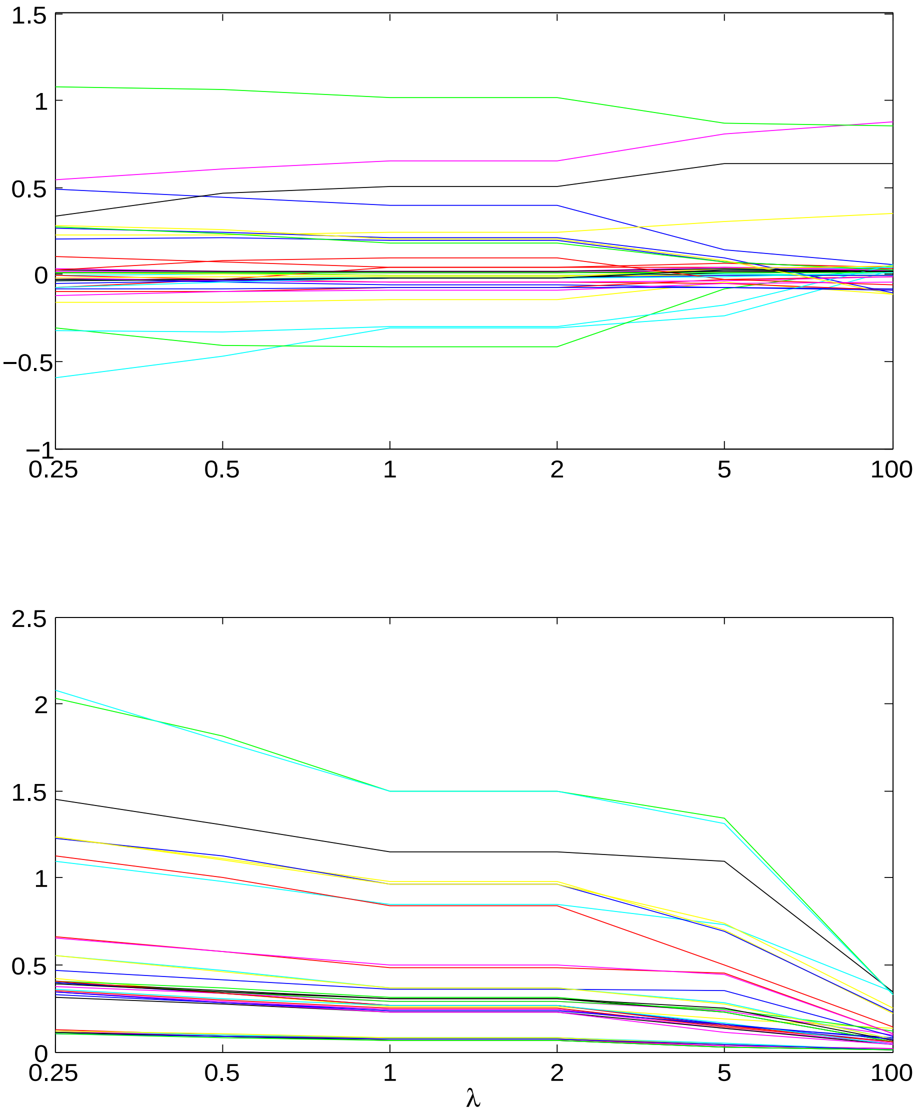


Figure 2: Posterior mean of Σ for increasing values of the prior weight λ.

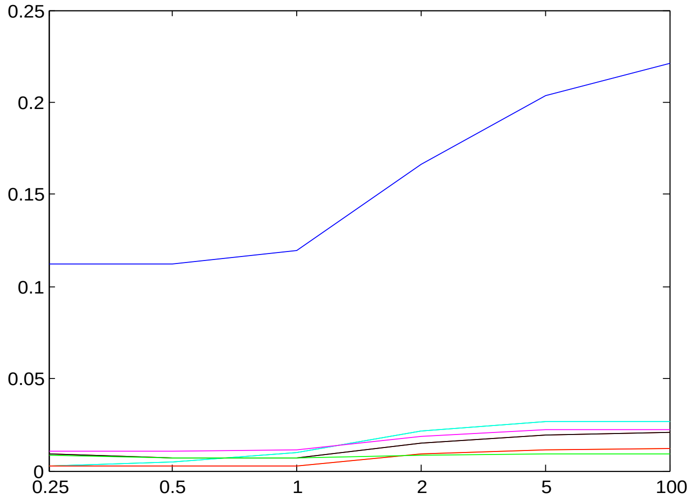


Figure 3: Response of factors to a contractionary monetary shock: $\lambda = 100$ and $M_0^{-1} = 0.5$

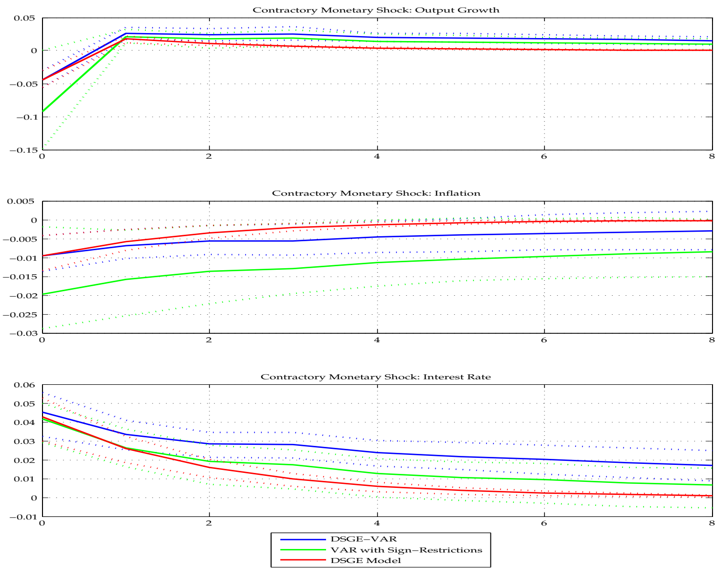


Figure 4: Response of factors to a contractionary monetary shock: $\lambda = 5$ and $M_0^{-1} = 0.5$

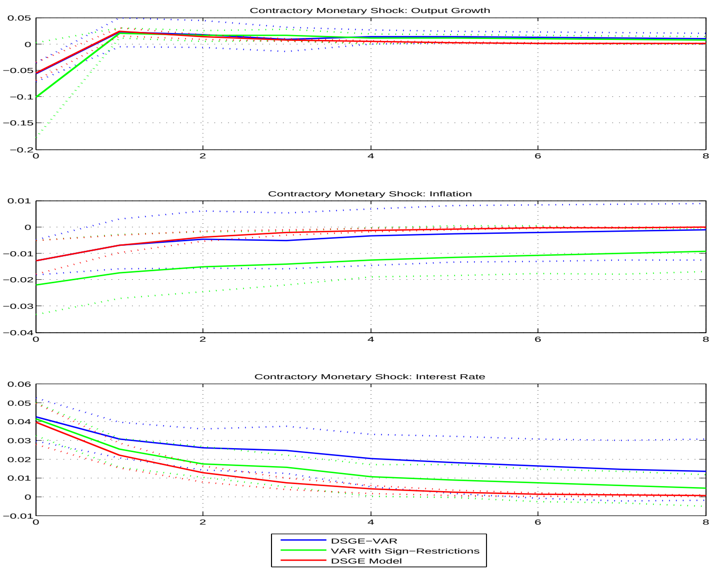


Figure 5: Response of factors to a contractionary monetary shock: $\lambda = 1$ and $M_0^{-1} = 0.5$

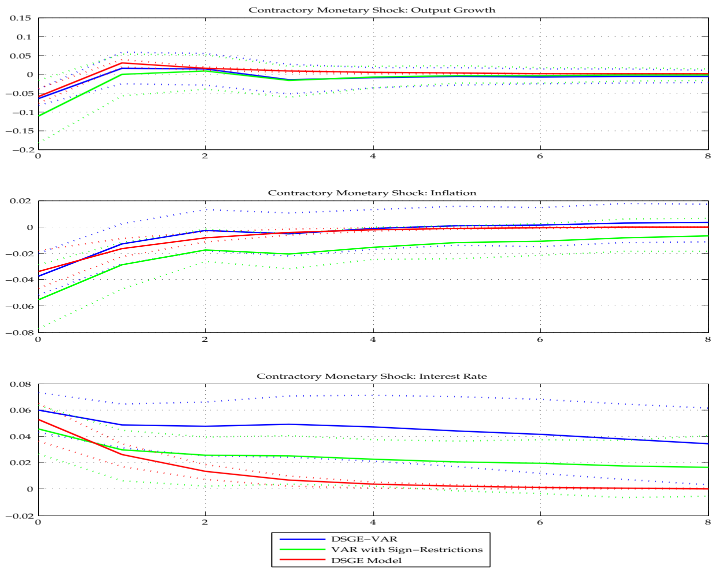


Figure 6: Response of factors to a contractionary monetary shock: $\lambda = 0.25$ and $M_0^{-1} = 0.5$

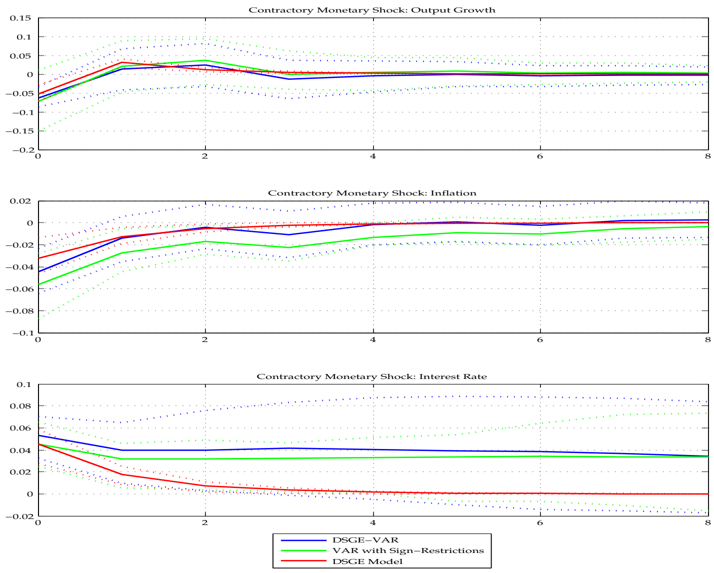


Figure 7: Cumulated ('Level') Response of Output Factor for $M_0^{-1} = 0.5$ and selected values for λ

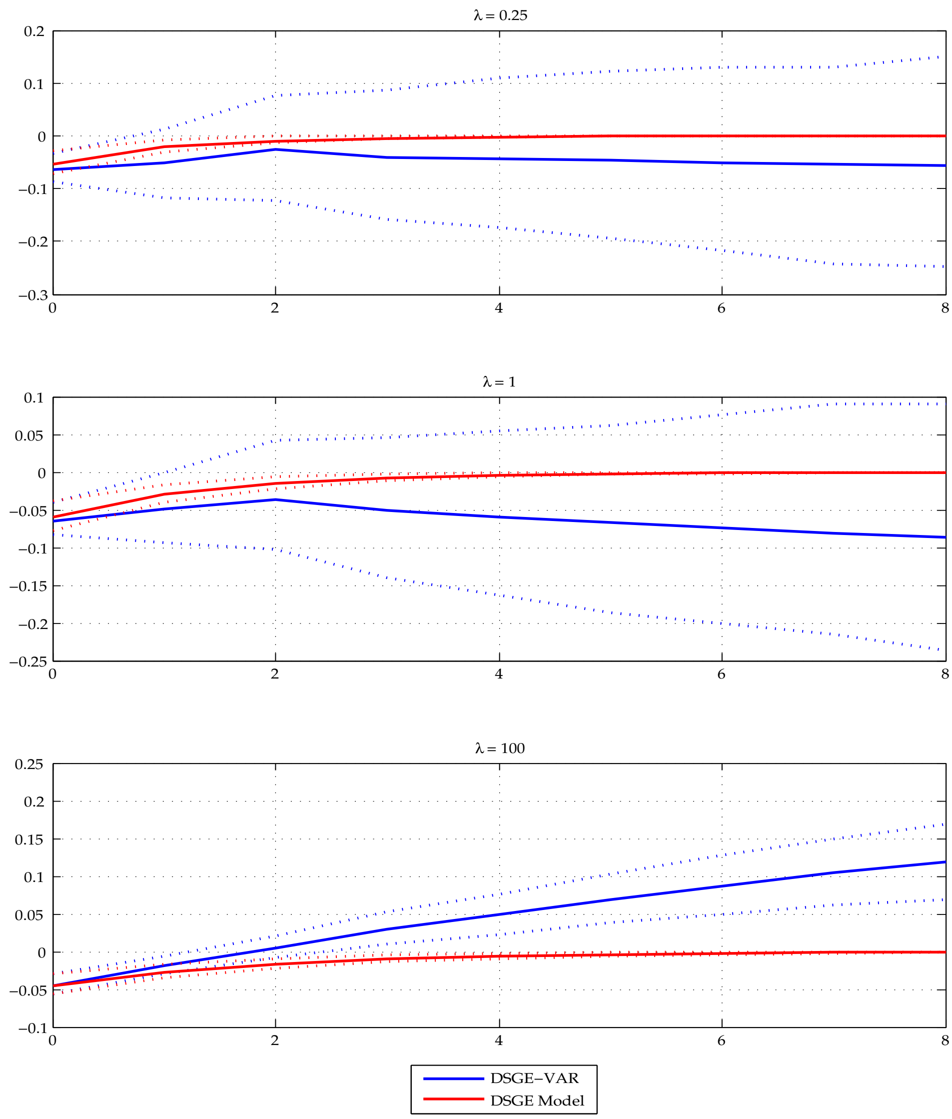


Figure 8: Response of selected observed output series with $\lambda = 1$ and $M_0^{-1} = 0.5$ (growth rates)

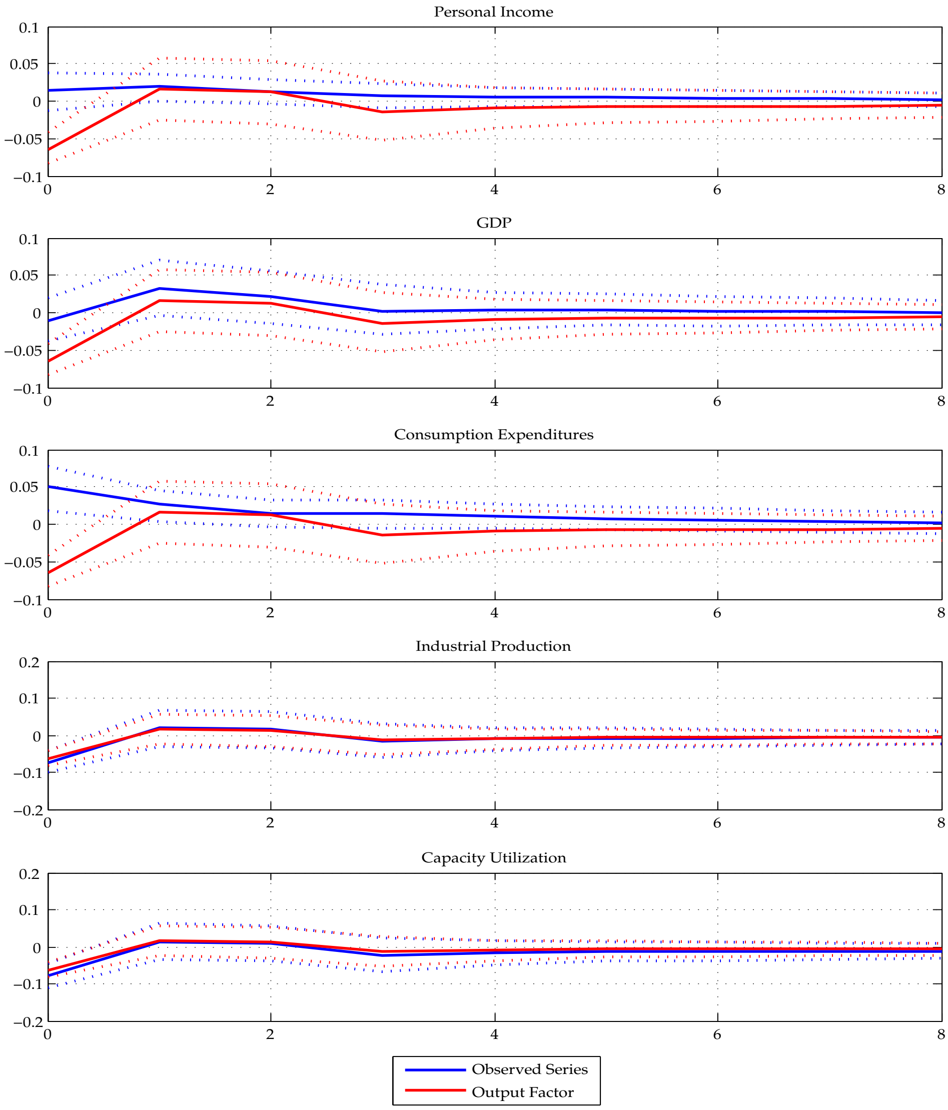


Figure 9: Responses of selected observed inflation series with $\lambda = 1$ and $M_0^{-1} = 0.5$

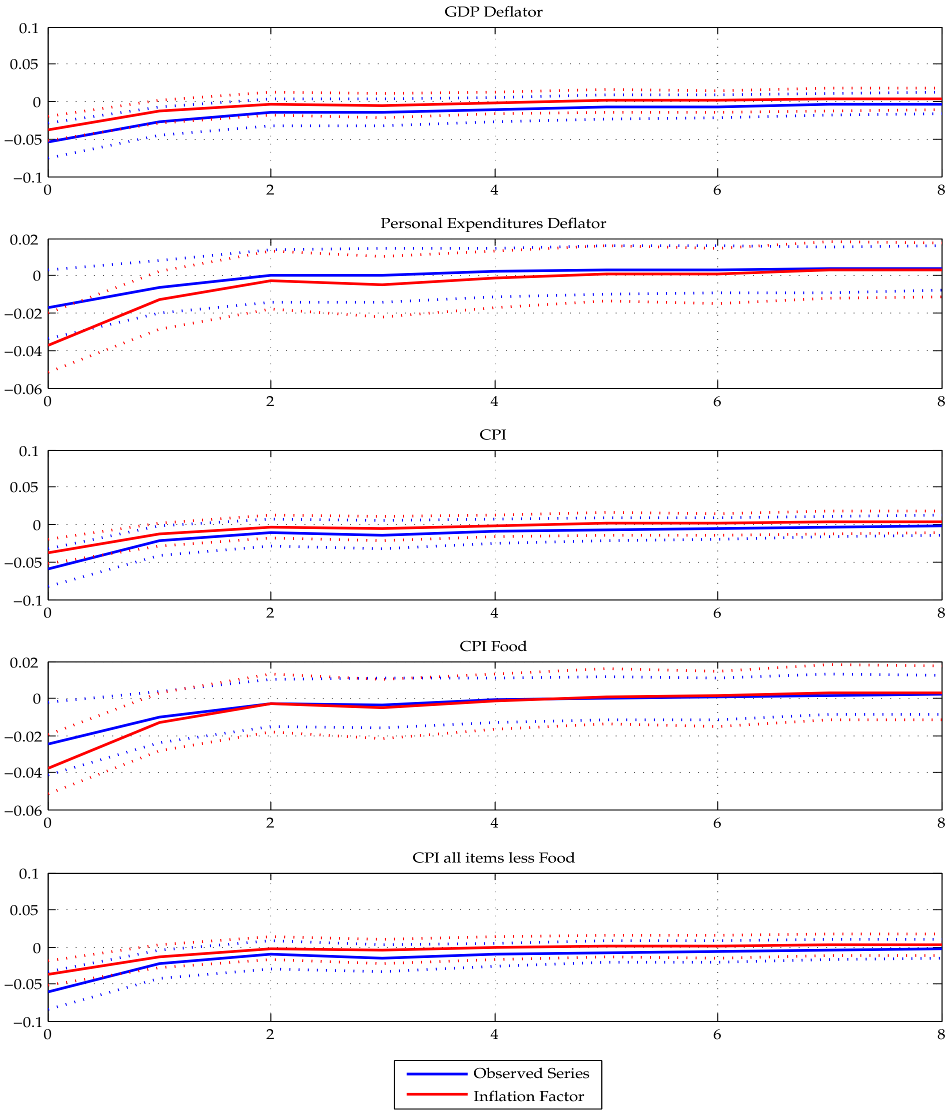
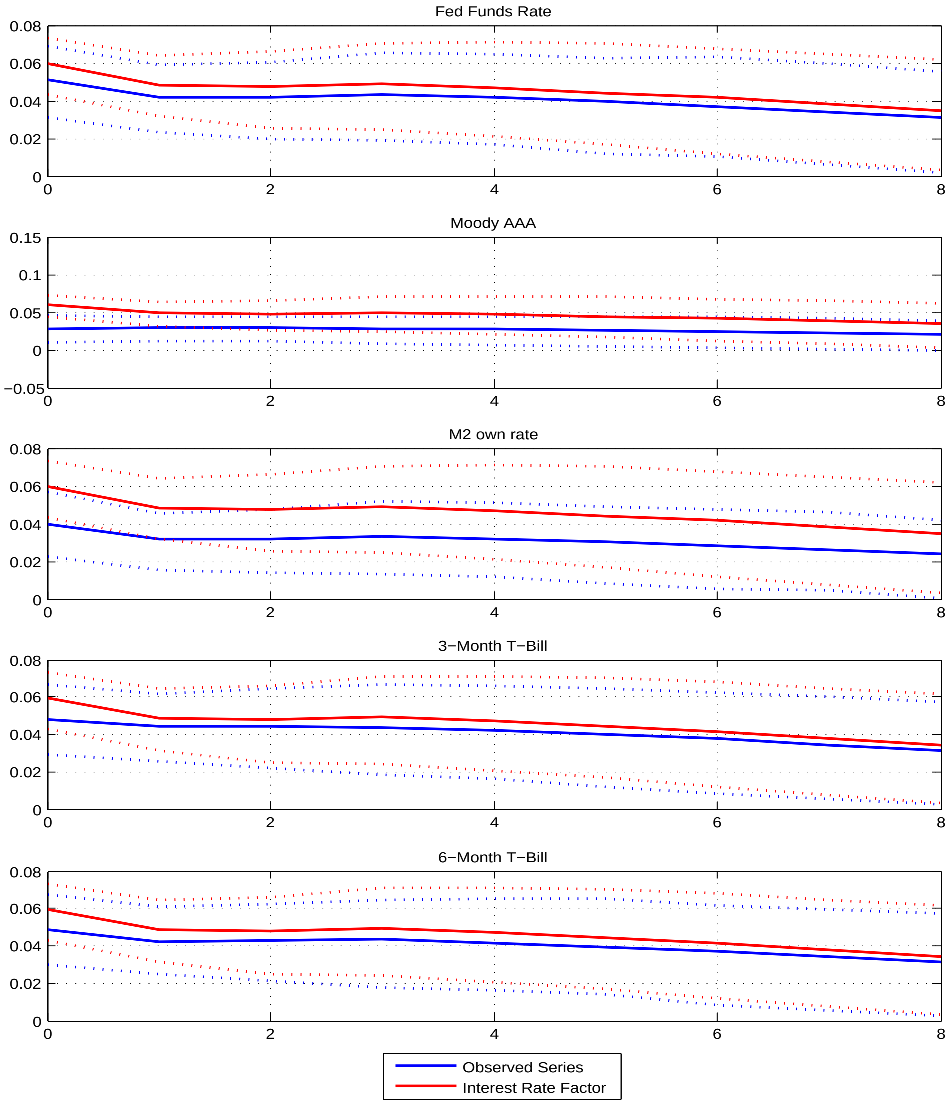


Figure 10: Responses of selected observed interest rates with $\lambda = 1$ and $M_0^{-1} = 0.5$



C DSGE model in generic form

$$S_t = [y_t, \pi_t, r_t, E_t[y_{t+1}], E_t[\pi_{t+1}], g_t, z_t]$$

$$\varepsilon_t = [\varepsilon_{r,t}, \varepsilon_{g,t}, \varepsilon_{z,t}]$$

$$\eta_t = [y_t - E_{t-1}[y_t], \pi_t - E_{t-1}[\pi_t]]$$

$$\theta_{DSGE} = [\psi_1, \psi_2, \rho_R, \beta, \kappa, \tau, \rho_g, \rho_z, \rho_{gz}, \sigma_R, \sigma_g, \sigma_z]$$

'Expectational' equations have to be added:

$$x_{t+1} = E_t(x_{t+1}) + \xi_t$$

So the model used in Lubik and Schorfheide (2004) can be written as

$$\Gamma_0 s_t = \Gamma_1 s_{t-1} + \Psi \varepsilon_t + \Pi \eta_t$$

$$\Gamma_0 = \begin{pmatrix} 1 & 0 & \tau & -1 & -\tau & -1 & 0 \\ -\kappa & 1 & 0 & 0 & -\beta & 0 & \kappa \\ -(1-\rho_R)\psi_2 & -(1-\rho_R)\psi_1 & 1 & 0 & 0 & 0 & (1-\rho_R)\psi_2 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\Gamma_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \rho_g & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \rho_z \end{pmatrix}$$

$$\Psi = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\Pi = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\Omega = \begin{pmatrix} \sigma_R^2 & 0 & 0 \\ 0 & \sigma_g^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{pmatrix}$$

D Kalman Filter with Autocorrelated Errors

For a reference, see Anderson and Moore (1979). It is assumed that the vector of data evolves according to the following state space system:

$$\begin{aligned} X_t &= \Lambda F_t + v_t \\ F_t &= \Phi(L) F_{t-1} + e_t \end{aligned}$$

F_t is a vector of unobserved dynamic factors with a small dimension M. X_t is a potentially high dimensional vector of N data series observed over T time periods. Each variables in X_t loads on at least one factor, Λ is the $N \times T$ matrix of factor loadings. Factors F_t are related to lagged values of the factors by $\Phi(L) = \Phi_1 L + \ldots + \Phi_p L^p$. The errors e_t and v_t are distributed as follows:

$$\begin{pmatrix} u_t = v_t - \Psi v_{t-1} \\ e_t \end{pmatrix} \sim iiN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} R & 0 \\ 0 & \Sigma \end{bmatrix} \right)$$

We assume that R and Ψ are diagonal.

The system can be rewritten into a system of order one by defining

$$\Phi = \begin{pmatrix} \Phi_1 & \Phi_2 & \dots & & \Phi_p \\ I_M & 0 & \dots & & 0 \\ 0 & I_M & \dots & & 0 \\ 0 & 0 & \ddots & & 0 \\ 0 & 0 & \dots & I_M & 0 \end{pmatrix}$$

$$\tilde{F}_t = \begin{pmatrix} F_t & F_{t-1} & \dots & F_{t-p} \end{pmatrix}'$$

$$\tilde{X}_t = X_t - \Psi X_{t-1}$$

with

$$X_{-1} = 0$$

$$\tilde{\Lambda} = \begin{pmatrix} \Lambda\Phi - \Psi\Lambda & \Lambda\Phi_1 & \dots & \Lambda\Phi_p \end{pmatrix}$$

For $p > 1$ we can write

$$\begin{aligned} \tilde{X}_t &= X_t - \Psi X_{t-1} \\ &= \Lambda F_t - \Psi\Lambda F_{t-1} + v_t - \Psi v_{t-1} \\ &= \Lambda F_t - \Psi\Lambda F_{t-1} + u_t \\ &= \begin{pmatrix} \Lambda & \vdots & -\Psi\Lambda & \vdots & 0_{(p-2)\times M} \end{pmatrix} \tilde{F}_t + u_t \end{aligned}$$

and

$$\tilde{F}_t = \Phi\tilde{F}_{t-1} + \tilde{e}_t$$

where

$$\tilde{\Omega} = Var(\tilde{e}_t) = \begin{pmatrix} \Omega & 0 \\ 0 & 0 \end{pmatrix}$$

In this case, the standard Kalman filter (see e.g. Hamilton (1994)) applies. For $p = 1$ the definitions imply[15]

[15] We could also enlarge the state vector with lagged values. But there the solution provided here keeps the dimension of the state vector small.

$$\begin{aligned}\tilde{X}_t &= \tilde{\Lambda} F_{t-1} + \Lambda e_t + u_t \\ F_t &= \Phi F_{t-1} + e_t\end{aligned}$$

So the Kalman Filter has to be adjusted as there is correlation between errors in the observation and state equation. Additionally, we have F_{t-1} instead of F_t in the observation equation. Note that conditional on the parameter Ψ, the sequence $\{X_t\}$ contains the same information as $\{\tilde{X}_t\}$ and therefore $E[F_t|\{\tilde{X}_k\}] = E[F_t|\{X_k\}]$. The joint distribution of $\tilde{X}_t$ and F_t is

$$\begin{pmatrix} \tilde{X}_t \\ F_t \end{pmatrix} |\{X_k\}_{k=1}^{t-1} \sim N\left(\begin{pmatrix} \tilde{X}_{t|t-1} \\ F_{t|t-1} \end{pmatrix}, \begin{pmatrix} h_t & c_t' \\ c_t & P_{t|t-1} \end{pmatrix} \right)$$

where:

$$\begin{aligned}
F_{t|t} &= \Phi F_{t-1|t-1} + c_t h_t^{-1}(\tilde{X}_t - \underbrace{\tilde{X}_{t|t-1}}_{\tilde{\Lambda} F_{t-1|t-1}}) \\
h_t &= E_{t-1}[(\tilde{X}_t - \tilde{X}_{t|t-1})(\tilde{X}_t - \tilde{X}_{t|t-1})'] \\
&= E_{t-1}[(\tilde{\Lambda} F_{t-1} + \Delta e_t + \varepsilon_t - \tilde{\Lambda} F_{t-1|t-1})(\tilde{\Lambda} F_{t-1} + \Delta e_t + \varepsilon_t - \tilde{\Lambda} F_{t-1|t-1})'] \\
&= \tilde{\Lambda} P_{t-1|t-1} \tilde{\Lambda}' + \Delta \Omega \Delta' + R \\
c_t &= E_{t-1}[(F_t - F_{t|t-1})(\tilde{X}_t - \tilde{X}_{t|t-1})'] \\
&= E_{t-1}[(\Phi(F_t - F_{t-1|t-1}) + e_t)(\tilde{\Lambda} F_{t-1} + \Delta e_t + \varepsilon_t - \tilde{\Lambda} F_{t-1|t-1})'] \\
&= \Phi P_{t-1|t-1} \tilde{\Lambda}' + \Omega \Delta' \\
P_{t|t-1} &= E_{t-1}[(F_t - F_{t|t-1})(F_t - F_{t|t-1})'] \\
&= E_{t-1}[(\Phi F_{t-1} + e_t - \Phi F_{t-1|t-1})(\Phi F_{t-1} + e_t - \Phi F_{t-1|t-1})'] \\
&= \Phi P_{t-1|t-1} \Phi' + \Omega \\
P_{t|t} &= \Phi P_{t-1|t-1} \Phi' + \Omega - c_t h_t^{-1} c_t'
\end{aligned}$$

Given initial values $P_{0|0} = E(F_t)(F_t)'$ and $F_{0|0} = 0$ we can iteratively calculate $P_{t|t}$ and $F_{t|t}$ for $t = 1 \ldots, T$. Note that the assumption $X_{-1} = 0$ ensures that the sample size is not reduced, even for $p > 1$.

E Inverse Wishart Distribution

The Wishart distribution is the multivariate version of the inverted Gamma distribution. Let Σ be a $n \times n$ positive definite random matrix. Σ has the inverted Wishart $IW(S, \nu)$ distribution if its density is of the form:

$$p(\Sigma|S,\nu) \propto |S|^{\nu/2}|\Sigma|^{-(\nu+n+1)/2} \exp(-\frac{1}{2}tr(\Sigma^{-1}S))$$

To sample Σ from this distribution, draw $n \times 1$ vectors $Z_1, \ldots, Z_\nu$ form a multivariate normal $N(0, S^{-1})$ and let

$$\Sigma = \left(\sum_{i=1}^{\nu} Z_i Z_i' \right)^{-1}$$

(see Bauwens, Lubrano, and Richard (1999)).

F Inverted Gamma Distribution

There is considerable confusion about the implementation of the inverted gamma distribution for maybe three reasons: First, some authors specify the priors for the variances, others for the standard deviations. Second, there are differences in the parametrization of the density function. Third, most authors report the mean and the standard deviations of the prior distribution, while the distribution is specified in terms of other hyperparameter. Ideally, the authors explicitly state the density function they use (as their mean and standard deviations as well as the hyperparameters, e.g. DelNegro and Schorheide (2004). However, most authors do not. A complete discussion is contained in Bauwens, Lubrano, and Richard (1999). The define the density of the variance $x = \sigma^2$ as

$$f_x(x) = \frac{1}{\Gamma(\frac{\nu}{2})} \left(\frac{s}{2}\right)^{\frac{-\nu}{2}} x^{-\frac{1}{2}(\nu+2)} e^{-\frac{s}{2x}}$$

We follow Bauwens, Lubrano, and Richard (1999) in that we refer to this density as 'inverted gamma-2 density'. We skip the indicator function for the variance and standard deviations here and in what follows. All the densities are equal to zero if its arguments are negative.

From variances to standard deviations To calculate the implied density for $y = \sigma = \sqrt{x}$, one needs to apply change of variable formula. Define $y = g(x) = \sqrt{x}$.

The inverse is $g^{-1}(y) = y^2$ and its first derivative $\frac{\partial g^{-1}(y)}{\partial y} = 2y$. Applying the formula yields

$$\begin{aligned} f_y(y) &= \frac{1}{\Gamma(\frac{\nu}{2})}\left(\frac{s}{2}\right)^{\frac{-\nu}{2}}(y^2)^{-\frac{1}{2}(\nu+2)}e^{-\frac{s}{2y^2}}|2y| \\ &= \frac{2}{\Gamma(\frac{\nu}{2})}\left(\frac{s}{2}\right)^{\frac{-\nu}{2}}y^{-(\nu+1)}e^{-\frac{s}{2y^2}} \end{aligned}$$

This density is called the 'inverted gamma-1 density'.

Different parametrizations In the this paper, we use the parametization of DelNegro and Schorheide (2004) (see notes to Table 1 in their paper):

$$f_y(y) = \frac{2}{\Gamma(\frac{\nu}{2})}\left(\frac{\nu s_*^2}{2}\right)^{\frac{-\nu}{2}}y^{-(\nu+1)}e^{-\frac{\nu s_*^2}{2y^2}}$$

Hence, the difference between the specifications is

$$s = \nu s_*^2$$

The values given in Table 1 refer to s_*. The parametrization of the inverted gamma-2 density follows Bauwens, Lubrano, and Richard (1999). The parameter ν is the same in either specification.

From parameters to moments For $\nu > 1$, the expected value of y is

$$\mathbb{E}(y) = \sqrt{\frac{s}{2}}\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}$$

and for $\nu > 2$, the variance is

$$\mathbb{V}(y) = \frac{s}{\nu-2} - \mathbb{E}(y)^2$$

There is no easy way to invert the system. Hence, the specification is terms of ν and s.
For $\nu > 2$, the expected value of x is

$$E(x) = \frac{s}{\nu-2}$$

and for $\nu > 4$, the variance is

$$\mathbb{V}(x) = \frac{2}{\nu - 4}\mathbb{E}(x)^2$$

It follows that

$$\nu = \left(\frac{2\mathbb{E}(x)^2}{\mathbb{V}(x)} + 4\right)$$

and

$$s = \mathbb{E}(x)(\nu - 2)$$