



CPB Netherlands Bureau for Economic  
Policy Analysis

CPB Discussion Paper | 270

# Reduced-rank time-varying vector autoregressions

Joris de Wind  
Luca Gambetti



# Reduced-rank time-varying vector autoregressions\*

Joris de Wind<sup>†</sup>      Luca Gambetti<sup>‡</sup>

March, 2014

## Abstract

The standard time-varying VAR workhorse suffers from overparameterization, which is a serious problem as it limits the number of variables and lags that can be incorporated in the model. As a solution for the overparameterization problem, we propose a new, more parsimonious time-varying VAR model setup with which we can reliably estimate larger models including more variables and/or more lags than was possible until now. The new model setup implies cross-equation restrictions on the time variation that are empirically supported, theoretically appealing, and make the Bayesian estimation procedure much faster.

## 1 Introduction

Thanks to the pioneer work of Sims (1980), vector autoregressions (VARs) have a long tradition in applied macroeconomics, and are widely used in policy analysis and forecasting. They are flexible time-series models that can capture complex dynamic interrelationships among macroeconomic variables—by allowing for responses of all variables to all variables at all lags. Nevertheless, in standard VARs these complex dynamic interrelationships are *not* allowed to change over time, which seriously limits policy analysis regarding economies that are subject to changes in the underlying structure.

---

\*We are very grateful to Wouter den Haan for his extensive comments and suggestions. Views expressed in this paper are our own and do not necessarily reflect those of the CPB.

<sup>†</sup>CPB Netherlands Bureau for Economic Policy Analysis, email: j.de.wind@cpb.nl.

<sup>‡</sup>Universitat Autònoma de Barcelona, email: luca.gambetti@uab.es.

A natural evolution of the literature was the development of time-varying VARs, in which the dynamic interrelationships among variables are—in contrast—allowed to change over time. Cogley and Sargent (2001) have developed a VAR with coefficients that are random walks with reflecting barriers that keep the VAR stable. Cogley and Sargent (2005) have extended their earlier work with stochastic volatility. Primiceri (2005) has developed a similar model to Cogley and Sargent (2005) but without the reflecting barriers and with a setup for stochastic volatility that suits better with structural analysis. The model setup of Primiceri (2005) can be considered as today’s standard time-varying VAR workhorse.

Currently, time-varying VARs are only used for small models—including only few variables and few lags. The aforementioned papers include, for instance, only three variables (inflation, nominal interest, and unemployment) and two lags. There are also papers that include more variables or more lags, albeit only marginally. Canova and Gambetti (2009, 2010) and Hofmann, Peersman, and Straub (2012) include, for example, four variables instead of three. There are, to the best of our knowledge, no papers in the literature with larger time-varying VARs without compromises made elsewhere in the model setup. For example, Gambetti, Pappa, and Canova (2008), in their model with five variables and two lags, assume that the covariance matrix of the innovations to the time-varying parameters is diagonal, while Baumeister and Peersman (2013), in their model with four variables and four lags, employ an unusual long training sample for determining their prior and their prior is also more informative than usual.

It is questionable whether we can obtain reliable inference for larger time-varying VARs, as the generality of the VAR model brings along a large number of parameters even for systems of moderate size. The number of parameters in a system with  $n$  variables and  $p$  lags is equal to  $k = n + pn^2$ , including an  $n \times 1$  vector of constants and for each lag an  $n \times n$  matrix of autoregressive coefficients. Note that with standard VARs inference is (just) about  $k$  fixed coefficients, but with time-varying VARs inference is about  $k$  *time paths* of coefficients. In addition, the covariance matrix of the innovations to the time-varying parameters, a  $k \times k$  matrix, also needs to be estimated. It should be clear that the number of objects to be estimated quickly gets large when more variables and/or more lags are added to the system.

Nevertheless, for many applications a large set of variables is vital, for example, to avoid omitted variable bias such as the so-called *price puzzle* or when using disaggregate information. Moreover, when using quarterly data it seems natural to include at least four lags, which is, for example, advocated by Blanchard and Perotti (2002) for the literature on the dynamic effects of fiscal policy. Altogether, we would like to be able to estimate larger systems including more variables and/or more lags than is possible with the standard time-varying VAR model setup of Primiceri (2005).

In this paper, we propose a new, more parsimonious time-varying VAR model setup with which we can estimate larger systems than is currently possible. The key distinctive feature of the new model setup is the covariance matrix of the innovations to the time-varying parameters which is now assumed to be of reduced rank instead of full rank. The implied cross-equation restrictions amount to a reduction in the number of underlying factors driving the time-varying parameters. In fact, in the new model setup (from now onwards “reduced-rank model”) there are  $q \leq k$  underlying factors driving the  $k$  time-varying parameters, while in the standard model setup (from now onwards “full-rank model”) there are as many underlying factors as time-varying parameters.

It turns out that the number of underlying factors included in the reduced-rank model can be chosen to be much smaller than the number of time-varying parameters, at least for the typical monetary VAR we have analyzed. Using the same data as Cogley and Sargent (2001, 2005), we find that the reduced-rank model with only a few factors is enough to capture the bulk of time variation that is present in the full-rank model. This was also to be expected given that results for the full-rank model already had suggested that the time variation in the parameters is highly structured. In particular, the first three principal components account together already for more than 95% of the variance of the innovations to the time-varying parameters, which is consistent with Cogley and Sargent (2005) who have presented a similar number.

In addition to the just-mentioned empirical support, the cross-equation restrictions in the time variation are also theoretically appealing. The learning model of Sargent (1999) predicts, for example, that reduced-form parameters should move in a highly structured way, because of the cross-equation restrictions associated with optimization and foresight. This was already mentioned by Cogley and Sargent (2005) who had suggested that a formal treatment of cross-equation restrictions in the time variation would be a priority for future research. Also Primiceri (2005) had mentioned this as clear direction for future research arguing that imposing an underlying factor structure would be essential for estimating larger systems.

Furthermore, the reduced-rank model also has practical advantages over the full-rank model. In particular, the computing time needed for the Bayesian estimation is much shorter as there are less parameters and underlying trajectories to be estimated in the reduced-rank model. The gain in terms of computing speed is, for example, about a factor four for the application of Cogley and Sargent (2001), and for applications with more variables and/or more lags the gains are potentially even larger. Moreover, (much) shorter Markov chains are sufficient to achieve convergence of the Gibbs sampler used for the

Bayesian estimation, and this can be exploited to speed up the Bayesian estimation even further. Note that running an out-of-sample forecasting exercise, such as the one by D'Agostino, Gambetti, and Giannone (2013), becomes much more practical as less computing time is needed for the recursive Bayesian estimation.

Altogether, the reduced-rank model proposed in this paper seems to be a useful new tool for studying macroeconomic questions. With the reduced-rank model we can estimate larger systems than is currently possible, which opens up the world of potential applications. For a follow-up paper we are working on an application about financial innovation regarding consumer loans, and include home mortgages and consumer credit as extra variables in an otherwise standard monetary VAR. The application is similar to Den Haan and Sterk (2011), who compare a time-invariant VAR estimated on an early sub-sample with a time-invariant VAR estimated on a later sub-sample. Herewith, they are simply imposing the timing and number of sub-samples rather than estimating a model (such as the reduced-rank model) in which this is endogenously determined. Another interesting application, which pushes the number of variables even further, is the transmission of shocks between G-7 countries, as analyzed by Canova and Ciccarelli (2009). They have developed a multi-country VAR model with time-varying coefficients and cross-equation restrictions, yet they are *imposing* particular cross-equation restrictions instead of estimating them, which is a crucial difference with the reduced-rank model. Furthermore, De Wind (2014) applies the reduced-rank model to analyze to what extent the dynamic effects of unanticipated changes in US tax policy have changed structurally over the post World War II period.

Besides, there are also papers that model time variation as discrete breaks, i.e. regime-switching models. For a general overview, see, for instance, Hamilton (1989) and Kim and Nelson (1999), and for an application regarding monetary policy, see, for instance, Sims and Zha (2006). As argued by Primiceri (2005), discrete breaks may well describe some rapid shifts in policy but are less suitable to capture changes in private sector behavior, where changes are smoothed out due to learning and aggregation. Moreover, given that VARs consist of reduced-form parameters that are a mixture of underlying policy and private sector behavior, even rapid shifts in policy might lead to gradually changing coefficients rather than discrete breaks. From a different angle, regime-switching models are not capable of capturing gradually changing coefficients, but time-varying VARs are, as demonstrated in a Monte Carlo study by Baumeister and Peersman (2013), capable of capturing discrete breaks.

Finally, we have no formal procedure for determining how many factors to include in the reduced-rank model. We have simply estimated the reduced-rank model for various ranks and examined whether the

results remain the same as more factors are added to the system. By all means, a formal procedure based on the marginal likelihood or a reversible jump Markov chain Monte Carlo procedure (a procedure previously used in the context of time-varying VARs by Primiceri (2005) for prior selection) is preferred, which remains for future research.

The organization of the rest of this paper is as follows. In section 2, we first review the time-varying VAR model of Primiceri (2005), i.e. the full-rank model, and then we introduce the reduced-rank model. In section 3, we outline the Bayesian estimation procedure for the reduced-rank model. In section 4, we show how many factors are actually needed to capture the bulk of time variation in the small macro model that was earlier analyzed by Cogley and Sargent (2001). Finally, section 5 concludes.

## 2 Model

### 2.1 Full-rank model

The full-rank model is basically identical to the time-varying VAR model of Primiceri (2005), and includes both time-varying parameters and stochastic volatility.<sup>1</sup> Consider the model with  $n$  variables and  $p$  lags

$$y_t = c_t + \sum_{i=1}^p B_{i,t} y_{t-i} + u_t \quad (1)$$

where  $y_t$  is an  $n \times 1$  vector of observed endogenous variables,  $c_t$  is an  $n \times 1$  vector of time-varying constants,  $\{B_{i,t}\}_{i=1}^p$  are  $n \times n$  matrices of time-varying autoregressive parameters, and  $u_t$  is an  $n \times 1$  vector of shocks. Shocks are assumed to be distributed according to the normal distribution  $u_t \sim N(0, \Sigma_t)$ . The time-varying parameters and the stochastic volatility are discussed in turn.

Let  $\theta_t \equiv \text{vec}([c_t, \{B_{i,t}\}_{i=1}^p]')$  denote the vector of time-varying parameters, where  $\text{vec}$  is the column stacking operator. The dimension of  $\theta_t$  is  $k$  by 1 with  $k = n + pn^2$ . The law of motion for  $\theta_t$  is assumed to be equal to

$$\theta_t = \theta_{t-1} + \nu_{\theta,t} \quad (2)$$

where  $\nu_{\theta,t}$  is a  $k \times 1$  vector of shocks. Shocks are assumed to be distributed according to the normal distribution  $\nu_{\theta,t} \sim N(0, Q_{\theta})$ .

Introducing stochastic volatility is a bit more involved, since this requires extra restrictions to make

---

<sup>1</sup>There is one minor difference between the full-rank model discussed here and the model of Primiceri (2005). For details, see footnote 2.

sure that the covariance matrix is always positive definite. Cholesky decompose the covariance matrix  $\Sigma_t = (A_t^{-1}\Omega_t)(A_t^{-1}\Omega_t)'$  where  $A_t^{-1}$  is a lower triangular matrix with ones on the main diagonal and  $\Omega_t$  is a diagonal matrix. Let  $\alpha_t$  be the vector of elements below the main diagonal of the matrix  $A_t$  stacked by rows. The dimension of  $\alpha_t$  is  $r$  by 1 with  $r = \frac{n(n-1)}{2}$ . The law of motion for  $\alpha_t$  is assumed to be equal to

$$\alpha_t = \alpha_{t-1} + \nu_{\alpha,t} \quad (3)$$

where  $\nu_{\alpha,t}$  is an  $r \times 1$  vector of shocks. Shocks are assumed to be distributed according to the normal distribution  $\nu_{\alpha,t} \sim N(0, Q_\alpha)$ . Here,  $Q_\alpha$  is a *dense* matrix, in contrast to Primiceri (2005).<sup>2</sup> Let  $\sigma_t$  denote the vector of diagonal elements of the matrix  $\Omega_t$ . The dimension of  $\sigma_t$  is  $n$  by 1. The law of motion for  $\sigma_t$  is assumed to be equal to

$$\log(\sigma_t) = \log(\sigma_{t-1}) + \nu_{\sigma,t} \quad (4)$$

where  $\nu_{\sigma,t}$  is an  $n \times 1$  vector of shocks. Shocks are assumed to be distributed according to the normal distribution  $\nu_{\sigma,t} \sim N(0, Q_\sigma)$ . Adding to this, the four vectors of shocks discussed so far are assumed to be mutually uncorrelated.

Finally, define the matrix of regressors  $X_t' \equiv I_n \otimes [1, \{y_{t-i}'\}_{i=1}^p]$ , where  $\otimes$  denotes the Kronecker product, and rewrite the VAR equation in concise matrix form

$$y_t = X_t'\theta_t + A_t^{-1}\Omega_t\varepsilon_t \quad (5)$$

where  $X_t$  is a  $k \times n$  matrix of regressors and  $\varepsilon_t$  is an  $n \times 1$  vector of standard normally distributed shocks. All the above can now be represented by the following non-linear state-space representation

$$y_t = X_t'\theta_t + A_t^{-1}\Omega_t\varepsilon_t \quad (6a)$$

$$\theta_t = \theta_{t-1} + \nu_{\theta,t} \quad (6b)$$

$$\alpha_t = \alpha_{t-1} + \nu_{\alpha,t} \quad (6c)$$

$$\log(\sigma_t) = \log(\sigma_{t-1}) + \nu_{\sigma,t} \quad (6d)$$

The first equation is the measurement equation and the other ones are the state equations. The dimensions

---

<sup>2</sup>Primiceri (2005) imposes a block-diagonal structure on  $Q_\alpha$  in order to simplify the sampling procedure and to increase efficiency, but by doing a bit more algebra than Primiceri (2005) the sampling procedure can be made as simple and as efficient even if  $Q_\alpha$  is dense. This will be explained in section 3 about Bayesian estimation.

of the measurement and state equations are summarized in tables 1 and 2, respectively. The time index  $t$  runs from 1 to  $\tau$ . Initial values of the observations are assumed to be given and starting values of the time-varying parameters are assumed to be distributed according to normal prior distributions, to be discussed in section 3 about the Bayesian estimation procedure.

	Vector		Matrix		
Data	$y_t$	$n \times 1$	$X_t$	$k \times n$	
Parameter	$\theta_t$	$k \times 1$	$A_t$	$n \times n$	$\Omega_t \quad n \times n$
Shock	$\varepsilon_t$	$n \times 1$			

**Table 1: Dimensions of measurement equation (with  $k = n + pn^2$ )**

	Parameter		Shock		Covariance	
Regression	$\theta_t$	$k \times 1$	$\nu_{\theta,t}$	$k \times 1$	$Q_\theta$	$k \times k$
Off-diagonal	$\alpha_t$	$r \times 1$	$\nu_{\alpha,t}$	$r \times 1$	$Q_\alpha$	$r \times r$
Diagonal	$\sigma_t$	$n \times 1$	$\nu_{\sigma,t}$	$n \times 1$	$Q_\sigma$	$n \times n$

**Table 2: Dimensions of state equations (with  $k = n + pn^2$  and  $r = \frac{n(n-1)}{2}$ )**

## 2.2 Reduced-rank model

### 2.2.1 Motivation

The full-rank model is likely to be overparameterized. As pointed out in the introduction, common sense and estimated spectral concentrations suggest that only a few factors are important in driving the time-varying parameters and stochastic volatility. This idea is implemented in the reduced-rank model via the underlying structure put on the covariance matrices. The covariance matrix driving the time-varying parameters ( $Q_\theta$ ) and the covariance matrices driving the stochastic volatility ( $Q_\alpha$  and  $Q_\sigma$ ) are now of reduced rank instead of full rank. This corresponds to a reduction in the number of underlying factors driving the time-varying parameters and stochastic volatility.

Although the rank is the only dimension in which the two models differ, the modification requires substantial changes in the Bayesian estimation procedure used for estimation. The Bayesian estimation procedure will be outlined in section 3, but the state-space representation will be discussed already in the next subsection.

### 2.2.2 State-space representation

Starting with the time-varying parameters, there are assumed to be  $q_\theta \leq k$  common factors driving the  $k$  time-varying parameters. This means that the covariance matrix  $Q_\theta$  has less than full rank. Decompose the covariance matrix  $Q_\theta = \Lambda_\theta \Lambda_\theta'$  where  $\Lambda_\theta$  is a  $k \times q_\theta$  matrix of factor loadings implying  $\text{rank}(Q_\theta) = q_\theta$ . The law of motion for the time-varying parameters can now be written as

$$\theta_t = \theta_{t-1} + \Lambda_\theta v_{\theta,t} \quad (7)$$

where  $v_{\theta,t}$  is a  $q_\theta \times 1$  vector of standard normally distributed shocks. This law of motion implies that (i)  $\Delta\theta_t$  is in the column space of  $\Lambda_\theta$  and (ii)  $\theta_t$  is not necessarily in the column space of  $\Lambda_\theta$ . Hence, the *changes* in the parameters are driven by the common factors, but that is not necessarily the case for the *levels* of the parameters. This makes sense since there are (much) more forces determining the macroeconomy than there are forces changing the macroeconomy.

The time-varying part of the parameters is in the column space of  $\Lambda_\theta$  and the time-invariant part of the parameters is in the left null space of  $\Lambda_\theta$ . A natural decomposition for the time-varying parameters is therefore

$$\theta_t = P_\theta \theta_t + M_\theta \theta_t \quad (8)$$

where  $P_\theta = \Lambda_\theta (\Lambda_\theta' \Lambda_\theta)^{-1} \Lambda_\theta'$  is the projection matrix onto the column space of  $\Lambda_\theta$  and  $M_\theta = I_k - P_\theta$  is the projection matrix onto the left null space of  $\Lambda_\theta$ . The first term on the right-hand side is in the column space of  $\Lambda_\theta$  and can thus be expressed in terms of the underlying factors. The second term on the right-hand side is in the left null space of  $\Lambda_\theta$  and is thus time-invariant. Hence, the  $q_\theta \times 1$  vector of underlying factors  $\tilde{\theta}_t$  can be defined implicitly by

$$\theta_t \equiv \Lambda_\theta \tilde{\theta}_t + M_\theta \theta_0 \quad (9)$$

Substitute the implicit definition of the factors in expression (7) and premultiply with  $(\Lambda_\theta' \Lambda_\theta)^{-1} \Lambda_\theta'$ . The law of motion is now expressed in terms of the underlying factors

$$\tilde{\theta}_t = \tilde{\theta}_{t-1} + v_{\theta,t} \quad (10)$$

The stochastic volatility is based on similar decompositions. The state-space representation can

therefore be given directly. The measurement equation is the same as in the full-rank model. Since the state equations are now expressed in terms of the underlying factors, the dimensions of the state equations are smaller than in the full-rank model. This speeds up the Bayesian estimation procedure, which is now based on the following non-linear state-space representation

$$y_t = X_t' \theta_t + A_t^{-1} \Omega_t \varepsilon_t \quad (11a)$$

$$\tilde{\theta}_t = \tilde{\theta}_{t-1} + v_{\theta,t} \quad \theta_t \equiv \Lambda_\theta \tilde{\theta}_t + M_\theta \theta_0 \quad (11b)$$

$$\tilde{\alpha}_t = \tilde{\alpha}_{t-1} + v_{\alpha,t} \quad \alpha_t \equiv \Lambda_\alpha \tilde{\alpha}_t + M_\alpha \alpha_0 \quad (11c)$$

$$\tilde{\sigma}_t = \tilde{\sigma}_{t-1} + v_{\sigma,t} \quad \log \sigma_t \equiv \Lambda_\sigma \tilde{\sigma}_t + M_\sigma \log(\sigma_0) \quad (11d)$$

The first equation is the measurement equation and the other ones are the state equations together with the mappings between the underlying factors and the time-varying parameters. The dimensions of the measurement and state equations are summarized in tables 1 and 3, respectively. The time index  $t$  runs from 1 to  $\tau$ . Initial values of the observations are assumed to be given and starting values of the time-varying parameters are assumed to be distributed according to normal prior distributions, to be discussed in section 3 about the Bayesian estimation procedure.

	Factor	Shock	Loading
Regression	$\tilde{\theta}_t$ $q_\theta \times 1$	$v_{\theta,t}$ $q_\theta \times 1$	$\Lambda_\theta$ $k \times q_\theta$
Off-diagonal	$\tilde{\alpha}_t$ $q_\alpha \times 1$	$v_{\alpha,t}$ $q_\alpha \times 1$	$\Lambda_\alpha$ $r \times q_\alpha$
Diagonal	$\tilde{\sigma}_t$ $q_\sigma \times 1$	$v_{\sigma,t}$ $q_\sigma \times 1$	$\Lambda_\sigma$ $n \times q_\sigma$

**Table 3: Dimensions of state equations (with  $k = n + pn^2$  and  $r = \frac{n(n-1)}{2}$ )**

### 3 Bayesian estimation procedure for reduced-rank model

The reduced-rank time-varying VAR model is estimated using Bayesian methods. Inference is based on the joint posterior distribution of the time-varying parameters<sup>3</sup> and the reduced-rank covariance matrices  $p(\theta^\tau, Q_\theta, \alpha^\tau, Q_\alpha, \sigma^\tau, Q_\sigma | y^\tau, X^\tau)$ ,<sup>4</sup> where the notation  $x^\tau$  is used to denote the complete history of  $x$  up

<sup>3</sup>In this subsection, the notion of time-varying parameters also refers to stochastic volatility.

<sup>4</sup>The marginal posterior distributions of the underlying factors and factor loadings are not informative since they are only separately identified up to an orthogonal transformation, see footnote 6 for a discussion. The marginal posterior distributions

to and including time  $\tau$ . The posterior distribution can be decomposed as

$$p(\theta^\tau, Q_\theta, \alpha^\tau, Q_\alpha, \sigma^\tau, Q_\sigma | y^\tau, X^\tau) \propto p(y^\tau, X^\tau | \theta^\tau, Q_\theta, \alpha^\tau, Q_\alpha, \sigma^\tau, Q_\sigma) p(\theta^\tau, Q_\theta, \alpha^\tau, Q_\alpha, \sigma^\tau, Q_\sigma) \quad (12)$$

where the first term on the right-hand side is the likelihood function and the second term on the right-hand side is the joint prior distribution. The joint prior distribution used in this paper is constructed from

$$p(\theta_0) p(Q_\theta) p(\alpha_0) p(Q_\alpha) p(\sigma_0) p(Q_\sigma) \quad (13)$$

where the priors for the reduced-rank covariance matrices are independent singular inverse-Wishart distributions—to be defined in section 3.2—and the priors for the *starting* values of the time-varying parameters are independent normal distributions. Note that the normal priors in (13) are only for the starting values and not for the entire sequences. But conditional on the reduced-rank covariance matrices, the normal priors for the starting values imply—via the random walk processes—normal priors for the entire sequences. Hence, the implied joint prior distribution is

$$p(\theta^\tau, Q_\theta, \alpha^\tau, Q_\alpha, \sigma^\tau, Q_\sigma) = \prod_{t=1}^{\tau} p(\theta_t | \theta_{t-1}, Q_\theta) \prod_{t=1}^{\tau} p(\alpha_t | \alpha_{t-1}, Q_\alpha) \prod_{t=1}^{\tau} p(\sigma_t | \sigma_{t-1}, Q_\sigma) \quad (14)$$

Analytical evaluation of the joint posterior distribution is not possible. Gibbs sampling is therefore used for numerical evaluation. For a general treatment of the Gibbs sampler, see, for example, Casella and George (1992). Here, only the implementation for the model under consideration is given:

0. Initialize the Gibbs sampler with  $\theta^{\tau,0}$ ,  $Q_\theta^0$ ,  $\alpha^{\tau,0}$ ,  $Q_\alpha^0$ ,  $\sigma^{\tau,0}$ ,  $Q_\sigma^0$ , and set  $i$  to 1.
1. Draw a history of regression parameters  $\theta^{\tau,i}$  from  $p(\theta^\tau | Q_\theta^{i-1}, \alpha^{\tau,i-1}, Q_\alpha^{i-1}, \sigma^{\tau,i-1}, Q_\sigma^{i-1}; y^\tau, X^\tau)$ .
2. Draw a reduced-rank covariance matrix  $Q_\theta^i$  from  $p(Q_\theta | \theta^{\tau,i}, \alpha^{\tau,i-1}, Q_\alpha^{i-1}, \sigma^{\tau,i-1}, Q_\sigma^{i-1}; y^\tau, X^\tau)$ .
3. Draw a history of off-diagonal elements  $\alpha^{\tau,i}$  from  $p(\alpha^\tau | \theta^{\tau,i}, Q_\theta^i, Q_\alpha^{i-1}, \sigma^{\tau,i-1}, Q_\sigma^{i-1}; y^\tau, X^\tau)$ .
4. Draw a reduced-rank covariance matrix  $Q_\alpha^i$  from  $p(Q_\alpha | \theta^{\tau,i}, Q_\theta^i, \alpha^{\tau,i}, \sigma^{\tau,i-1}, Q_\sigma^{i-1}; y^\tau, X^\tau)$ .
5. Draw a history of diagonal elements  $\sigma^{\tau,i}$  from  $p(\sigma^\tau | \theta^{\tau,i}, Q_\theta^i, \alpha^{\tau,i}, Q_\alpha^i, Q_\sigma^{i-1}; y^\tau, X^\tau)$ .
6. Draw a reduced-rank covariance matrix  $Q_\sigma^i$  from  $p(Q_\sigma | \theta^{\tau,i}, Q_\theta^i, \alpha^{\tau,i}, Q_\alpha^i, \sigma^{\tau,i}; y^\tau, X^\tau)$ .
7. Increment  $i$  by 1, and go back to step 1 until  $i > g$ .

---

of the time-varying parameters and reduced-rank covariance matrices are, on the contrary, separately identified, and hence it makes sense to work with their joint posterior distribution, i.e.  $p(\theta^\tau, Q_\theta, \alpha^\tau, Q_\alpha, \sigma^\tau, Q_\sigma | y^\tau, X^\tau)$ .

8. Throw away the first  $b$  draws as burn-in and keep the remaining  $g - b$  draws for inference.

Note that at each step one of the conditioning variables is updated. Subject to regularity conditions (see, for example, Roberts and Smith, 1994), the Gibbs sampler generates a Markov chain with  $p(\theta^\tau, Q_\theta, \alpha^\tau, Q_\alpha, \sigma^\tau, Q_\sigma | y^\tau, X^\tau)$  as stationary distribution. It is important to check convergence of the Markov chain, to be discussed concurrently with the applications. The various drawing steps are outlined in the remainder of this section.<sup>5</sup>

Also note that the full-rank model is nested within the more general reduced-rank model and can thus be estimated by the same procedure. Yet, some of the steps needed for the reduced-rank model are not deemed necessary for the full-rank model, and hence it is easier to estimate the full-rank model by the procedure outlined in Primiceri (2005).

### 3.1 Draw a history of regression parameters $\theta^\tau$

Recall that  $\theta_t$  is the  $k \times 1$  vector of time-varying parameters including both the constants and the autoregressive parameters. In this step, a history of  $\theta^\tau$  is sampled from the conditional distribution  $p(\theta^\tau | Q_\theta^{i-1}, \alpha^{\tau, i-1}, Q_\alpha^{i-1}, \sigma^{\tau, i-1}, Q_\sigma^{i-1}; y^\tau, X^\tau)$ . This conditional distribution does depend on  $\alpha^{\tau, i-1}$  and  $\sigma^{\tau, i-1}$  only via  $\Sigma^{\tau, i-1}$  and does not depend on  $Q_\alpha^{i-1}$  and  $Q_\sigma^{i-1}$ , which implies that the conditional distribution simplifies to  $p(\theta^\tau | Q_\theta^{i-1}, \Sigma^{\tau, i-1}; y^\tau, X^\tau)$ . In the rest of the discussion, dependence on  $i$  is suppressed for notational convenience. Sampling is based on the following state-space representation

$$y_t = X_t' \theta_t + u_t \tag{15a}$$

$$\tilde{\theta}_t = \tilde{\theta}_{t-1} + v_{\theta, t} \quad \theta_t \equiv \Lambda_\theta \tilde{\theta}_t + M_\theta \theta_0 \tag{15b}$$

where  $u_t \sim N(0, \Sigma_t)$  and  $v_{\theta, t} \sim N(0, I_{q_\theta})$ . Note that the sampling procedure is conditional on  $Q_\theta$  while the state-space representation is based on  $\Lambda_\theta$ . The conditional distribution of  $\theta^\tau$  does not depend on the particular decomposition of  $Q_\theta$  used which is crucial because the decomposition  $\Lambda_\theta \Lambda_\theta' = Q_\theta$  is not

---

<sup>5</sup>The Bayesian estimation procedure simplifies considerably when the stochastic volatility feature is left out. In particular, steps 3–6 of the Gibbs sampling procedure are replaced by one single step to draw the covariance matrix of the VAR forecast errors ( $\Sigma$ ), which is now time-invariant. The sampling procedure is similar to the procedure as outlined in section 3.2, yet with full-rank residual sum-of-squares and prior scale matrices. To be precise, when the prior distribution for the covariance matrix of the VAR forecast errors is specified as  $IW(\bar{\Sigma}, \tau_{\Sigma, 0})$ , the sampling distribution is  $IW(S_u + \bar{\Sigma}, \tau + \tau_{\Sigma, 0})$ , where  $S_u$  is the residual sum-of-squares matrix of the VAR forecast errors.

unique.<sup>6</sup> Here, the eigenvalue decomposition is used, that is

$$\Lambda_\theta = V_\theta D_\theta \quad (16)$$

where  $D_\theta$  is a  $q_\theta \times q_\theta$  diagonal matrix with the square roots of the non-zero eigenvalues of  $\Lambda_\theta$  on the diagonal and  $V_\theta$  is a  $k \times q_\theta$  matrix with the corresponding eigenvectors (normalized to unit length). A history of regression parameters  $\theta^\tau$  is sampled indirectly by combining (i) a draw for the time-varying part of the regression parameters  $\Lambda_\theta \tilde{\theta}^\tau$  with (ii) a draw for the time-invariant part of the regression parameters  $M_\theta \theta_0$ . These two steps—actually two separate Gibbs steps, as explained below—are discussed in turn.

### 3.1.1 Draw a history of the regression parameters' time-varying part $\Lambda_\theta \tilde{\theta}^\tau$

The procedure is to first draw a history of underlying factors  $\tilde{\theta}^\tau$  from the conditional distribution  $p(\tilde{\theta}^\tau | M_\theta^{i-1} \theta_0^{i-1}, \Lambda_\theta^{i-1}, \Sigma^{\tau, i-1}; y^\tau, X^\tau)$ , and then to premultiply this history by  $\Lambda_\theta^{i-1}$ . Note that the aforementioned distribution is also conditional on the previous draw for the regression parameters' time-invariant part  $M_\theta^{i-1} \theta_0^{i-1}$ , which implies that the two steps (to draw a history of regression parameters) are, in fact, two separate Gibbs steps. The conditioning is on the regression parameters' time-invariant part  $M_\theta^{i-1} \theta_0^{i-1}$  rather than on the starting values  $\theta_0^{i-1}$  because otherwise there would have been unnecessary autocorrelation in the Gibbs sampler, which would have been fatal for its convergence.<sup>7</sup> Again, in the rest of the discussion, dependence on  $i$  is suppressed for notational convenience.

First, to sample the underlying factors, substitute the implicit definition of the underlying factors in the measurement equation and rewrite. This gives

$$y_t^* = X_t^* \tilde{\theta}_t + u_t \quad (17a)$$

$$\tilde{\theta}_t = \tilde{\theta}_{t-1} + v_{\theta,t} \quad (17b)$$

where  $y_t^* = y_t - X_t' M_\theta \theta_0$  and  $X_t^* = X_t' \Lambda_\theta$ . This state-space representation is linear and Gaussian. The

---

<sup>6</sup>Indeed, any orthogonal transformation  $\tilde{\Lambda}_\theta = \Lambda_\theta Z$ , where  $Z$  is an orthogonal matrix, gives an alternative decomposition  $\tilde{\Lambda}_\theta \tilde{\Lambda}_\theta' = Q_\theta$ . Yet, the column space of  $\tilde{\Lambda}_\theta$  is identical to the column space of  $\Lambda_\theta$ , since the projection matrix onto the column space of  $\tilde{\Lambda}_\theta$  is equal to the projection matrix onto the column space of  $\Lambda_\theta$ . The column space is thus invariant to the particular orthogonal transformation used, which is the only requisite for uniquely determining the conditional distribution of  $\theta^\tau$ .

<sup>7</sup>If the conditioning had instead been on  $\theta_0^{i-1}$ , the distribution of  $\tilde{\theta}_0$  would have been degenerate. Then, in the next step of the Gibbs sampler, the distribution of  $\theta_0$  would have been degenerate with respect to the left null space of  $\Lambda_\theta$ . This circle would have been fatal for the convergence of the Gibbs sampler.

underlying factors can therefore be sampled via the Kalman filter and a backward recursion as described in Carter and Kohn (1994). The Kalman filter is given by

$$P_{t|t-1} = P_{t-1|t-1} + I_{q_\theta} \quad (18a)$$

$$K_t = P_{t|t-1} X_t^{*'} (X_t^* P_{t|t-1} X_t^{*'} + \Sigma_t)^{-1} \quad (18b)$$

$$P_{t|t} = P_{t|t-1} - K_t X_t^* P_{t|t-1} \quad (18c)$$

$$\tilde{\theta}_{t|t} = \tilde{\theta}_{t-1|t-1} + K_t (y_t^* - X_t^* \tilde{\theta}_{t-1|t-1}) \quad (18d)$$

where the notation  $x_{\cdot|t}$  is used to condition on the information set up to and including time  $t$ . The initialization of the recursions follows from the normal prior distribution on the starting values of the time-varying parameters  $\theta_0 \sim N(\bar{\theta}, \bar{P}_\theta)$ . Via the implicit definition of the underlying factors it follows that  $\tilde{\theta}_{0|0} = R_\theta \bar{\theta}$  and  $P_{0|0} = R_\theta \bar{P}_\theta R_\theta'$  where  $R_\theta = (\Lambda_\theta' \Lambda_\theta)^{-1} \Lambda_\theta'$ . Draws for  $\tilde{\theta}_T$  can be obtained by drawing from  $N(\tilde{\theta}_{T|T}, P_{T|T})$ , and successive draws for  $\tilde{\theta}_t$  can be obtained by drawing from  $N(\tilde{\theta}_{t|t+1}, P_{t|t+1})$  with

$$\tilde{\theta}_{t|t+1} = \tilde{\theta}_{t|t} + P_{t|t} P_{t+1|t}^{-1} (\tilde{\theta}_{t+1} - \tilde{\theta}_{t|t}) \quad (19a)$$

$$P_{t|t+1} = P_{t|t} - P_{t|t} P_{t+1|t}^{-1} P_{t|t} \quad (19b)$$

Finally, premultiply the history of underlying factors  $\tilde{\theta}^\tau$  by  $\Lambda_\theta^{i-1}$  to obtain a history of the regression parameters' time-varying part.

### 3.1.2 Draw a vector with the regression parameters' time-invariant part $M_\theta \theta_0$

To draw a vector with the regression parameters' time-invariant part  $M_\theta \theta_0$  from the conditional distribution  $p(M_\theta \theta_0 | \tilde{\theta}^{\tau, i}, \Lambda_\theta^{i-1}, \Sigma^{\tau, i-1}; y^\tau, X^\tau)$ —the underlying factors were just updated—, substitute the implicit definition of the underlying factors in the measurement equation and reorganize. This gives

$$y_t^* = X_t' M_\theta \theta_0 + u_t \quad (20)$$

where  $y_t^* = y_t - X_t' \Lambda_\theta \tilde{\theta}_t$ , and again dependence on  $i$  is suppressed for notational convenience. Stack the regressands in  $y^*$ , the regressors in  $X^*$ , and the shocks in  $u^*$ . Define  $\Sigma^* \equiv \text{diag}(\{\Sigma_t\}_{t=1}^\tau)$  and rewrite the above as

$$y^* = X^* M_\theta \theta_0 + u^* \quad u^* \sim N(0, \Sigma^*) \quad (21)$$

This equation can be interpreted as restricted linear regression, where  $y^*$  is projected on  $X^*$  under the restriction that the regression parameters are in the left null space of  $P_\theta$  (column space of  $M_\theta$ ). A draw for the regression parameters, from now on denoted by  $\vartheta$  instead of  $M_\theta\theta_0$ , should thus satisfy

$$0 = P_\theta\vartheta \quad (22)$$

This formulation amounts to  $k$  equations with only  $q_\theta = \text{rank}(P_\theta)$  underlying restrictions. The restricted linear regression model can be written more compactly as

$$y^* = X^*\vartheta + u^* \quad u^* \sim N(0, \Sigma^*) \quad (23a)$$

$$0 = R_\theta\vartheta \quad (23b)$$

where, again,  $R_\theta = (\Lambda'_\theta\Lambda_\theta)^{-1}\Lambda'_\theta$ . Formulae for the Bayesian estimation of the restricted linear regression model are developed in appendix A. Given the normal prior distribution on the starting values of the time-varying parameters  $\theta_0 \sim N(\bar{\theta}, \bar{P}_\theta)$ , the sampling distribution is normal

$$\vartheta \sim N(\vartheta_\mu, \vartheta_\sigma) \quad (24a)$$

$$\vartheta_\mu = (I_k - \theta_\sigma R'_\theta (R_\theta\theta_\sigma R'_\theta)^{-1} R_\theta)\theta_\mu \quad (24b)$$

$$\vartheta_\sigma = (I_k - \theta_\sigma R'_\theta (R_\theta\theta_\sigma R'_\theta)^{-1} R_\theta)\theta_\sigma \quad (24c)$$

where the mean  $\vartheta_\mu$  and variance  $\vartheta_\sigma$  of the restricted sampling distribution are functions of the mean  $\theta_\mu$  and variance  $\theta_\sigma$  of the unrestricted (standard Bayesian regression) sampling distribution

$$\theta_\mu = (X^{*\prime}\Sigma^{*-1}X^* + \bar{P}_\theta^{-1})^{-1}(X^{*\prime}\Sigma^{*-1}y^* + \bar{P}_\theta^{-1}\bar{\theta}) \quad (25a)$$

$$\theta_\sigma = (X^{*\prime}\Sigma^{*-1}X^* + \bar{P}_\theta^{-1})^{-1} \quad (25b)$$

### 3.1.3 Draw a history of regression parameters $\theta^\tau$

Finally, combine the draw with the regression parameters' time-varying part and the draw with the regression parameters' time-invariant part using  $\theta_t \equiv \Lambda_\theta\tilde{\theta}_t + M_\theta\theta_0$ , yielding a history of regression parameters  $\theta^\tau$ . This completes the sampling from the conditional distribution  $p(\theta^\tau|Q_\theta^{i-1}, \Sigma^{\tau, i-1}; y^\tau, X^\tau)$ , which—as explained above—is, in fact, based on two separate Gibbs steps.

### 3.2 Draw a reduced-rank covariance matrix $Q_\theta$

Recall that  $Q_\theta$  is the reduced-rank covariance matrix driving the time-varying parameters. In this step, a reduced-rank covariance matrix  $Q_\theta$  is sampled from the conditional distribution  $p(Q_\theta | \theta^{\tau,i}, \alpha^{\tau,i-1}, Q_\alpha^{i-1}, \sigma^{\tau,i-1}, Q_\sigma^{i-1}; y^\tau, X^\tau)$ . The conditional distribution of  $Q_\theta$  does depend only on  $\theta^{\tau,i-1}$ , which implies that the conditional distribution simplifies to  $p(Q_\theta | \theta^{\tau,i})$ . Again, in the rest of the discussion, dependence on  $i$  is suppressed for notational convenience. Sampling is based on the following state equation

$$\theta_t = \theta_{t-1} + \nu_{\theta,t} \quad \nu_{\theta,t} \sim N(0, Q_\theta) \quad (26)$$

The complete history of residuals  $\nu_\theta^\tau$  is observed conditional on  $\theta^\tau$ . Stack the residuals columnwise in a  $k \times \tau$  matrix  $\nu_\theta^*$ . The residual sum-of-squares matrix is  $S_\theta = \nu_\theta^* \nu_\theta^{*'}.$  Note that  $\text{rank}(S_\theta) = q_\theta$  implying that  $S_\theta$  is a reduced-rank matrix. The prior on the covariance matrix  $Q_\theta$  is assumed to be an inverse-Wishart distribution with  $k \times k$  scale matrix  $\bar{Q}_\theta$  with singular rank  $(\bar{Q}_\theta) = q_\theta$  and  $\tau_{\theta,0}$  degrees of freedom

$$SIW(\bar{Q}_\theta, \tau_{\theta,0}, k, q_\theta) \quad (27)$$

The singular inverse-Wishart prior distribution is conjugate and the sampling distribution of  $Q_\theta$  is the singular inverse-Wishart distribution with  $k \times k$  scale matrix  $S_\theta + \bar{Q}_\theta$  with rank  $(\bar{Q}_\theta) = q_\theta$  and  $\tau + \tau_{\theta,0}$  degrees of freedom

$$SIW(S_\theta + \bar{Q}_\theta, \tau + \tau_{\theta,0}, k, q_\theta) \quad (28)$$

The procedure to draw from a singular inverse-Wishart distribution is explained in appendix B.

### 3.3 Draw a history of off-diagonal elements $\alpha^\tau$

Recall that  $\alpha_t$  is the  $r \times 1$  vector of elements below the main diagonal of the matrix  $A_t$  stacked by rows, i.e. the off-diagonal elements. In this step, a history  $\alpha^\tau$  is sampled from the conditional distribution  $p(\alpha^\tau | \theta^{\tau,i}, Q_\theta^i, Q_\alpha^{i-1}, \sigma^{\tau,i-1}, Q_\sigma^{i-1}; y^\tau, X^\tau)$ . The conditional distribution of  $\alpha^\tau$  does not depend on  $Q_\theta^i$  and  $Q_\sigma^{i-1}$ , which implies that the conditional distribution simplifies to  $p(\alpha^\tau | \theta^{\tau,i}, Q_\alpha^{i-1}, \sigma^{\tau,i-1}; y^\tau, X^\tau)$ . Again, in the rest of the discussion, dependence on  $i$  is suppressed for notational convenience. Sampling is based

on the following state-space representation

$$y_t = X_t' \theta_t + A_t^{-1} \Omega_t \varepsilon_t \quad (29a)$$

$$\tilde{\alpha}_t = \tilde{\alpha}_{t-1} + v_{\alpha,t} \quad \alpha_t \equiv \Lambda_\alpha \tilde{\alpha}_t + M_\alpha \alpha_0 \quad (29b)$$

where  $\varepsilon_t \sim N(0, I_n)$  and  $v_{\alpha,t} \sim N(0, I_{q_\alpha})$ . The above state-space representation is non-linear, yet a linear state-space representation can be uncovered with several steps of algebra. Rewrite the measurement equation as

$$A_t \hat{y}_t = \Omega_t \varepsilon_t \quad (30)$$

where  $\hat{y}_t = y_t - X_t' \theta_t$ . Since  $A_t$  is a lower triangular matrix with ones on the main diagonal, it can be decomposed as  $A_t = I_n + A_t^*$  where  $A_t^*$  is a lower triangular matrix with non-zero elements only below the main diagonal (equal to the corresponding elements in  $A_t$ ). The measurement equation can now be written as

$$\hat{y}_t = -A_t^* \hat{y}_t + \Omega_t \varepsilon_t \quad (31)$$

Rewrite this expression using the vec operator and the rule  $A = BCD \Rightarrow \text{vec}(A) = (B \otimes D') \text{vec}(C')$  with  $B = I_n$ . The measurement equation then becomes

$$\hat{y}_t = -(I_n \otimes \hat{y}_t') \text{vec}(A_t^{*'}) + \Omega_t \varepsilon_t \quad (32)$$

The vector  $\text{vec}(A_t^{*'})$  consists of many zero elements, namely the elements  $\{(i-1)n + j\}_{j=i}^n$  for  $i = 1, 2, \dots, n$ . These elements can be deleted from  $\text{vec}(A_t^{*'})$  leaving behind exactly  $\alpha_t$ . But then also the corresponding columns must be deleted from  $-(I_n \otimes \hat{y}_t')$  defining  $Z_t$ . The state space can now be represented as

$$\hat{y}_t = Z_t \alpha_t + \Omega_t \varepsilon_t \quad (33a)$$

$$\tilde{\alpha}_t = \tilde{\alpha}_{t-1} + v_{\alpha,t} \quad \alpha_t \equiv \Lambda_\alpha \tilde{\alpha}_t + M_\alpha \alpha_0 \quad (33b)$$

This state-space representation is linear and Gaussian. Note that  $Z_t$  is not predetermined at time  $t$ , which suggests that conditional distributions cannot be computed using the standard Kalman filter. Yet, conditional distributions can still be computed as usual using the standard Kalman filter, because of the special nature of  $Z_t$ , i.e. the  $i$ th element of  $\hat{y}_t$  does not depend on the  $j$ th element of  $\hat{y}_t$  for  $i = 1, 2, \dots, n$

and  $j > i$ .<sup>8</sup> A history of off-diagonal elements  $\alpha^\tau$  can thus be sampled via the same procedure as outlined in section 3.1.

### 3.4 Draw a reduced-rank covariance matrix $Q_\alpha$

Recall that  $Q_\alpha$  is the reduced-rank covariance matrix driving the off-diagonal elements. The procedure to draw a reduced-rank covariance matrix  $Q_\alpha$  from  $p(Q_\alpha | \theta^{\tau,i}, Q_\theta^i, \alpha^{\tau,i}, \sigma^{\tau,i-1}, Q_\sigma^{i-1}; y^\tau, X^\tau) = p(Q_\alpha | \alpha^{\tau,i})$  is identical to the procedure as outlined in section 3.2.

### 3.5 Draw a history of diagonal elements $\sigma^\tau$

Recall that  $\sigma_t$  is the  $n \times 1$  vector of diagonal elements of the matrix  $\Omega_t$ , i.e. the diagonal elements. In this step, a history  $\sigma^\tau$  is sampled from the conditional distribution  $p(\sigma^\tau | \theta^{\tau,i}, Q_\theta^i, \alpha^{\tau,i}, Q_\alpha^i, Q_\sigma^{i-1}; y^\tau, X^\tau)$ . The conditional distribution of  $\sigma^\tau$  does not depend on  $Q_\theta^i$  and  $Q_\alpha^i$ , which implies that the conditional distribution simplifies to  $p(\sigma^\tau | \theta^{\tau,i}, \alpha^{\tau,i}, Q_\sigma^{i-1}; y^\tau, X^\tau)$ . Again, in the rest of the discussion, dependence on  $i$  is suppressed for notational convenience. Sampling is based on the following state-space representation

$$y_t = X_t' \theta_t + A_t^{-1} \Omega_t \varepsilon_t \quad (34a)$$

$$\tilde{\sigma}_t = \tilde{\sigma}_{t-1} + v_{\sigma,t} \quad \log(\sigma_t) \equiv \Lambda_\sigma \tilde{\sigma}_t + M_\sigma \log(\sigma_0) \quad (34b)$$

where  $\varepsilon_t \sim N(0, I_n)$  and  $v_{\sigma,t} \sim N(0, I_{q_\sigma})$ . The above state-space representation is non-linear, yet an *approximate* linear Gaussian state-space representation can be uncovered with several steps of algebra. Rewrite the measurement equation as

$$y_t^* = \Omega_t \varepsilon_t \quad (35)$$

where  $y_t^* = A_t(y_t - X_t' \theta_t)$ . The measurement equation can be converted into a linear measurement equation by squaring and taking logs. The state space can now be represented as

$$y_t^{**} = 2 \log(\sigma_t) + \log(\varepsilon_t^2) \quad (36a)$$

$$\tilde{\sigma}_t = \tilde{\sigma}_{t-1} + v_{\sigma,t} \quad \log(\sigma_t) \equiv \Lambda_\sigma \tilde{\sigma}_t + M_\sigma \log(\sigma_0) \quad (36b)$$

---

<sup>8</sup>Primiceri (2005)—with the same measurement equation—argues that conditional distributions can be computed by running the Kalman filter equation by equation, but not by running the standard Kalman filter. This is not true, i.e. conditional distributions can still be computed as usual using the standard Kalman filter. This can be checked numerically by running both recursions and observing that the generated numbers are the same.

where  $y_t^{**} = \log(y_t^{*2} + \bar{c})$ .<sup>9</sup> This state-space representation is linear but non-Gaussian.

All the shocks in the measurement equation are distributed according to independent  $\log \chi^2(1)$  distributions. The  $\log \chi^2(1)$  distribution can be approximated by a mixture of seven normal distributions as proposed by Kim, Shephard, and Chib (1998) and as implemented by Primiceri (2005). The mixture probabilities  $m_j^p$ , the means  $m_j^\mu$ , and the variances  $m_j^v$  are given in table 4. These numbers are chosen in order to match a number of moments of the  $\log \chi^2(1)$  distribution (after accounting for the offset as indicated in the note under the table). The hierarchical nature of the mixture is exploited to sample a history of diagonal elements  $\sigma^\tau$  in two steps. The two steps—actually two separate Gibbs steps—are discussed in turn.<sup>10</sup>

First, let  $S$  be the matrix indicating which of the seven normal distributions is used for which shock. All the elements of the matrix  $S$  have a value from  $\{1, 2, \dots, 7\}$ . The dimension of  $S$  is  $n$  by  $\tau$  since there are  $n$  shocks in the measurement equation and there are  $\tau$  time periods. Conditional on a history of diagonal elements  $\sigma^\tau$ , the shocks in the measurement equations are observed. The shocks are independent across equations and time periods. All the elements of the matrix  $S$  can therefore be sampled independently. The conditional distribution of an element  $S_{i,t}$  is discrete with point mass at  $\{1, 2, \dots, 7\}$ . The probability mass function is proportional to

$$\Pr(S_{i,t} = j | y_{i,t}^{**}, \sigma_{i-1,t}) \propto f_N(y_{i,t}^{**} - 2 \log(\sigma_{i-1,t}) | m_j^\mu - 1.2704, m_j^v) m_j^p \quad j = 1, 2, \dots, 7 \quad (37)$$

where the mixture probabilities  $m_j^p$  are updated with the (Gaussian) likelihoods of observing a particular realization of the shock  $\log(\varepsilon_{i,t}^2) = y_{i,t}^{**} - 2 \log(\sigma_{i-1,t})$ . The normalizing constant is the reciprocal of the sum of the right-hand sides over  $j = 1, 2, \dots, 7$ .

Second, conditional on a matrix  $S^i$ , the state-space representation is linear and Gaussian, and a history of diagonal elements  $\sigma^\tau$  can now simply be sampled via the same procedure as outlined in section 3.1.

---

<sup>9</sup>The procedure is numerically more robust by adding an offset constant  $\bar{c}$  after taking squares and before taking logs. The offset constant was introduced into the stochastic volatility literature by Fuller (1996) with the intention to avoid taking logs of very small numbers, at the expense of creating a small upward bias. In practice, an offset constant of 0.001 works alright for standard macroeconomic series such as inflation, nominal interest, and unemployment when scaled as percentages.

<sup>10</sup>Del Negro and Primiceri (2013) have pointed out that the ordering of the two Gibbs steps is important and, in particular, that the ordering that was used by Primiceri (2005) is incorrect. The ordering presented in this paper is the correct one.

Component ( $j$ )	Probability ( $m_j^p$ )	Mean ( $m_j^\mu$ )	Variance ( $m_j^v$ )
1	0.00730	-10.12999	5.79596
2	0.10556	-3.97281	2.61369
3	0.00002	-8.56686	5.17950
4	0.04395	2.77786	0.16735
5	0.34001	0.61942	0.64009
6	0.24566	1.79518	0.34023
7	0.25750	-1.08819	1.26261

**Table 4: Mixture of seven normals, from Kim, Shephard, and Chib (1998)**

Note: The above numbers imply a mean of zero and must still be offset by adding the mean of the  $\log \chi^2(1)$  distribution, which is  $-1.2704$ .

### 3.6 Draw a reduced-rank covariance matrix $Q_\sigma$

Recall that  $Q_\sigma$  is the reduced-rank covariance matrix driving the diagonal elements. The procedure to draw a reduced-rank covariance matrix  $Q_\sigma$  from  $p(Q_\sigma | \theta^{\tau,i}, Q_\theta^i, \alpha^{\tau,i}, Q_\alpha^i, \sigma^{\tau,i}; y^\tau, X^\tau) = p(Q_\sigma | \sigma^{\tau,i})$  is identical to the procedure as outlined in section 3.2.

## 4 Number of factors needed in a small macro model

The techniques just described are applied to the estimation of a small macro model. In particular, this section first estimates the full-rank model on the same data as used in Cogley and Sargent (2001) and then estimates the reduced-rank model for the purpose of analyzing how many factors are actually needed to capture the time variation in the full-rank model. The main finding is that the reduced-rank model with only a few factors is enough to capture the bulk of time variation that is present in the full-rank model, which includes the full set of 21 factors. This suggests that the model of Cogley and Sargent (2001) could have been parameterized more efficiently.

### 4.1 Model description

The model of Cogley and Sargent (2001) is designed for describing time variation in the inflation-unemployment dynamics in the US after World War II, and includes a short-term interest rate, un-

employment, and inflation. There are two lags included in the VAR. The sample period is from 1948Q1 up to and including 2000Q4. Details on the application and the data can be found in Cogley and Sargent (2001). Here, we only discuss details on the model setup.

The model setup used here is different from Cogley and Sargent (2001) in two dimensions. First, they include a covariance term between the VAR forecast errors ( $u_t$ ) and the parameter innovations ( $\nu_{\theta,t}$ ). Second, they use a prior that rules out trajectories of time-varying parameters ( $\theta^T$ ) that imply explosive dynamics. Both features are left out in this paper since they are non-standard in the literature.<sup>11</sup> There is also a difference between the model setup used here and the full-rank model as described in section 2.1, i.e. the stochastic volatility feature is left out, as in Cogley and Sargent (2001).<sup>12</sup>

## 4.2 Prior

The strategy for choosing the prior is similar to Cogley and Sargent (2001)—similar but not identical since the model setup used here is slightly different and besides, additional steps have to be taken for the reduced-rank model. The period from 1948Q1 up to and including 1958Q4 is used as training sample, which is a period of eleven years. The first quarter is used to construct the inflation series and the next two quarters are used to initialize the VAR with two lags. In the end, 41 quarters are used to estimate a time-invariant VAR from which the prior is constructed. There are 34 degrees of freedom in the training sample since there are seven parameters per equation in the trivariate VAR with two lags and constant. The various prior components are discussed in turn.

First, the marginal prior on the covariance matrix of the VAR forecast errors ( $\Sigma$ ) is assumed to be an inverse-Wishart distribution with  $n \times n$  scale matrix  $\bar{R}$  and  $n + 1$  degrees of freedom. The scale matrix  $\bar{R}$  is chosen to be equal to  $n + 1$  times the time-invariant estimate of the covariance matrix of the VAR forecast errors in the training sample. Note that the multiplication with the degrees of freedom is because the scale matrix has a sum-of-squares interpretation in the inverse-Wishart distribution.

Second, the marginal prior on the starting values of the time-varying parameters ( $\theta_0$ ) is assumed to

---

<sup>11</sup>Apart from the differences regarding the model setup, Cogley and Sargent (2001) use filtered estimates for their reporting rather than smoothed estimates, which is the standard in the literature. This was met with criticism from Sims (2001), who argued that filtered estimates also contain a component of variation that is learning rather than actual time variation in the behavior of the economy. In fact, by using filtered estimates rather than smoothed estimates, Cogley and Sargent (2001) may have overreported the amount of actual time variation. In later work, Cogley and Sargent (2005) have changed their reporting to smoothed estimates.

<sup>12</sup>The Bayesian estimation procedure for the model setup without stochastic volatility is explained in footnote 5.

be a normal distribution with mean  $\bar{\theta}$  and covariance matrix  $\bar{P}_\theta$ . The mean  $\bar{\theta}$  is chosen to be equal to the time-invariant estimate of the vector of VAR parameters in the training sample and the covariance matrix  $\bar{P}_\theta$  is chosen to be equal to four times the asymptotic covariance matrix of the time-invariant estimate of the vector of VAR parameters in the training sample.

Finally, the marginal prior on the covariance matrix of the shocks driving the time-varying parameters ( $Q_\theta$ ) needs a bit more explanation since it depends on the chosen rank. For the full-rank model, the prior is identical to the prior of Cogley and Sargent (2001). The prior is assumed to be an inverse-Wishart distribution with  $k \times k$  scale matrix  $\bar{Q}_\theta^{\text{full}}$ —with full rank—and  $\tau_{\theta,0}^{\text{full}}$  degrees of freedom. The scale matrix  $\bar{Q}_\theta^{\text{full}}$  is chosen to be proportional to the asymptotic covariance matrix of the time-invariant estimate of the vector of VAR parameters in the training sample, and the constant-of-proportionality is chosen to be equal to 1E-4 multiplied with the degrees of freedom  $\tau_{\theta,0}^{\text{full}}$ . The degrees of freedom  $\tau_{\theta,0}^{\text{full}}$  are chosen to be equal to the degrees of freedom in the training sample, which is 34.

For the reduced-rank model, the prior is likewise assumed to be an inverse-Wishart distribution with  $k \times k$  scale matrix  $\bar{Q}_\theta^{\text{red}}$  but now with singular rank( $\bar{Q}_\theta^{\text{red}}$ ) =  $q_\theta$ . The singular scale matrix  $\bar{Q}_\theta^{\text{red}}$  is constructed from the non-singular scale matrix  $\bar{Q}_\theta^{\text{full}}$  by selecting the  $q_\theta$  most important eigenvalues, i.e. by replacing the  $k - q_\theta$  smallest eigenvalues by zeros in the eigenvalue decomposition. The prior for the reduced-rank model is supposed to imply the same amount of time variation as the prior for the full-rank model. This requires an upscaling since the sum of the eigenvalues—which measures the amount of time variation—has decreased by the rank reduction. This is exactly offset by multiplying the scale matrix with the sum of all the eigenvalues divided by the sum of the  $q_\theta$  largest eigenvalues. Moreover, the degrees of freedom are chosen to be equal to  $\tau_{\theta,0}^{\text{red}} = \tau_{\theta,0}^{\text{full}} - k + q_\theta$  in order to match exactly the expected sum of the eigenvalues in the reduced-rank model with the full-rank model.<sup>13</sup>

### 4.3 Results

We want to determine how many factors are needed in the (typical) small macro model given above. For this purpose, we have generated results for all possible ranks in the reduced-rank model as well as for the full-rank model. First, we present results for the full-rank model and show how structured the time variation is, which suggests that the reduced-rank model is a useful new tool for studying macroeconomic

---

<sup>13</sup>Note that the reduced-rank scale matrix  $\bar{Q}_\theta^{\text{red}}$  is constructed from the full-rank scale matrix  $\bar{Q}_\theta^{\text{full}}$ , which is based on the degrees of freedom in the full-rank model  $\tau_{\theta,0}^{\text{full}}$ . The way in which the prior for the reduced-rank model is constructed implies the same amount of time variation as the prior for the full-rank model.

questions. Next, we present results for the reduced-rank model and show that the reduced-rank model with only a few factors is indeed enough to capture the bulk of time variation that is present in the full-rank model.

The full-rank model is estimated using four Markov chains of 100,000 draws each of which every fourth draw is kept for posterior inference and the first 50% is thrown away as burn-in period. For our application, this is more than enough for the Markov chains to converge to the posterior distribution, which we have confirmed by comparing the results of various Markov chains that have started from different initial conditions.

The posterior medians of the trajectories of time-varying parameters are presented in figure 1, together with the posterior distribution of the trace of the covariance matrix driving the time-varying parameters. It is evident from the figure that most of the time variation is concentrated in particular elements of the vector of time-varying parameters. Figure 2 zooms in on the selection of parameters that display most of the time variation. The comovement between the various trajectories stands out in the figure, which suggests that the time variation is highly structured. The comovement is especially strong between parameters with the same regressand and the same regressor but with different lag.

The principal components of the posterior median of the covariance matrix driving the time-varying parameters confirm that the time variation is indeed highly structured—the contributions of the seven most important principal components (out of 21) are presented in table 5. The first principal component already accounts for about 70% of the variance of the shocks driving the time-varying parameters. The first two principal components account together for more than 90% and the first three principal components account together for even more than 95%. These numbers suggest that the reduced-rank model is a useful new tool for studying macroeconomic questions. Besides, Cogley and Sargent (2005) as well as Primiceri (2005) have already suggested that a formal treatment of cross-equation restrictions in the time variation would be a priority for future research. The reduced-rank model developed in this paper is, to the best of our knowledge, the first tool in this direction.

Principal component	Variance	Cumulative percentage of total variance
1	0.00168	70.2%
2	0.00050	91.0%
3	0.00011	95.8%
4	0.00003	97.0%
5	0.00003	98.2%
6	0.00002	98.9%
7	0.00001	99.2%

**Table 5: Principal components of the covariance matrix driving the time-varying parameters**

We have estimated the reduced-rank model for all possible ranks (using the same settings regarding the Markov chains as for the full-rank model). As mentioned above, the prior for the reduced-rank model is supposed to imply the same amount of time variation as the prior for the full-rank model, no matter the rank. The results are first compared on the basis of the key object for determining the amount of time variation in the parameters, which is the covariance matrix of the shocks driving the time-varying parameters.<sup>14</sup> The total amount of time variation can be measured by the trace of this matrix—this is the sum of the variances driving the time-varying parameters—and the contribution of the most important driving force can be measured by the largest eigenvalue. The posterior medians of these two statistics are presented in table 6 for the reduced-rank model for all possible ranks as well as for the full-rank model.

We did not only generate results for the VAR with two lags but also for a version with only one lag.<sup>15</sup> There are twelve parameters in the VAR with one lag, yet it is evident from table 6 that we need much

<sup>14</sup>The comparison based on the covariance matrix of the shocks driving the time-varying parameters can nicely be represented in one single table with the results for all possible ranks, which is not generally possible for comparisons based on other objects. Below we also present several figures in which we compare the results for the full-rank model with the results for the reduced-rank model for some particular ranks.

<sup>15</sup>The prior for the version with one lag is constructed in the same way as the prior for the version with two lags. Again, the degrees of freedom are chosen to be equal to the degrees of freedom in the training sample, which is now 38 since there are less parameters to be estimated. Moreover, the constant-of-proportionality is now based on 1E-3 rather than 1E-4, which more or less offsets the fact that the parameters are more precisely estimated in the version with one lag. Furthermore, the results for the version with one lag are generated with shorter Markov chains than the results for the version with two lags, as indicated below table 6.

less underlying factors to capture the bulk of time variation that is present in the full-rank model. With only five factors, the trace and largest eigenvalue are already the same as in the full-rank model (for the reported decimals). And with less than five factors, the differences are actually quite small—and besides, note that the deviations from the full-rank model are not necessarily a bad thing. There is no clear reason why we should believe that the full-rank numbers are the true numbers. In fact, the reduced-rank numbers with only a few factors could be more accurate as these numbers are less prone to overfitting. It would make sense to only include the most important and robust driving forces for the time-varying parameters and to leave out the other ones.

Apart from the argument about overfitting, there are also other arguments in favor of including only a few factors. The first one is about the computing time needed for the Bayesian estimation. The computing time is about one third less for the reduced-rank model with a few factors than for the full-rank model. For the VAR with two lags, the gain in terms of computing speed is even a factor four. In general, the larger the model, the larger the potential gain in computing speed. The second argument is also related to the computing time needed for the Bayesian estimation, yet in an indirect manner. Since there are less parameters and underlying trajectories to be estimated in the reduced-rank model, (much) shorter Markov chains are sufficient to achieve convergence.<sup>16</sup> This makes the reduced-rank model much more practical than the full-rank model, especially when one needs to include more variables and/or more lags.

A similar picture emerges for the main version with two lags. There are 21 parameters in this case, yet again it is evident from table 6 that we need much less underlying factors to capture the bulk of time variation that is present in the full-rank model. With only three factors, the trace and largest eigenvalue are already the same as in the full-rank model (for the reported decimals), which is even less than in the version with only one lag. This is good news as this means that including an additional lag in the VAR does not require many extra factors, if any. The results confirm—without any doubt—the principal component analysis that was presented in table 5 for the full-rank model. Again, this suggests that the model of Cogley and Sargent (2001) could have been parameterized more efficiently.

Of course, the comparison should not only rely on the covariance matrix driving the time-varying parameters, since other objects *might* give a different picture. However, this turns out not to be the case. In figure 3, we present, for various ranks, the posterior medians of the trajectories of the selection of parameters that display most of the time variation. For the parameters in the inflation equation, three factors are indeed enough to capture almost all of the time variation. But for the parameters in the

---

<sup>16</sup>This is not exploited for generating the table. The gains in computing speed are thus potentially much larger.

unemployment equation, three factors turn out not to be sufficient to capture the time variation present in the full-rank model. Yet, with four factors the pattern is already quite close, with six factors also the magnitude is getting closer, and with twelve factors the trajectories are almost indistinguishable.

Furthermore, only a few factors are needed to capture most of the time variation in the spectrum of inflation. Figures 4 and 5 present the evolution of the spectrum of inflation for the full-rank model and the reduced-rank model with three factors, respectively. With only three factors, the evolution is already quite close to the full-rank model, not only in terms of pattern but also in terms of magnitude. The remaining deviation in terms of magnitude disappears slowly when more and more factors are added to the reduced-rank model. Yet, note again that the deviations from the full-rank model are not necessarily a bad thing, as the reduced-rank results are less prone to overfitting.

## 5 Concluding remarks

Altogether, the main finding is that the time variation present in the full-rank model is highly structured and that the reduced-rank model with only a few factors is enough to capture the bulk of time variation. This is very good news since this allows us to use the new tool developed in this paper, which in comparison with the full-rank model (i) suffers much less from overparameterization, (ii) achieves convergence with much shorter Markov chains, and (iii) runs much faster on the computer. Moreover, the results make us feel confident that the reduced-rank model is also well-suited for larger models including more variables and/or more lags. If the time variation had, on the contrary, not been highly structured, it would have been difficult to estimate larger models as the number of parameters and underlying trajectories to be estimated would quickly have become large. But given that we can confine on the number of factors, we can potentially estimate larger models since the number of parameters and underlying trajectories to be estimated are kept under control.

$q_\theta$	$p = 1$			$p = 2$		
	$\text{tr}(Q_\theta)$	$\max(\text{eig}(Q_\theta))$	comp. time	$\text{tr}(Q_\theta)$	$\max(\text{eig}(Q_\theta))$	comp. time
1	0.0031	0.0031	361s	0.0021	0.0021	808s
2	0.0029	0.0026	365s	0.0022	0.0018	800s
3	0.0027	0.0024	366s	0.0025	0.0019	806s
4	0.0031	0.0028	376s	0.0025	0.0019	834s
5	0.0032	0.0029	387s	0.0025	0.0019	846s
6	0.0032	0.0029	403s	0.0025	0.0019	883s
7	0.0032	0.0029	423s	0.0025	0.0019	936s
8	0.0032	0.0029	451s	0.0025	0.0019	964s
9	0.0032	0.0029	481s	0.0025	0.0019	1079s
10	0.0032	0.0029	516s	0.0025	0.0019	1163s
11	0.0032	0.0029	569s	0.0025	0.0019	1276s
12	0.0032	0.0029	623s	0.0025	0.0018	1377s
13				0.0025	0.0018	1462s
14				0.0025	0.0018	1720s
15				0.0025	0.0018	1893s
16				0.0025	0.0018	1905s
17				0.0025	0.0018	2391s
18				0.0025	0.0019	2629s
19				0.0025	0.0019	2939s
20				0.0025	0.0019	3254s
21				0.0025	0.0019	3590s
full	0.0032	0.0029	542s	0.0025	0.0019	3323s

**Table 6: The posterior median of the trace and largest eigenvalue of the covariance matrix driving the time-varying parameters ( $Q_\theta$ ) for various number of factors ( $q_\theta$ ), for the version with one lags and the version with two lags—how many factors are needed?**

Note: The results for the version with one lag are based on four Markov chains of 50,000 draws each of which every second draw is kept for posterior inference and the results for the version with two lags are based on four Markov chains of 100,000 draws each of which every fourth draw is kept for posterior inference. Moreover, 50% is thrown away as burn-in period. The results are generated with an Intel(R) Core(TM) i7-860 processor running (slightly overclocked) at 3.5GHz and with 12GB of internal memory. The Matlab version used is R2008b 64-bit with Parallel Computing Toolbox and the operating system used is Windows 7 64-bit.

## A Bayesian estimation of restricted linear regression model

In this appendix, the sampling distribution of the restricted linear regression model is developed. To simplify the discussion, the least-squares estimator of the univariate restricted linear regression model is derived as preliminary result.

### Least-squares estimator restricted univariate linear regression model

Consider the univariate linear regression model

$$y = X\theta + u \quad (38)$$

where  $y$  is the  $n \times 1$  vector of regressands,  $X$  is the  $n \times k$  matrix of regressors,  $\theta$  is the  $k \times 1$  vector of parameters, and  $u$  is the  $n \times 1$  vector of shocks. Shocks are distributed according to the normal distribution  $u \sim N(0, \sigma^2 I_n)$ . The ordinary least-squares estimator for  $\theta$  is given by

$$\hat{\theta}_{ols} = (X'X)^{-1} X'y \quad (39)$$

Now consider the linear restriction

$$0 = R\theta \quad (40)$$

where  $R$  is a  $q \times k$  matrix with  $\text{rank}(R) = q \leq k$ . The restricted least-squares estimator for  $\theta$  can be derived by pooling the linear regression model together with the linear restriction

$$y = X\theta + u \quad (41a)$$

$$0 = R\theta + v \quad (41b)$$

Here,  $v$  is a  $q \times 1$  vector of shocks to the linear restriction. These shocks are distributed according to the normal distribution  $v \sim N\left(0, \frac{\sigma^2}{\lambda} I_q\right)$ . Of course,  $v$  should be equal to the zero vector. The idea is to use generalized least-squares on the pooled regression model for given  $\lambda$  and then let  $\lambda \rightarrow \infty$ . The generalized least-squares estimator for  $\theta$  for given  $\lambda$  is given by

$$\hat{\theta}_{gls}(\lambda) = (X'X + \lambda R'R)^{-1} X'y \quad (42)$$

Before letting  $\lambda \rightarrow \infty$ , rewrite this expression using the matrix inversion lemma.

**Lemma 1** *matrix inversion lemma.*  $(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$ .<sup>17</sup>

Choose  $A = X'X$ ,  $B = R'$ ,  $C = \lambda I_q$ , and  $D = R$ . Applying the matrix inversion lemma results in

$$\hat{\theta}_{gls}(\lambda) = \left( (X'X)^{-1} - (X'X)^{-1}R' \left( I_q \lambda^{-1} + R(X'X)^{-1}R' \right)^{-1} R(X'X)^{-1} \right) X'y \quad (43)$$

The generalized least-squares estimator can be expressed in terms of the ordinary least-squares estimator

$$\hat{\theta}(\lambda) = \left( I_k - (X'X)^{-1}R' \left( I_q \lambda^{-1} + R(X'X)^{-1}R' \right)^{-1} R \right) \hat{\theta}_{ols} \quad (44)$$

Now let  $\lambda \rightarrow \infty$ . The term  $I_q \lambda^{-1}$  simply drops out, because the matrix  $R(X'X)^{-1}R'$  has full rank. This results in the well-known restricted least-squares estimator

$$\hat{\theta}_{res} = \left( I_k - (X'X)^{-1}R' \left( R(X'X)^{-1}R' \right)^{-1} R \right) \hat{\theta}_{ols} \quad (45)$$

### Sampling distribution restricted linear regression model

Consider the multivariate linear regression model

$$y = X\theta + u \quad (46)$$

where  $y$  is the  $nT \times 1$  vector of regressands,  $X$  is the  $nT \times k$  matrix of regressors,  $\theta$  is the  $k \times 1$  vector of parameters, and  $u$  is the  $nT \times 1$  vector of shocks. Shocks are distributed according to the normal distribution  $u \sim N(0, \Omega)$  with  $\Omega = I_T \otimes \Sigma$ . Consider the normal conjugate prior distribution

$$\theta \sim N(\bar{\theta}, \bar{P}_\theta) \quad (47)$$

where  $\bar{\theta}$  is the  $k \times 1$  vector with prior means and  $\bar{P}_\theta$  is the  $k \times k$  matrix with prior covariances. The

---

<sup>17</sup>The matrix inversion lemma, also known as Woodbury formula, can be proved directly by checking that  $A + BCD$  times the right-hand side gives the identity matrix.

posterior distribution of  $\theta$  is given by

$$\theta \sim N(\theta_\mu, \theta_\sigma) \quad (48a)$$

$$\theta_\mu = (X'\Omega^{-1}X + \bar{P}_\theta^{-1})^{-1}(X'\Omega^{-1}y + \bar{P}_\theta^{-1}\bar{\theta}) \quad (48b)$$

$$\theta_\sigma = (X'\Omega^{-1}X + \bar{P}_\theta^{-1})^{-1} \quad (48c)$$

Now consider the linear restriction

$$0 = R\theta \quad (49)$$

where  $R$  is a  $q \times k$  matrix with  $\text{rank}(R) = q \leq k$ . The posterior distribution of  $\theta$  can be derived by pooling the multivariate linear regression model together with the prior information and the linear restriction

$$y = X\theta + u \quad (50a)$$

$$\bar{\theta} = I_k\theta + \zeta \quad (50b)$$

$$0 = R\theta + v \quad (50c)$$

Here,  $v$  is a  $q \times 1$  vector of shocks to the linear restriction. These shocks are distributed according to the normal distribution  $v \sim N(0, \frac{1}{\lambda}I_q)$ . Of course,  $v$  should be equal to the zero vector. The idea is to derive the posterior distribution of  $\theta$  for given  $\lambda$  and then let  $\lambda \rightarrow \infty$ . The posterior distribution of  $\theta$  for given  $\lambda$  is given by

$$\theta(\lambda) \sim N(\theta_\mu(\lambda), \theta_\sigma(\lambda)) \quad (51a)$$

$$\theta_\mu(\lambda) = (X'\Omega^{-1}X + \bar{P}_\theta^{-1} + \lambda R'R)^{-1}(X'\Omega^{-1}y + \bar{P}_\theta^{-1}\bar{\theta}) \quad (51b)$$

$$\theta_\sigma(\lambda) = (X'\Omega^{-1}X + \bar{P}_\theta^{-1} + \lambda R'R)^{-1} \quad (51c)$$

Before letting  $\lambda \rightarrow \infty$ , rewrite  $\theta_\sigma(\lambda)$  using the matrix inversion lemma given above. In particular, choose  $A = X'\Omega^{-1}X + \bar{P}_\theta^{-1}$ ,  $B = R'$ ,  $C = \lambda I_q$ , and  $D = R$ . Applying the matrix inversion lemma results in

$$\theta_\sigma(\lambda) = \frac{(X'\Omega^{-1}X + \bar{P}_\theta^{-1})^{-1} - (X'\Omega^{-1}X + \bar{P}_\theta^{-1})^{-1} R' (I_q \lambda^{-1} + R (X'\Omega^{-1}X + \bar{P}_\theta^{-1})^{-1} R')^{-1} R (X'\Omega^{-1}X + \bar{P}_\theta^{-1})^{-1}}{(X'\Omega^{-1}X + \bar{P}_\theta^{-1})^{-1} R' (I_q \lambda^{-1} + R (X'\Omega^{-1}X + \bar{P}_\theta^{-1})^{-1} R')^{-1} R (X'\Omega^{-1}X + \bar{P}_\theta^{-1})^{-1}} \quad (52)$$

The posterior distribution of  $\theta$  for given  $\lambda$  can be expressed in terms of the unrestricted (standard Bayesian

regression) posterior distribution of  $\theta$

$$\theta(\lambda) \sim N(\theta_\mu(\lambda), \theta_\sigma(\lambda)) \quad (53a)$$

$$\begin{aligned} \theta_\mu(\lambda) &= \left( \theta_\sigma - \theta_\sigma R' (I_q \lambda^{-1} + R \theta_\sigma R')^{-1} R \theta_\sigma \right) (X' \Omega^{-1} y + \bar{P}_\theta^{-1} \bar{\theta}) \\ &= \left( I_k - \theta_\sigma R' (I_q \lambda^{-1} + R \theta_\sigma R')^{-1} R \right) \theta_\mu \end{aligned} \quad (53b)$$

$$\theta_\sigma(\lambda) = \left( I_k - \theta_\sigma R' (I_q \lambda^{-1} + R \theta_\sigma R')^{-1} R \right) \theta_\sigma \quad (53c)$$

Now let  $\lambda \rightarrow \infty$ . The term  $I_q \lambda^{-1}$  simply drops out, because the matrix  $R \theta_\sigma R'$  has full rank. This results in the posterior distribution of  $\theta$

$$\theta_{res} \sim N(\theta_{res,\mu}, \theta_{res,\sigma}) \quad (54a)$$

$$\theta_{res,\mu} = \left( I_k - \theta_\sigma R' (R \theta_\sigma R')^{-1} R \right) \theta_\mu \quad (54b)$$

$$\theta_{res,\sigma} = \left( I_k - \theta_\sigma R' (R \theta_\sigma R')^{-1} R \right) \theta_\sigma \quad (54c)$$

## B Procedure to draw from singular inverse-Wishart distribution

The procedure to draw from a singular inverse-Wishart distribution is explained in this appendix. Consider an  $m \times m$  scale matrix  $S$  with  $\text{rank}(S) = q$  and  $d$  degrees of freedom.

1. Draw a  $m \times q$  matrix  $X$  from the standard normal distribution. QR factorize  $X = QR$  where  $Q$  is an orthogonal matrix and  $R$  is an upper triangular matrix.
2. Eigenvalue decompose  $S = (VD)(VD)'$  where  $D$  is a diagonal matrix with the square roots of the  $q$  non-zero eigenvalues of  $S$  on the diagonal and  $V$  is a matrix with the corresponding eigenvectors.
3. Now  $(VR^{-1})(VR^{-1})'$  is a draw from the singular inverse-Wishart distribution.

## References

- BAUMEISTER, C., AND G. PEERSMAN (2013): “Time-Varying Effects of Oil Supply Shocks on the US Economy,” *American Economic Journal: Macroeconomics*, 5(4), pp. 1–28.
- BLANCHARD, O., AND R. PEROTTI (2002): “An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output,” *The Quarterly Journal of Economics*, 117(4), pp. 1329–1368.
- CANOVA, F., AND M. CICCARELLI (2009): “Estimating Multicountry VAR Models,” *International Economic Review*, 50(3), pp. 929–959.
- CANOVA, F., AND L. GAMBETTI (2009): “Structural changes in the US economy: Is there a role for monetary policy?,” *Journal of Economic Dynamics and Control*, 33(2), pp. 477–490.
- (2010): “Do Expectations Matter? The Great Moderation Revisited,” *American Economic Journal: Macroeconomics*, 2(3), pp. 183–205.
- CARTER, C. K., AND R. KOHN (1994): “On Gibbs Sampling for State Space Models,” *Biometrika*, 81(3), pp. 541–553.
- CASELLA, G., AND E. I. GEORGE (1992): “Explaining the Gibbs Sampler,” *The American Statistician*, 46(3), pp. 167–174.
- COGLEY, T., AND T. J. SARGENT (2001): “Evolving Post-World War II U.S. Inflation Dynamics,” *NBER Macroeconomics Annual*, 16, pp. 331–373.
- (2005): “Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US,” *Review of Economic Dynamics*, 8(2), pp. 262–302.
- D’AGOSTINO, A., L. GAMBETTI, AND D. GIANNONE (2013): “Macroeconomic Forecasting and Structural Change,” *Journal of Applied Econometrics*, 28(1), pp. 82–101.
- DE WIND, J. (2014): “Time variation in the dynamic effects of unanticipated changes in tax policy,” CPB Discussion Paper No. 271, CPB Netherlands Bureau for Economic Policy Analysis.
- DEL NEGRO, M., AND G. PRIMICERI (2013): “Time-varying structural vector autoregressions and monetary policy: a corrigendum,” Staff Report No. 619, Federal Reserve Bank of New York.

- DEN HAAN, W. J., AND V. STERK (2011): "The Myth of Financial Innovation and the Great Moderation," *The Economic Journal*, 121(553), pp. 707–739.
- FULLER, W. A. (1996): *Introduction to Statistical Time Series*. John Wiley, New York.
- GAMBETTI, L., E. PAPPA, AND F. CANOVA (2008): "The Structural Dynamics of U.S. Output and Inflation: What Explains the Changes?," *Journal of Money, Credit and Banking*, 40(2-3), pp. 369–388.
- HAMILTON, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57(2), pp. 357–384.
- HOFMANN, B., G. PEERSMAN, AND R. STRAUB (2012): "Time variation in U.S. wage dynamics," *Journal of Monetary Economics*, 59(8), pp. 769–783.
- KIM, C., AND C. R. NELSON (1999): *State-Space Models With Regime Switching*. MIT Press, Cambridge, Massachusetts.
- KIM, S., N. SHEPHARD, AND S. CHIB (1998): "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *The Review of Economic Studies*, 65(3), pp. 361–393.
- PRIMICERI, G. E. (2005): "Time Varying Structural Vector Autoregressions and Monetary Policy," *The Review of Economic Studies*, 72(3), pp. 821–852.
- ROBERTS, G. O., AND A. F. M. SMITH (1994): "Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms," *Stochastic Processes and their Applications*, 49(2), pp. 207–216.
- SARGENT, T. J. (1999): *The Conquest of American Inflation*. Princeton University Press, Princeton.
- SIMS, C. A. (1980): "Macroeconomics and Reality," *Econometrica*, 48(1), pp. 1–48.
- (2001): "Evolving Post-World War II U.S. Inflation Dynamics: Comment," *NBER Macroeconomics Annual*, 16, pp. 373–379.
- SIMS, C. A., AND T. ZHA (2006): "Were There Regime Switches in U.S. Monetary Policy?," *The American Economic Review*, 96(1), pp. 54–81.

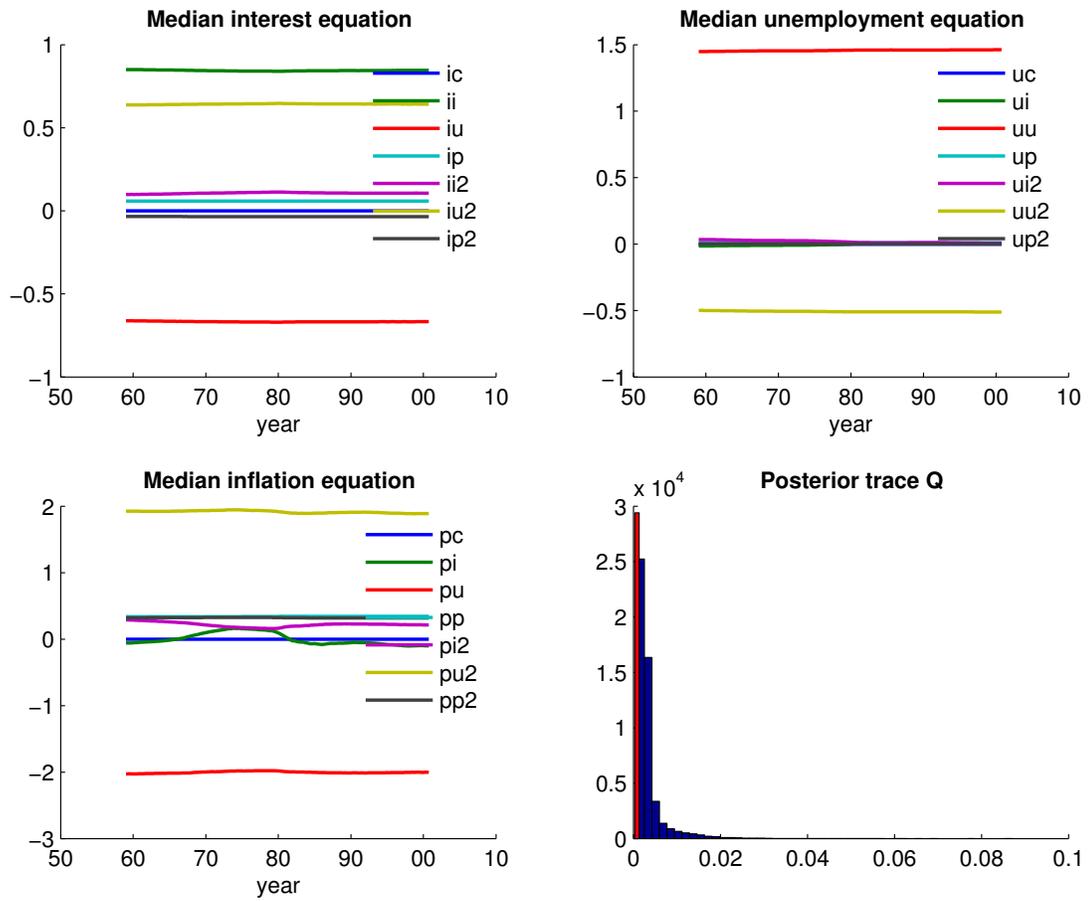
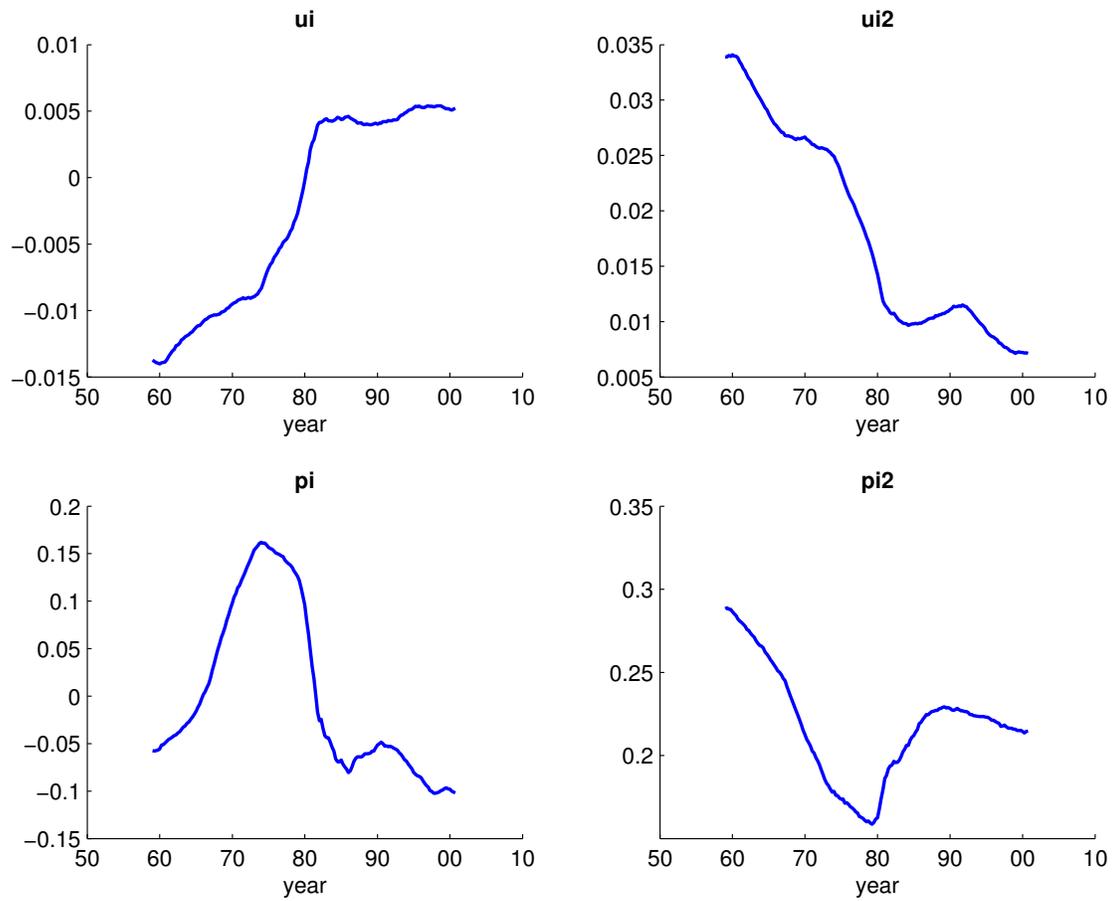


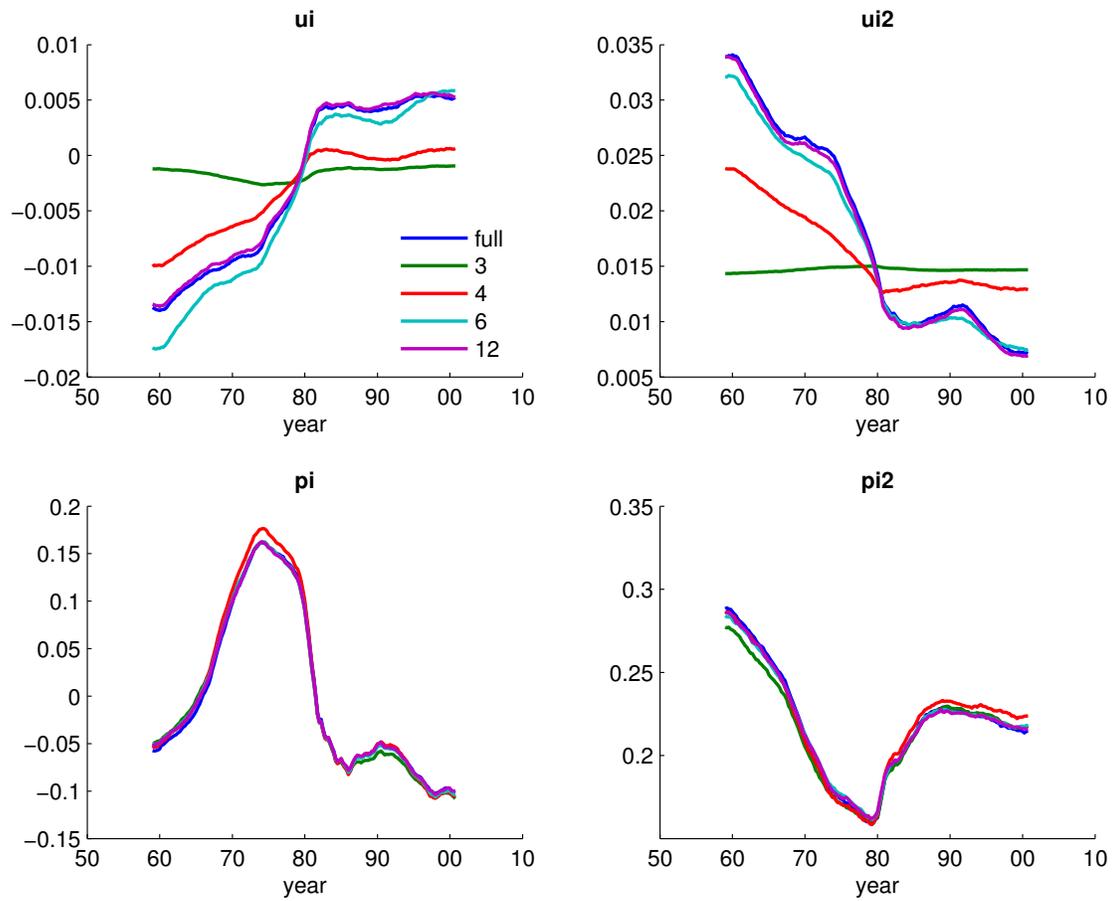
Figure 1: Posterior results for the full-rank model

Note: In the legends, i refers to interest, u refers to unemployment, p refers to inflation, and c refers to constant. The first letter in the abbreviations refers to the regressand, the second letter refers to the regressor, and a 2 indicates whether it concerns the second lag of the regressor.



**Figure 2: Median trajectories for selection of time-varying parameters in full-rank model**

Note: In the titles, i refers to interest, u refers to unemployment, p refers to inflation, and c refers to constant. The first letter in the abbreviations refers to the regressand, the second letter refers to the regressor, and a 2 indicates whether it concerns the second lag of the regressor.



**Figure 3: Median trajectories for selection of time-varying parameters for various ranks**

Note: In the titles, i refers to interest, u refers to unemployment, p refers to inflation, and c refers to constant. The first letter in the abbreviations refers to the regressand, the second letter refers to the regressor, and a 2 indicates whether it concerns the second lag of the regressor.

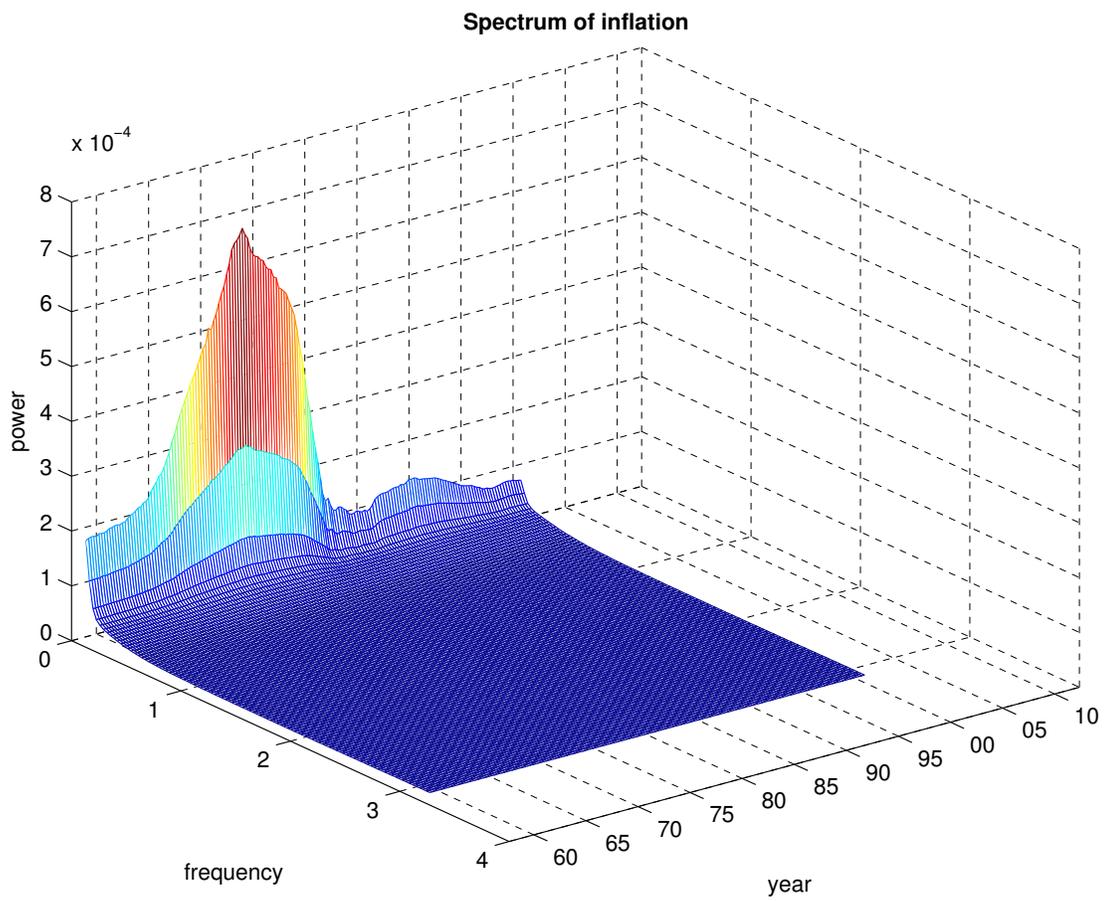


Figure 4: Spectrum of inflation in the full-rank model

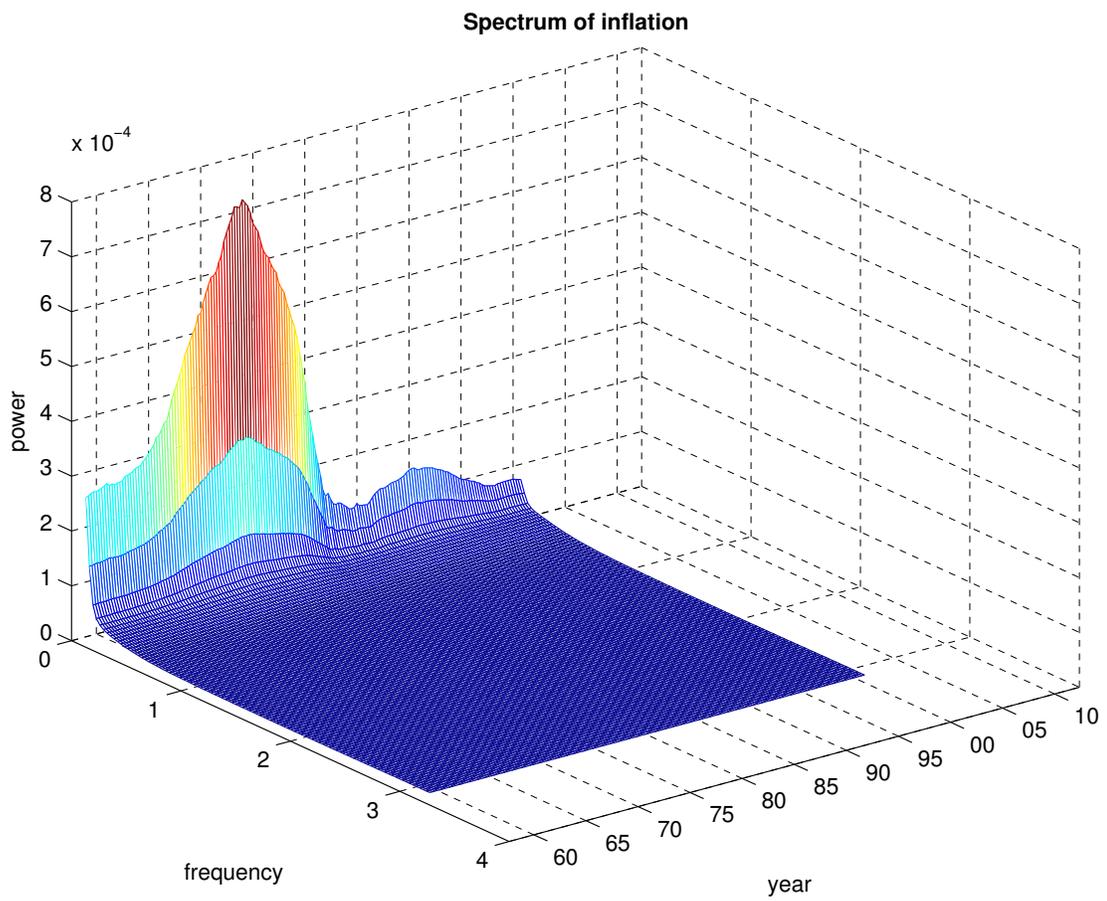


Figure 5: Spectrum of inflation in the reduced-rank model with three factors



Publisher:

CPB Netherlands Bureau for Economic Policy Analysis  
P.O. Box 80510 | 2508 GM The Hague  
T (070) 3383 380

March 2014 | ISBN 978-90-5833-635-4