

Structural Analysis with Multivariate Autoregressive Index Models*

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

Bocconi University, IGIER and CEPR

massimiliano.marcellino@unibocconi.it

May 29, 2015

Abstract

We address the issue of parameter dimensionality reduction in Vector Autoregressive models (VARs) for many variables by imposing specific reduced rank restrictions on the coefficient matrices that simplify the VARs into Multivariate Autoregressive Index (MAI) models. We derive the Wold representation implied by the MAIs and show that it is closely related to that associated with dynamic factor models. Next, we describe classical and Bayesian estimation of large MAIs, and discuss methods for the rank determination. Then, the theoretical analysis is extended to the case of general rank restrictions on the VAR coefficients. Finally, the performance of the MAIs is compared with that of large Bayesian VARs in the context of Monte Carlo simulations and two empirical applications, on the transmission mechanism of monetary policy and the propagation of demand, supply and financial shocks..

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

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

*We are grateful to the Editor Jeorg Breitung, three anonymous Referees, Luc Bauwens, Siddhartha Chib, Todd Clark, Herman van Dijk, Simon Potter, Frank Schorfheide, and participants at the Esobe Conference, seminars at Univeristy of Pennsylvania, Erasmus University Rotterdam, DIW-Berlin and the Federal Reserve Bank of St. Louis and of Cleveland, for helpful comments and suggestions.

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 dimensions. 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 VARs, see e.g. Lutkepohl (2007).

Classical methods 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, e.g. Bernanke et al. (2005), Kose et al. (2005), Del Negro and Otrok (2008), Baumeister and Mumtaz (2010), Eickmeier, Lemke, Marcellino (2014).

More recently, large Bayesian VARs (BVARs) were proposed as an alternative modelling device to factor models, e.g. De Mol, 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 and 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 (see also Lutkepohl (1984)). Furthermore, structural identification in factor models is in principle rather easy but in practice often complex, so that few empirical applications have been produced (e.g., Forni and Gambetti (2010)). In addition, 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.

In this paper, we suggest to use a model that bridges BVARs and FAVARs. Specifically, we propose to impose reduced rank restrictions on the parameter matrices of 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. The resulting specification is a Multivariate Autoregressive Index (MAI) model, originally introduced by Reinsel (1983) within a classical context.

From a theoretical point of view, we build on Reinsel (1983) and extend his work in four directions. First, we derive asymptotic results for classical estimation of MAI models for large N . Second, we provide conditional posterior distributions and an efficient MCMC algorithm for Bayesian estimation of large MAI models. Third, we introduce a moving average representation of the MAI model that is particularly useful for identifying structural shocks and their dynamic propagation. Finally, we extend the theoretical analysis to general reduced rank VAR models, finding however a substantial increase in computational costs, which makes them less attractive than MAI for economic applications based on large datasets.

From an applied perspective, we assess the relative performance of large MAI and BVAR models both in extensive simulation experiments and in two empirical applications.

The paper is structured as follows. In Section 2 we introduce the MAI model, where 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 MAI. We then derive alternative moving average representations of the MAI, where each variable is driven either by the N original MAI 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 MAI. Reduced rank regressions have been introduced by Anderson (1951), and the specific case of reduced rank autoregressions and MAI models has been studied in detail by Reinsel (1983) (see also Velu, Reinsel and Wichern (1986) and Reinsel and Velu (1998)). As mentioned, we 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 the coefficients.

In Section 4 we discuss classical and Bayesian methods for the determination of the rank, r , of the MAI. 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 the maximum of the posterior density of r , assuming a flat prior. We suggest to approximate the marginal data density numerically by using Rao-Blackwellisation combined with the harmonic mean estimator proposed by Gelfand and Dey (1994) and Geweke's (1999).

In Section 5 we discuss a more general reduced rank VAR model, which nests the MAI as a special case, as well as the relationship of the MAI model with the reduced rank multivariate regression studied in a Bayesian setting by Geweke (1996).

In Section 6 we perform a set of Monte Carlo exercises, which show that the MAI estimated with Bayesian methods systematically outperforms the classical MAI, as well as an unrestricted BVAR when the data generating process contains rank reduction in the conditional mean parameters.

In Section 7 we illustrate the theoretical proposals by means of two empirical applications. First, we replicate in the MAI context the BVAR analysis of the transmission of US monetary policy shocks conducted by Banbura, Giannone and Reichlin (2010), using an updated dataset. We use the N -shock MA representation of the MAI and obtain responses that are economic sensible and sometimes different from those resulting from the full rank BVAR approach of Banbura et al. (2010). We also show that the Bayesian procedure produces more reasonable impulse responses than the classical ones, and reduces estimation uncertainty. Second, we assess the effects of demand, supply, and financial / monetary shocks. In this case we use the FAVAR-style MA representation of the MAI and assume that the factors reflect movements in real, financial, and price variables, where the shocks associated with these 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 similar.

Finally, in Section 8 we summarize the main results of the paper and propose directions for additional research in this area.

2 MAI Model Specification and Moving Average Representations

2.1 The MAI-model

We assume that the N -dimensional zero mean weakly stationary process $Y_t = (y_{1,t}, \dots, y_{N,t})'$ admits the representation:

$$Y_t = \Phi(L)Y_t + \epsilon_t, \quad (1)$$

where $t = 1, \dots, T$, $\Phi(L) = \Phi_1 L + \dots + \Phi_p L^p$ is a polynomial of order p , and ϵ_t are i.i.d. $N(0, \Sigma)$.

Following Reinsel (1983), we further assume that $\Phi(L)$ can be factorized as $\Phi(L) = A(L)B_0$, where $A(L) = A_1 L + \dots + A_p L^p$, each matrix A_u is of dimension $N \times r$, $u = 1, \dots, p$, and B_0 is of dimension $r \times N$ and full row rank. The resulting model, labeled Multivariate Autoregressive Index (MAI) model by Reinsel (1983), is:¹

$$Y_t = \sum_{u=1}^p A_u B_0 Y_{t-u} + \epsilon_t. \quad (2)$$

If r is much smaller than N there are much fewer parameters in the MAI model in (2) than in the corresponding unrestricted VAR in (1). For example, in our empirical application, we have $T = 360$, $N = 20$, $p = 13$ and $r = 3$, so that there are $N(Np - r(p+1)) = 4360$ parameters less in the MAI in (2) than in the corresponding unrestricted VAR in (1). The total number of parameters in (1) and (2) is, respectively, $N^2 p = 5200$ and $Nr(p+1) = 20 \cdot 3 \cdot 14 = 840$.

From an economic point of view, the MAI model in (2) implies that all the variables are driven by a (possibly much) smaller number of indicators, the r variables $B_0 Y_{t-u}$, which can be labeled as "indexes" (as in Reinsel, 1983) or as "factors", as in the factor literature. We prefer the latter denomination and therefore define the factors F_t as:

$$F_t = B_0 Y_t. \quad (3)$$

Using (3), it is straightforward to rewrite the MAI in (2) as:

$$Y_t = \sum_{u=1}^p A_u F_{t-u} + \epsilon_t = A(L)F_t + \epsilon_t. \quad (4)$$

¹More general reduced rank models are considered in Section 5. Error correction models for cointegrated variables are also a special class of reduced rank models, see e.g. Johansen (1995 and Koop et al. (2006) in, respectively, classical and Bayesian contexts. See also George et al. (2005) for a Bayesian stochastic search approach to selecting restrictions for VAR models.

As in the case of the factor model, the "loadings" A_u and the factor weights B_0 are not uniquely identified in a MAI model. Without any loss of generality, we assume that $B_0 = (I_r, \tilde{B}_0)$. We will come back to the relationship between MAI and factor models in the last subsection.

An important characteristic of the MAI model is that the linear combinations $B_0 Y_t$ in (3) have a closed form $VAR(p)$ representation, while in general when Y_t follows an unrestricted 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 (2) and use (3) to get:

$$F_t = B_0 \sum_{u=1}^p A_u F_{t-u} + B_0 \epsilon_t = C(L)F_t + u_t, \quad (5)$$

with $C(L) = B_0 A(L) = B_0 A_1 L + B_0 A_2 L^2 + \dots + B_0 A_p L^p$ and with $u_t = B_0 \epsilon_t$ being an i.i.d. Gaussian process with mean zero and variance $\Omega = B_0 \Sigma B_0'$.

2.2 Moving average representations

In order to use the MAI model for structural impulse response analysis, we need to derive its moving average (MA) representation. We consider three alternative representations.

Inverting equation (2), under the weak stationarity assumption, provides a first moving average representation:

$$Y_t = (I - A(L)B_0)^{-1} \epsilon_t. \quad (6)$$

From this expression it is easy to derive optimal forecasts and impulse response functions by using standard techniques, see e.g. Lutkepohl (1990, 2007).

A second moving average representation is:

$$Y_t = (A(L)(I - B_0 A(L))^{-1} B_0 + I) \epsilon_t. \quad (7)$$

This expression is obtained by first deriving the moving average representation for F_t from equation (5):

$$F_t = (I - C(L))^{-1} u_t = (I - B_0 A(L))^{-1} B_0 \epsilon_t, \quad (8)$$

and then inserting it into equation (4). The two alternative moving average representations

for Y_t in (6) and (7) are of course equivalent:²

$$(I - A(L)B_0)^{-1} = A(L)(I - B_0A(L))^{-1}B_0 + I. \quad (9)$$

A third moving average representation is particularly convenient for structural analysis. Let us introduce the $(N - r) \times N$ full row 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 $B_0B'_0$ and $B_0\Sigma B'_0$ have full rank (as we assumed B_0 has full row rank) and we have the following decomposition (see Johansen 1995, p.39, and Centoni and Cubadda 2003, p.48):³

$$\Sigma B'_0(B_0\Sigma B'_0)^{-1}B_0 + B'_{0\perp}(B_{0\perp}\Sigma^{-1}B'_{0\perp})^{-1}B_{0\perp}\Sigma^{-1} = I_N. \quad (10)$$

This key identity can now be inserted into the Wold representation in (7) to yield:

$$Y_t = (\Sigma B'_0(B_0\Sigma B'_0)^{-1} + A(L)(I - B_0A(L))^{-1})B_0\epsilon_t + B'_{0\perp}(B_{0\perp}\Sigma^{-1}B'_{0\perp})^{-1}B_{0\perp}\Sigma^{-1}\epsilon_t. \quad (11)$$

Since $B_0\epsilon_t = u_t$, $\Omega = B_0\Sigma B'_0$, and defining $B_{0\perp}\Sigma^{-1}\epsilon_t = \xi_t$, we have:

$$Y_t = (\Sigma B'_0\Omega^{-1} + A(L)(I - B_0A(L))^{-1})u_t + B'_{0\perp}(B_{0\perp}\Sigma^{-1}B'_{0\perp})^{-1}\xi_t. \quad (12)$$

The representation in (12) 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 linear combinations of ξ_t . Since

$$E(u_t\xi'_t) = E(B_0\epsilon_t\epsilon'_t\Sigma^{-1}B'_{0\perp}) = 0, \quad (13)$$

$$E(u_{t-i}\xi'_t) = 0, \quad E(u_t\xi'_{t-i}) = 0, \quad i > 0, \quad (14)$$

u_t and ξ_t are uncorrelated at all leads and lags.

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. (2014). For example, the simplest option is the Cholesky decomposition

$$v_t = Pu_t, \quad (15)$$

²To derive this result note that $A(L)^{-1} = B_0 + (I - B_0A(L))A(L)^{-1}$, and premultiply both sides of this equation by $(I - B_0A(L))^{-1}$ to obtain $(I - B_0A(L))^{-1}A(L)^{-1} = (I - B_0A(L))^{-1}B_0 + A(L)^{-1}$. The term on the LHS $(I - B_0A(L))^{-1}A(L)^{-1}$ can be rewritten as $[A(L)(I - B_0A(L))]^{-1} = [(I - A(L)B_0)A(L)]^{-1} = A(L)^{-1}(I - A(L)B_0)^{-1}$, therefore we have that

$$A(L)^{-1}(I - A(L)B_0)^{-1} = (I - B_0A(L))^{-1}B_0 + A(L)^{-1},$$

and (9) is obtained by multiplying both sides of the equation above by $A(L)$.

³We are grateful to an anonymous Referee for pointing out the decomposition in (10)

where P is a lower triangular matrix such that $\Omega = P^{-1}P^{-1'}$, which implies that v_t are structural shocks with $\text{Var}(v_t) = P\Omega P' = I$. Hence, Combining (15) with (12) yields

$$Y_t = (\Sigma B_0' \Omega^{-1} + A(L)(I - B_0 A(L))^{-1})P^{-1}v_t + B_{0\perp}'(B_{0\perp}\Sigma^{-1}B_{0\perp}')^{-1}\xi_t, \quad (16)$$

from which impulse response functions can be easily computed.

Note that, since $u_t = B_0\epsilon_t$, the structural shocks v_t are also related to the ϵ_t errors in the Wold representations in (6) or (7), via the relationship $v_t = PB_0\epsilon_t$. However, from a structural point of view, there is an important difference between the representations in (6) or (7) and that in (12). In the former case there can be as many structural shocks as variables, namely N , while in (12) we are explicitly assuming that there is a reduced number of structural shocks, r , which drive all the factors F_t . In principle, there could be other $N - r$ structural shocks that drive the $(N - r)$ errors ξ_t in (12), but in practice these are never considered in the factor literature.⁴

2.3 Relationship with factor models

The MAI model is clearly 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 . Moreover, the "common component" of the MAI model, $(\Sigma B_0' \Omega^{-1} + A(L)(I - B_0 A(L))^{-1})u_t$ in (12), is uncorrelated at all leads and lags with the error terms $B_{0\perp}'(B_{0\perp}\Sigma^{-1}B_{0\perp}')^{-1}\xi_t$.

However, there are also important differences between MAI and factor models. In particular, in the MAI model only lags of the "factors" F_t affect the variables while in factor models there can be contemporaneous effects as well. Moreover, the errors $B_{0\perp}'(B_{0\perp}\Sigma^{-1}B_{0\perp}')^{-1}\xi_t$ in (12) can be in general correlated among themselves, while in factor models they must

⁴There is a case where shocking the factors or shocking the variables produces the same responses and this happens when the factors are equal to a subset of the variables and we shock one of the variables in this subset. Formally, suppose that $\tilde{B}_0 = 0$ in $B_0 = (I_r, \tilde{B}_0)$, 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

$$Y_{1t} = C(L)Y_{1t} + u_t = C_1Y_{1t-1} + C_2Y_{1t-2} + \dots + C_pY_{1t-p} + \epsilon_{1t}, \quad (17)$$

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

$$Y_t = \begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = A(L)Y_t + \epsilon_t = A_1Y_{1t-1} + A_2Y_{1t-2} + \dots + A_pY_{1t-p} + \epsilon_t, \quad (18)$$

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

be assumed to be either uncorrelated (exact factor models) or at most admit some limited dependence (approximate factor models) to make sure that the idiosyncratic component is not confounded with the common part. This separation between common and idiosyncratic components also requires conditions on the loadings, ensuring that the factors affect almost all variables (see e.g. Stock and Watson (2002a, 2002b)).

Importantly, 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 a MAI context it is possible to consistently estimate the factors F_t even when N is finite (and without having to impose conditions on the loadings or the error terms). Furthermore, testing specific hypotheses on the factors F_t , such as equality of a factor to a specific economic variable, is much simpler in the MAI context (by imposing restrictions on B_0) than in a factor context (see Bai and Ng (2006b)). Finally, in general, factors estimated by principal components do not admit an exact VAR representation (see Dufour and Stevanovic (2010), while as is clear from equation (5) this is the case within the MAI model.

Overall, with respect to the factor approach, the MAI model 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 (2) as:

$$Y_t = AZ_{t-1} + \epsilon_t, \quad (19)$$

where $Z'_{t-1} = (F'_{t-1}, \dots, F'_{t-p}) = (Y'_{t-1}B'_0, \dots, Y'_{t-p}B'_0) = (Y'_{t-1}, \dots, Y'_{t-p})(I_p \otimes B'_0)$ and is a $1 \times rp$ vector, and where $A = (A_1, \dots, A_p)$ is a $N \times rp$ matrix. As for all j s, $A_j B_0 = A_j Q^{-1} Q B_0$ for any nonsingular matrix Q , we add the identification restriction $B_0 = (I_r, \tilde{B}_0)$. Defining $Y = (Y_1, \dots, Y_T)'$ and $Z = (Z_0, Z_1, \dots, Z_{T-1})'$ and $E = (\epsilon_1, \dots, \epsilon_T)'$, stacking the equations in (19) for $t = 1, \dots, T$ we have

$$Y = ZA' + E, \quad (20)$$

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

3.1 Estimation via Maximum Likelihood

Reinsel (1983) studied estimation of the model in (19) via Maximum Likelihood (ML). 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:

$$-0.5T \log |\Sigma| - 0.5 \sum_{t=1}^T (Y_t - AZ_{t-1})' \Sigma^{-1} (Y_t - AZ_{t-1}). \quad (21)$$

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

$$\hat{\Sigma} = (Y - ZA')(Y - ZA')/T. \quad (22)$$

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

$$AZ_{t-1} = \text{vec}(Z'_{t-1}A') = (I_N \otimes Z'_{t-1})\text{vec}(A') \quad (23)$$

and the corresponding first order conditions are given by:

$$\frac{\partial l}{\partial \text{vec}(A')} = \sum_{t=1}^T (I_N \otimes Z_{t-1})' \Sigma^{-1} (I_N \otimes Z_{t-1}) = 0. \quad (24)$$

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

$$AZ_{t-1} = \sum_{j=1}^p A_j Y_{1,t-j} + \sum_{j=1}^p (A_j \otimes Y'_{2,t-1}) \text{vec}(\tilde{B}_0), \quad (25)$$

where $Y'_{2,t}$ comes from partitioning Y'_t in the first r and last $N - r$ components: $Y'_t = (Y'_{1,t}, Y'_{2,t})$. The corresponding first order conditions are given by:

$$\frac{\partial l}{\partial \text{vec}(\tilde{B}_0)} = \sum_{t=1}^T U_{t-1} A' \Sigma^{-1} \{Y_t - (I_N \otimes Z'_{t-1})\alpha\} = 0, \quad (26)$$

where $U_{t-1} = (I_r \otimes Y_{2,t-1}, \dots, I_r \otimes Y_{2,t-p})$.

Reinsel (1983) suggested to solve in turn equations (22), (24) and (26) until convergence is achieved, and established consistency and asymptotic normality of this estimator for fixed N . Of course these consistency and asymptotic normality results can be coupled with the standard impulse response analysis for finite dimensional VAR models to produce standard errors for such impulse responses (see, e.g., Section 3.7 of Lutkepohl (2007)). Also, specific hypotheses on the parameters, and in particular on B_0 , can be tested using likelihood ratio statistics.

Reinsel's proof of the consistency of the MLE estimator (Reinsel (1983), pp. 148-149) is for a finite number of variables, and we now want to extend it to the case where N possibly diverges. This is undertaken in the Appendix A. In particular, Assumption 3, of Appendix A, for (2) implies Lemma 6 which in turn implies Theorems 2 and 3. In turn, these imply the following Theorem, where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix.

Theorem 1 *Let Assumption 3 hold for (2). Define $\theta_{N0} = (\text{vec}(A)', \text{vec}(B)')'$ to be the true value of the parameters and $\hat{\theta}_n$ its MLE estimator. Then,*

$$\left\| \hat{\theta}_N - \theta_{N0} \right\|_F = O_p \left(\frac{N^{5/2}}{T^{1/2}} \right) \quad (27)$$

Our rate derivations require bounds that may not be as sharp as possible. Therefore, we conjecture that a faster rate may, in fact, hold. However, given our Monte Carlo and empirical results where large values of N seem to lead to a deteriorated performance, it may be the case that this rate is close to the best possible. It is worth noting that similar arguments can be used to prove the properties of MLE estimators of the more general models of Section 5.

3.2 Priors and Estimation via Markov Chain Monte Carlo

In this Subsection we elicit the priors for the parameters of the MAI model in (19), derive the conditional posterior distributions, and provide an MCMC algorithm for Bayesian estimation.

3.2.1 Priors

The model (20) has three sets of parameters, contained respectively in the matrices A' , \tilde{B}_0 , and Σ . We elicit a natural conjugate Normal-Inverse Wishart prior for A' and Σ :

$$A' | \Sigma \sim N(A_0, \Sigma \otimes V_0), \quad \Sigma \sim IW(S_0, v_0). \quad (28)$$

This prior features a Kronecker structure that restricts somehow the way shrinkage can be imposed, but ensures conjugacy and dramatically improves the computational time.

In our empirical application, the prior moments are set as follows. The prior mean of the coefficients is set to $A_0 = 0$ (a $rp \times N$ matrix of zeros). The prior variance V_0 is set to a diagonal $rp \times rp$ matrix:

$$V_0 = \tau \cdot \text{diag}(V_0^1, V_0^2, \dots, V_0^p), \quad (29)$$

where each V_0^k for $k = 1, \dots, p$ is a r -dimensional vector with all the entries equal to $1/k^2$. This choice for the variance matrix shrinks more towards the prior means the coefficients attached to lags which are more far in time (at a quadratic decay), in line with the Minnesota prior. The hyperparameter τ provides the overall shrinkage and it is chosen optimally by maximizing the marginal data density of the model over a grid, more details can be found below. The prior scale matrix S_0 is set to a diagonal matrix with entries given by the sum of squared residuals resulting from least squares estimation of simple AR(1) models for each

of the N variables, based on a pre-sample of 84 observations. The priors degrees of freedom are set to $v_0 = N + 2$ to ensure that the prior on the error variance is as diffuse as possible while remaining proper.⁵

We now consider the elicitation of priors on the matrix \tilde{B}_0 . This matrix contains the weights that each variable has in the composition of each of the factors, for example the element in row j and column i of \tilde{B}_0 measures the weight variable i has in the composition of factor j . To set the prior on \tilde{B}_0 we use an auxiliary model estimated on a pre-sample. Using a pre-sample of 84 observations we compute r factors using principal components.⁶ Then we regress each of the factors for $j = 1, \dots, r$ onto each individual variable y_{it} , $i = 1, \dots, N$ and we use the resulting point estimate and standard deviation of the regression coefficient as prior means and standard deviation for the element in row j and column i of \tilde{B}_0 . Table 6 contains the prior mean and standard deviations resulting from this prior elicitation strategy. These are the values used in our empirical application.

3.2.2 Posteriors and MCMC algorithm

The joint posterior distribution $p(A', \tilde{B}_0, \Sigma | Y)$ has not a known form, but it can be simulated by using a Gibbs sampler drawing in turn from the conditional posterior distributions $p(A', \Sigma | \tilde{B}_0, Y)$ and $p(\tilde{B}_0 | A', \Sigma, Y)$.

Drawing from the conditional posterior $p(A', \Sigma | \tilde{B}_0, Y)$ is straightforward. Given knowledge of \tilde{B}_0 and Y , the variable Z_{t-1} is known, and (19) is a simple multivariate regression model as the one described in Zellner (1973). Then, under the natural conjugate prior described by (28), the conditional posterior distributions are:

$$A' | \Sigma, \tilde{B}_0, Y \sim N(\bar{A}, \Sigma \otimes \bar{V}), \quad \Sigma | \tilde{B}_0, Y \sim IW(\bar{S}, \bar{v}), \quad (30)$$

where:

$$\bar{V} = (V_0^{-1} + Z'Z)^{-1} \quad (31)$$

$$\bar{A} = \bar{V}(V_0^{-1}A_0 + Z'Z\hat{A}) \quad (32)$$

$$\bar{S} = S_0 + Y'Y + A'_0V_0^{-1}A_0 - \bar{A}'\bar{V}^{-1}\bar{A} \quad (33)$$

$$\bar{v} = v_0 + T \quad (34)$$

⁵Given that in the empirical application the ratio of number of observations T to number variables N is quite large (about 460 to 20) it is possible to use such a diffuse prior for the error variance. However for completeness we have also experimented with an informative version of the prior, setting $v_0 = 42$ and $v_0 = 84$. Both of these setups produced similar posterior estimates for the VAR coefficients but a slightly inferior mixing.

⁶The principal components estimates are appropriately rescaled in order to ensure the normalization and identification restrictions $B_0 = (I_r, \tilde{B}_0)$ are satisfied.

and where \hat{A} are the Ordinary Least Squares estimates of the matrix A' . Draws from $p(A', \Sigma | \tilde{B}_0, Y)$ can be easily obtained by generating a sequence of M draws $\{\Sigma_m\}_{m=1}^M$ from $\Sigma | Y \sim IW(\bar{S}, \bar{v})$ and then for each m drawing from $A' | \Sigma, Y \sim N(\bar{A}, \Sigma_m \otimes \bar{V})$, which provides the sequence $\{A'_m, \Sigma_m\}_{m=1}^M$.

Drawing from $p(\tilde{B}_0 | A', \Sigma, Y)$ is less straightforward, as B_0 contains restrictions and enters the model in a nonlinear way. To draw \tilde{B}_0 conditional on A' and Σ we use a Random Walk Metropolis step. To improve the mixing in performing this step we use multiple blocks, and specifically we draw each element in the matrix \tilde{B}_0 separately. Let \tilde{B}_{0ji} denote the element in row j and column i in the matrix \tilde{B}_0 , and let \tilde{B}_{0ji-} denote the set of all the remaining elements of \tilde{B}_0 . At iteration m , a candidate \tilde{B}_{0ji}^* is drawn, conditional on A', Σ , and the remaining elements \tilde{B}_{0ji-} , using a random walk proposal:

$$\tilde{B}_{0ji}^* = \tilde{B}_{0ji}^{m-1} + c\eta_t, \quad (35)$$

where η_t is a standard Gaussian i.i.d. process and c is a scaling factor calibrated in order to have a rejection rate of about 65%-70%.⁷ The candidate draw is then accepted with probability

$$\alpha_k = \min \left\{ 1, \frac{p(\tilde{B}_{0ji}^* | \tilde{B}_{0ji-}, A', \Sigma, Y)}{p(\tilde{B}_{0ji}^{m-1} | \tilde{B}_{0ji-}, A', \Sigma, Y)} \right\}. \quad (36)$$

If the draw is accepted then \tilde{B}_{0ji}^m is set equal to the candidate \tilde{B}_{0ji}^* , otherwise it is set equal to the previous draw \tilde{B}_{0ji}^{m-1} . The procedure is repeated for all the elements of \tilde{B}_0 , i.e. for $j = 1, \dots, r$ and $i = 1, \dots, N$.

Drawing in turn from $p(A', \Sigma | \tilde{B}_0, Y)$ and $p(\tilde{B}_0 | A', \Sigma, Y)$ provides a sequence of M draws $\{A'_j, \Sigma_j, \tilde{B}_0\}_{m=1}^M$ from the joint posterior distribution of A', Σ, \tilde{B}_0 . Each draw can be then inserted into equation (16), 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 r columns and rows to an identity matrix). In Appendix B 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 Appendix B shows that the algorithm has good convergence properties and it is not affected by problems related to the nonlinearity. A more detailed discussion of the role of normalization

⁷To choose the scaling constant c , which is the standard deviation of the proposal density, we use the standard deviation of the prior density for each individual coefficient. As described in Section 3.2.1 these prior densities are obtained using auxiliary univariate AR models on a pre-sample of 84 observations. We then set c to 4 times the prior standard deviation, as this multiple ensures the desired rejection rates for all the coefficients.

in reduced rank models can be found in Hamilton et al. (2007) and Kleibergen and van Dijk (1994, 1998).

4 Determining the Rank

4.1 Classical approach

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, Kapetanios, Smith and Weale (2003) discusses this problem in detail. There are two main approaches. The first uses information criteria. This approach simply estimates (19) 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 is 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 (19). 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.⁸

The first procedure, proposed by Cragg and Donald (1996), is based on the transformation of the matrix A using Gaussian elimination with complete pivoting⁹. Performing r steps of Gaussian elimination with full pivoting on matrix A amounts to the following operations:

$$Q_{r^*} R_{r^*} Q_{r^*-1} R_{r^*-1} \dots Q_1 R_1 A C_1 \dots C_{r^*-1} C_{r^*} = \begin{bmatrix} A_{11}(r^*) & A_{12}(r^*) \\ \mathbf{0} & A_{22}(r^*) \end{bmatrix}$$

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

⁸ A simple alternative to tests of rank maybe the use of a sequence of LR tests for the models with different rank orders. However, tests of rank have well established asymptotic and finite properties in many contexts, as detailed in Camba-Mendez and Kapetanios (2009), whereas the finite sample properties of the sequence of LR tests is not known.

⁹ 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.

to \mathbf{A} to obtain the following relation

$$R_{r^*} R_{r^*-1} \dots R_1 A C_1 \dots C_{r^*-1} C_{r^*} = R A C = F = \begin{bmatrix} F_{11}(r^*) & F_{12}(r^*) \\ F_{21}(r^*) & F_{22}(r^*) \end{bmatrix}$$

where F is partitioned accordingly, i.e. $F_{11}(r^*)$ is of dimension $r^* \times r^*$. Note that in this case $F_{11}(r^*)$ has full rank, under the null hypothesis that $\rho[\mathbf{A}] = r^*$. 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 $\sqrt{T} \text{vec}(\hat{A} - A) \xrightarrow{d} N(0, V)$ where V has full rank, the following result can be shown, under H_0 .

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

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

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

where $\hat{\mathbf{\Gamma}}$ and \hat{V} are the sample estimates of $\mathbf{\Gamma}$ and \mathbf{V} and χ_l^2 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 $\mathbf{\Upsilon} A \mathbf{\Pi} A'$ where $\mathbf{\Upsilon}$ and $\mathbf{\Pi}$ are positive definite matrices, is considered. It follows that $\rho[\mathbf{A}] = \rho[\mathbf{\Upsilon} A \mathbf{\Pi} A'] = r^*$, 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 $\hat{\mathbf{\Upsilon}} \hat{A} \hat{\mathbf{\Pi}} \hat{A}'$ in descending order, $\hat{\mathbf{\Upsilon}}$ and $\hat{\mathbf{\Pi}}$ are estimates of $\mathbf{\Upsilon}$ and $\mathbf{\Pi}$ respectively. Under the null hypothesis, the above statistic converges in distribution to a weighted sum of independent χ_1^2 random variables. The weights are given by the eigenvalues of $(D_{r^*}' \otimes C_{r^*}') V (D_{r^*} \otimes C_{r^*})$, τ_i , $i = 1, \dots, (m-r^*)(n-r^*)$. 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 $\mathbf{\Pi} A' \mathbf{\Upsilon} A$ and $\mathbf{\Upsilon} A \mathbf{\Pi} A'$ 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 Υ and Π 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 Υ and Π 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 approach

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

$$p_r(Y) = \int p(Y|\theta)p(\theta)d\theta, \quad (37)$$

where $\theta = (A, \Sigma, \tilde{B}_0)$ contains all the coefficients of the model. The optimal rank for the system is associated with the model featuring the highest data density:

$$r^* = \arg \max_r p_r(Y). \quad (38)$$

Even though the number of coefficients in the MAI model is large, the density $p_r(Y)$ can be efficiently approximated numerically by using Rao-Blackwellisation and the harmonic mean estimator proposed by Gelfand and Dey (1994). In particular, we have that given M simulated posterior draws $\{\tilde{B}_0\}_{m=1}^M$:

$$\hat{p}_r(Y) = \left[\frac{1}{M} \sum_{m=1}^M \frac{1}{p(Y|\tilde{B}_0^m)p(\tilde{B}_0^m)} f(\tilde{B}_0^m) \right]^{-1}, \quad (39)$$

where $f(\cdot)$ is a truncated multivariate normal distribution calibrated using the moments of the simulated posterior draws (see Geweke's 1999). The term $p(Y|\tilde{B}_0^m)$ is the integrating constant of the conditional posterior distribution $p(A, \Sigma|Y, \tilde{B}_0)$ and is available in closed form, because conditional on \tilde{B}_0^m the model is a multivariate regression with a naturally

conjugate prior:

$$p(Y|\tilde{B}_0^m) = \pi^{\frac{-TN}{2}} \times |(I + Z^m V_0 Z^{.m'})^{-1}|^{\frac{N}{2}} \times |S_0|^{\frac{v_0}{2}} \times \frac{\Gamma_N(\frac{v_0+T}{2})}{\Gamma_N(\frac{v_0}{2})} \times |S_0 + (Y - Z^m A_0')(I + Z^m V_0 Z^{.m'})^{-1}(Y - Z^m A_0')|^{-\frac{v}{2}}. \quad (40)$$

where $\Gamma_N(\cdot)$ is denoting the N -variate gamma function and where the conditioning on \tilde{B}_0^m is implicit in the conditioning on $Z^m = (Z_0^m, Z_1^m, \dots, Z_{T-1}^m)'$ because $Z_{t-1}^{m'} = (Y_{t-1}', \dots, Y_{t-p}') (I_p \otimes \tilde{B}_0^{m'})$. The result (40) for a general multivariate regression dates back to Zellner (1971), and a straightforward derivation based on theorem A.19 in Bauwens, Lubrano and Richard (1999) can be found in Carriero, Kapetanios, and Marcellino (2010).

The marginal data density (39) can also be used to select the optimal lag length p and optimal shrinkage hyperparameter τ .

5 General Reduced Rank VAR and Multivariate Reduced Rank Regression

The model we considered so far is a special case of a more general reduced rank specification, which also nests the reduced rank models of, e.g., Anderson (1951) and Geweke (1996). In this Section we discuss this more general model, as well as the relationship of the MAI model with the model studied by Geweke (1996).

5.1 General reduced rank VAR

Let us again consider the VAR model in (1):

$$Y_t = \Phi(L)Y_t + \epsilon_t. \quad (41)$$

As before, we assume that $\Phi(L)$ can be factorized as $\Phi(L) = A(L)B(L)$, where $A(L) = A_1L + \dots + A_{p_1}L^{p_1}$ and each A_u is of dimension $N \times r$. However, we now assume a more general specification for $B(L)$, namely $B(L) = B_0 + B_1L + \dots + B_{p_2}L^{p_2}$ where each B_v is full rank of dimension $r \times N$. Furthermore we have $p_1 + p_2 = p$, $p_1 \geq 1$, $p_2 \geq 0$. This gives the following more general reduced rank VAR specification:

$$Y_t = A(L)B(L)Y_t + \epsilon_t = \sum_{u=1}^{p_1} \sum_{v=0}^{p_2} A_u B_v Y_{t-u-v} + \epsilon_t. \quad (42)$$

In this more general model, the factors or indexes are the r -dimensional vectors of

variables:

$$F_t = B(L)Y_t = B_0Y_t + B_1Y_{t-1} + \dots + B_{p_2}Y_{t-p_2}. \quad (43)$$

With respect to the MAI model considered so far, there is more flexibility in the specification of the autoregressive matrices, which need not have all rank equal to r . For example, for the case $p_1 = 2$ and $p_2 = 1$ it is

$$A(L)B(L) = (A_1L + A_2L^2)(B_0 + B_1L) = A_1B_0L + (A_1B_1 + A_2B_0)L^2 + A_2B_1L^3,$$

so that $\text{rank}(A_1B_0) \leq r$, $\text{rank}(A_2B_1) \leq r$ but $\text{rank}(A_1B_1 + A_2B_0)$ can be larger than r . There is also more flexibility in the specification of the factors, compare (3) with (43). On the other hand, the factors no longer follow a finite order VAR but rather a VARMA, as it is:

$$F_t = B(L)A(L)F_t + B(L)\epsilon_t. \quad (44)$$

The moving average representation associated with (42) is

$$Y_t = (I - A(L)B(L))^{-1}\epsilon_t, \quad (45)$$

which is the counterpart of (6). The second moving average representation is:

$$Y_t = (A(L)(I - B(L)A(L))^{-1}B(L) + I)\epsilon_t, \quad (46)$$

which is the counterpart of (7) and is obtained by inserting the moving average representation for F_t , which is

$$F_t = (I - B(L)A(L))^{-1}B(L)\epsilon_t, \quad (47)$$

into the equation for Y_t in (42). For the extension of the third moving average representation of the MAI model, we define

$$D(L) = (I - B(L)A(L))^{-1} \quad (48)$$

so that we can compactly rewrite (46) and (47) as $Y_t = (A(L)D(L)B(L) + I)\epsilon_t$ and $F_t = D(L)\epsilon_t$. Then, given the decomposition

$$D(L) = ((D'_0D_0)^{-1}D'_0 + D_1(D'_0D_0)^{-1}D'_0L + D_2(D'_0D_0)^{-1}D'_0L^2 + \dots)D_0 = D^*(L)D_0, \quad (49)$$

we can introduce the $(N - r) \times N$ full row rank matrix $D_{0\perp}$ that is orthogonal to D_0 , i.e. $D_0D'_{0\perp} = 0$, such that the rank of $(D'_0, D'_{0\perp})$ is N . We then insert the following

decomposition:

$$\Sigma D'_0(D_0 \Sigma D'_0)^{-1} D_0 + D'_{0\perp}(D_{0\perp} \Sigma^{-1} D'_{0\perp})^{-1} D_{0\perp} \Sigma^{-1} = I_N, \quad (50)$$

into the Wold representation in (46) to yield:

$$Y_t = (\Sigma D'_0(D_0 \Sigma D'_0)^{-1} + A(L)D^*(L))D_0\epsilon_t + D'_{0\perp}(D_{0\perp} \Sigma^{-1} D'_{0\perp})^{-1} D_{0\perp} \Sigma^{-1} \epsilon_t. \quad (51)$$

Defining $D_0\epsilon_t = u_t$ and $D_{0\perp} \Sigma^{-1} \epsilon_t = \xi_t$, we have:

$$Y_t = (\Sigma D'_0(D_0 \Sigma D'_0)^{-1} + A(L)D^*(L))u_t + D'_{0\perp}(D_{0\perp} \Sigma^{-1} D'_{0\perp})^{-1} \xi_t, \quad (52)$$

where, as in the MAI case, u_t and ξ_t are uncorrelated at all leads and lags.

We should point out that the analytical derivation of $D(L)$ is complex even in the case where $A(L)$ and $B(L)$ are known, which, for structural analysis, gives a substantial computational advantage to the MAI specification.

To conclude, we need to discuss estimation of the general reduced rank model in (42). It is convenient to rewrite it as:

$$Y_t = A_1 Z_{t-1} + A_2 Z_{t-2} + \dots A_{p_1} Z_{t-p_1} + \epsilon_t, \quad (53)$$

where $Z'_{t-i} = (F'_{t-i} + \dots + F'_{t-p_2-i}) = (Y'_{t-i} B'_0 + \dots + Y'_{t-p_2-i} B'_{p_2}) = (Y'_{t-i}, \dots, Y'_{t-p_2-i})B$ is of dimension $1 \times r$, for $i = 1, \dots, p_1$ and where $B = (B_0, \dots, B_{p_2})'$ is of dimension $Np_2 \times r$. We add the identification restrictions $B_j = (I_r, \tilde{B}_j)$, $j = 0, \dots, p_2$. Next, the system (53) can be written more compactly as

$$Y_t = A W_t + \epsilon_t, \quad (54)$$

where $A = (A_1, \dots, A_{p_1})$ is of dimension $N \times p_1 r$ and $W_t = (Z'_{t-1}, \dots, Z'_{t-p_1+1})'$ is of dimension $p_1 r \times 1$. Defining $Y = (Y_1, \dots, Y_T)'$ of dimension $T \times N$, $W = (W_0, W_1, \dots, W_{T-1})'$ of dimension $T \times p_1 r$, and $E = (\epsilon_1, \dots, \epsilon_T)'$, stacking the equations in (54) for $t = 1, \dots, T$ we have

$$Y = W A' + E, \quad (55)$$

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

As equations (54) and (55) are similar to (19) and (20), the same classical and Bayesian estimation procedures described in Section 3 can be used, though the computational complexity increases substantially, providing another reason for the use of the MAI specification. We will see in the next subsection that a substantial simplification occurs for the case $p_1 = 1$, $p_2 = p - 1$.

5.2 Multivariate reduced rank regression and relation with Geweke (1996)

In this section we focus on another special case of the general reduced rank VAR in (42), which is obtained by setting $p_1 = 1$, $p_2 = p - 1$. Defining $X_t = (Y'_{t-1}, \dots, Y'_{t-p})'$, the resulting model can be written as:

$$\underset{N \times 1}{Y_t} = \underset{N \times r}{A_1} [\underset{r \times Np}{B_0, \dots, B_{p-1}}] \underset{Np \times 1}{X_t} + \underset{N \times 1}{\epsilon_t}, \quad (56)$$

which is a multivariate reduced rank regression. This model was studied by Anderson (1951), Velu et al. (1986) in a classical context and Geweke (1996) in a Bayesian context, among others. It is useful to compare (56) with the MAI model:

$$\underset{N \times 1}{Y_t} = [\underset{N \times rp}{A_1, \dots, A_p}] (\underset{rp \times Np}{I_p \otimes B'_0})' \underset{Np \times 1}{X_t} + \underset{N \times 1}{\epsilon_t}. \quad (57)$$

where recognizing that $(I_p \otimes B'_0)' X_t = Z_{t-1}$ leads to expression (19). As is clear from comparison of (56) with (57) the reduced rank VAR in (56) has only one A_1 matrix of dimension $N \times r$ and p matrices B_0, \dots, B_{p-1} each of dimension $r \times N$, while the MAI in (57) has p matrices A_1, \dots, A_p each of dimension $N \times r$ and only one B_0 matrix of dimension $r \times N$.

The main advantage of the specification in (56) is that A_1 has full rank r , therefore it is possible to premultiply the system by the generalized inverse $A^+ = (A'_1 A_1)^{-1} A'_1$, and -conditional on A_1 - to derive a closed form posterior distribution for $[B_0, \dots, B_{p-1}]$, which can then be easily simulated using a Gibbs sampling step (details can be found in Geweke 1996). Instead the matrix $[A_1, \dots, A_p]$ appearing in the MAI in (57) is not full rank, which is the reason why B_0 can only be simulated using a Metropolis step.¹⁰

The main advantage of specification (57) is that, as we have discussed in Section 2, premultiplication of (56) by B_0 provides a VAR specification for the factors $B_0 Y_t$, while in (56) the factors do not admit a finite order VAR representation. Therefore, the MAI model is more suited for structural economic analysis as it implies that all the variables are driven by a limited number of r "factors" and their lags, $B_0 y_{t-1}, \dots, B_0 y_{t-p}$, which can have different effects over time and across variables, and $B_0 y_t$ admit a VAR representation. Instead, in the multivariate reduced rank model the large set of Np factors have a changing composition over time, $B_0 y_{t-1}, B_1 y_{t-2}, \dots, B_{p-1} y_{t-p}$, and require a large VARMA specification.

To summarize, estimation of the multivariate reduced rank model in (57) is easier than estimation of the MAI model in (56), but the MAI model allows to derive a finite order VAR

¹⁰Blocking the system in p different blocks and deriving conditional posteriors is also not feasible in the MAI model. Indeed, while the regressor matrix $I_p \otimes B_0$ has a block-diagonal structure, each of the blocks in this matrix is equal to the same matrix B_0 , and this cross-equations restriction precludes the derivation of conditional posteriors for each of the p blocks.

representations for a set of r factors. For these reasons, specification (57) can be preferable when the interest is in forecasting (see e.g. Carriero, Kapetanios and Marcellino 2011 for an application with a large dataset), while specification (56) is better suited for structural analysis.

6 Monte Carlo Evaluation

In this section we present an extensive Monte Carlo study focusing on the properties of the MAI model, which will be later used also in empirical applications.

We produce artificial data from two alternative Data Generating Processes (DGP). We recall equation (1) and rewrite it as:

$$Y_t = \sum_{u=1}^p \Phi_u Y_{t-u} + \epsilon_t, \quad \epsilon_t \sim i.i.d.N(0, \Sigma). \quad (58)$$

The first DGP (DGP1) is an unrestricted VAR, so it uses (58) without imposing any further restriction. The second DGP (DGP2) is the MAI, so it imposes the rank reduction restriction:

$$\Phi_u = A_u B_0 \quad (59)$$

with $u = 1, \dots, p$. To set up the parameters Σ and Φ_1, \dots, Φ_p we use the estimates obtained from our empirical application (which is extensively discussed in the next Section, along with a description of the data). For DGP1 we estimate (58) using a standard Bayesian approach,¹¹ which provides us with the estimated values $\hat{\Sigma}$ and $\hat{\Phi}_1, \dots, \hat{\Phi}_p$.¹² Similarly, for DGP2 we estimate (58) again but this time under the restriction (59), with rank set to $r = 3$, and using the estimation approach described in Section 3.2, which provides us with the estimated values $\check{\Sigma}$ and $\check{\Phi}_1, \dots, \check{\Phi}_p$.

To simulate artificial data from the two alternative DGPs we set Σ and Φ_1, \dots, Φ_p to $\hat{\Sigma}$ and $\hat{\Phi}_1, \dots, \hat{\Phi}_p$ (under DGP1) or $\check{\Sigma}$ and $\check{\Phi}_1, \dots, \check{\Phi}_p$ (under DGP2), draw 100 different disturbances vectors from $\epsilon_t \sim i.i.d.N(0, \hat{\Sigma})$ and project forward (58), which provides 100 different realizations of the process Y_t under DGP1 and 100 different realizations under DGP2.

Finally, for each of the two DGPs we estimate three alternative models: i) the MAI under the Bayesian approach, described in Section 3.2; ii) the MAI under the classical approach,

¹¹To save space we do not spell out the Bayesian VAR estimation details here. Our implementation follows Kadiyala and Karlsson (1996) and Carriero, Clark and Marcellino (2015) to which the reader is referred for further details.

¹²Note that while it is true that in a large sample the unrestricted BVAR estimates would eventually capture a rank reduction such as the one in (59), the dimension of the system is such that this does not happen with the sample size we are working with. Also, we have carefully checked that the values that we are using in this MC experiment for DGP1 always involve full rank Φ_1, \dots, Φ_p matrices.

described in Section 3.1; iii) an unrestricted BVAR, estimated as in the baseline specification of Carriero, Clark, and Marcellino (2015).

To ascertain the properties of the different models under the different DGPs we focus on the Root Mean Squared Error (*RMSE*) and Mean Absolute Error (*MAE*) arising from estimation of the conditional mean parameters. As the number of coefficients is very large rather than looking at individual *RMSEs* and *MAEs* for each of the coefficients appearing in Φ_1, \dots, Φ_p we focus on the average *RMSE* and *MAEs* over all the estimated coefficients. Moreover, to facilitate comparisons, for the MAI models we provide results in relative terms with respect to those obtained by the standard unrestricted *BVAR*. We evaluate the performance along various dimensions, considering different values for the total number of variables N , the number of observations T , and the system rank r .

We start with the results obtained under DGP2, i.e. the data generating process which does feature rank reduction. Results are displayed in Table 1. The Table is divided into two panels. Panel A displays results for different combinations of sample size and cross sectional size, in particular $N = 5, 10, 15, 20$ and $T = 300, 460, 720$.¹³ Panel B displays results for fixed N and T (20 and 460 respectively, which are the dimensions of our empirical application) and for different values of the rank of the DGP, $r = 1, 2, 3, 4, 5$ in (59).¹⁴ The entries of the table show the *RMSE* and *MAE* of the MAI model estimated with the Bayesian and classical approaches, relative to the *RMSE* and *MAE* obtained with the unrestricted *BVAR*, therefore a figure below one in the entries of the table signals that the *MAI* model is performing better than the *BVAR* benchmark.

The Bayesian MAI performance is systematically better than the classical MAI performance, with gains decreasing with the sample size T but remaining very large (over 100%) even with $T=720$. The Bayesian MAI performs also much better than BVAR, in particular when $N=15, 20$, while the classical MAI is never better than the BVAR, which suggests that this model is still too overparametrized to be effectively handled by maximum likelihood estimation.

The considerations above are still valid when looking at results for different system ranks, with the Bayesian MAI outperforming consistently the BVAR. Interestingly, the classical MAI performs better than the BVAR only for $r=1$, which again points towards the idea that

¹³It should be noted that while comparing results for increasing T and fixed N involves looking at the same DGP estimated with an increasing number of observations, comparing results for increasing N and fixed T is not as straightforward. Indeed, since the DGPs are calibrated using estimates obtained in a preliminary step based on actual data, qualitative differences in the data used for the DGP calibration enter the picture and should be kept in mind. Another potential difference lies in the fact that the overall shrinkage for the Bayesian approaches is kept fixed as N increase, while in theory the shrinkage parameter should be chosen optimally for each cross-sectional size and typically should decrease as the number of variables decreases.

¹⁴Recall that the rank restriction is imposed in the preliminary estimation step that provides us with the values of the DGP parameters. To obtain different ranks in the DGP it is sufficient to impose restriction (59) with the desired alternative values for the rank.

for maximum likelihood estimation to work well one needs a rather small system, while for larger N and r the use of the Bayesian approach is preferable. It is worth noting that the best RMSE is achieved by a rank of 2, while the model with rank equal to 3 (which is the true rank in the DGP) performs slightly worse.

We now turn to the results obtained under DGP1, i.e. the data generating process which does not feature rank reduction. Results for this case are displayed in Table 2. Looking at the results in panel A, the Bayesian MAI remains systematically better than classical MAI, with gains decreasing with the sample size T but remaining very large even with $T = 720$. As expected, the Bayesian MAI is imposing a restriction which is not holding true in the data so it underperforms the BVAR under this data generating process. However it is interesting to note that the cost of the rank reduction is decreasing as the dimension of the system increases, being about 50% worse than the BVAR for $N = 5$, but only less than 20% worse when $N = 20$, a result driven by the fact that the bias gets compensated by substantial improvements in efficiency as N increases.

Looking at results for different ranks, the Classical MAI remains systematically worse than the BVAR and the classical MAI, and losses increase with r . With $r = 1$ and $N = 20$, the Bayesian MAI performs even slightly better than the BVAR in terms of *RMSE* (but not for *MAE*). With higher r both *RMSE* and *MAE* increase, which is related to the increased complexity of the model.

Overall, the Monte Carlo experiments suggest that Bayesian estimation of the MAI model is systematically better than classical estimation. The ranking of the MAI and full rank BVAR models is instead not clear-cut. However, even with a full rank BVAR DGP, the MAI does reasonably well, in particular when N is large and/or r is small.

7 Empirical Applications

In this section we illustrate how the MAI model can be used for structural analysis. We begin with describing the data and then we move on to select the optimal rank, lag length and shrinkage of the model. Finally, we use the selected optimal model to study two alternative examples of structural shocks.

7.1 Data and selection of optimal model

We use the "medium" dataset of Banbura, Giannone, Reichlin (2010, BGR), which includes the 20 variables described in Table 3. The sample is at monthly frequency and we have extended it to cover the period January 1974 to December 2013.

Since the Monte Carlo experiments have shown that the Bayesian approach produces much more reliable estimates than the classical approach, we focus the discussion on the

former, but we will provide some results for the classical MAI when we analyze the effects of monetary policy shocks.

Before using the model for structural analysis, we proceed to select some of its key features such as lag length, rank, and shrinkage hyperparameter. We consider lag lengths of 1 to 13 lags, a possible rank of the system ranging from 1 to 5, and different values of the shrinkage hyperparameter τ in the grid $\sqrt{\tau} \in \{0.01, 0.02, 0.03, 0.04, 0.05, 0.075, 0.1\}$. This provides a total of 455 alternative specifications. We estimate all these specifications and rank them according to the marginal data density computed as shown in equation (39).

In Table 4 we provide results for the best 20 specifications. The first three columns contain the rank-lags-shrinkage combination that uniquely identifies a specification. Columns 4 and 5 contain the MDD of the MAI and the BVAR (note that the BVAR MDD can be obtained in closed form using a formula similar to (40) and is of course insensitive to the rank). For reference, columns 6 and 7 contain the Potential Scale Reduction Factors for the MAI model.¹⁵ Columns 8 and 9 contain the Bayesian Information Criterion computed for the MAI and the BVAR. As is clear from the table, the best specification selects 13 lags, a rank of 3, and a shrinkage parameter $\sqrt{\tau} = 0.02$. With this combination of rank and lag length the MAI features a MDD of -9444. However it is important to note that instead the best BVAR model is obtained by setting $p = 13$ and with an overall shrinkage of 0.1 (a combination not shown in the table) which produces a MDD of -8956, a BIC of 81.35 and a trace adjusted R-squared of -28.6.¹⁶ As is apparent from the Table, specifications with $r = 1$ or $r = 2$ are also providing good results in terms of MDD and especially in terms of BIC. However, as the MDD criterion is more in line with the Bayesian approach, we rely on it and select a rank of 3 for our application. Moreover, while the BIC tends to favour more parsimonious specifications, lag exclusion Wald tests performed with a rank of 3 rejected the null that all the coefficients attached to the third factors were equal to 0.

Having chosen the system rank to be $r = 3$, we further restrict the B_0 matrix in order to identify some economically relevant 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 5. The resulting factors and their components are plotted in Figure 1. Once this set of restrictions is imposed, we compute again the marginal data density and find that its value increases from -9444 to -9380, providing support for the restrictions.¹⁷ Table 6 shows the prior and posterior mean and standard deviation of the elements of the matrix B_0 under this optimal specification, which is the one we use for the structural analysis

¹⁵The PSRFs provide an easy diagnostic tool for the convergence of the algorithm. Values below 1.1 are considered an indicator of good mixing and convergence properties of the algorithm.

¹⁶When comparing the MAI impulse responses with the BVAR impulse responses we use this optimal specification for the BVAR.

¹⁷We also re-compute the optimal shrinkage and lag-length under this restricted specification and the resulting optimal values are $\sqrt{\tau} = 0.1$ and $p = 13$.

discussed in the next subsections

7.2 Structural analysis

To illustrate how to conduct empirically structural analysis using the MAI, we first replicate in the MAI 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 MAI. 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 MAI.

7.2.1 Monetary policy shock

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.¹⁸ Formally, the impulse responses are based on the representation:

$$Y_t = \{A(L)[I - B_0A(L)]^{-1}B_0 + I\}\Lambda^{-1}\epsilon_t^* \quad (60)$$

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:

$$\Psi_s = A_1B_0\Psi_{s-1} + \dots + A_{\min(s,p)}B_0\Psi_{s-\min(s,p)}; \quad s > 0 \quad (61)$$

with

$$\Psi_0 = \{A(0)[I - B_0A(0)]^{-1}B_0 + I\}\Lambda^{-1} = \Lambda^{-1}. \quad (62)$$

We simulate the distribution of the impulse responses using 40000 draws¹⁹ and plot the median responses together with the 16th and 84th quantiles in Figure 2 and Figure 3. In Figure 2 the Bayesian MAI impulse responses are overlayed with those obtained with a classical estimation of the MAI, while in Figure 3 they are overlayed with the responses obtained using the unrestricted BVAR approach of Banbura, Giannone, and Reichlin (2010).

As is clear from the figures, the impulse responses of the Bayesian MAI model are reasonable from an economic point of view. Following an increase in the federal funds rate industrial production, capacity utilization, employment, consumption and housing starts decline, while unemployment increases. There is a negative reaction also in CPI, PPI, PCE deflator, and earnings. Money and reserves decrease, while the exchange rate appreciates

¹⁸Other approaches are of course possible, see e.g. Lanne and Lutkepohl (2008).

¹⁹The 40000 draws are obtained by running 2 parallel chains of 25000 draws. For each chain we produce 25000 draws and discard the first 5000 for burn-in.

and the reaction of the stock market is close to zero.

Comparing these responses to those obtained with maximum likelihood estimation displayed in Figure 2 it emerges that while the classical and Bayesian responses are overall similar at short horizons, in the long run they diverge because some classical responses tend to explode (black lines in Figure 2), which we attribute to the problem of overparameterization in the classical set-up that makes the Bayesian approach more suited (and also in line with the Monte Carlo results).

Comparing these responses to those obtained with an unrestricted BVAR displayed in Figure 3 it emerges that the MAI model produces more reasonable responses for the real variables. Indeed the BVAR specification of Banbura, Giannone and Reichlin (2010) (blue lines in Figure 3) implies a puzzling reaction for the real variables in the first 6 to 12 months, with variables such as employment, industrial production, capacity utilization and housing starts initially increasing and unemployment initially decreasing after a contractionary shock, which is at odd with economic intuition.

7.2.2 Demand, supply and financial shocks

In this subsection we analyze the effects of demand, supply and financial shocks. More precisely, recall that we identified an output factor, a price factor, and a financial / monetary factor by imposing restrictions on the matrix B_0 , as detailed in Table 5 and Figure 1.

The s -period ahead responses on the VAR equations are based on the representation (16) and are:

$$\Psi_s = A_1\Pi_{s-1} + \dots + A_{\min(s,p)}\Pi_{s-\min(s,p)}; \quad s > 0 \quad (63)$$

with

$$\Psi_0 = \{\Sigma B_0'\Omega^{-1} + A(0)[I - B_0A(0)]^{-1}\}P^{-1} = \Sigma B_0'P' \quad (64)$$

where the second equality follows from $\Omega^{-1} = P'P$.

We simulate the distribution of the impulse responses using 40000 draws and plot the median responses together with the 16th and 84th quantiles. Specifically, Figures 4, 5, and 6 show the responses of the 20 macroeconomic variables to a demand, supply, and financial shock, respectively.

The effects of a (positive) demand shock are illustrated in Figure 4. This shock is modelled as a shock to the first factor. 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 a (negative) supply shock are presented in Figure 5. This shock is modelled as a shock to the second factor. 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.

The effects of a (negative) financial shock are presented in Figure 6. This shock is modelled as a shock to the third factor. The shock is similar to the monetary policy shock analyzed in the previous subsection but now not only the short term but all the interest rates increase on impact, and in addition there is a decrease in the amount of money, reserves and the stock market index. Hence, qualitatively the responses of the real and price variables are similar to those reported in Figure 2 and Figure 3 but with a clearer negative effect already in the first periods after the financial shock.

Overall, these empirical applications illustrate how the MAI can be easily used to conduct structural analysis, along the lines of either the structural VAR and BVAR approaches or the FAVAR methodology. The two possibilities lead to similar results in the case of a monetary/financial shock, with even more sensible responses from an economic point of view.

8 Conclusions

In this paper we address the issue of parameter dimensionality reduction in Vector Autoregressive models (VARs) for many variables by using the Multivariate Autoregressive Index (MAI) model of Reinsel (1986), which imposes reduced rank restrictions on the coefficient matrices.

As we are particularly interested in the use of MAI models for structural analysis, we derive alternative Wold representations for them. We focus on a representation that highlights the similarities of MAI and dynamic factor models, a competing approach to model large datasets, but also discuss the differences in the two methods.

Next, we review classical estimation of the MAI model, and we extend the asymptotic results to the case of N diverging. Moreover, we provide the conditional posteriors and an MCMC algorithm for Bayesian estimation of the model. We then extend the representation and estimation results to general reduced rank VARs.

We assess the finite sample performance of the MAI estimation methods in Monte Carlo experiments. The results show that Bayesian estimation of the MAI model performs much better than classical maximum likelihood, due to the overparameterization when N is large. The Bayesian MAI also provides relevant gains against an unrestricted Bayesian VAR when

the true data generating process features less than full rank in the conditional mean matrices.

Finally, structural analysis with the MAI is 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. The results are quite sensible from an economic point of view, often more than those from unrestricted BVARs.

Overall, the method is general, simple, and well performing. It could be also extended in several directions, for example to allow for non-normal errors or Markov Switching changes in the parameters as e.g. in, respectively, Lanne and Lutkepohl (2010) and Lanne, Lutkepohl and Maciejowska (2010). Hence, the MAI model is promising as an alternative tool for structural analysis using information in large datasets.

Appendix A: properties of MLE estimation

In this appendix we set out a framework for analysing Maximum Likelihood estimation in the presence of a large dataset modelled through the use of a parametric model. We first provide a general analysis of consistency and rates of convergence for the estimator and we then proceed to prove Theorem 1 by verifying the conditions needed for the general result.

Consider a random matrix of dimension $T \times N$, $Y = (Y_1, \dots, Y_T)'$ with density $F(Y, \theta_{N0})$ depending on a parameter $\theta_{N0} \in \Theta_N \subseteq \mathbb{R}^{kN}$ for some sequence of finite constants k_N . We assume that N is a function of T . Y is an array. Let θ_N be an arbitrary element of Θ_N , and let $L(\theta_N) = F(Y, \theta_N) = \prod_{i=p}^T F_i(Y_i | Y_{1:i-1}, \theta_{N0})$ denote the assumed likelihood function of Y , for some $p > 1$, where $Y_{1:i-1} = (Y_1, \dots, Y_{i-1})'$. Then, $l(\theta_N) = \log L(\theta_N) = \sum_{i=p}^T f_i(Y_i | Y_{1:i-1}, \theta_{N0})$ denotes the log-likelihood function where $f_i(Y_i | Y_{1:i-1}, \theta_{N0}) = \log F_i(Y_i | Y_{1:i-1}, \theta_{N0})$. We also define the $kN \times 1$ score vector by

$$s(Y, \theta_N) = s(\theta_N) = \partial l / \partial \theta_N,$$

and the Hessian

$$H(Y, \theta_N) = H(\theta_N) := \frac{1}{T} \sum_{i=1}^T H_i(\theta_N) := \frac{1}{T} \sum_{i=1}^T \frac{\partial^2 f_i(Y_i, \theta_N)}{\partial \theta_N \partial \theta_N'}.$$

When $l(\theta_N)$ is differentiable, the MLE $\hat{\theta}_N$ satisfies

$$s(Y, \hat{\theta}_N) = 0. \tag{65}$$

In the general analysis that follows, we assume the following set of regularity conditions:

Assumption 1 (RC1) *The support of F , $S = \{Y \in \mathbb{R}^{T \times N} : F(Y, \theta_N) > 0\}$, is independent*

of θ_N ; (RC2) $F(Y, \theta_N)$ is twice continuously differentiable with respect to θ_N ; (RC3) The matrix $J(\theta) := \mathbb{E} \left(\frac{\partial^2 \log f(Y, \theta_N)}{\partial \theta_N \partial \theta_N'} \right)$ has finite elements and is negative definite at $\theta_N = \theta_{N0}$.

We first consider consistency of MLE estimation. We make the following assumption:

Assumption 2 (C1) $F(Y, \theta)$ is continuous w.r.t. θ ; (C2) Θ_N is a compact subset of $\mathbb{R}^{k_N N}$ for all N ; (C3) θ_{N0} lies on the interior of Θ_N and is the unique maximiser of $\mathbb{E}l(\theta_N)$ over Θ_N . In other words,

$$\kappa(\theta_N) := \mathbb{E}l(\theta_{N0}) - \mathbb{E}l(\theta_N) > 0 \quad (66)$$

for all $\theta_N \in \Theta \setminus \{\theta_{N0}\}$; (C4) $l(\theta_N)$ satisfies a uniform law of large numbers over Θ_N :

$$\max_{\theta_N \in \Theta_N} \left| \frac{1}{T} \sum_{i=p}^T f_i(Y_i | Y_{1:i-1}, \theta_{N0}) - \mathbb{E} f_i(Y_i | Y_{1:i-1}, \theta_{N0}) \right| \rightarrow_p 0 \text{ as } N \rightarrow \infty. \quad (67)$$

Theorem 2 (Consistency) Under (C1)-(C4),

$$\left\| \hat{\theta}_N - \theta_{N0} \right\| \rightarrow_p 0 \text{ as } T \rightarrow \infty.$$

Proof. For arbitrary $\delta > 0$, consider an open neighbourhood around θ_{N0} of radius δ :

$$N(\delta) := \{\theta_N \in \Theta_N : \|\theta_N - \theta_{N0}\| < \delta\}.$$

Both $N(\delta)$ and its complement

$$\overline{N(\delta)} = \{\theta_N \in \Theta_N : \|\theta_N - \theta_{N0}\| \geq \delta\}$$

are subsets of Θ_N which, in turn, is a subset of $\mathbb{R}^{k_N N}$. It is easy to see that $N(\delta)$ is an open set, so $\overline{N(\delta)}$ is a closed set. $\overline{N(\delta)}$ is also bounded, since it is a subset of the bounded set Θ_N (see (C2)). We conclude that $\overline{N(\delta)}$ is a closed and bounded subset of $\mathbb{R}^{k_N N}$, so $\overline{N(\delta)}$ is compact by the Heine-Borel theorem. Since, by (C1), $\mathbb{E}l(Y, \theta_N)$ is a continuous function w.r.t. θ_N , the maximiser of $\mathbb{E}l(Y, \theta_N)$ over $\overline{N(\delta)}$ belongs to $\overline{N(\delta)}$: denoting this maximiser by $\theta_{\delta N}$, we conclude that there exists $\theta_{\delta N} \in \overline{N(\delta)}$ satisfying

$$\mathbb{E}l(Y, \theta_{\delta N}) \geq \mathbb{E}l(Y, \theta_N) \quad \text{for all } \theta_N \in \overline{N(\delta)}. \quad (68)$$

We next note that (67) is equivalent to the event

$$A(\varepsilon) := \left\{ \omega : \max_{\theta_N \in \Theta_N} \left| \left(\frac{1}{T} l(Y(\omega), \theta_N) - \mathbb{E}[l(Y(\omega), \theta_N)] \right) \right| < \frac{\varepsilon}{2} \right\}$$

occurring with probability tending to 1 for arbitrary $\varepsilon > 0$, i.e., $\lim_{T \rightarrow \infty} P(A(\varepsilon)) = 1$. If

we can show the inequality

$$P(A(\varepsilon)) \leq P\left(\left\{\omega : \hat{\theta}_N(\omega) \in N(\delta)\right\}\right), \quad (69)$$

for *arbitrary* $\delta > 0$ and *some* $\varepsilon > 0$, consistency of $\hat{\theta}_N$ will follow immediately since

$$P\left(\left\{\omega : \hat{\theta}_N(\omega) \in N(\delta)\right\}\right) = P\left(\left\{\omega : \left\|\hat{\theta}_N(\omega) - \theta_{N0}\right\| < \delta\right\}\right),$$

and the right hand side tends to 1 as $T \rightarrow \infty$ as $P(A(\varepsilon)) \rightarrow 1$ for arbitrary $\varepsilon > 0$. It remains to show (69). Using the identity $|x| < r \Leftrightarrow -r < x < r$, we obtain for all $\theta_N \in \Theta_N$:

$$\omega \in A(\varepsilon) \Rightarrow \begin{cases} \mathbb{E}\left[\frac{1}{T}l(Y(\omega), \theta_N)\right] - \frac{\varepsilon}{2} < \frac{1}{T}l(Y(\omega), \theta_N) \\ \frac{1}{T}l(Y(\omega), \theta_N) < \mathbb{E}\left[\frac{1}{T}l(Y(\omega), \theta_N)\right] + \frac{\varepsilon}{2}. \end{cases} \quad (70)$$

Since $\theta_{N0} \in \Theta_N$ and $\hat{\theta}_N \in \Theta_N$ (by compactness of Θ_N and continuity of the log-likelihood) (70) will apply for $\theta_N = \theta_{N0}$ for the top inequality and $\theta_N = \hat{\theta}_N$ for the bottom inequality. Since $l(Y(\omega), \theta_N) \leq l(Y(\omega), \hat{\theta}_N)$ by definition of the MLE, (70) implies the chain of inequalities

$$\begin{aligned} & \mathbb{E}\left[\frac{1}{T}l(Y(\omega), \theta_{N0})\right] - \frac{\varepsilon}{2} < \frac{1}{T}l(Y(\omega), \theta_{N0}) \\ & \leq \frac{1}{T}l(Y(\omega), \hat{\theta}_N) < \mathbb{E}\left[\frac{1}{T}l(Y(\omega), \hat{\theta}_N)\right] + \frac{\varepsilon}{2}. \end{aligned}$$

We conclude that

$$\omega \in A(\varepsilon) \Rightarrow \mathbb{E}\left[\frac{1}{T}l(Y(\omega), \theta_{N0})\right] < \mathbb{E}\left[\frac{1}{T}l(Y(\omega), \hat{\theta}_N)\right] + \varepsilon \quad (71)$$

for arbitrary $\varepsilon > 0$. Since (71) holds for arbitrary $\varepsilon > 0$, we may choose

$$\varepsilon = \kappa(\theta_{\delta N}) = \mathbb{E}\left[\frac{1}{T}l(Y(\omega), \theta_{N0})\right] - \mathbb{E}\left[\frac{1}{T}l(Y(\omega), \theta_{\delta N})\right]$$

in the notation of (66). This choice is possible since $\theta_{\delta N} \in \overline{N(\delta)} \subseteq \Theta \setminus \{\theta_0\}$ so positivity of $\kappa(\theta_{\delta N})$ is guaranteed by (C2). Imposing the choice of $\varepsilon = \kappa(\theta_{\delta N})$ in (71), we obtain

$$\begin{aligned} \omega \in A(\kappa(\theta_{\delta N})) & \Rightarrow \mathbb{E}\left[\frac{1}{T}l(Y(\omega), \theta_{\delta N})\right] < \mathbb{E}\left[\frac{1}{T}l(Y(\omega), \hat{\theta}_N)\right] \\ & \Rightarrow \hat{\theta}_N \notin \Theta_N \cap \overline{N(\delta)} \quad \text{by (68)}. \end{aligned}$$

Therefore $\omega \in A(\kappa(\theta_{\delta N})) \Rightarrow \hat{\theta}_N(\omega) \in N(\delta)$, i.e. $P[A(\kappa(\theta_{\delta N}))] \leq P[\omega : \hat{\theta}_N(\omega) \in N(\delta)]$, establishing (69). \square ■

We next consider the rates of convergence for the MLE estimator. We define the Frobenius norm of a matrix A as $\|A\|_F$. Then we have the following Theorem.

Theorem 3 (Rates) *Under (C1)-(C4), if (T1) $\mathbb{E}[H_N(\theta_{N0})]$ is invertible, $\hat{\theta}_N$ is consistent and*

$$(T2) \sup_i \text{Var} \left(\sum_{t=1}^T y_{i,t} - \mathbb{E}(y_{i,t}) \right) = O(T^{-1}),$$

then

$$\|\hat{\theta}_N - \theta_{N0}\|_F = O_p \left(\frac{N^{5/2}}{T^{1/2}} \right) + O_p \left(\frac{N^{3/2}}{T^{1/2}} \right)$$

Proof.

$$\begin{aligned} \|\hat{\theta}_N - \theta_{N0}\|_F &= \left\| \left[\frac{1}{T} \sum_{i=1}^T H_i(\theta_N^*) \right]^{-1} \frac{1}{T} \sum_{i=1}^T z_i(\theta_{N0}) \right\|_F = \\ &\left\| \left[\frac{1}{T} \sum_{i=1}^T H_i(\theta_N^*) \right]^{-1} \frac{1}{T} \sum_{i=1}^T z_i(\theta_0) - \left(\frac{1}{T} \sum_{i=1}^T \mathbb{E}[H_i(\theta_0)] \right)^{-1} \frac{1}{T} \sum_{i=1}^T z_i(\theta_0) + \right. \\ &\quad \left. \left(\frac{1}{T} \sum_{i=1}^T \mathbb{E}[H_i(\theta_0)] \right)^{-1} \frac{1}{T} \sum_{i=1}^T z_i(\theta_0) \right\|_F \leq \end{aligned}$$

$$\begin{aligned} &\left\| \left[\frac{1}{T} \sum_{i=1}^T H_i(\theta_N^*) \right]^{-1} \frac{1}{T} \sum_{i=1}^T z_i(\theta_0) - \left(\frac{1}{T} \sum_{i=1}^T \mathbb{E}[H_i(\theta_0)] \right)^{-1} \frac{1}{T} \sum_{i=1}^T z_i(\theta_0) \right\|_F + \\ &\left\| \left(\frac{1}{T} \sum_{i=1}^T \mathbb{E}[H_i(\theta_0)] \right)^{-1} \frac{1}{T} \sum_{i=1}^T z_i(\theta_0) \right\|_F \end{aligned}$$

We have

$$\left\| \left[\frac{1}{T} \sum_{i=1}^T H_i(\theta_N^*) \right]^{-1} \frac{1}{T} \sum_{i=1}^T z_i(\theta_0) - \left(\frac{1}{T} \sum_{i=1}^T \mathbb{E}[H_i(\theta_0)] \right)^{-1} \frac{1}{T} \sum_{i=1}^T z_i(\theta_0) \right\|_F =$$

$$\begin{aligned} &\left\| \left(\left[\frac{1}{T} \sum_{i=1}^T H_i(\theta_N^*) \right]^{-1} - \left(\frac{1}{T} \sum_{i=1}^T \mathbb{E}[H_i(\theta_0)] \right)^{-1} \right) \frac{1}{T} \sum_{i=1}^T z_i(\theta_0) \right\|_F \leq \\ &\left\| \left[\frac{1}{T} \sum_{i=1}^T H_i(\theta_N^*) \right]^{-1} - \left(\frac{1}{T} \sum_{i=1}^T \mathbb{E}[H_i(\theta_0)] \right)^{-1} \right\|_F \left\| \frac{1}{T} \sum_{i=1}^T z_i(\theta_0) \right\|_F \leq \end{aligned}$$

$$\frac{\left\| \left(\frac{1}{T} \sum_{i=1}^T \mathbb{E}[H_i(\theta_0)] \right)^{-1} \right\|_F^2 \left\| \frac{1}{T} \sum_{i=1}^T H_i(\theta_N^*) - \frac{1}{T} \sum_{i=1}^T \mathbb{E}[H_i(\theta_0)] \right\|_F \left\| \frac{1}{T} \sum_{i=1}^T z_i(\theta_0) \right\|_F}{1 - \left\| \left(\frac{1}{T} \sum_{i=1}^T \mathbb{E}[H_i(\theta_0)] \right)^{-1} \right\|_F \left\| -\frac{1}{T} \sum_{i=1}^T H_i(\theta_N^*) - \frac{1}{T} \sum_{i=1}^T \mathbb{E}[H_i(\theta_0)] \right\|_F}$$

We examine $\left\| \left(\frac{1}{T} \sum_{i=1}^T \mathbb{E}[H_i(\theta_0)] \right)^{-1} \right\|_F$, $\left\| \frac{1}{T} \sum_{i=1}^T H_i(\theta_N^*) - \frac{1}{T} \sum_{i=1}^T \mathbb{E}[H_i(\theta_0)] \right\|_F$ and $\left\| \frac{1}{T} \sum_{i=1}^T z_i(\theta_0) \right\|_F$. We have that

$$\left\| \mathbb{E}[H(\theta_{N0})]^{-1} \right\|_F = \sqrt{\sum_{i=1}^N \lambda_{\mathbb{E}[H(\theta_{N0})],i}^{-1}} = O(\sqrt{N}) \quad (72)$$

where $\lambda_{A,i}$ denotes the i -th eigenvalue of A , in order of magnitude in absolute value, as long as all eigenvalues of $\mathbb{E}[H_N(\theta_{N0})]$ are bounded away from zero, as assumed. By consistency of $\hat{\theta}_N$, and twice differentiability of f it follows that every element of $H(\theta_N^*)$ converges to the respective element of $\mathbb{E}[H(\theta_{N0})]$ uniformly over all elements. Therefore,

$$\left\| \frac{1}{T} \sum_{i=1}^T H_i(\theta_N^*) - \mathbb{E}[H(\theta_{N0})] \right\|_F = O_p(N) \quad (73)$$

Further, by (T2)

$$\left\| \frac{1}{N} \sum_{i=1}^N z_i(\theta_{N0}) \right\|_F = O_p\left(\frac{N^{1/2}}{T^{1/2}}\right)$$

Overall noting that, by (73) and (T1), $\frac{1}{1 - \left\| \mathbb{E}[H(\theta_{N0})]^{-1} \right\|_F \left\| \frac{1}{T} \sum_{i=1}^T H_i(\theta_N^*) - \mathbb{E}[H_N(\theta_0)] \right\|_F}$ is bounded in probability, we have

$$\begin{aligned} \left\| \hat{\theta}_N - \theta_{N0} \right\| &\leq \frac{\left\| \mathbb{E}[H_N(\theta_0)]^{-1} \right\|^2 \left\| -\frac{1}{T} \sum_{i=1}^T H_i(\theta_N^*) - \mathbb{E}[H_N(\theta_0)] \right\|}{1 - \left\| \mathbb{E}[H_N(\theta_0)]^{-1} \right\| \left\| -\frac{1}{T} \sum_{i=1}^T H_i(\theta_N^*) - \mathbb{E}[H_N(\theta_0)] \right\|} \\ &\left\| \frac{1}{T} \sum_{i=1}^T z_i(\theta_0) \right\| + \left\| \mathbb{E}[H(\theta_0)]^{-1} \frac{1}{N} \sum_{i=1}^N z_i(\theta_0) \right\|_F = O_p\left(\frac{N^{5/2}}{T^{1/2}}\right) + O_p\left(\frac{N^{3/2}}{T^{1/2}}\right) \end{aligned}$$

□ ■

Remark 4 *Given the above general results, we need to prove the conditions needed for Theorems 2 and 3 to hold for the MAI model (19). For ease of reference, we recall the general VAR model*

$$Y_t = \Phi(L)Y_t + \epsilon_t,$$

and its MAI specialisation

$$Y_t = A(L)B_0Y_t + \epsilon_t = \sum_{u=1}^p A_u B_0 Y_{t-u} + \epsilon_t, \quad (74)$$

We will provide results for (74) while using the relation $\Phi(L) = A(L)B_0$ and noting that straightforward extensions would make (42) also amenable to our analysis. The general notation introduced in this Appendix for the MLE estimation will apply for the parametric reduced rank model with the obvious adjustments. In particular, we consider the concentrated likelihood as presented in page 147 of Reinsel (1983) and therefore $\theta_N = (\text{vec}(A)', \text{vec}(B)')'$ where $A = (A_1, \dots, A_p)$.

We make the following assumption:

Assumption 3 (i) All roots of $\Phi(L)$ are bounded away from the unit circle uniformly over N . (ii) ϵ_t is an iid sequence, which has a continuous, twice differentiable and bounded probability density function and finite $2 + \zeta$ moments for some $\zeta > 0$ and (iii) $\mathbb{E}[H_n(\theta_{n0})]$ is invertible.

Remark 5 Assumption 3 ensures that Assumption 1 and Assumption 2 (i)-(iii) hold. Therefore we only need to prove Assumption 2 (iv) and Condition (T2) of Theorem 3. This result is provided by Lemma 6.

Lemma 6 Under Assumption 3 and if $N = o(T^{1/2})$ the following hold

$$\sup_i \text{Var} \left(\sum_{t=1}^T y_{i,t} - \mathbb{E}(y_{i,t}) \right) = O(T^{-1}). \quad (75)$$

$$\max_{\theta_N \in \Theta_N} \left| \frac{1}{T} \sum_{i=p}^T f_i(Y_i | Y_{1:i-1}, \theta_{N0}) - \mathbb{E} f_i(Y_i | Y_{1:i-1}, \theta_{N0}) \right| \rightarrow_p 0 \text{ as } N \rightarrow \infty, \quad (76)$$

Proof. To show (75) we use Theorem 18.5.3 of Ibragimov and Linnik (1971). Then, it is sufficient to prove that

$$\sup_i \mathbb{E} y_{i,t}^{2+\zeta} < \infty, \text{ for some } \zeta > 0, \quad (77)$$

and

$$\sup_i \sum_{m=1}^{\infty} \alpha_{i,m}^{\zeta/(2+\zeta)} < \infty, \quad (78)$$

where $\alpha_{i,m}$ are the strong mixing coefficients of $y_{i,t}$. By assumption, the eigenvalues of the companion form matrix obtained from $\Phi(L)$ are bounded away from 1 in absolute value. This implies that $\alpha_{i,m} = \xi_i^m$, $\xi_i > 0$, where $\sup_i \xi_i < 1$, which implies (78) for all $\zeta > 0$. Further, the above eigenvalue assumption implies that $\sup_i \mathbb{E} y_{i,t}^2 < \infty$, which implies that $\sup_i \sum_{m=1}^{\infty} \mu_{i,m}^2 < \infty$, where $\mu_{i,m}$ are the coefficients of the univariate MA representation

of $y_{i,t}$. This, coupled with the Marcinkiewicz–Zygmund inequality gives

$$\sup_i \mathbb{E} y_{i,t}^{2+\zeta} \leq \sup_i \left(\sum_{m=1}^{\infty} \mu_{i,m}^2 \right)^{1+\zeta/2} E \epsilon_{i,t}^{2+\zeta} < \infty, \quad (79)$$

proving (77). Using the above we now derive the properties of $f(Y_i, \theta)$. By the above analysis we have that every element of $Y_t - AZ_{t-1}$ has finite variance for all values of A that satisfy Assumption 3 (i), where $Z_{t-1} = (Y'_{t-1}B'_0, \dots, Y'_{t-p}B'_0)'$. Then, for all $\theta_N \in \Theta_N$,

$$\mathbb{E} \left((f(Y_t, Z_{t-1}, \theta))^2 \right) \leq \|A\|_F^2 \|B\|_F^2 \sup_i \mathbb{E} y_{i,t}^2 = O(N^2)$$

Then,

$$\frac{1}{T} \sum_{i=p}^T f_i(Y_i | Y_{1:i-1}, \theta_{N0}) - \mathbb{E} f_i(Y_i | Y_{1:i-1}, \theta_{N0}) = O_p \left(NT^{-1/2} \right) = o_p(1) \quad (80)$$

To prove (76) we use (80) and note that

$$\sup_{\theta_N \in \Theta_N} \mathbb{E} \left\| N^{-2} \partial l / \partial \theta_N \right\|_F^2 \leq N^{-2} \sup_{\theta_N \in \Theta_N} \|A\|_F^2 \sup_{\theta_N \in \Theta_N} \|B\|_F^2 \sup_i \mathbb{E} y_{i,t}^2 = O(1)$$

and so

$$\sup_{\theta_N \in \Theta_N} \left\| N^{-2} \partial l / \partial \theta_N \right\|_F = O_p(1)$$

Then, stochastic equicontinuity follows by Theorem 21.10 and (21.57) of Davidson (1994) proving (76). \square ■

Appendix B: convergence diagnostics

In this section we discuss convergence of the algorithm used in the paper. The results in the paper are based on 40000 draws from the simulated posterior, obtained by drawing 2 parallel chains of 25000 draws each and discarding the first 5000 draws for burn-in.

We assess convergence by looking at the Inefficiency Factor (IF) and the Potential Scale Reduction Factor (PSRF). The IF are related to the autocorrelation functions and measure how efficient the sampler is, in reference to i.i.d. sampling. An IF of 1 denotes that the draws produced by the algorithm are virtually i.i.d. Typically, an IF below 20 is considered satisfactory for an MCMC sampler. The 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 below 1.1, this is taken as indication of convergence of the algorithm. Results can be found in Table 7 and show that the algorithm is efficient and reaches convergence.

References

- [1] Anderson, T. W., 1951. "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions," *Annals of Mathematical Statistics* 22, 327-351.
- [2] Bai, J. and Ng, S., 2006a. "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions," *Econometrica*, 74(4), 1133-1150.
- [3] Bai, J. and Ng, S., 2006b. "Evaluating Latent and Observed Factors in Macroeconomics and Finance", *Journal of Econometrics*, 113(1-2), 507-537.
- [4] Banbura M., Giannone D. and Reichlin, L., 2010. "Large Bayesian VARs", *Journal of Applied Econometrics*, 25(1), 71-92.
- [5] Baumeister, C., Liu, P. and Mumtaz, H. 2010. "Changes in the transmission of monetary policy: Evidence from a time-varying factor-augmented VAR", Bank of England Working Paper.
- [6] Bernanke, B., J. Boivin, P. Eliasziw, 2005, Measuring the effects of monetary policy: a Factor-Augmented Vector Autoregressive (FAVAR) approach, *The Quarterly Journal of Economics*, 120(1), 387-422.
- [7] 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.
- [8] Camba-Mendez, G., Kapetanios, G., 2009. "Statistical Tests and Estimators of the Rank of a Matrix and Their Applications in Econometric Modelling ", *Econometric Reviews*, 28, 581-611.
- [9] Carriero, C., Clark, T., and Marcellino, M., 2015. "Bayesian VARs: specification choices and forecast accuracy", *Journal of Applied Econometrics*, 30(1), 46-73, 2015.
- [10] Carriero, C., Kapetanios, G. and Marcellino, M., 2011. "Forecasting Large Datasets with Bayesian Reduced Rank Multivariate Models", *Journal of Applied Econometrics*, 26, 736-761.
- [11] 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.
- [12] Davidson, J., 1994. *Stochastic Limit Theory*, Oxford University Press.

- [13] 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.
- [14] Doan, T., R. Litterman, and C. A. Sims, 1984. "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews*, 3, 1-100.
- [15] Del Negro, M., C. Otrok, 2008. "Dynamic factor models with time-varying parameters: Measuring changes in international business cycles", mimeo.
- [16] 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.
- [17] Dufour, J.M. and Stevanovic, D., 2010. "Factor-Augmented VARMA Models: Identification, Estimation, Forecasting and Impulse Responses", mimeo.
- [18] Eickmeier, S., W. Lemke, M. Marcellino, 2014. "Classical time-varying FAVAR models - estimation, forecasting and structural analysis", *Journal of the Royal Statistical Society*, forthcoming.
- [19] 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.
- [20] Forni and Gambetti, 2010. "The dynamic effects of monetary policy: A structural factor model approach", *Journal of Monetary Economics*, 57, 203-216.
- [21] Gelman, A., and Rubin, D., 1992. "Inference from Iterative Simulation using Multiple Sequences," *Statistical Science*, 7, 457-511.
- [22] Gelfand, A.E. and Dey, D., 1994. "Bayesian model choice: asymptotic and exact calculations", *Journal Royal Statistical Society B*, 56, 501-514.
- [23] George, E.I., Sun, D. and Ni, S., 2008. "Bayesian stochastic search for VAR model restrictions", *Journal of Econometrics*, 142, 553-580.
- [24] 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).
- [25] 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.

- [26] Geweke, J., 1996. "Bayesian Reduced Rank Regression in Econometrics", *Journal of Econometrics* 75, 121–146.
- [27] Geweke, J., 1999. "Using simulation methods for Bayesian econometric models: inference, development, and communication," *Econometric Reviews*, 18(1), 1-73.
- [28] 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.
- [29] Hamilton, J.D., Waggoner, D.F. and Zha, T. 2007. "Normalization in Econometrics", *Econometric Reviews*, 26 pp.221–252.
- [30] Ibragimov, I. A. and Linnik, Y. V., 1971. "Independent and stationary sequences of random variables ", Wolters-Noordhoff.
- [31] Kleibergen, F., van Dijk H.K., 1994. "On the Shape of the Likelihood/Posterior in Cointegration Models", *Econometric Theory* 10, 514-551
- [32] Kleibergen, F., van Dijk H.K., 1998. "Bayesian Simultaneous Equations Analysis Using Reduced Rank Structures", *Econometric Theory* 14, 701-743.
- [33] Koop, G.M., Strachan, R.W., Van Dijk, H. and M Villani, 2006. "Bayesian approaches to cointegration" in: *The Palgrave Handbook of Theoretical Econometrics*, Palgrave Macmillan, 871-898.
- [34] 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.
- [35] Johansen, S., 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford: Oxford University Press.
- [36] Lanne, M. and Lutkepohl, H., 1984. "Identifying Monetary Policy Shocks via Changes in Volatility", *Journal of Money, Credit and Banking*, 40, 1131-1149.
- [37] Lanne, M. and Lutkepohl, H., 2010. "Structural Vector Autoregressions with Nonnormal Residuals", *Journal of Business & Economic Statistics*, 28, 159-168.
- [38] Lanne, M., Lutkepohl, H. and Maciejowska K., 2010. "Structural Vector Autoregressions with Markov Switching, *Journal of Economic Dynamics and Control*, 34, 121-131.
- [39] Lutkepohl, H., 1984. "Linear Transformations of Vector ARMA Processes", *Journal of Econometrics*, 26, 283-293.

- [40] Lutkepohl, H., 1990. "Asymptotic Distributions of Impulse Response Functions and Forecast Error Variance Decompositions of Vector Autoregressive Models", *Review of Economics and Statistics*, 72, 116-125.
- [41] Lutkepohl, H., 2007. *New Introduction to Multiple Time Series Analysis*, Oxford University Press.
- [42] Reinsel, G., 1983. "Some Results on Multivariate Autoregressive Index Models," *Biometrika*, 70, 145–156.
- [43] Reinsel, G. C., and Velu, R. P., 1998. "Multivariate Reduced Rank Regression," *Lecture Notes in Statistics* 136. New York: Springer-Verlag.
- [44] Robin, J. M., and Smith, R. J., 2000, "Tests of Rank," *Econometric Theory*, 16, 151-175.
- [45] Stock, J. H., and M.W. Watson, 1989. "New indexes of coincident and leading economic indicators", *NBER Macroeconomic Annual*, 351-94.
- [46] Stock, J. H. and M. W. Watson, 2002a. "Macroeconomic Forecasting Using Diffusion Indexes", *Journal of Business and Economic Statistics*, 20, 147-62.
- [47] 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.
- [48] Velu, R. P., Reinsel, G. C., and Wichern, D. W., 1986. "Reduced Rank Models for Multiple Time Series," *Biometrika*, 73, 105–118.

Table 1. MC results under the MAI DGP

PANEL A- $r=3$, increasing N and T						
	N=5			N=10		
	T=300	T=460	T=720	T=300	T=460	T=720
RMSE						
Bayesian MAI	0.76	0.74	0.74	0.74	0.74	0.74
Classical MAI	6.23	4.75	3.64	4.69	3.56	2.70
BVAR (benchmark)	0.009	0.010	0.009	0.011	0.011	0.011
MAE						
Bayesian MAI	0.90	0.89	0.88	0.84	0.84	0.84
Classical MAI	5.94	4.53	3.49	4.25	3.23	2.48
BVAR (benchmark)	0.008	0.008	0.008	0.009	0.009	0.009
	N=15			N=20		
	T=300	T=460	T=720	T=300	T=460	T=720
RMSE						
Bayesian MAI	0.53	0.48	0.43	0.49	0.44	0.39
Classical MAI	4.99	3.64	2.80	4.28	2.89	2.08
BVAR (benchmark)	0.010	0.010	0.010	0.010	0.010	0.010
MAE						
Bayesian MAI	0.52	0.48	0.43	0.48	0.43	0.39
Classical MAI	4.47	3.30	2.55	3.86	2.63	1.88
BVAR (benchmark)	0.008	0.008	0.008	0.008	0.008	0.008
PANEL B- N=20, T=460, increasing r						
	r=1	r=2	r=3	r=4	r=5	
RMSE						
Bayesian MAI	0.50	0.41	0.43	0.47	0.51	
Classical MAI	0.87	2.10	2.93	3.41	3.38	
BVAR (benchmark)	0.012	0.010	0.010	0.010	0.012	
MAE						
Bayesian MAI	0.47	0.40	0.43	0.45	0.49	
Classical MAI	0.77	1.90	2.65	3.11	3.03	
BVAR (benchmark)	0.010	0.008	0.008	0.009	0.010	

For the Bayesian and Classical MAI the entries show the RMSE and MAE relative to the BVAR (i.e. ratios). The BVAR entries are the RMSE and MAE (levels).

Table 2. MC results under the VAR DGP

PANEL A- r=3, increasing N and T						
	N=5			N=10		
	T=300	T=460	T=720	T=300	T=460	T=720
RMSE						
Bayesian MAI	1.45	1.43	1.51	1.33	1.38	1.37
Classical MAI	4.84	3.86	3.22	4.57	3.51	2.82
BVAR (benchmark)	0.012	0.011	0.011	0.011	0.010	0.010
MAE						
Bayesian MAI	1.52	1.57	1.65	1.38	1.44	1.48
Classical MAI	4.48	3.65	3.14	4.22	3.33	2.74
BVAR (benchmark)	0.010	0.010	0.009	0.009	0.009	0.008
	N=15			N=20		
	T=300	T=460	T=720	T=300	T=460	T=720
RMSE						
Bayesian MAI	1.21	1.22	1.19	1.19	1.17	1.16
Classical MAI	5.58	3.88	2.87	4.91	3.35	2.53
BVAR (benchmark)	0.010	0.010	0.010	0.010	0.010	0.009
MAE						
Bayesian MAI	1.25	1.30	1.29	1.24	1.26	1.27
Classical MAI	5.04	3.61	2.77	4.49	3.17	2.48
BVAR (benchmark)	0.009	0.008	0.008	0.008	0.008	0.008
PANEL B- N=20, T=460, increasing r						
	r=1	r=2	r=3	r=4	r=5	
RMSE						
Bayesian MAI	0.98	1.08	1.16	1.23	1.20	
Classical MAI	2.10	2.74	3.35	4.12	4.43	
BVAR (benchmark)	0.010	0.010	0.010	0.010	0.010	
MAE						
Bayesian MAI	1.13	1.18	1.24	1.30	1.29	
Classical MAI	1.97	2.59	3.16	3.89	4.18	
BVAR (benchmark)	0.008	0.008	0.008	0.008	0.008	

For the Bayesian and Classical MAI the entries show the RMSE and MAE relative to the BVAR (i.e. ratios). The BVAR entries are the RMSE and MAE (levels).

Table 3: Data

Variable	FRED code
Employees on nonfarm payroll	PAYEMS
Average hourly earnings	AHETPI
Personal income	A229RX0
Real Consumption	PCE÷PCEPI
Industrial Production Index	INDPRO
Capacity Utilization	TCU
Unemployment rate	UNRATE
Housing starts	HOUST
CPI all items	CPIAUCSL
Producer Price Index (finished goods)	PPIFGS
Implicit price deflator for personal cons. exp.	PCEPI
PPI ex food and energy	PPILFE
Federal Funds, effective	FEDFUNDS
M1 money stock	M1SL
M2 money stock	M2SL
Total reserves of depository institutions	TOTRESNS
Nonborrowed reserves of depository institutions	NONBORRES
S&P's common stock price index	S&P
Interest rate on treasury bills, 10 year constant maturity	GS10
Effective Exchange rate	CCRETT01USM661N

The sample is at monthly frequency and covers the period January 1974 to December 2013

Table 4: Model selection

rank	lag	shrinkage	MDD MAI	MDD BVAR	PSRF A	PSRF B	BIC MAI	BIC BVAR
3	13	0.02	-9444.7093	-9515.2808	1.0002	1.0059	20.0427	88.6515
1	13	0.01	-9450.9542	-9640.5519	1.0001	1.0054	12.3536	90.1107
3	13	0.01	-9455.7154	-9640.5519	1.0001	1.0057	20.8802	90.1107
2	13	0.02	-9459.1787	-9515.2808	1.0002	1.0064	15.9236	88.6515
2	13	0.01	-9466.1494	-9640.5519	1.0214	1.128	16.9676	90.1107
3	13	0.03	-9466.5513	-9381.8707	1.0002	1.0064	19.7179	87.1783
2	13	0.03	-9473.9637	-9381.8707	1.0003	1.0068	15.6196	87.1783
2	13	0.04	-9496.0593	-9266.832	1.0004	1.0078	15.4534	85.9316
4	13	0.01	-9499.3865	-9640.5519	1.0099	1.0809	25.1342	90.1107
3	13	0.04	-9508.7044	-9266.832	1.0004	1.0075	19.5049	85.9316
4	13	0.02	-9511.0926	-9515.2808	1.0002	1.0051	24.1133	88.6515
2	13	0.05	-9521.6541	-9174.9242	1.0004	1.0085	15.3483	84.8981
3	10	0.02	-9543.7343	-9606.8359	1.0001	1.0055	17.4953	69.6937
4	13	0.03	-9545.1576	-9381.8707	1.0002	1.0056	23.8075	87.1783
3	10	0.01	-9553.7817	-9733.2675	1.0001	1.0052	18.2257	71.1452
2	10	0.02	-9556.1981	-9606.8359	1.0002	1.0064	14.2989	69.6937
1	10	0.01	-9562.3531	-9733.2675	1.0001	1.0053	11.8007	71.1452
1	13	0.02	-9566.8502	-9515.2808	1.0001	1.0055	12.2667	88.6515
2	10	0.03	-9566.8998	-9472.8706	1.0004	1.008	14.071	68.2408
3	10	0.03	-9567.0252	-9472.8706	1.0002	1.0062	17.2419	68.2408

The table displays the top-20 MAI specifications (in terms of MDD) we found over the total 455 specifications we searched over. The first three columns contain the rank-lags-shrinkage combination that uniquely identifies a specification. Columns 4 and 5 contain the value of the Marginal Data Density of the MAI and the BVAR. Columns 6 and 7 contain the Potential Scale Reduction Factors for the MAI model, for the parameters in the matrices A and B respectively. Columns 8 and 9 contain the Bayesian Information Criterion for the MAI and the BVAR.

Table 5: Composition of factors

Variable	F1	F2	F3
Employees on nonfarm payroll	1	0	0
Average hourly earnings	$b_{\{1,2\}}$	0	0
Personal income	$b_{\{1,3\}}$	0	0
Real Consumption	$b_{\{1,4\}}$	0	0
Industrial Production Index	$b_{\{1,5\}}$	0	0
Capacity Utilization	$b_{\{1,6\}}$	0	0
Unemployment rate	$b_{\{1,7\}}$	0	0
Housing starts	$b_{\{1,8\}}$	0	0
CPI all items	0	1	0
Producer Price Index (finished goods)	0	$b_{\{2,10\}}$	0
Implicit price deflator for personal cons. exp.	0	$b_{\{2,11\}}$	0
PPI ex food and energy	0	$b_{\{2,12\}}$	0
Federal Funds, effective	0	0	1
M1 money stock	0	0	$b_{\{3,14\}}$
M2 money stock	0	0	$b_{\{3,15\}}$
Total reserves of depository institutions	0	0	$b_{\{3,16\}}$
Nonborrowed reserves of depository institutions	0	0	$b_{\{3,17\}}$
S&P's common stock price index	0	0	$b_{\{3,18\}}$
Interest rate on treasury bills, 10 year constant maturity	0	0	$b_{\{3,19\}}$
Effective Exchange rate	0	0	$b_{\{3,20\}}$

In the table, the notation $b_{\{j,i\}}$ denotes the element in the j -th row and i -th column of the matrix B_0 . The index j runs through different factors $j=1,...,3$ and the index i runs through different variables $i=1,...,N$.

Table 6: Prior and posterior moments of B_0

Variable	Prior						Posterior					
	Mean			Std.			Mean			Std.		
	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
Employees nonfar	1.000	-	-	-	-	-	1.000	-	-	-	-	-
Avg hourly earn.	0.286	-	-	0.077	-	-	0.263	-	-	0.075	-	-
Personal income	0.328	-	-	0.113	-	-	0.186	-	-	0.099	-	-
Consumption	0.455	-	-	0.097	-	-	0.550	-	-	0.092	-	-
Industrial Prod	1.125	-	-	0.026	-	-	1.118	-	-	0.027	-	-
Capacity utiliz	1.102	-	-	0.030	-	-	1.086	-	-	0.030	-	-
Unemp. Rate	-0.739	-	-	0.074	-	-	-0.741	-	-	0.072	-	-
Housing Starts	0.324	-	-	0.105	-	-	0.383	-	-	0.101	-	-
CPI	-	1.000	-	-	-	-	-	1.000	-	-	-	-
PPI	-	0.703	-	-	0.068	-	-	0.670	-	-	0.066	-
PCE deflator	-	1.063	-	-	0.028	-	-	1.071	-	-	0.028	-
PPI ex food	-	1.108	-	-	0.036	-	-	1.100	-	-	0.036	-
FedFunds	-	-	1.000	-	-	-	-	-	1.000	-	-	-
M1	-	-	-0.155	-	-	0.102	-	-	-0.140	-	-	0.091
M2	-	-	-0.345	-	-	0.066	-	-	-0.260	-	-	0.062
Reserves, tot.	-	-	0.082	-	-	0.107	-	-	-0.146	-	-	0.092
Reserves, nonbor.	-	-	-0.036	-	-	0.351	-	-	-2.101	-	-	0.180
SP500	-	-	-1.482	-	-	0.035	-	-	-1.449	-	-	0.035
10.00 yr T-bond	-	-	0.709	-	-	0.071	-	-	0.548	-	-	0.068
Ex. rate	-	-	0.911	-	-	0.057	-	-	0.841	-	-	0.055

Table 7: Inefficiency Factors and Potential Scale Reduction Factors

	B		A		A-B	
	IF	PSRF	IF	PSRF	IF	PSRF
mean	7.614	1.000	0.898	1.000	1.128	1.000
median	5.649	1.000	0.826	1.000	0.918	1.000
10% quan:	2.810	1.000	0.460	1.000	0.491	1.000
90% quan:	11.235	1.001	1.436	1.000	1.851	1.000
min	2.336	1.000	0.177	1.000	0.151	1.000
max	31.694	1.001	2.334	1.000	13.397	1.000

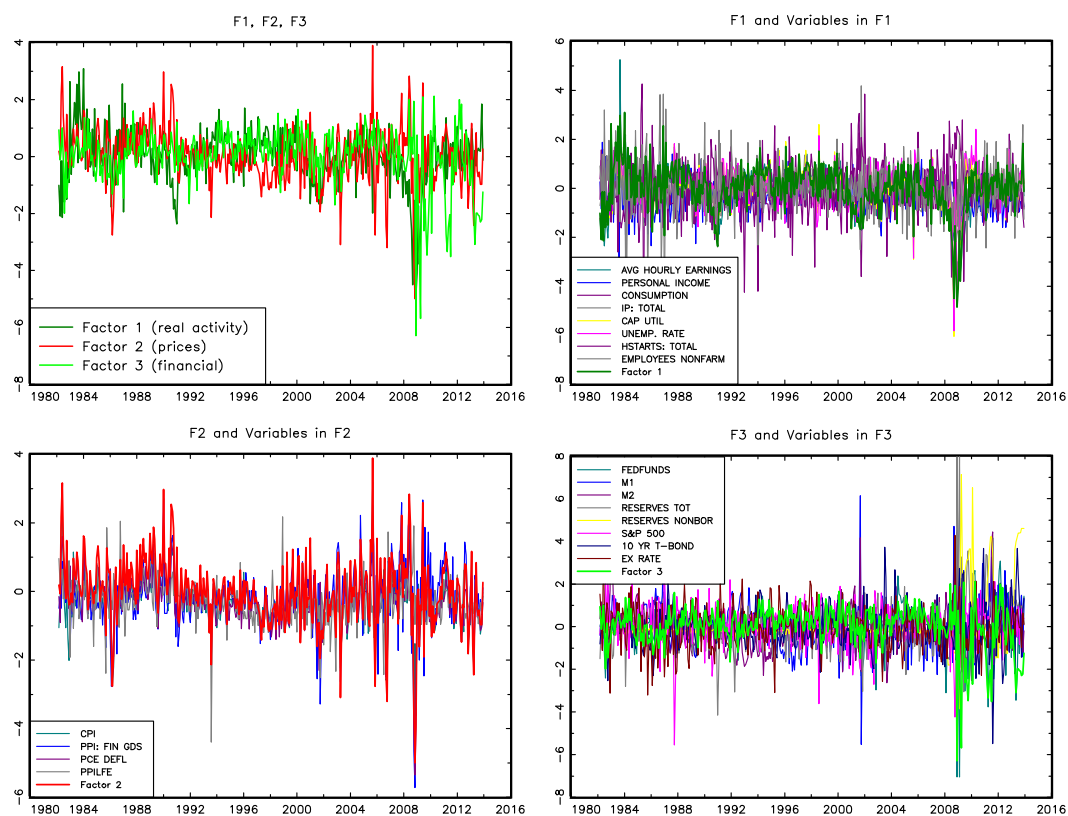


Figure 1: Factors and their components

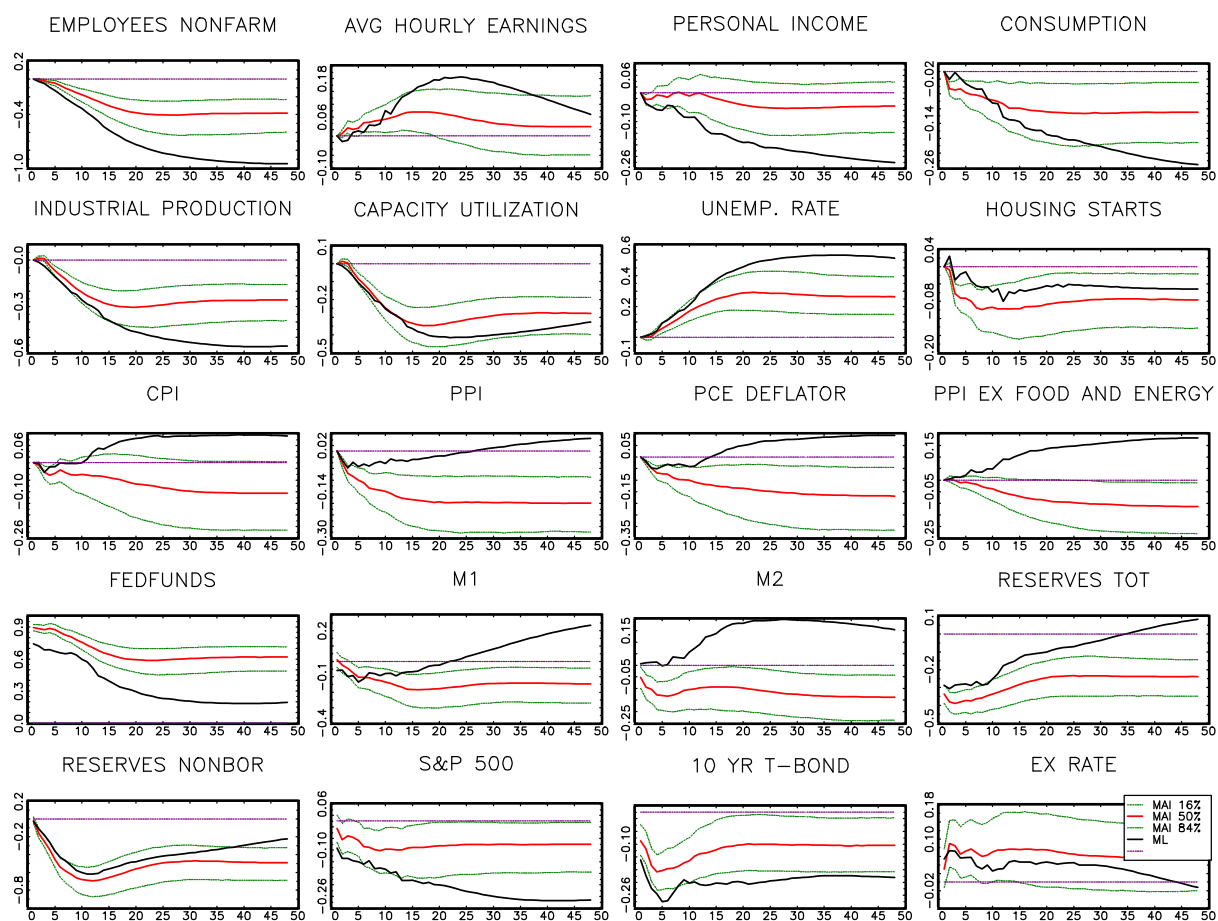


Figure 2: Bayesian vs Classical MAI. Responses to a permanent shock to the Federal Funds rate. Red solid line and green dashed lines are the median and 16%-84% quantiles of the Bayesian MAI impulse responses. The solid black line represents the responses computed using maximum likelihood estimation.

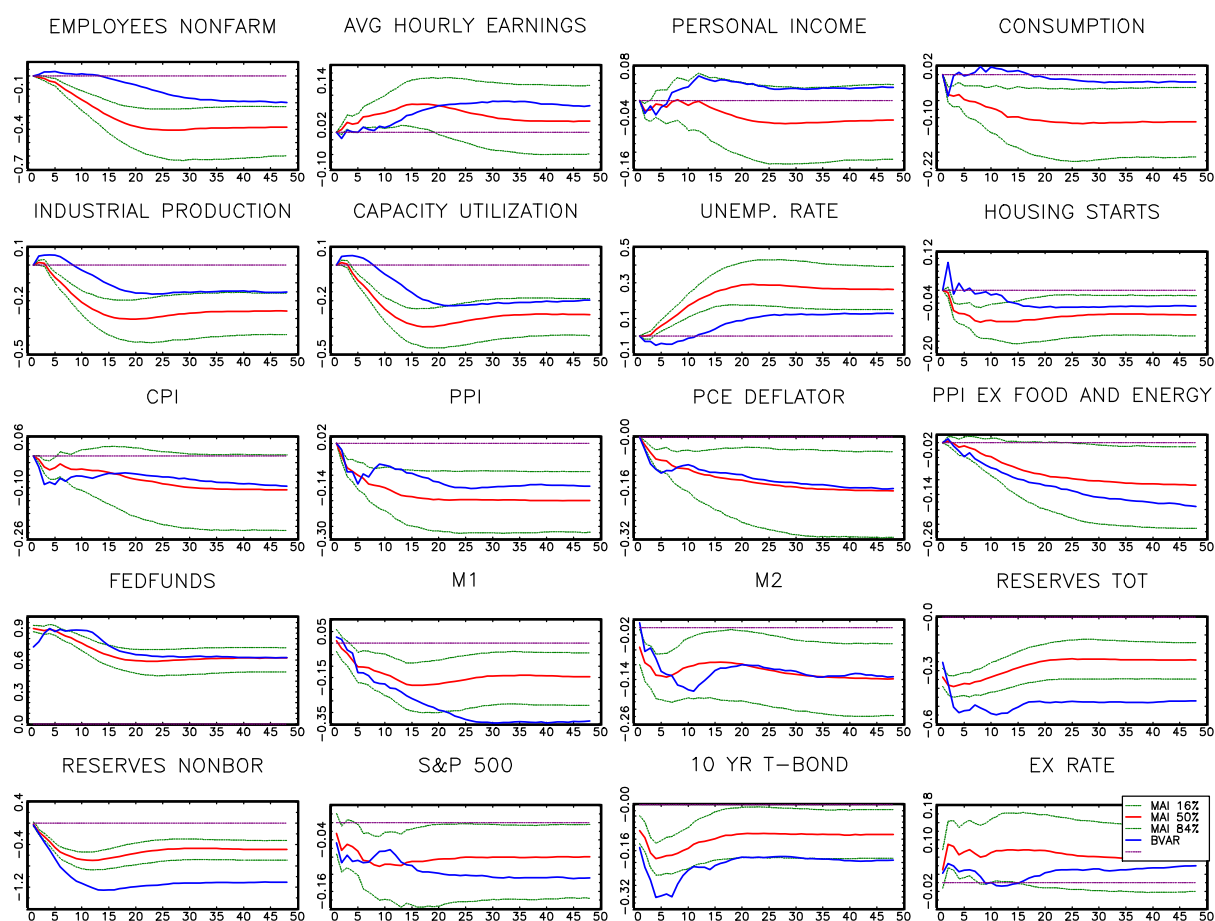


Figure 3: Bayesian MAI vs BVAR. Responses to a permanent shock to the Federal Funds rate. Red solid line and green dashed lines are the median and 16%-84% quantiles of the Bayesian MAI impulse responses. The solid blue line represents the responses computed using the unrestricted BVAR.

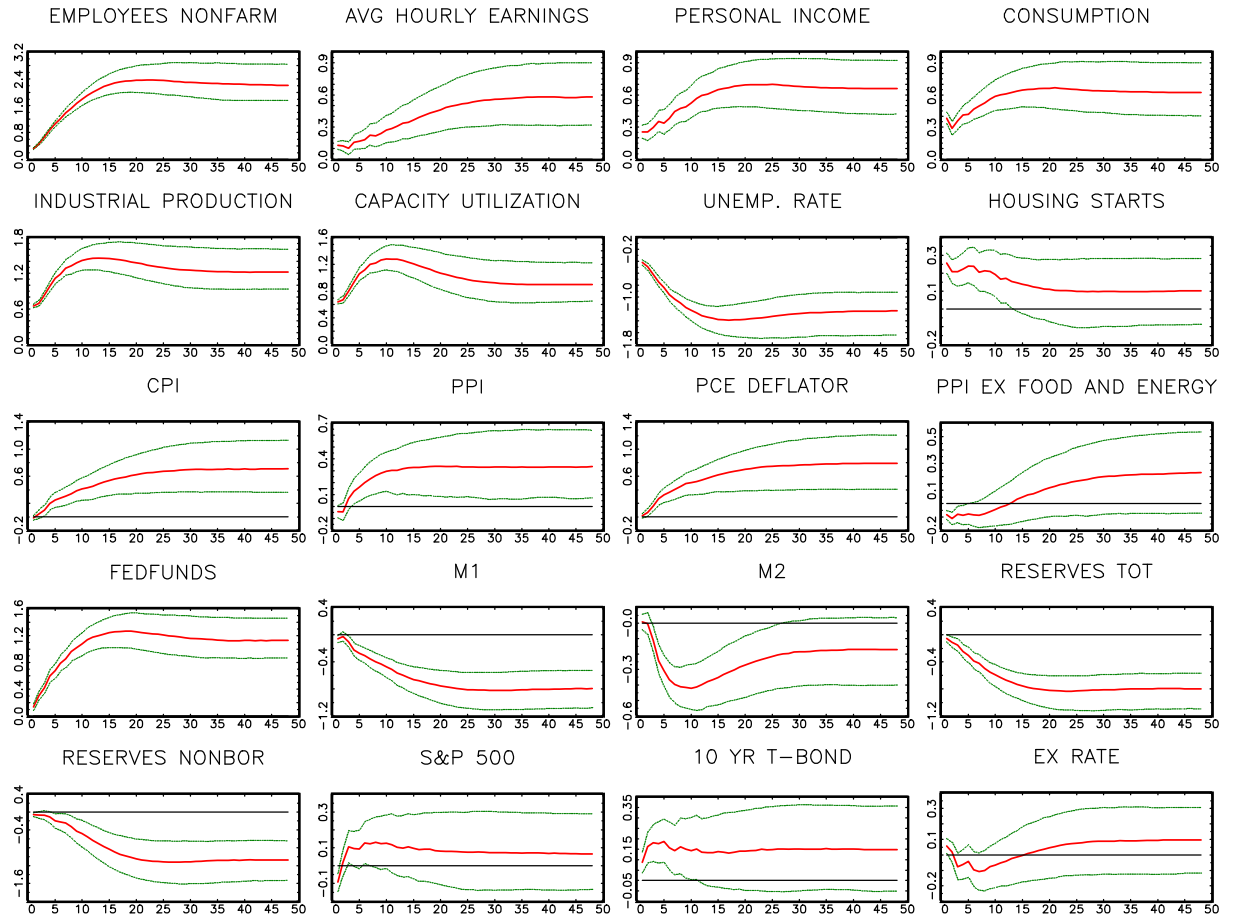


Figure 4: Demand Shock. Responses to a permanent shock to factor 1. Red solid line and green dashed lines are the median and 16%-84% quantiles of the Bayesian MAI impulse responses.

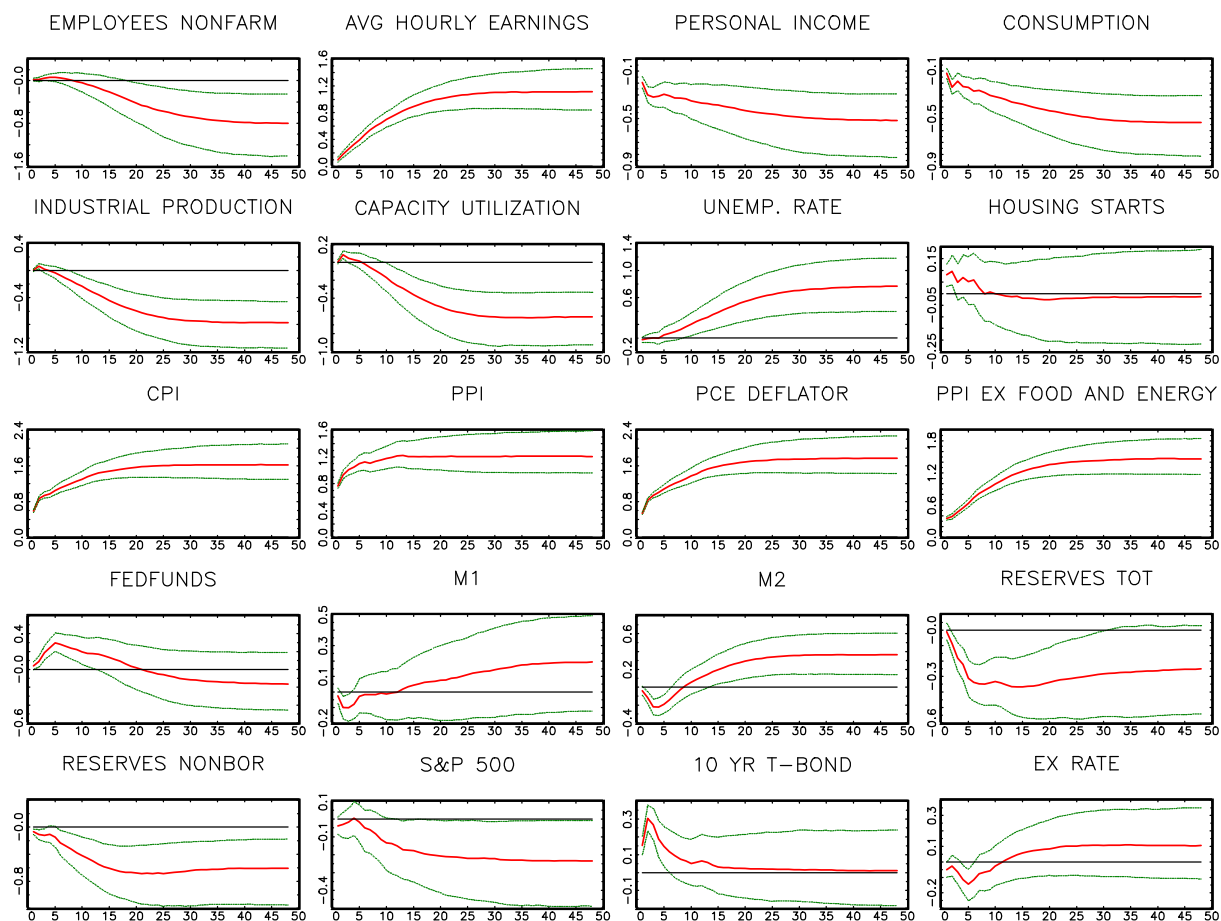


Figure 5: Supply shock. Responses to a permanent shock to factor 2. Red solid line and green dashed lines are the median and 16%-84% quantiles of the Bayesian MAI impulse responses.

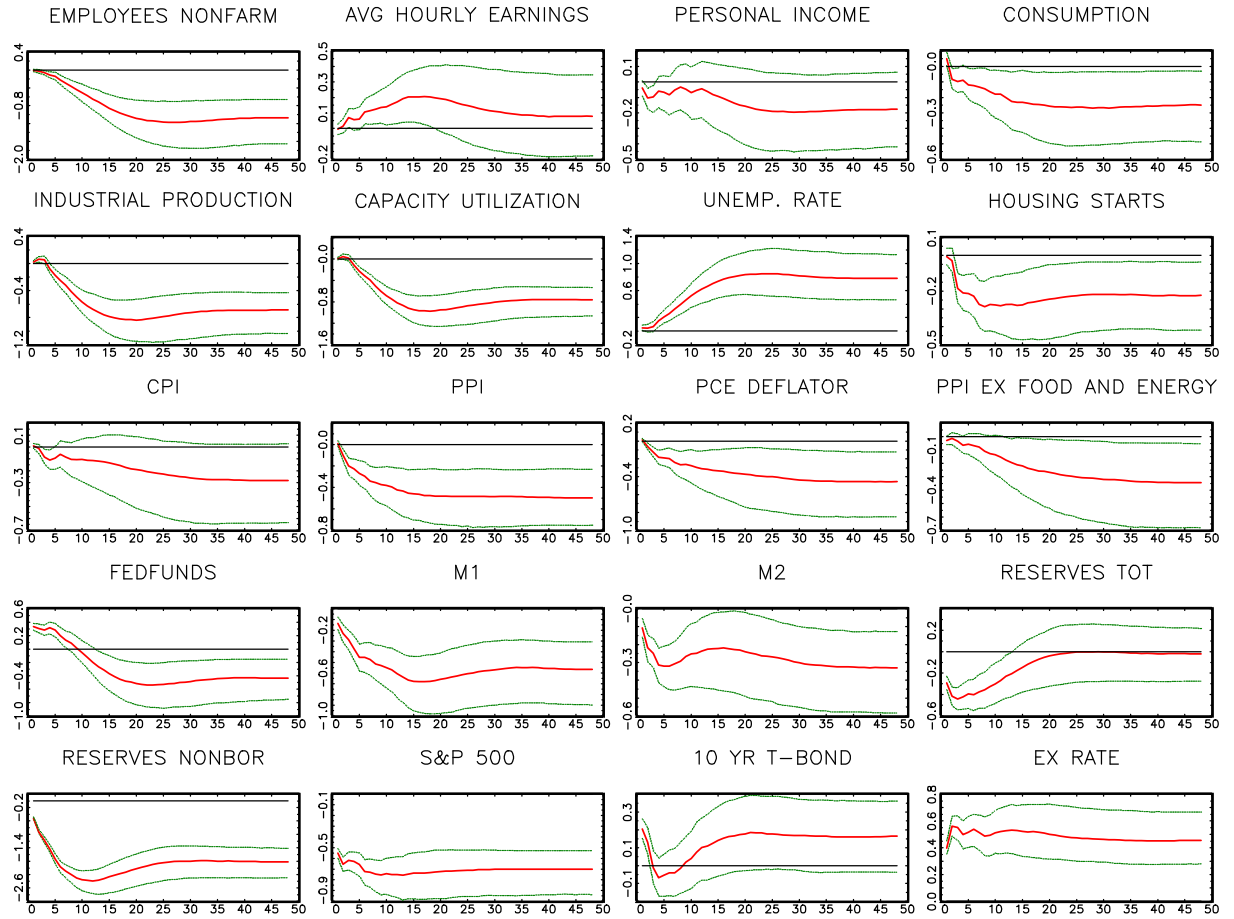


Figure 6: Financial shock. Responses to a permanent shock to factor 3. Red solid line and green dashed lines are the median and 16%-84% quantiles of the Bayesian MAI impulse responses.