Structural Analysis with Classical and Bayesian Large Reduced Rank VARs*

Andrea Carriero
Queen Mary, University of London
a.carriero@qmul.ac.uk

George Kapetanios
Queen Mary, University of London
g.kapetanios@qmul.ac.uk

Massimiliano Marcellino
European University Institute, Bocconi University and CEPR
massimiliano.marcellino@eui.eu

This draft: September 2012

Abstract

We address the issue of parameter dimensionality reduction in Vector Autoregressive models (VARs) for many variables by imposing reduced rank restrictions on the coefficient matrices. We derive the Wold representation implied by reduced rank VARs (RR-VARs) and show that it is closely related to that associated with dynamic factor models. Next we describe classical and Bayesian estimation of large RR-VARs, and illustrate how the estimated models can be used for structural analysis. The performance of the RR-VARs is then compared with that of factor models and classical and Bayesian VARs by means of a set of simulation experiments. Finally, a similar comparison is conducted in the context of an empirical application on the transmission mechanism of monetary policy.

Keywords: Large datasets, Reduced Rank Regressions, Bayesian VARs, Factor Models, Forecasting, Structural Analysis.

J.E.L. Classification: C11, C13, C33, C53.

*We are grateful to Luc Bauwens, Siddartha Chib, Herman van Dijk, Simon Potter and participants in the 2011 Esobe Conference for helpful comments and discussions.

1 Introduction

The recent theoretical and applied econometric literature has paid substantial attention to modelling in the presence of datasets with a large cross-sectional and temporal dimension. The two main approaches are factor models and Vector Autoregressive models (VARs). Both approaches started in a small dataset context, and typically relied on Bayesian methods to overcome the curse of dimensionality problem (Geweke (1977), Doan et al. (1984)). However, classical methods quickly replaced the Bayesian ones, see e.g. Stock and Watson (1989) in the factor context and the large literature on structural VARs.

Classical method were also used in the early large datasets developments of factor techniques, often combined with non-parametric procedures for factor estimation, see e.g. Stock and Watson (2002a, 2002b), Forni et al. (2000). Parametric and sometimes Bayesian approaches emerged later, in the structural factor augmented VAR (FAVAR) literature, in particular in particular to allow for parameter time variation, e.g. Kose et al. (2003), Del Negro and Otrok (2008), Baumeister and Mumtaz (2010), but also Eickmeier, Lemke, Marcellino (2010).

More recently, large Bayesian VARs (BVARs) were proposed as an alternative modelling device to factor models, e.g. Doz, Giannone and Reichlin (2006) and Banbura, Giannone, Reichlin (2010). Large classical VARs are not feasible, unless constraints are imposed in order to substantially reduce the number of free parameters, see e.g., Carriero, Kapetanios, Marcellino (2011)).

Both FAVARs and BVARs have pros and cons. The FAVARs nicely capture the idea of few key shocks or variables as drivers of the entire economy. However, they often rely on a two-step approach (estimate factors, then treat them as known in subsequent analyses), though full Kalman filter based estimation has been also developed, see e.g. Doz, Giannone and Reichlin (2011)). In both cases, the number of variables, N, must diverge in order to get consistent factor estimators, and the speed of divergence must be faster than that of the temporal dimension, T, in order to avoid generated regressors problems in subsequent analyses, see e.g. Bai and Ng (2006a). Moreover, it is unclear why the factors are modelled as a VAR in FAVARs, in particular when they are estimated as the static or dynamic principal components of the variables, e.g., Dufour and Stevanovic (2010) demonstrate that a VARMA representation is more appropriate, though more complex. And structural identification is in principle rather easy but in practice often complex, so that few empirical applications have been produced. Moreover, testing hypotheses on the factors, e.g. whether they are equal to specific macroeconomic or financial variables, is quite complex, see Bai and Ng (2006b).

The BVARs are overall easier to handle than FAVARs in terms of (Bayesian) estimation

and inference. However, estimation remains computationally demanding due to the curse of dimensionality, and the fact of having one shock for each variable, each of them equally important, is not so attractive from an economic point of view.

Hence, in this paper we suggest to use a model that bridges BVARs and FAVARs. Specifically, we propose to impose reduced rank restrictions in a BVAR that, as we will see, makes it similar to a factor model in terms of having a smaller set of key shocks or variables, but preserves the attractive features of a BVAR, substantially reducing its parameter dimensionality. Reduced rank BVARs (RR-BVARs) have been previously considered in the literature, see e.g. Geweke (1996) and Carriero et al. (2011) in a small and large datasets context, respectively, but not for structural analysis. Hence, our main contribution is to show how an RR-BVAR can be used to compute the response functions to structural shocks, for which we introduce and study the properties of (both classical and Bayesian) methods for estimation, inference and rank determination.

In Section 2 we derive alternative moving average representations for RR-BVARs, discussing their relative merits. More specifically, we show that in an RR-BVAR each variable is driven by a limited number of specific linear combinations of the other variables, say r, with r much smaller than N. Since these combinations are the counterpart of the factors in the factor literature, we also refer to them as "factors". We also show that these factors admit an exact VAR representation, whose coefficients can be analytically derived from those of the RR-BVAR. It is then possible to have moving average representations of the RR-BVAR where each variable is driven either by the N original RR-BVAR (1-step ahead forecast) errors, or by the r errors in the VAR for the factors (common to all variables) plus $N - r$ other errors, orthogonal to the factor errors. The former representation is similar to the one used in the BVAR literature, the latter to the one used in the FAVAR literature. We do not prefer either representation, we suggest to use the one that is more suited to address the specific empirical problem under analysis.

In Section 3 we introduce classical and Bayesian estimation methods for the RR-BVAR. The classical method mostly relies on Reinsel (1983), see also Velu, Reinsel and Wichern (1986) and Reinsel and Velu (1998). Our main original contribution is to show that this technique can be also implemented when N diverges, under some regularity conditions. In the Bayesian context, we derive the conditional distributions of the parameters under standard assumptions on the priors, and provide a new MCMC algorithm to handle the model non-linearity.

In Section 4 we discuss classical and Bayesian methods for the determination of the rank r of the RR-BVAR. In a classical context, rank determination can be determined either by information criteria or by sequential testing methods. We briefly review them and discuss their applicability in a large N context. In a Bayesian framework, we propose to select the

rank associated with the highest data density, which also corresponds to maximum of the posterior density of r, assuming a flat prior. We suggest to approximate the marginal data density numerically by using Geweke's (1999) modified harmonic mean estimator.

In Section 5 we assess the performance in finite samples of the classical and Bayesian procedures for estimation and rank determination, and compare the performance of the RR-BVAR with that of a FAVAR.

In Section 6 we illustrate the theoretical proposals by means of two empirical applications. First, we replicate in the RR-BVAR context the BVAR analysis of the transmission of US monetary policy shocks conducted by Banbura, Giannone and Reichlin (2010). We use the N-shock MA representation of the RR-BVAR and obtain both classical and Bayesian responses that are economic sensible and similar to those of Banbura et al. (2010). The Bayesian procedure produces though more accurate estimates than the classical ones. Second, we assess the effects of demand, supply, and financial / monetary shocks. In this case we use the FAVAR-style MA representation of the RR-BVAR and assume that the factors reflect movements in slow, fast and price variables, where the shocks associated with the slow, fast and price factors are interpreted as, respectively, demand, financial / monetary and supply shocks. Again, the resulting responses are very sensible from an economic point of view. In addition, the responses to the monetary shock resulting from the two exercises are very similar.

Finally, in Section 7 we summarize the main results of the paper and propose directions for additional research in this area.

2 Specification

2.1 The RR-VAR

Consider the following representation for the N-dimensional weakly stationary process $Y_t = (y_{1,t}, y_{2,t}, ..., y_{N,t})'$:

$$Y_t = \Phi(L)Y_t + \epsilon_t, \tag{1}$$

where $\Phi(L) = \Phi_1 L + ... + \Phi_p L^p$ is a polynomial of order p and ϵ_t are i.i.d. $N(0, \Sigma)$.

We assume that $\Phi(L)$ can be factorised as $\Phi(L) = A(L)B(L)$, where $A(L) = A_1 L + ... + A_{p_1} L^{p_1}$ and each A_i is of dimension $N \times r$, and $B(L) = B_0 + B_1 L + ... + B_{p_2} L^{p_2}$ and each B_i is of dimension $r \times N$, with $p_1 + p_2 = p$, $p_1 \geq 1$, $p_2 \geq 0$. We obtain

$$Y_t = A(L)B(L)Y_t + \epsilon_t = \sum_{u=1}^{p_1} \sum_{v=0}^{p_2} A_u B_v Y_{t-u-v} + \epsilon_t \tag{2}$$

If r is much smaller than N there are much fewer parameters in the RR-VAR in (2)

than in the corresponding unrestricted VAR in (1). For example, if $N = 50$, $p = 2$ and $r = 2$, there are $N(Np - r(p+1)) = 4700$ parameters less in the reduced rank VAR than in the corresponding unrestricted VAR, the total number of parameters in (1) and (2) being, respectively, $N^2p = 5000$ and $Nr(p+1) = 300$.

The moving average representation associated with (2) is

$$Y_t = (I - A(L)B(L))^{-1}\epsilon_t, \tag{3}$$

and starting from this expression it is easy to derive optimal forecasts and impulse response functions by using standard techniques.

2.2 RR-VAR and factor models

Let us now consider the relationship between the RR-VAR and the factor model. The r-dimensional vector of variables:

$$F_t = B(L)Y_t = B_0Y_t + B_1Y_{t-1} + ... + B_{p_2}Y_{t-p_2} \tag{4}$$

can be interpreted as a set of r factors common to all the variables under analysis. Indeed the model in (2) can be re-written again as:

$$Y_t = A(L)F_t + \epsilon_t = \sum_{u=1}^{p_1} A_u F_{t-u} + \epsilon_t, \tag{5}$$

which shows that as long as a rank reduction can be imposed on the VAR in (1), then it can be written as a factor model as in (5).

A further simplification is obtained by setting $p_2 = 0$. In this case (2) reduces to:

$$Y_t = A(L)B_0Y_t + \epsilon_t = \sum_{u=1}^{p} A_u B_0 Y_{t-u} + \epsilon_t, \tag{6}$$

and (4) simplifies to $F_t = B_0Y_t$. Here B_0 is assumed to have full rank r. As in the case of the factor model, the "loadings" A_u and the factor weights B_0 are not uniquely identified in an RR-VAR. Without any loss of generality, we will assume that $B_0 = (I_r,\ \widetilde{B}_0)$. The number of parameters to be estimated in (6) is then $Nrp + r(N-r)$, while (1) contains N^2p parameters.[1]

An important characteristic of (6) is that the resulting linear combinations $F_t = B_0Y_t$ have a closed form $VAR(p)$ representation, while in general when Y_t follows an unrestricted

[1] Another particular case of interest is when $p_1 = 1$, $p_2 = p - 1$. The resulting model is that studied by Geweke (1996) in a small N context and by Carriero et al. (2011) in a large N context.

VAR linear combinations of Y_t are complicated $VARMA$ processes, see e.g. Lutkepohl (2007). To see this, it is sufficient to pre-multiply by B_0 both sides of equation (6) and use $F_t = B_0 Y_t$ to get:

$$F_t = B_0 \sum_{u=1}^{p} A_u F_{t-u} + B_0 \epsilon_t = C(L) F_t + u_t \tag{7}$$

with $C(L) = B_0 A(L) = B_0 A_1 L + B_0 A_2 L^2 + ... + B_0 A_p L^p$ and $u_t = B_0 \epsilon_t$ is i.i.d. $N(0, \Omega)$ with $\Omega = B_0 \Sigma B_0'$.

We can then group together (6) and (7) to form the system

$$\begin{cases} Y_t = A(L) F_t + \epsilon_t = A_1 F_{t-1} + A_2 F_{t-2} + ... + A_p F_{t-p} + \epsilon_t \\ F_t = C(L) F_t + u_t = C_1 F_{t-1} + C_2 F_{t-2} + ... + C_p F_{t-p} + u_t \end{cases} \tag{8}$$

It is now evident that if we want Y_t to be decomposed into a common and a fully idiosyncratic component, we need to restrict the variance covariance matrix of ϵ_t, Σ, to be either diagonal (corresponding to the exact factor model case) or anyway with only a limited amount of non-zero covariances (as in an approximate factor model). Instead orthogonality between $A(L)F_t$ and ϵ_t is guaranteed by the assumption of independence of ϵ_t.

The moving average representation associated with (8) is:

$$Y_t = \left(A(L) \left[(I - B_0 A(L))^{-1} \right] B_0 + I \right) \epsilon_t. \tag{9}$$

This expression is obtained by inserting the moving average representation for F_t, which is

$$F_t = (I - C(L))^{-1} u_t = (I - B_0 A(L))^{-1} B_0 \epsilon_t \tag{10}$$

into the equation for Y_t in (8).

At this point we have two alternative moving average representations for Y_t, (3) and (9), and we can easily show that, naturally, they are equivalent when $p_2 = 0$, namely[2]

$$(I - A(L) B_0)^{-1} = A(L) \left[(I - B_0 A(L))^{-1} \right] B_0 + I. \tag{11}$$

[2] In fact, we have:

$$A(L)^{-1} = B_0 + [I - B_0 A(L)] A(L)^{-1}$$
$$[I - B_0 A(L)]^{-1} A(L)^{-1} = [I - B_0 A(L)]^{-1} B_0 + A(L)^{-1}$$

using:

$$[I - B_0 A(L)]^{-1} A(L)^{-1} = [A(L)(I - B_0 A(L))]^{-1} = [(I - A(L) B_0) A(L)]^{-1} = A(L)^{-1} (I - A(L) B_0)^{-1}$$

we get:

$$A(L)^{-1} (I - A(L) B_0)^{-1} = [I - B_0 A(L)]^{-1} B_0 + A(L)^{-1}$$

and (11) is obtained by multiplying by $A(L)$.

A third moving average representation is particularly convenient for structural analysis. Let us introduce the $(N-r)\times N$ full raw rank matrix $B_{0\perp}$ that is orthogonal to B_0, i.e. $B_0B_{0\perp}'=0$, such that the rank of $(B_0',B_{0\perp}')$ is N. Note that B_0B_0' has full rank (as we assumed B_0 has full rank) and we have the following decomposition (see Johansen (1995, p.39) for a similar decomposition)):

$$B_0'(B_0B_0')^{-1}B_0+B_{0\perp}'(B_{0\perp}B_{0\perp}')^{-1}B_{0\perp}=I_N. \tag{12}$$

This key identity can now be inserted into the Wold representation in (9) to yield:

$$\begin{aligned}Y_t&=\left(A(L)\left[(I-B_0A(L))^{-1}\right]B_0+(B_0'(B_0B_0')^{-1}B_0+B_{0\perp}'(B_{0\perp}B_{0\perp}')^{-1}B_{0\perp})\right)\epsilon_t\\&=(B_0'(B_0B_0')^{-1}+A(L)\left[(I-B_0A(L))^{-1}\right])B_0\epsilon_t+B_{0\perp}'(B_{0\perp}B_{0\perp}')^{-1}B_{0\perp}\epsilon_t,\end{aligned}$$

and, since $B_0\epsilon_t=u_t$, we have:

$$Y_t=(B_0'(B_0B_0')^{-1}+A(L)\left[(I-B_0A(L))^{-1}\right])u_t+B_{0\perp}'(B_{0\perp}B_{0\perp}')^{-1}B_{0\perp}\epsilon_t. \tag{13}$$

The representation in (13) shows that each element of Y_t is driven by a set of r common errors, the u_t that are the drivers of the factors F_t, and by an element of $B_{0\perp}\epsilon_t$, where u_{t-i} and $B_{0\perp}\epsilon_t$ are orthogonal for each $i\geq 0$.

The recovery of the structural shocks v_t driving F_t starting from the reduced form errors u_t can be achieved using any technique adopted in the structural VAR and structural FAVAR literatures, see e.g. Bernanke et al. (2005) or Eickmeier et al. (2011). For example, the simplest option is the Choleski decomposition $\Omega=P^{-1}SP^{-1'}$, where P is a lower triangular matrix and S is a diagonal matrix containing the variances of the structural shocks. Hence,

$$v_t=Pu_t. \tag{14}$$

Combining the previous equation with (13) yields

$$Y_t=(B_0'(B_0B_0')^{-1}+A(L)\left[(I-B_0A(L))^{-1}\right])P^{-1}v_t+B_{0\perp}'(B_{0\perp}B_{0\perp}')B_{0\perp}\epsilon_t \tag{15}$$

from which impulse response functions can be easily computed.

Note that, since $u_t=B_0\epsilon_t$, the structural shocks are also related to the ϵ_t errors in the Wold representations in (3) or (9), via the relationship $v_t=Pu_t=PB_0\epsilon_t$. However, from a structural point of view, there is an important difference between the representations in (3) or (9) and that in (13). In the former case there can be as many structural shocks as variables, namely N, while in (13) we are explicitly assuming that there is a reduced number of structural shocks, r, which drive all the factors $F_t=B_0Y_t$, which in turn drive

the common components of all the variables in Y_t. In principle there could be other $N-r$ structural shocks that drive the $(N-r)$ errors $B_{0\perp}\epsilon_t$ in (13), but in practice these are never considered in the factor literature.

There is a case where shocking the factors or shocking the variables produces the same responses and this is, perhaps not surprisingly, when the factors are equal to a subset of the variables and we shock one of the variables in this subset. Formally, suppose that $\widetilde{B}_0 = 0$ in $B_0 = (I_r,\ \widetilde{B}_0)$, so that $B_{0\perp} = (0,\ I_{N-r})$, and split Y_t into the first r variables $Y_{1t} = F_t$ and the remaining $N-r$ variables Y_{2t}. Similarly, ϵ_t is split into ϵ_{1t} and ϵ_{2t}, where ϵ_{1t} and ϵ_{2t} are orthogonal. Then, the model for the factors becomes

$$Y_{1t} = C(L)Y_{1t} + u_t = C_1 Y_{1t-1} + C_2 Y_{1t-2} + ... + C_p Y_{1t-p} + \epsilon_{1t},$$

which also coincides with the first r equations (those for Y_{1t}) in the model for Y_t:

$$Y_t = A(L)Y_{1t} + \epsilon_t = A_1 Y_{1t-1} + A_2 Y_{1t-2} + ... + A_p Y_{1t-p} + \epsilon_t,$$

and in addition ϵ_{1t} and ϵ_{2t} are orthogonal. Hence, identifying the shocks in the model for the factors or in the full system for the variables is equivalent.

To conclude, the equations (8) and (13) highlight that the RR-VAR is similar to the generalized dynamic factor model of Stock and Watson (2002a, 2002b) and Forni et al. (2000), and even more to the parametric versions of these models later adopted in the structural Factor augmented VAR (FAVAR) literature, e.g. Bernanke et al. (2005) and Doz et al. (2011). The similarities increase when the unobservable factors are estimated by static principal components, since in this case the estimated factors end up being linear combinations of the variables, exactly like the elements of F_t. However, there are also possibly important differences between RR-VAR and factor models. In particular, in the factor literature the factors are unobservable and can be consistently estimated only when N diverges. As we will see in the next section, within an RR-VAR context it is possible to consistently estimate the "factors" F_t even when N is finite. Moreover, testing specific hypotheses on the "factors" F_t, such as equality of one factor to a specific economic variable, is much simpler in the RR-VAR context (restrictions on B) than in a factor context (see Bai and Ng (2006b)). Next, in general factors estimated by principal components do not admit an exact VAR representation (see Dufour and Stevanovic (2010), while from (8) this is the case within the RR-VAR context. Finally, in the factor literature precise conditions have to be imposed on the idiosyncratic components of each variable, to make sure that it is not confounded with the common part, and on the factor loadings, to make sure that the factors affect almost all variables (see e.g. Stock and Watson (2002a, 2002b)). In our context these conditions would impose constraints on the variance covariance matrix of $B_{0\perp}\epsilon_t$ in (13),

and on $A(L)$ in (8). However, no constraints are needed to obtain consistent parameter estimators for the RR-VAR model, as we will see in the next Section. Given all this, with respect to the factor approach, the RR-VAR seems to provide an easier, less constrained and theoretically more consistent framework for parametric modelling of large datasets.

3 Estimation

For estimation it is convenient to compactly rewrite (6) as in Reinsel (1983):

$$Y_t = AZ_{t-1} + \epsilon_t, \tag{16}$$

where $Z'_{t-1} = (F'_{t-1}..., F'_{t-p}) = (Y'_{t-1}B'_{0,}..., Y'_{t-p}B'_{0,}) = (Y'_{t-1,}..., Y'_{t-p})(I_p \otimes B'_0)$ and $A = (A_1, ..., A_p)$. As for all js, $A_jB_0 = A_jQ^{-1}QB_0$ for any nonsingular matrix Q we add the identification restriction $B_0 = (I_r,\ \widetilde{B}_0)$. Defining $Y = (Y_1, ..., Y_T)'$ and $Z = (Z_0, Z_1, ..., Z_{T-1})'$ and $E = (\epsilon_1, ..., \epsilon_T)'$, stacking the equations in (16) for $t = 1, ..., T$ we have

$$Y = ZA + E, \tag{17}$$

where $VAR(E) = (I_T \otimes \Sigma)$.

3.1 Estimation via Maximum Likelihood

Reinsel (1983) studied classical estimation of the model in (16) via Maximum Likelihood. In particular, he showed that ML estimates can be obtained by iterating over the first order conditions on the maximization problem. The likelihood function is:

$$-0.5T \log|\Omega| - 0.5\Sigma_{t=1}^{T}(Y_t - AZ_{t-1})'\Omega^{-1}(Y_t - AZ_{t-1}) \tag{18}$$

For any A and $\tilde{B}_0$ the maximization with respect to Ω yields:

$$\hat{\Omega} = (Y - ZA')'(Y - ZA')/T \tag{19}$$

The partial derivatives with respect to A (given $\tilde{B}_0$ and Ω) can be obtained by noting that:

$$AZ_{t-1} = vec(Z'_{t-1}A') = (I_N \otimes Z'_{t-1})vec(A') \tag{20}$$

and the corresponding first order conditions are given by:

$$\frac{\partial l}{\partial vec(A')} = \Sigma_{t=1}^{T}(I_N \otimes Z_{t-1})'\Omega^{-1}(I_N \otimes Z_{t-1}) = 0 \tag{21}$$

The partial derivatives with respect to $\tilde{B}_0$ (given A and Ω) can be obtained by noting that:

$$AZ_{t-1} = \Sigma_{t=1}^{p} A_j Y_{1,t-j} + \Sigma_{t=1}^{p} (A_j \otimes Y'_{2,t-1}) vec(\tilde{B}_0), \tag{22}$$

and the corresponding first order conditions are given by:

$$\frac{\partial l}{\partial vec(\tilde{B}_0)} = \Sigma_{t=1}^{T} U_{t-1} A' \Omega^{-1} \{Y_t - (I_N \otimes Z'_{t-1})\alpha\} = 0 \tag{23}$$

where $U_{t-1} = (I_r \otimes Y_{2,t-1}, ..., I_r \otimes Y_{2,t-p})$ and $Y_{2,t}^{'}$ comes from partitioning $Y_t^{'}$ in the first r and last $N - r$ components: $Y_t^{'} = (Y_{1,t}^{'}, Y_{2,t}^{'})$.

Reinsel suggested to solve in turn equations (19), (21) and (23) until convergence is achieved, and established consistency and asymptotic normality of this estimator. Of course these consistency and asymptotic normality results can be coupled with the standard impulse reponse analysis for finite dimensional VAR models to produce standard errors for such impulse responses (see, e.g., Section 3.7 of Lutkerpohl (2007)).

To conclude, it is worth noting that specific hypotheses on the parameters, and in particular on B_0, can be tested using likelihood ratio statistics.

3.2 Estimation via Markov Chain Monte Carlo

In this subsection we derive the conditional distributions and provide a new MCMC algorithm for the estimation of the model in (16)

The model contains three sets of parameters, contained respectively in the matrices A, $\widetilde{B}_0$, and Σ. The joint posterior distribution $p(A, \widetilde{B}_0, \Sigma|Y)$ has not a known form, but it can be simulated by drawing in turn from the conditional posterior distributions $p(A|\widetilde{B}_0, \Sigma, Y)$, $p(\widetilde{B}_0|A, \Sigma, Y)$, and $p(\Sigma|A, \widetilde{B}_0, Y)$.

Deriving the conditional posteriors of A and Σ is straightforward. Under the knowledge of $\widetilde{B}_0$ and Y the variable Z_{t-1} is known, and (16) constitutes a simple multivariate regression model. Under a Normal-Inverse Wishart prior for A and Σ :

$$A|\Sigma \sim N(A_0, \Sigma \otimes \Omega_0),\ \Sigma \sim IW(S_0, v_0), \tag{24}$$

the model features the following conditional posterior distributions:

$$A|\Sigma, \widetilde{B}_0, Y \sim N(\bar{A}, \Sigma \otimes \bar{\Omega}),\ \Sigma|\widetilde{B}_0, Y \sim IW(\bar{S}, \bar{v}). \tag{25}$$

where:

$$\bar{A} = \bar{\Omega}(\Sigma_0^{-1}\Phi_0 + Z'Y) \tag{26}$$

$$\bar{\Omega} = (\Sigma_0^{-1} + Z'Z)^{-1} \tag{27}$$

$$\bar{S} = \hat{E}'\hat{E} - (A_0 - \bar{A})'(\Sigma_0^{-1} + (Z'Z)^{-1})^{-1}(A_0 - \bar{A}) \tag{28}$$

$$\bar{v} = v_0 + T \tag{29}$$

and where $\hat{\Phi}$ and $\hat{E}$ are the OLS estimates. This prior features a Kronecker structure that restricts somehow the way shrinkage can be imposed, but dramatically improves the computation time. Simulating draws from $p(A, \Sigma|\widetilde{B}_0, Y)$ can therefore be easily performed by generating a sequence of m draws $\{\Sigma_j\}_{j=1}^m$ from $\Sigma|Y \sim IW(\bar{S}, \bar{v})$ and then for each j drawing from $A|\Sigma, Y \sim N(\bar{A}, \Sigma_j \otimes \bar{\Omega})$, which provides with the sequence $\{A_j\}_{j=1}^m$.

Drawing from $p(\widetilde{B}_0|A, \Sigma, Y)$ is less straightforward, as B_0 contains identification restrictions and also enters the model in a nonlinear way: $Y_t = A(L)B_0Y_t$. To draw $\widetilde{B}_0$ conditional on A, Σ one can use a Metropolis step. In particular, for each draw in the sequence $\{A_j, \Sigma_j\}_{j=1}^m$ a candidate $\widetilde{B}_{0j}^*$ for each element of $\widetilde{B}_0$ is drawn by sampling from the following random walk:

$$vec(\widetilde{B}_{0j}^*) = vec(\widetilde{B}_{0j-1}) + cV\eta_t \tag{30}$$

where η_t is a standard normal i.i.d. process and V is the Cholesky factor of the maximum likelihood estimate of the variance of $vec(\widetilde{B}_0)$. The candidate draw is then accepted with probability:

$$\alpha_k = \min\left\{\frac{p(\widetilde{B}_{0j}^*|\Sigma, \widetilde{B}_0, Y)}{p(\widetilde{B}_{0j-1}|\Sigma, \widetilde{B}_0, Y)}, 1\right\}$$

If the draw is accepted, then $\widetilde{B}_{0j} = \widetilde{B}_{0j}^*$, otherwise $\widetilde{B}_{0j} = \widetilde{B}_{0j-1}$. The scalar c is scaling factor calibrated in order to have about 35% rejections.

Drawing in turn from $p(A|\widetilde{B}_0, \Sigma, Y)$, $p(B_0|A, \Sigma, Y)$ and $p(\Sigma|A, \widetilde{B}_0, Y)$ provides a sequence of m draws $\left\{A_j, \Sigma_j, \widetilde{B}_0\right\}_{j=1}^m$ from the joint posterior distribution of $\widetilde{B}_0, A, \Sigma$. Each draw can be then inserted into equation (15), which can be used to derive the impulse response functions for any horizon.

Given that the parameters in $A(L)$ and B_0 interact nonlinearly, there is a potential concern about convergence if elements in either $A(L)$ or B_0 get close to 0. This potential problem is dramatically mitigated by the normalization choice we make for B_0 (setting the first r columns to an identity matrix). In the Appendix we provide a series of convergence checks on the draws of $A(L)$, B_0 and their product $A(L)B_0$. The analysis provided in the Appendix shows that the algorithm has good convergence properties and it is not affected by problems related to the nonlinearity.

4 Determining the rank of the system

4.1 Classical

The matter of determining the rank of the coefficient matrix in reduced rank VAR models has been analyzed extensively in the literature. A paper by Camba-Mendez et al. (2003) discusses this problem in detail. There are two main approaches. The first uses information criteria. This approach simply estimates (6) for all possible values of r and chooses the one that minimizes an information criterion (IC) that uses the fit of the model penalized by a penalty term that depends on the number of free parameters associated with every possible value of r. Standard information criteria can be used such as the Akaike IC or the Bayesian IC. An attractive feature of the use of ICs that both r and the number of lags can be jointly determined in a single search.

The second approach is based on sequential testing Starting with the null hypothesis of $r = 1$, a sequence of tests is performed. If the null hypothesis is rejected, r is augmented by one and the test is repeated. When the null cannot be rejected, r is adopted as the estimate of the rank of each matrix A_i in (16). Here, A must be estimated in an unrestricted way, i.e. without imposing a given rank. Then, standard tests of rank can be used on estimates of A. So this approach boils down to a repeated application of a test of rank. We review two such tests.

The first procedure, proposed by Cragg and Donald (1996), is based on the transformation of the matrix A using Gaussian elimination with complete pivoting[3]. r steps of Gaussian elimination with full pivoting on matrix A amounts to the following operations:

$$Q_{r^*}R_{r^*}Q_{r^*-1}R_{r^*-1}\dots Q_1R_1AC_1\dots C_{r^*-1}C_{r^*} = \begin{bmatrix} A_{11}(r^*) & A_{12}(r^*) \\ \mathbf{0} & A_{22}(r^*) \\ & \end{bmatrix}$$

where R_i and C_i are pivoting matrices for step i and Q_i are Gauss transformation matrices. The pivoting matrices used to perform the first r^* steps of Gaussian elimination are applied to $\boldsymbol{A}$ to obtain the following relation

$$R_{r^*}R_{r^*-1}\dots R_1AC_1...C_{r^*-1}C_{r^*} = RAC = F = \begin{bmatrix} F_{11}(r^*) & F_{12}(r^*) \\ F_{21}(r^*) & F_{22}(r^*) \\ & \end{bmatrix}$$

where F is partitioned accordingly, i.e. $F_{11}(r^*)$ is of dimension $r^* \times r^*$. Note that in this case

[3]The foundations behind this strategy follow the work of Gill and Lewbel (1992). The asymptotic distribution of the test suggested by Gill and Lewbel (1992) was incorrect, nonetheless, it provided researchers with an ingenious strategy to test for the rank.

$F_{11}(r^*)$ has full rank, under the null hypothesis that $\rho[\boldsymbol{A}] = r^*$. It then follows, (see Cragg and Donald (1996)), that $F_{22}(r^*) - F_{21}(r^*)F_{11}^{-1}(r^*)F_{12}(r^*) = 0$. The estimated counterpart of the above relation, i.e. $\hat{F}_{22} - \hat{F}_{21}\hat{F}_{11}^{-1}\hat{F}_{12} = \hat{\Lambda}_{22}(r^*)$, may be used as a test statistic of the hypothesis that the rank of A is r^*. Under regularity conditions, including the requirement that $\sqrt{T}vec(\hat{A} - A) \xrightarrow{d} N(0, V)$ where V has full rank, the following result can be shown, under H_0.

$$\sqrt{T}vec(\hat{\Lambda}_{22}(r^*)) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma} V \boldsymbol{\Gamma}')$$

where $\boldsymbol{\Gamma} = \Phi_2 \otimes \Phi_1$ and $\Phi_1 = \left[-\boldsymbol{F}_{21}\boldsymbol{F}_{11}^{-1} \;\; \boldsymbol{I}_{m-r^*}\right] R$, $\Phi_2 = \left[-\boldsymbol{F}_{12}'\boldsymbol{F}_{11}^{-1'} \;\; \boldsymbol{I}_{n-r^*}\right] C'$ and $\xrightarrow{d}$ denotes convergence in distribution. Then,

$$GE = T vec\ \hat{\Lambda}_{22}(r^*)'(\hat{\boldsymbol{\Gamma}}\hat{V}\hat{\boldsymbol{\Gamma}}')^{-1} vec\ \hat{\Lambda}_{22}(r^*) \xrightarrow{d} \chi^2_{(m-r^*)(n-r^*)}$$

where $\hat{\boldsymbol{\Gamma}}$ and $\hat{V}$ are the sample estimates of $\boldsymbol{\Gamma}$ and $\boldsymbol{V}$ and χ^2_l denotes the χ^2 distribution with l degrees of freedom.

The second testing procedure, suggested by Robin and Smith (2000), focuses on the eigenvalues of quadratic forms of A. The quadratic form $\boldsymbol{\Upsilon} A \boldsymbol{\Pi} A'$ where $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Pi}$ are positive definite matrices, is considered. It follows that $\rho[\boldsymbol{A}] = \rho[\boldsymbol{\Upsilon}\boldsymbol{A}\boldsymbol{\Pi}\boldsymbol{A}'] = r^*$, and therefore this quadratic form has $\min(m,n) - r^*$ zero eigenvalues. Additionally, the eigenvalues of the estimator of the above quadratic form converge in probability to their population counterparts. Robin and Smith (2000) consider the statistic

$$CRT = T \sum_{i=r^*+1}^{\min(m,n)} \hat{\lambda}_i$$

where $\hat{\lambda}_i$ are the eigenvalues of $\hat{\boldsymbol{\Upsilon}}\hat{A}\hat{\boldsymbol{\Pi}}\hat{A}'$ in descending order, $\hat{\Upsilon}$ and $\hat{\Pi}$ are estimates of Υ and Π respectively. Under the null hypothesis, the above statistic converges in distribution to a weighted sum of independent χ^2_1 random variables. The weights are given by the eigenvalues of $(D_{r^*}' \otimes C_{r^*}')V(D_{r^*} \otimes C_{r^*})$, τ_i, $i = 1, \ldots, (m-r^*)(n-r^*)$. D_{r^*} and C_{r^*} are $n \times (n-r^*)$ and $m \times (m-r^*)$ matrices containing the eigenvectors corresponding to the $n - r^*$ and $m - r^*$ smallest eigenvalues of $\Pi A' \Upsilon A$ and $\Upsilon A \Pi A'$ respectively. The sample counterparts of the above matrices may be obtained straightforwardly to estimate the asymptotic distribution of the test statistic. A few comments are in order for this test. Choices for $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Pi}$ are not discussed in much detail by Robin and Smith (2000). This choice can depend crucially on the application considered. An obvious choice that can be made irrespective of application is to set both $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Pi}$ equal to the identity matrix. Robin and Smith (2000) also consider another choice for their Monte Carlo but they do not elaborate on their motivation. Finally, it is worth noting that Robin and Smith (2000) claim that a big advantage of their test is

that neither full nor known rank for V is needed or, therefore, assumed.

The above tests of rank and the theoretical results that justify them relate to the case where N is finite. To the best of our knowledge, there are no extensions to the case where N is large. However, we expect that for moderately large values of N they can provide a useful guide for setting the value of r.

4.2 Bayesian

A natural way to choose the rank of the system is to compute the marginal data density as a function of the chosen r. Such density is given by:

$$p(Y) = \int p(Y|\theta)p(\theta)d\theta$$

and the optimal rank for the system is associated with the model featuring the highest data density:

$$r^* = \arg\max_r \ p(Y) \tag{31}$$

The probability $p(Y) = \int p(Y|\theta)p(\theta)d\theta$ is approximated numerically by using Geweke's (1999) modified harmonic mean estimator. In particular, by collecting all the coefficients in the vector $\theta = (A, \Sigma, \widetilde{B}_0)$ and considering the simulated posterior $\{\theta_j, \}_{j=1}^m = \left\{A_j, \Sigma_j, \widetilde{B}_0\right\}_{j=1}^m$ the estimator is:

$$\hat{p}(Y) = \left[\frac{1}{d}\sum_{m=1}^{d}\frac{f(\theta^m)}{p(Y|\theta^m)p(\theta^m)}\right]^{-1},$$

where $f(\cdot)$ is a truncated multivariate normal distribution calibrated using the moments of the simulated posterior draws $\{\theta\}_{m=1}^d$. See Geweke (1999) for details.

The determination of p_1 and p_2, i.e. the lag length of $A(L)$ and $B(L)$, could be also based on the marginal likelihood. However, in practice computing the marginal likelihood for this model is quite demanding because of the extremely high number of parameters. Hence, the use of information criteria seems more promising.

5 Monte Carlo evaluation

To be completed

6 Empirical applications

To illustrate how to conduct empirically structural analysis using RR-VARs, we first replicate in the RR-VAR context the BVAR analysis of the transmission of US monetary policy

shocks conducted by Banbura, Giannone and Reichlin (2010), using the N-shock MA representation of the RR-VAR. Then, we assess the effects of demand, supply and financial / monetary shocks, modelling the same dataset but with the FAVAR-style MA representation of the RR-VAR.

6.1 The effects of monetary policy shocks

We use the "medium" dataset of Banbura, Giannone, Reichlin (2010, BGR), which includes the 20 variables described in Table 1. The table also reports the classification of each indicator into "slow" and "fast", where slow moving variables can only react with a delay to the monetary policy shock, while fast variables can react contemporaneously. Among slow variables, there are nominal indicators such CPI, PPI and the PCE deflator, and real indicators, such as industrial production, employment, and personal income. The fast indicators are typically related to the financial sector, including interest rates, money aggregates and the S&P's stock price index. The sample is at monthly frequency and covers the period January 1959 to December 2003.

The variables are modelled with a RR-VAR with 13 lags, in line with BGR. We set the rank of the system to 3. We do so because while information criteria such as BIC and AIC selected a rank of 1 and 2 respectively, having a rank of 3 offers a particularly convenient way to identify the factors, a feature that we will exploit in the next empirical application we consider. Lag exclusion Wald tests performed with a rank of 3 could not reject the null that all the coefficients attached to the three factors were different from 0.

In line with the literature, the monetary policy shock is identified with a Cholesky scheme where the federal funds rate is ordered after the slow moving variables and before the fast ones. Formally, the impulse responses are based on the representation:

$$Y_t = \{A(L)[I - B_0 A(L)]^{-1} B_0 + I\}\Lambda^{-1}\epsilon_t^* \tag{32}$$

where ϵ_t^* are the structural shocks and Λ^{-1} is the Cholesky factor of the reduced form shocks ϵ_t. The resulting s-period ahead response is:

$$\Psi_s = A_1 B_0 \Psi_{s-1} + ... + A_{\min(s,p)} B_0 \Psi_{s-\min(s,p)};\ s > 0 \tag{33}$$

with $\Psi_0 = \{A(0)[I - B_0 A(0)]^{-1} B_0 + I\}\Lambda^{-1} = \Lambda^{-1}$.

We simulate the distribution of the impulse responses using 5000 draws and plot in Figure 1 the median, 5th, 16th, 84th, and 95th quantiles. For comparison, we also compute and report the impulse responses using classical ML estimation. We do not report classical bands, since bootstrapping yielded very large intervals.

The results are overall in line with those reported by BGR. In particular, after a few months with close to zero reaction, industrial production, capacity utilization, employment, consumption and housing starts decline, while unemployment increases. There is a delayed negative reaction also in CPI, PPI, PCE deflator, and earnings. Money and reserves decrease, while the exchange rate appreciates and the reaction of the stock market is close to zero. The classical and (median) Bayesian responses are overall very similar.

6.2 Demand, supply and financial / monetary shocks

Using the same RR-VAR as in the previous example, we now analyze the effects of shocks to demand, supply, and financial / monetary "factors". More precisely, we identify an output factor, a price factor, and a financial / monetary factor by imposing restrictions on the matrix B_0, as detailed in Table 2. The resulting factors and their components are graphed in Figure 2.

The impulse responses are now based on the representation:

$$Y_t = \{B_0'(B_0B_0')^{-1} + A(L)[I - B_0A(L)]^{-1}\}P^{-1}v_t + B_{0\perp}'(B_{0\perp}B_{0\perp}')B_{0\perp}\epsilon_t, \tag{34}$$

where P^{-1} is the Cholesky factor of the reduced form shocks u_t in the VAR representation of the factors, so that $P^{-1}SP^{-1'} = \Omega = B_0\Sigma B'$. The s-period ahead responses on the factor equations are:

$$\Pi_s = C_1\Pi_{s-1} + ... + C_{\min(s,p)}\Pi_{s-\min(s,p)};\ s > 0 \tag{35}$$

with $\Pi_0 = [I - B_0A(0)]^{-1}P^{-1} = P^{-1}$. The s-period ahead responses on the VAR equations are:

$$\Psi_s = A_1\Pi_{s-1} + ... + A_{\min(s,p)}\Pi_{s-\min(s,p)};\ s > 0 \tag{36}$$

with $\Psi_0 = \{B_0'(B_0B_0')^{-1} + A(0)[I - B_0A(0)]^{-1}\}P^{-1} = B_0'(B_0B_0')^{-1}P^{-1}$.

We simulate the distribution of the impulse responses using 16000 draws and plot the median responses together with the 5th, 16th, 84th, and 95th quantiles in Figure 6-9. Specifically, Figure 3 reports the responses of the factors, while Figures 4-6 those of the 20 variables to each of the three shocks.

Starting from Figure 3 and associating a shock to F1 with a demand shock, with see that both F2 (prices) and F3 (financial / monetary) increase. With a supply shock (on F2), the output factor (F1) decreases and the financial / monetary factor increases (F3). With a financial / monetary shock (on F3), both the output (F1) and the price (F2) factors decrease, though with a delay, along the lines of the delayed reactions of real and nominal variables to the monetary policy shock we have observed in the previous example.

Let us now move to the effects of the single structural shocks on each of the 20 variables

under analysis, and start with the demand shock, whose effects are illustrated in Figure 4. All the real variables react positively, and the prices also increase. As a consequence, the federal fund rate increases substantially, as well as the 10 year rate, with a drop in monetary indicators and in the stock market index and an appreciation of the effective exchange rate. The effects are generally statistically significant.

The effects of the supply shock are presented in Figure 5. Now all the real variables deteriorate, and all the price variables increase. The latter effect is more marked than the former, so that there is an increase in the federal fund rate, though much smaller than in the case of the demand shock. The 10 year rate also increases, and there is a drop in the monetary indicators and in the stock market index and a depreciation of the effective exchange rate, followed by an appreciation that starts about one year after the shock. The effects are generally statistically significant, in particular at short horizons in the case of the fast variables.

Finally, Figure 6 illustrates the consequences of a financial / monetary shock. The responses are very similar to those we have obtained in the previous subsection as a reaction to a monetary policy shock. In particular, there is a deterioration in all the real variables and a decrease in all the price variables, sometimes with a delay of a few months. The federal funds rate increases, as well as the 10 year rate, while the monetary indicators drop and the exchange rate appreciates. The main difference with respect to the analysis in the previous subsection is in the reaction of the stock index, which is now negative and significant, at least in the short run, rather than close to zero and positive.

Overall, these empirical applications illustrate how the RR-BVAR can be easily used to conduct structural analysis, along the lines of either the BVAR approach or the FAVAR methodology. The two possibilities lead to very similar results in the case of a monetary shock, with sensible responses from an economic point of view.

7 Conclusions

In this paper we address the issue of parameter dimensionality reduction in Vector Autoregressive models (VARs) for many variables by imposing reduced rank restrictions on the coefficient matrices.

We derive the Wold representation implied by reduced rank VARs (RR-VARs), and show that it is also closely related to that associated with dynamic factor models.

Next, we describe classical and Bayesian estimation methods for large RR-VARs, and illustrate how the estimated models can be used for structural analysis.

Structural analysis with the RR-VARs is then illustrated with empirical applications on the transmission mechanism of monetary policy, and of demand, supply and financial

shocks, in a model that includes 20 key macroeconomic variables for the US.

Overall, the method is general, simple, and well performing. Hence, it is promising as an alternative tool for structural analysis using information in large datasets.

Appendix: convergence diagnostics

In this section we discuss convergence of the algorithm used in the paper. The results in the paper are based on 2000 draws from the simulated posterior, obtained as follows. We draw 10 parallel chains of 2400 draws each. For each chain we discard the first 400 draws and keep each 10-th draw. Then the draws are from the various chains are mixed together.

We assess convergence by looking at the Potential Scale Reduction Factor (PSRF), the autocorrelation function and the recursive means of the coefficients. The Potential Scale Reduction Factor (PSRF), proposed by Gelman and Rubin (1992) is a measure of convergence based within-chain and between-chain variance of the draws. When the PRSF is high (greater than 1.1 or 1.2), this is taken as indication that one should run the chains longer to improve convergence to the stationary distribution. The autocorrelation function is directly related to the efficiency of the algorithm (see e.g. Geweke (1992)).

We have computed convergence diagnostics for all the coefficients, but we report only the results related to the more problematic coefficients to save space. In particular, we only report results for the estimation of the second empirical application on demand, supply and financial shocks. This is computationally more challenging because of the restrictions imposed on the matrix B_0.

Figure 7 reports the results of the convergence checks and is organized as follows. The first row of graphs contains results for the parameters in $A(L)$, the second row of graphs contains results for the parameters in $\tilde{B}_0$, the third row contains results for the product $A(L)B_0$. Then, in the first column we report the distribution of the PRSF across all the coefficients. As is clear, the distribution of the PRSF is concentrated around the optimal value of 1, signaling convergence. In the second and third column we report results related to the autocorrelation functions. As many coefficients are included in the matrices $A(L)$, $\tilde{B}_0$ and $A(L)B_0$, we report in the second column of graphs the *maximum* autocorrelation coefficient at all lag orders up to 50. So for example in the panel at the center of Figure 7 we report the maximum autocorrelation of the coefficients in $\tilde{B}_0$. Similarly, in the last columns we report the *median* autocorrelation coefficient at all lag orders up to 50. As is clear, the autocorrelation are small and decay very rapidly, especially for the coefficients in $A(L)$ and $A(L)B_0$.

As the draws from $\tilde{B}_0$ are more problematic, because these parameter enter the model nonlinearly and are drawn using a Metropolis step, we also provide graphs documenting the

autocorrelation function and recursive means for the 17 coefficients entering the matrix $\tilde{B}_0$ (listed in Table 2) in our empirical application on demand, supply and financial shocks. The autocorrelation functions are depicted in Figure 8 and show that for all coefficients there is a quick decay to 0. The recursive means for each of the 10 chains considered are depicted in Figure 9, and as is clear each of the 10 chains converges to the mean of the stationary target distribution.

References

[1] Bai, J. and Ng, S., 2006a. "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions," Econometrica, 74(4), 1133-1150.

[2] Bai, J. and Ng, S., 2006b. "Evaluating Latent and Observed Factors in Macroeconomics and Finance", Journal of Econometrics, 113(1-2), 507-537.

[3] Banbura M., Giannone D. and Reichlin, L., 2010. "Large Bayesian VARs", Journal of Applied Econometrics, 25(1), 71-92.

[4] Baumeister, C., Liu, P. and Mumtaz, H. 2010, "Changes in the transmission of moneetary policy: Evidence from a time-varying factor-augmented VAR", Bank of England Working Paper.

[5] Bernanke, B., J. Boivin, P. Eliasz‘, 2005, Measuring the effects of monetary policy: a Factor-Augmented Vector Autoregressive (FAVAR) approach, The Quarterly Journal of Economics, 120(1), 387-422.

[6] Camba-Mendez, G., Kapetanios, G., Smith, R.J., and Weale, M.R., 2003, "Tests of Rank in Reduced Rank Regression Models", Journal of Business and Economic Statistics, 21, 145-155.

[7] Carriero, C., Kapetanios, G. and Marcellino, M., 2011, “Forecasting Large Datasets with Bayesian Reduced Rank Multivariate Models”, Journal of Applied Econometrics, 26, 736-761.

[8] Cragg, J. G., and Donald, S. G. (1996), "On the Asymptotic Properties of LDU-Based Tests of the Rank of a Matrix," Journal of the American Statistical Association, 91, 1301-1309.

[9] De Mol, C., Giannone, D. and Reichlin, L., 2006. "Forecasting Using a Large Number of Predictors. Is Bayesian Regression a Valid Alternative to Principal Components?", Journal of Econometrics, 146(2), 318-328.

[10] Doan, T., R. Litterman, and C. A. Sims, 1984, "Forecasting and Conditional Projection Using Realistic Prior Distributions," Econometric Reviews, 3, 1-100.

[11] Del Negro, M., C. Otrok (2008), Dynamic factor models with time-varying parameters: Measuring changes in international business cycles, mimeo.

[12] Doz, C., Giannone, D. and Reichlin, L., 2011. "A two-step estimator for large approximate dynamic factor models based on Kalman filtering," Journal of Econometrics, 164(1), 188-205.

[13] Dufour, J.M. and Stevanovic, D. (2010), "Factor-Augmented VARMA Models: Identification, Estimation, Forecasting and Impulse Responses", mimeo.

[14] Eickmeier, S., W. Lemke, M. Marcellino, 2011, "Classical time-varying FAVAR models - estimation, forecasting and structural analysis", Bundesbank Discussion Paper 4/2011

[15] Forni, M., Hallin, M., Lippi, M. & Reichlin, L., 2000. The generalized factor model: identification and estimation, *The Review of Economics and Statistics* 82(4), 540-554.

[16] Gelman, A„ and Rubin, D. (1992), "Inference from Iterative Simulation using Multiple Sequences," Statistical Science, 7, 457-5 11.

[17] Geweke, J., 1977 "The dynamic factor analysis of economic time series", in Dennis J. Aigner and Arthur S. Goldberger (eds.) Latent Variables in Socio-Economic Models (Amsterdam: North-Holland).

[18] Geweke, J., 1992, "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments." In Bayesian Statistics 4 (eds. J.M. Bernardo, J. Berger, A.P. Dawid and A.F.M. Smith), Oxford: Oxford University Press, 169-193.

[19] Geweke, J., 1996, Bayesian Reduced Rank Regression in Econometrics, Journal of Econometrics 75, 121–146.

[20] Geweke, J., 1999. "Using simulation methods for bayesian econometric models: inference, development,and communication," Econometric Reviews, 18(1), 1-73.

[21] Gill, L., and Lewbel, A., 1992, "Testing the Rank and Definiteness of Estimated Matrices With Applications to Factor, State-Space and ARMA Models," Journal of the American Statistical Association, 87, 766-776.

[22] Kose, M. A.,C. Otrok, C. H. Whiteman, 2003, "International business cycles: world, region and country-specific factors ", The American Economics Review, 93(4), 1216-1239.

[23] Johansen, S. (1995), Likelihood-Based Inference in Cointegrated Vector Autoregressive Models, Oxford: Oxford University Press.

[24] Lutkepohl, H., 2007, "New Introduction to Multiple Time Series Analysis," Oxford University Press.

[25] Reinsel, G., 1983, "Some Results on Multivariate Autoregressive Index Models,"Biometrika, 70, 145–156.

[26] Reinsel, G. C., and Velu, R. P., 1998, "Multivariate Reduced Rank Regression," Lecture Notes in Statistics 136. New York: Springer-Verlag.

[27] Robin, J. M., and Smith, R. J., 2000, "Tests of Rank," Econometric Theory, 16, 151-175.

[28] Stock, J. H., and M.W. Watson, 1989, "New indexes of coincident and leading economic indicators", NBER Macroeconomic Annual, 351-94.

[29] Stock, J. H. and M. W. Watson, 2002a, "Macroeconomic Forecasting Using Diffusion Indexes", Journal of Business and Economic Statistics, 20, 147-62.

[30] Stock, J. H. and M. W. Watson, 2002b, "Forecasting Using Principal Components from a Large Number of Predictors", Journal of the American Statistical Association, 97, 1167–1179.

[31] Velu, R. P., Reinsel, G. C., and Wichern, D. W., 1986, "Reduced Rank Models for Multiple Time Series," Biometrika, 73, 105–118.

Table 1: Variables used in the empirical analysis

Variable	**Short name**	**Type**
Employees on nonfarm payroll	EMP: TOTAL	slow
CPI, all items	CPI-U: ALL	slow
Index of sensitive material prices	SENS MAT'LS PRICE	slow
Personal income	PI	slow
Real Consumption	CONSUMPTION	slow
Industrial Production Index	IP: TOTAL	slow
Capacity Utilization	CAP UTIL	slow
Unemployment rate	U: ALL	slow
Housing starts	HSTARTS: TOTAL	slow
Producer Price Index (finished goods)	PPI: FIN GDS	slow
Implicit price deflator for personal consumption expenditures	PCE DEFL	slow
Average hourly earnings	AHE: GOODS	slow
Federal Funds, effective	FEDFUNDS	fast
M1 money stock	M1	fast
M2 money stock	M2	fast
Total reserves of depository institutions	RESERVES TOT	fast
Nonborrowed reserves of depository institutions	RESERVES NONBOR	fast
S&P's common stock price index	S&P 500	fast
Interest rate n treasury bills, 10 year constant maturity	10 YR T-BOND	fast
Effective Echange rate	EX RATE: AVG	fast

Table 2: Identification of the "factors" in the RR-VAR

Variable	**Short name**	**F1**	**F2**	**F3**
Employees on nonfarm payroll	EMP: TOTAL	1	0	0
CPI, all items	CPI-U: ALL	0	1	0
Index of sensitive material prices	SENS MAT'LS PRICE	0	$b_{2,3}$	0
Personal income	PI	$b_{1,4}$	0	0
Real Consumption	CONSUMPTION	$b_{1,5}$	0	0
Industrial Production Index	IP: TOTAL	$b_{1,6}$	0	0
Capacity Utilization	CAP UTIL	$b_{1,7}$	0	0
Unemployment rate	U: ALL	$b_{1,8}$	0	0
Housing starts	HSTARTS: TOTAL	$b_{1,9}$	0	0
Producer Price Index (finished goods)	PPI: FIN GDS	0	$b_{2,10}$	0
Implicit price deflator for personal consumption expenditures	PCE DEFL	0	$b_{2,11}$	0
Average hourly earnings	AHE: GOODS	$b_{1,12}$	0	0
Federal Funds, effective	FEDFUNDS	0	0	1
M1 money stock	M1	0	0	$b_{3,14}$
M2 money stock	M2	0	0	$b_{3,15}$
Total reserves of depository institutions	RESERVES TOT	0	0	$b_{3,16}$
Nonborrowed reserves of depository institutions	RESERVES NONBOR	0	0	$b_{3,17}$
S&P's common stock price index	S&P 500	0	0	$b_{3,18}$
Interest rate n treasury bills, 10 year constant maturity	10 YR T-BOND	0	0	$b_{3,BaiandNg(2009)}$
Effective Echange rate	EX RATE: AVG	0	0	$b_{3,20}$

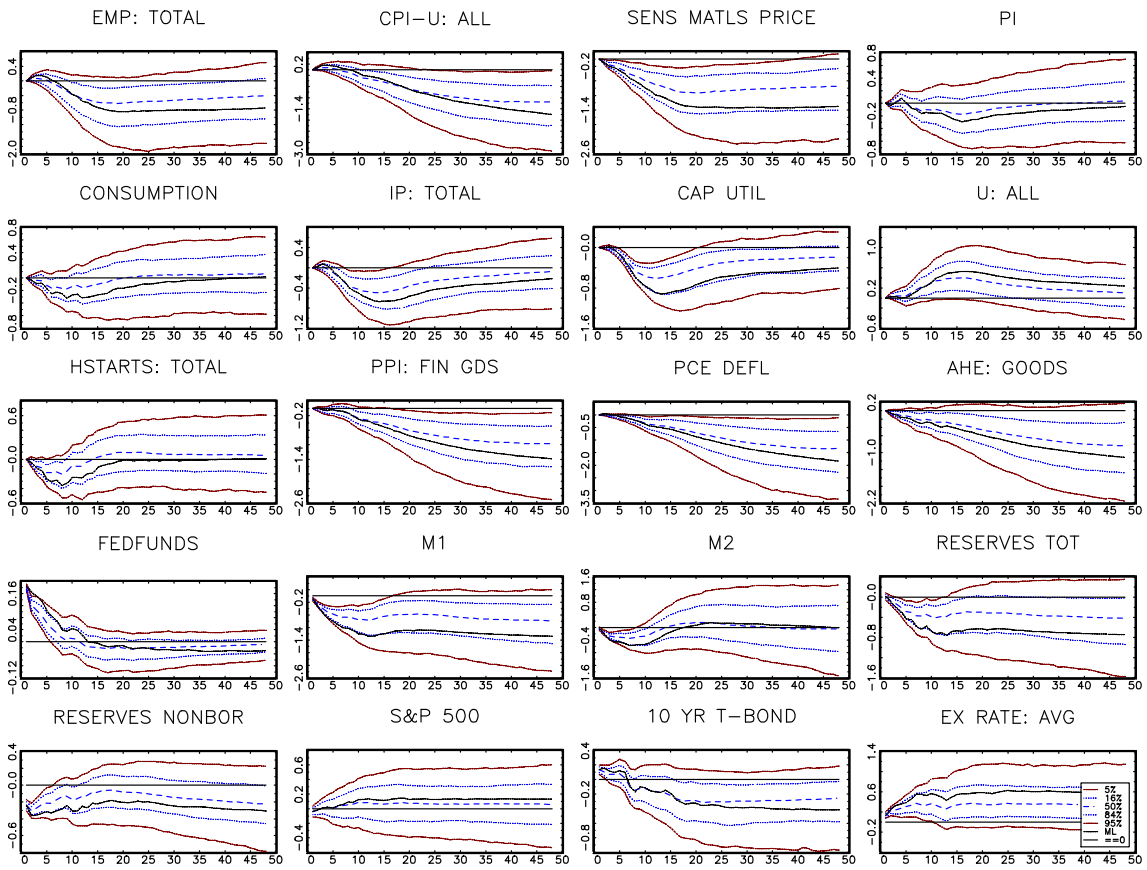

Figure 1: Bayesian and classical impulse responses to monetary policy shock based on RR-BVAR

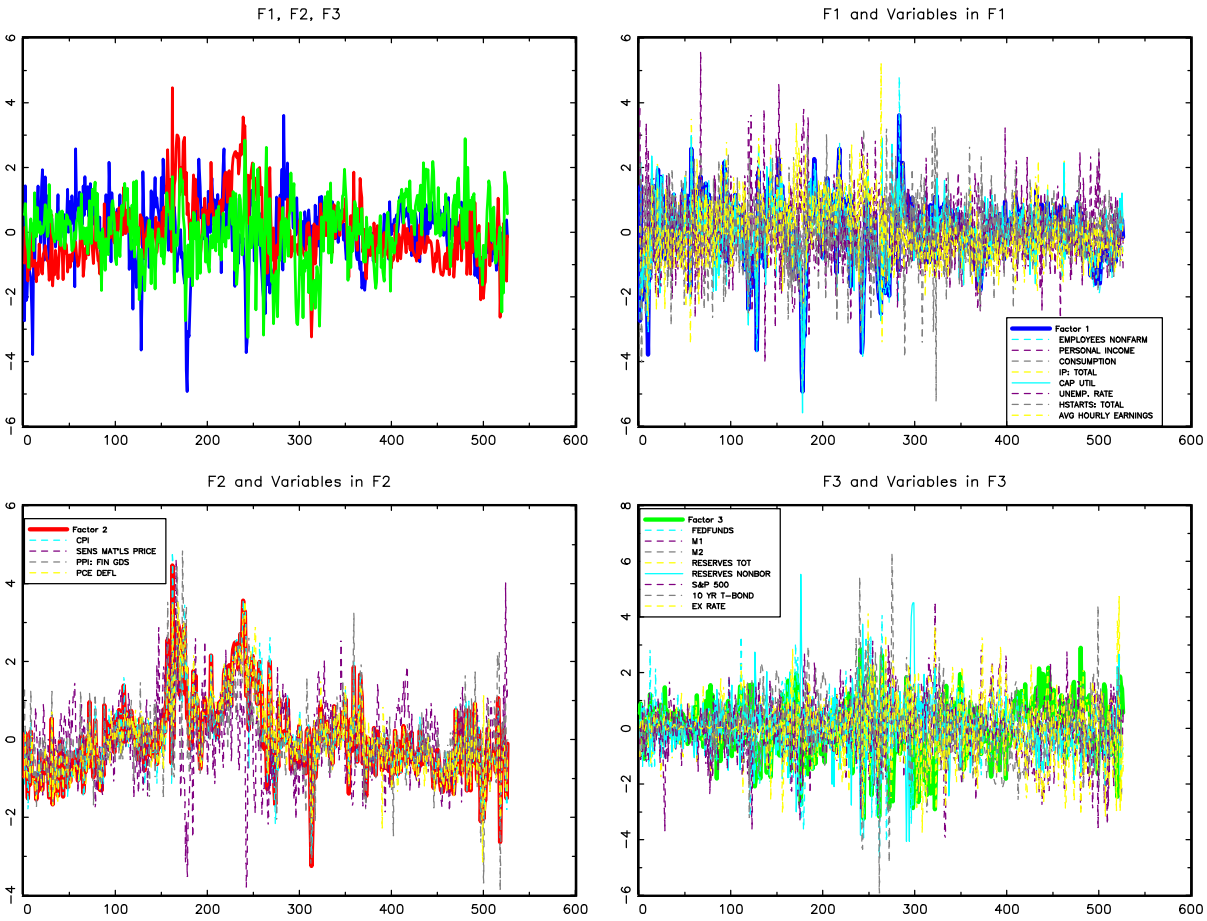

Figure 2: The RR-BVAR "factors" and their components

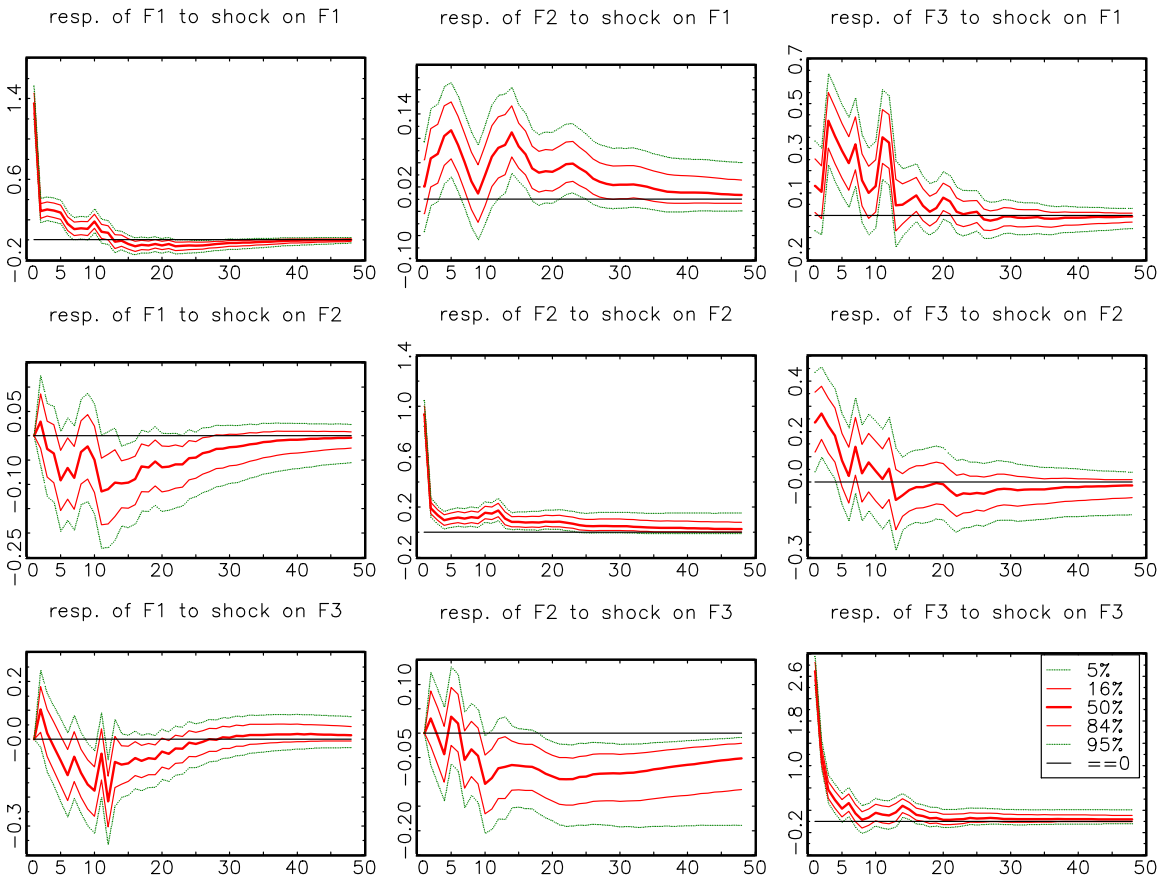

Figure 3: Responses of the RR-BVAR structural "factors"

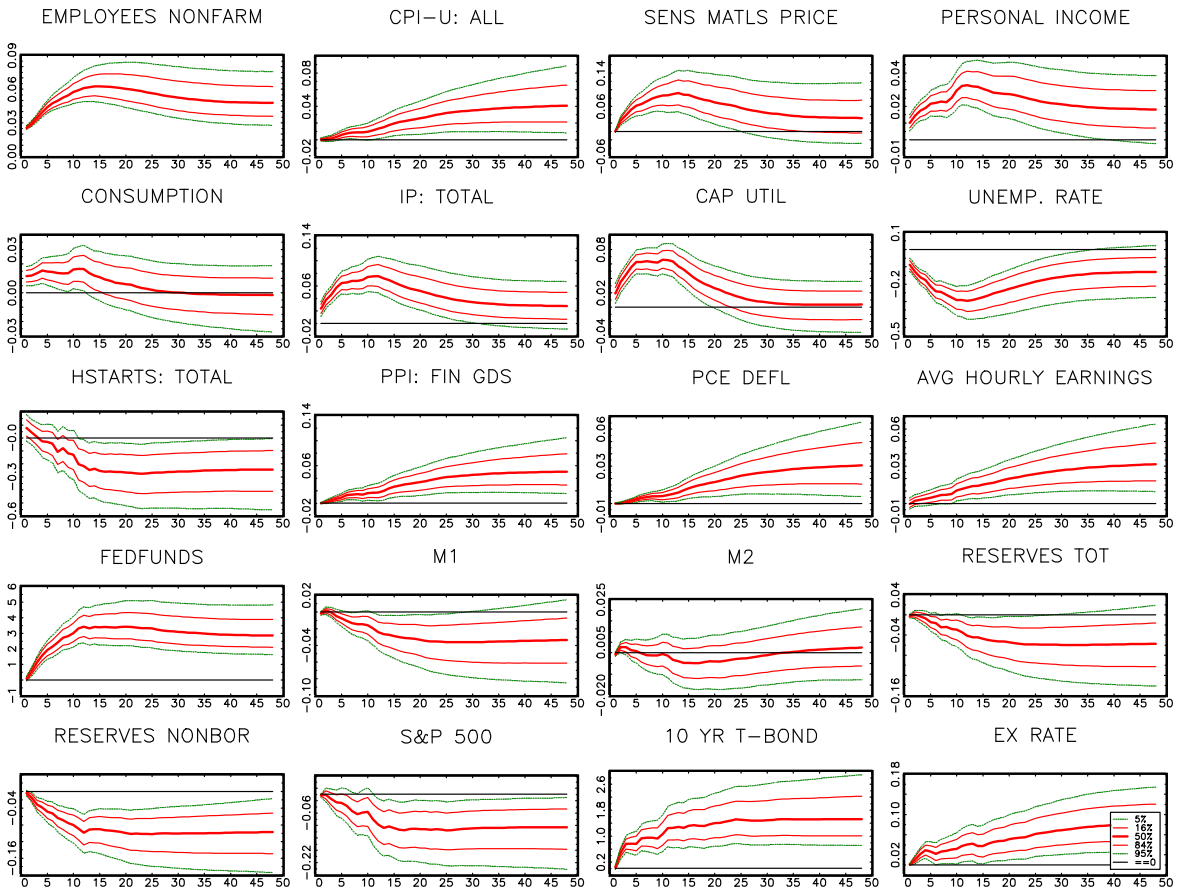

Figure 4: Effects of a demand shock computed from the factor representation of the RR-BVAR

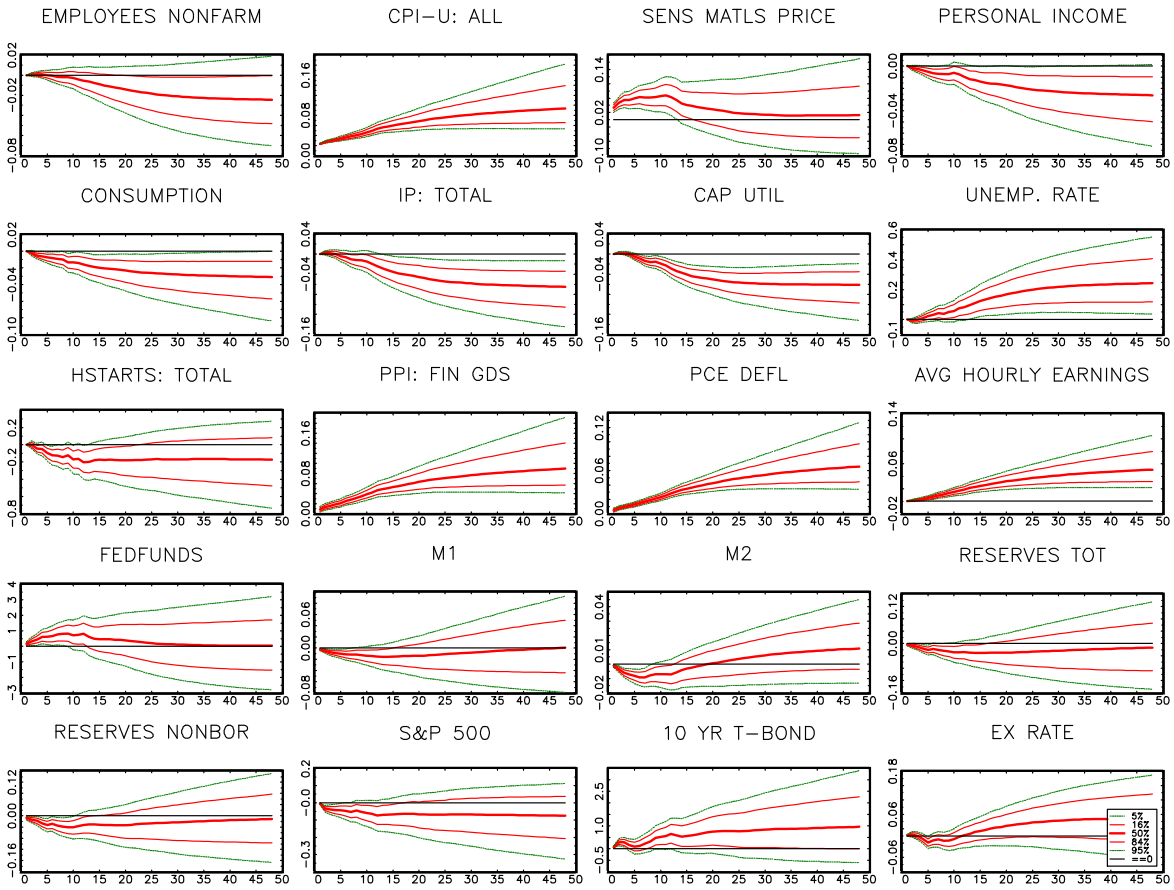

Figure 5: Effects of a supply shock computed from the factor representation of the RR-BVAR

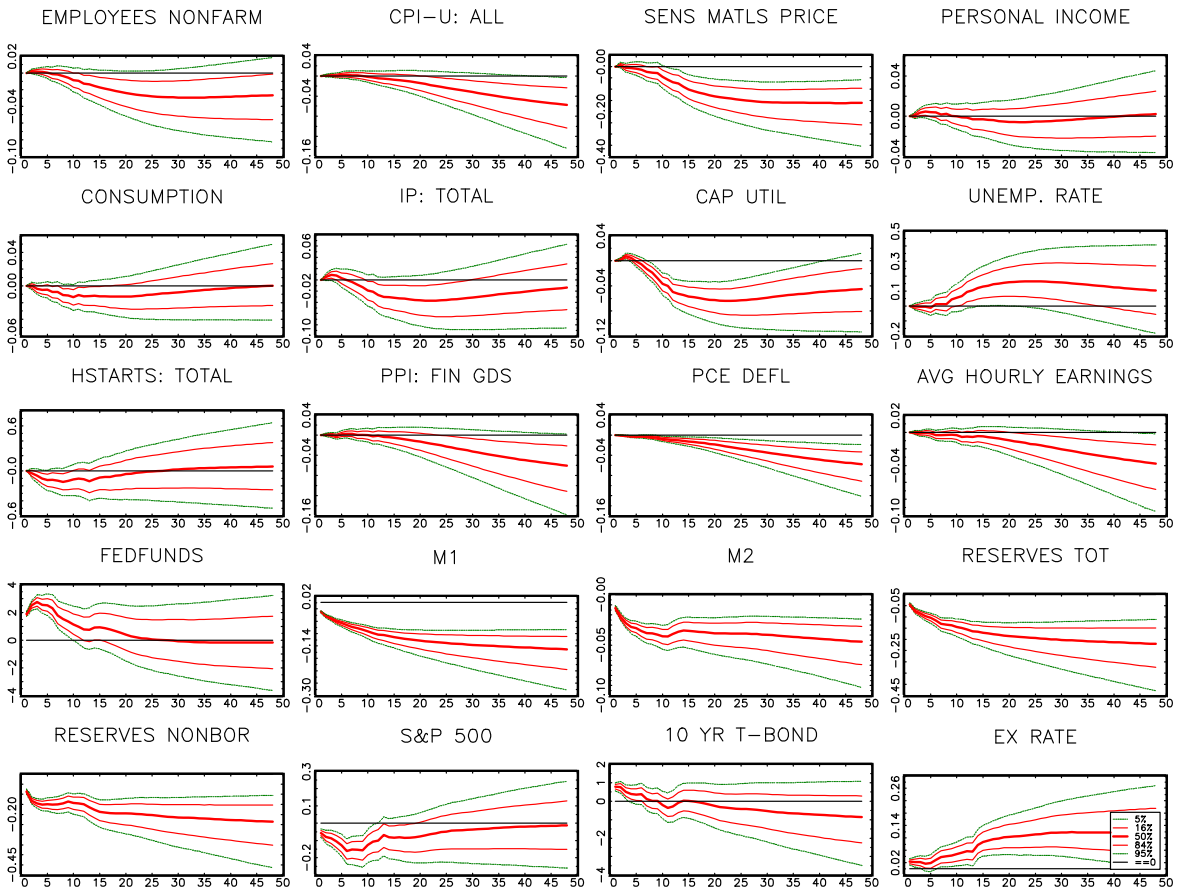

Figure 6: Effects of a financial / interest rate shock computed from the factor representation of the RR-BVAR

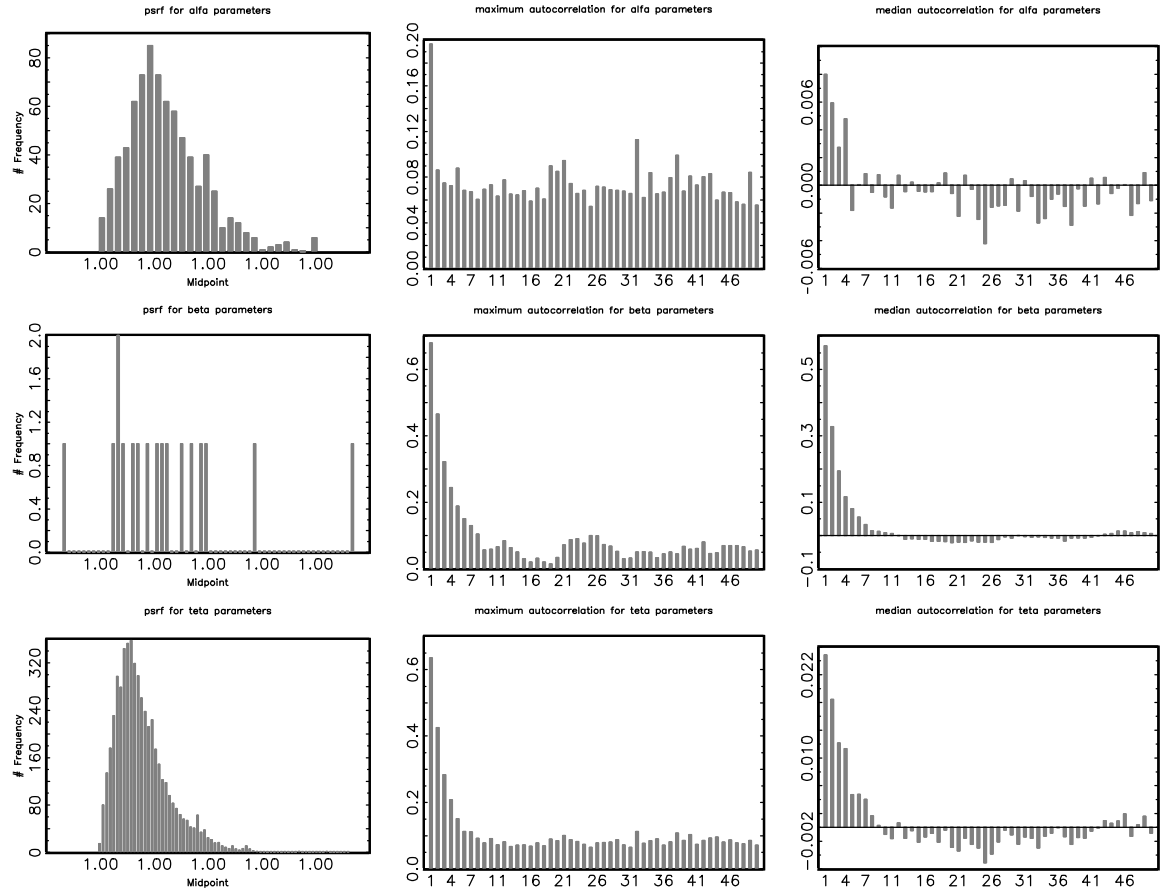

Figure 7: Convergence Diagnostics. The first row of graphs contains results for the parameters in $A(L)$, the second row of graphs contains results for the parameters in $\tilde{B}_0$, the third row contains results for the product $A(L)B_0$. Panels in the first column contain the distribution of the PRSF across all the coefficients , panels in the second column contain the maximum autocorrelation coefficient, panels in the third colon contain the median autocorrelation coefficient.

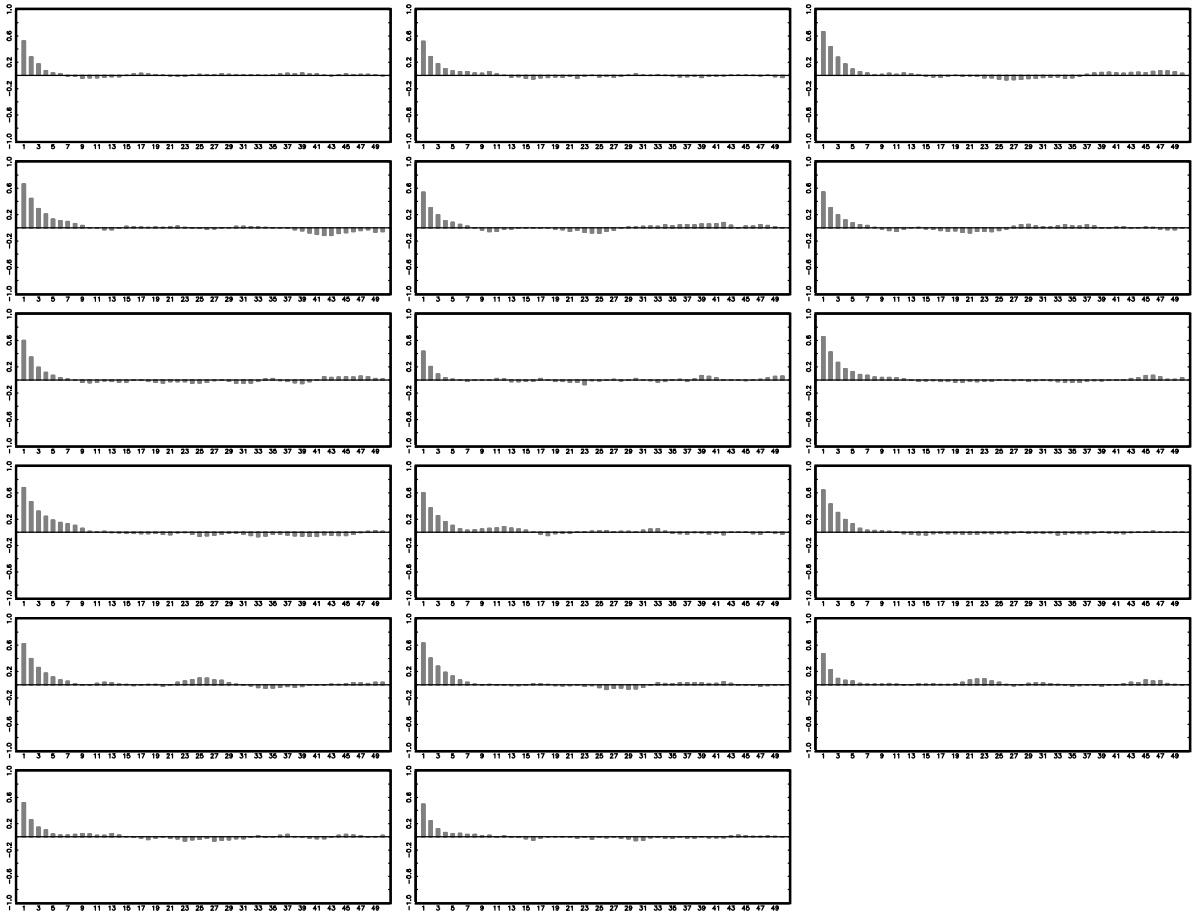

Figure 8: Autocorrelation function of the coefficients in $\tilde{B}_0$.

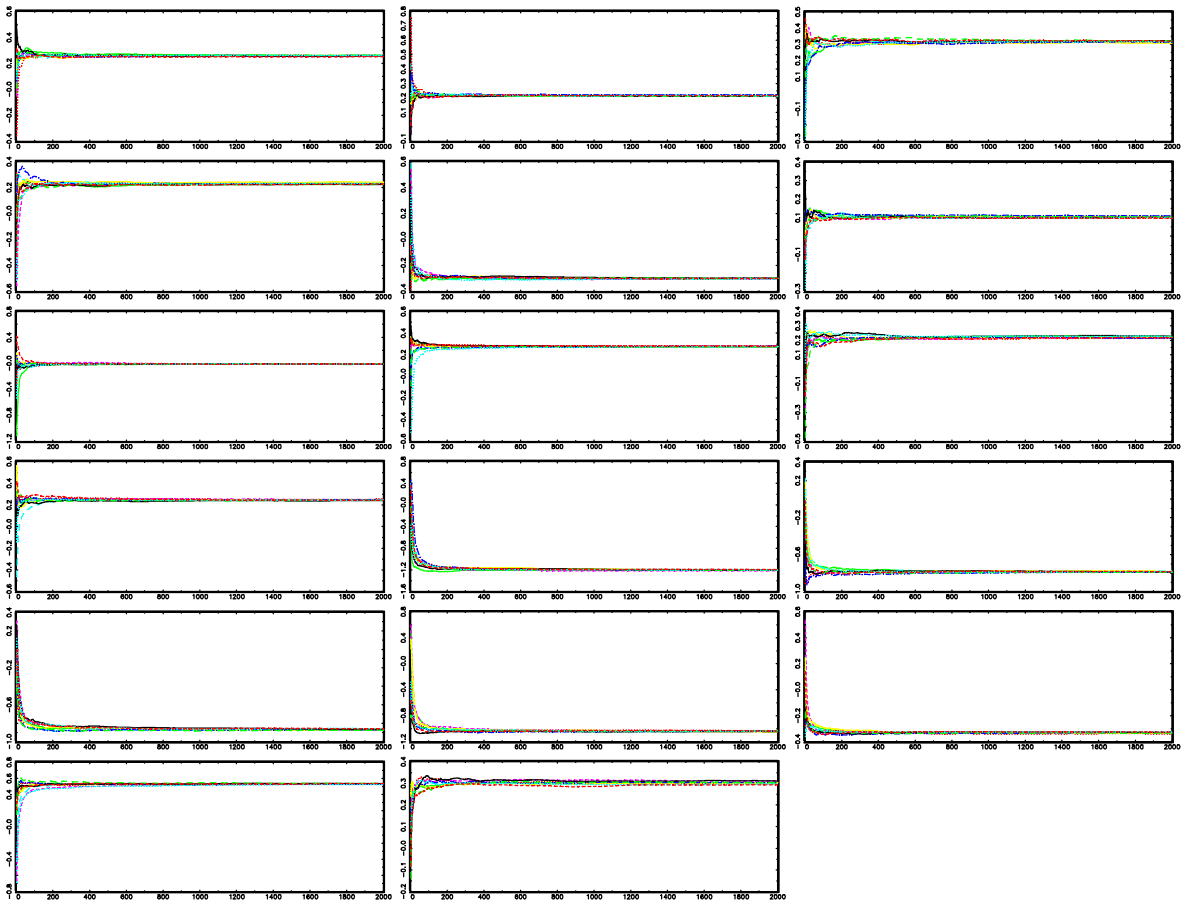

Figure 9: Recursive means of the coefficients in $\tilde{B}_0$. 10 independent chains.