Structural analysis with multivariate autoregressive index models

(Article begins on next page)

20 April 2024

Accepted Manuscript

Structural analysis with Multivariate Autoregressive Index models

Andrea Carriero, George Kapetanios, Massimiliano Marcellino

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

November 9, 2015

Abstract

We address the issue of parameter dimensionality reduction in Vector Autoregressive models (VARs) for many variables by imposing speci…c reduced rank restrictions on the coe¢cient 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. Then, the theoretical analysis is extended to the case of general rank restrictions on the VAR coe¢cients. Next, we describe classical and Bayesian estimation of large MAIs, and discuss methods for rank determination. 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 on the propagation of demand and supply shocks.

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

J.E.L. Classi…cation: C11, C13, C33, C53.

1   Introduction

The recent theoretical and applied econometrics literature has paid substantial attention to modelling in the presence of datasets with a large cross-sectional dimension. In estimating models for large datasets, the econometrician is faced with the so-called curse of dimensionality problem, namely the fact that the number of parameters that need to be estimated grows to be so large that it rapidly exhausts the degrees of freedom offered by the available observations. The macroeconometrics literature has tackled this problem using two main approaches.

The first approach uses a small number of factors to summarize the information contained in a large dataset. Factor models were introduced in economics by Geweke (1977) in a small dataset context. Larger systems (typically of more than a hundred variables) were estimated using semi-parametric techniques by Stock and Watson (1989, 2002a, 2002b), Forni et al. (2000). Parametric approaches to large datasets emerged later, mainly in the structural factor augmented VAR (FAVAR) literature, e.g. Bernanke et al. (2005), Kose et al. (2005), Del Negro and Otrok (2008), Baumeister, Liu and Mumtaz (2010), Doz et al. (2011), Eickmeier, Lemke, Marcellino (2014), and the review in Lütkepohl (2014).

The second approach uses Bayesian Vector Autoregressive models (BVARs). BVARs were introduced by Doan, Litterman, and Sims (1984) and Litterman (1986). These models handle large datasets by exploiting Bayesian shrinkage as a regularization method, and have been successfully used in estimation of increasingly large models in studies such as Leeper, Sims and Zha (1996), Banbura, Giannone, and Reichlin (2010), and Koop (2013). Large classical VARs are not a viable alternative, 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. Dufour and Stevanovic (2013) demonstrate that a VARMA representation is more appropriate, though more complex (see also Lütkepohl (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 (e.g. Bai and Ng (2006b)).

BVARs are overall easier to handle than FAVARs in terms of (Bayesian) estimation and inference. However, with the exception of some speci…c prior speci…cations guaranteeing conjugacy, their estimation remains computationally demanding (see Carriero, Clark and Marcellino (2015) for a discussion). Moreover, the fact of having one shock for each variable, each of them equally important, means that these models are somewhat less attractive than FAVAR models to perform structural analysis.

In this paper, we suggest to use a model that bridges BVARs and FAVARs. Speci…cally, we propose to impose reduced rank restrictions on the parameter matrices of a BVAR that, while preserving its attractive features, reduces its dimensionality and makes it similar to a factor model in terms of having a smaller set of key shocks or variables. The resulting speci…cation 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 e¢cient 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, …nding 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 speci…c 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 coe¢cients 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 speci…c empirical problem under analysis. In Section 2 we also discuss the relationship of the MAI model with factor models, more general reduced rank VAR models, and with the reduced rank multivariate regression studied in a Bayesian setting by Geweke (1996).

In Section 3 we introduce classical and Bayesian estimation methods for the MAI. Reduced rank regressions have been introduced by Anderson (1951), and the speci…c case of

2

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 coe¢cients.

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. In a Bayesian framework, we propose to select the rank associated with the highest marginal data density, which can be e¢ciently approximated numerically using Rao-Blackwellization combined with the harmonic mean estimator of Gelfand and Dey (1994) and Geweke's (1999), as suggested e.g. in Fuentes-Albero and Melosi (2013).

In Section 5 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 6 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 economically sensible and sometimes di¤erent 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. Second, we assess the e¤ects of demand and supply shocks. In this case we use the FAVAR-style MA representation of the MAI and assume that the factors re‡ect movements in real, …nancial, and price variables, where the shocks associated with these factors are interpreted as, respectively, demand, …nancial / monetary, and supply shocks. Again, the resulting responses are very sensible from an economic point of view.

Finally, Section 7 concludes and proposes directions for further research.

2   The MAI Model: Speci…cation and MA Representations

2.1   The MAI-model

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

$$Y_t = \Pi(L)Y_t + \varepsilon_t; \tag{1}$$

where $t = 1; \ldots; T$, $\Pi(L) = \Pi_1 L + \ldots + \Pi_p L^p$ is a polynomial of order p, and ε_t are i.i.d. $N(0; \Sigma)$.

3

Following Reinsel (1983), we further assume that $\Pi(L)$ can be factorized as $\Pi(L) = A(L)B_0$, where $A(L) = A_1L + \ldots + A_pL^p$, each matrix A_u is of dimension $N \times r$, $u = 1, \ldots, p$, 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]

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

If r is much smaller than N and N is large, 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 $T = 460$, $N = 20$, $p = 13$ 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, $N^2p = 5200$ and $Nr(p+1) = 20 \times 3 \times 14 = 840$.

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_0Y_{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:

$$F_t = B_0Y_t.$$ (3)

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

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

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 subsection 2.3.

An important characteristic of the MAI model is that the linear combinations B_0Y_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. Lütkepohl (2007). To see this, it is sufficient to pre-multiply by B_0 both sides of equation (2) and use (3) to get:

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

with $C(L) = B_0A(L) = B_0A_1L + B_0A_2L^2 + \ldots + B_0A_pL^p$ and with $u_t = B_0\varepsilon_t$ being an

_____
[1] More general reduced rank models are considered in Section 2.4. 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.

4

i.i.d. Gaussian process with mean zero and variance $\Sigma = B_0 B_0^0$

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 …rst moving average representation:

$$Y_t = (I \quad A(L)B_0)^{-1} \varepsilon_t: \tag{6}$$

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

A second moving average representation is:

$$Y_t = (A(L)(I \quad B_0 A(L))^{-1}B_0 + I)\varepsilon_t: \tag{7}$$

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

$$F_t = (I \quad C(L))^{-1}u_t = (I \quad B_0 A(L))^{-1}B_0\varepsilon_t; \tag{8}$$

and then inserting it into equation (4). The two alternative moving average representations for Y_t in (6) and (7) are of course equivalent:

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

A third moving average representation is particularly convenient for structural analysis. Let us introduce the $(N \quad r) \quad N$ full row rank matrix $B_{0?}$ that is orthogonal to B_0, i.e. $B_0 B_{0?}^0 = 0$, such that the rank of $(B_0^0 : B_{0?}^0)$ is N. Note that $B_0 B_0^0$ and $B_0 B_0^0$ 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)):[2]

$$B_0^0 (B_0 B_0^0)^{-1}B_0 + B_{0?}^0 (B_{0?} \quad {}^1B_{0?}^0)^{-1}B_{0?} \quad {}^1 = I_N: \tag{10}$$

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

$$Y_t = (\quad B_0^0 (B_0 \quad B_0^0)^{-1} + A(L)(I \quad B_0 A(L))^{-1})B_0\varepsilon_t + B_{0?}^0 (B_{0?} \quad {}^1B_{0?}^0)^{-1}B_{0?} \quad {}^1\varepsilon_t \cdot \tag{11}$$

Since $B_0\varepsilon_t = u_t$, $\Sigma = B_0 B_0^0$, and de…ning $B_{0?} \quad {}^1\varepsilon_t = \eta_t$, we have:

$$Y_t = (\quad B_0^0 \quad {}^{-1} + A(L)(I \quad B_0 A(L))^{-1})u_t + B_{0?}^0 (B_{0?} \quad {}^1B_{0?}^0)^{-1}\eta_t: \tag{12}$$

The representation in (12) shows that each element of Y_t is driven by a set of r common

---

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

errors, the u_t that are the drivers of the factors F_t, and by linear combinations of ξ_t. Since

$$E(u_t \xi_t^0) = E(B_0 \eta_t \xi_t^0 \xi_{0?}^1 B^0) = 0, \tag{13}$$

$$E(u_t \xi_{t-i}^0) = 0; \quad E(u_{t-i} \xi_t^0) = 0; \quad i > 0, \tag{14}$$

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

$$v_t = P\,u_t; \tag{15}$$

where P is a lower triangular matrix such that $\Sigma = P^{-1}P^{-1^0}$, which implies that v_t are structural shocks with $Var(v_t) = P\,\Sigma\,P^0 = I$. Hence, combining (15) with (12) yields

$$Y_t = (\,B_0^0{}^{-1} + A(L)(I - B_0 A(L))^{-1})P^{-1}v_t + B_{0?}^0 (B_{0?} {}^{-1}B_{0?}^0)^{-1} \xi_t; \tag{16}$$

from which impulse response functions can be easily computed.

Note that, since $u_t = B_0 \eta_t$, the structural shocks v_t are also related to the ξ_t errors in the Wold representations in (6) or (7), via the relationship $v_t = P B_0 \xi_t$. 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.

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, $(\,B_0^0{}^{-1} + A(L)(I - B_0 A(L))^{-1})u_t$ in (12), is uncorrelated at all leads and lags with the error terms $B_{0?}^0 (B_0{}^{-1}B_{0?}^0)^{-1}\xi_t$.

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 $B_{0?}^0 (B_0{}^{-1}B_{0?}^0)^{-1}\xi_t$

in (12) can be in general correlated among themselves, while in factor models they must 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 (2013)), 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.

2.4   Relationship with general reduced rank VARs

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). Let us again consider the VAR model in (1) and, as before, assume that (L) can be factorized as (L) = A(L)B(L), where $A(L) = A_1 L + \cdots + A_{p_1} L^{p_1}$ and each A_u is of dimension N   r, $u = 1, \ldots, p_1$. However, we now assume a more general specification for B(L), namely $B(L) = B_0 + B_1 L + \cdots + B_{p_2} L^{p_2}$ where each B_v is full rank of dimension r   N, $v = 1, \ldots, p_2$. 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:

$$Y_t = A(L)B(L)Y_t + \epsilon_t = \sum_{u=1}^{p_1}\sum_{v=0}^{p_2} A_u B_v Y_{t-u-v} + \epsilon_t. \qquad (17)$$

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

$$F_t = B(L)Y_t = B_0 Y_t + B_1 Y_{t-1} + \cdots + B_{p_2} Y_{t-p_2}. \qquad (18)$$

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

$$A(L)B(L) = (A_1 L + A_2 L^2)(B_0 + B_1 L) = A_1 B_0 L + (A_1 B_1 + A_2 B_0)L^2 + A_2 B_1 L^3,$$

7

so that $\text{rank}(A_1 B_0) \quad r$, $\text{rank}(A_2 B_1) \quad r$ but $\text{rank}(A_1 B_1 + A_2 B_0)$ can be larger than r. There is also more ‡exibility in the speci…cation of the factors, compare (3) with (18). On the other hand, the factors no longer follow a …nite order VAR but rather a VARMA, as it is:

$$F_t = B(L)A(L)F_t + B(L) \epsilon_t: \tag{19}$$

For a description of the moving average representations and estimation details of this model we refer the reader to the working paper version of this article, Carriero, Kapetanios and Marcellino (2015). We should point out that the analytical derivation the moving average representation corresponding to (12) for this more general model 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 speci…cation. We will see in the next subsection that a substantial simpli…cation occurs for the case $p_1 = 1$, $p_2 = p \quad 1$.

2.5 Relationship with multivariate reduced rank regression

In this section we focus on another special case of the general reduced rank VAR in (17), which is obtained by setting $p_1 = 1$, $p_2 = p \quad 1$. De…ning $X_t = (Y_{t\ 1}^{0}, \cdots ; Y_{t\ p}^{0})^{0}$, the resulting model can be written as:

$$\underset{N\ 1}{Y_t} = \underset{N\ r}{A_1} \underset{r\ Np}{[B_0; \cdots ; B_{p\ 1}]} \underset{Np\ 1}{X_t} + \underset{N\ 1}{\epsilon_t}; \tag{20}$$

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 (20) with the MAI model written as:

$$\underset{N\ 1}{Y_t} = \underset{N\ rp}{[A_1; \cdots ; A_p]} \underset{rp\ Np}{(I_p \quad B_0^{0})^{0}} \underset{Np\ 1}{X_t} + \underset{N\ 1}{\epsilon_t}: \tag{21}$$

As is clear from comparison of (20) with (21), the reduced rank VAR in (20) has only one A_1 matrix of dimension $N \quad r$ and p matrices $B_0; \cdots ; B_{p\ 1}$ each of dimension $r \quad N$, while the MAI in (21) has p matrices $A_1; \cdots ; A_p$ each of dimension $N \quad r$ and only one B_0 matrix of dimension $r \quad N$.

The main advantage of the speci…cation in (20) is that A_1 has full rank r, therefore it is possible to premultiply the system by the generalized inverse $A^+ = (A_1^{0} A_1)^{1} A_1^{0}$, and -conditional on A_1- to derive a closed form posterior distribution for $[B_0; \cdots ; B_{p\ 1}]$, which can then be easily simulated using a Gibbs sampling step (details can be found in Geweke (1996)). Instead, the matrix $[A_1; \cdots ; A_p]$ appearing in the MAI in (21) is not full rank, which is the reason why B_0 can only be simulated using a Metropolis step, as we will see in more details in the next Section.[3] The main advantage of speci…cation (21) is

_____

[3] Blocking the system in p di¤erent blocks and deriving conditional posteriors is also not feasible in the MAI model. Indeed, while the regressor matrix $I_p \quad B_0^{0}$ has a block-diagonal structure, each of the blocks in

that, as we have discussed in Section 2.1, premultiplication of (20) by B_0 provides a VAR speci…cation for the factors $B_0 Y_t$, while in (20) the factors do not admit a …nite 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, $B_0 y_{t\ 1}; \ldots; B_0 y_{t\ p}$, which can have di¤erent e¤ects over time and across variables, and $B_0 y_t$ admits a VAR representation. Instead, in the multivariate reduced rank model the large set of Np factors have a changing composition over time, $B_0 y_{t\ 1}; B_1 y_{t\ 2}; \ldots; B_{p\ 1} y_{t\ p}$, and require a large VARMA speci…cation.

To summarize, estimation of the multivariate reduced rank model in (20) is easier than estimation of the MAI model in (21), but the MAI model allows to derive a …nite order VAR representations for a set of r factors. For these reasons, speci…cation (20) 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 speci…cation (20) is better suited for structural analysis.

3   Estimation of Large MAI Models

For estimation it is convenient to compactly rewrite (2) as:

$$Y_t = A Z_{t\ 1} + {}_t; \tag{22}$$

where $Z^0_{t\ 1} = (F^0_{t\ 1} \ldots; F^0_{t\ p}) = (Y^0_{t\ 1} B^0_0; \ldots; Y^0_{t\ p} B^0_0) = (Y^0_{t\ 1}, \ldots; Y^0_{t\ p})(I_p \quad B^0_0)$, ${}_t$ is a 1 rp vector, and $A = (A_1; \ldots; A_p)$ is a N rp matrix. As, for all js, $A_j B_0 = A_j Q^{\ 1} Q B_0$ for any nonsingular matrix Q, we add the identi…cation restriction $B_0 = (I_r; \ B_0)$. De…ning $Y = (Y_1; \ldots; Y_T)^0$, $Z = (Z_0; Z_1; \ldots; Z_{T\ 1})^0$, $E = ({}_1; \ldots; {}_T)^0$ and stacking the equations in (22) for $t = 1; \ldots; T$, we have:

$$Y = Z A^0 + E; \tag{23}$$

where E has a matricvariate normal distribution with variance $(\quad I_T)$.

3.1   Estimation via Maximum Likelihood

Reinsel (1983) studied estimation of the model in (22) via Maximum Likelihood (ML). In particular, he showed that ML estimators can be obtained by iterating over the …rst order conditions of the maximization problem. The likelihood function is:

$$0:5T \log j \quad j \quad 0:5 \sum_{t=1}^{T} (Y_t \quad A Z_{t\ 1})^0 \quad {}^1 (Y_t - A Z_{t\ 1}): \tag{24}$$

---

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.

For any A and B_0 the maximization with respect to Σ yields:

$$\hat{\Sigma} = (Y - ZA^0)'(Y - ZA^0)/T:\tag{25}$$

The partial derivatives with respect to A (given B_0 and Σ) can be obtained by noting that $AZ_{t-1} = vec(Z^0_{t-1}A^0) = (I_N \otimes Z^0_{t-1})vec(A^0)$, and the corresponding …rst order conditions are given by:

$$\frac{@l}{@vec(A^0)} = \sum_{t=1}^{T}(I_N \otimes Z_{t-1})\Sigma^{-1}\{Y_t - (I_N \otimes Z_{t-1})vec(A^0)\}g = 0:\tag{26}$$

The partial derivatives with respect to B_0 (given A and Σ) can be obtained by noting that $AZ_{t-1} = \sum_{j=1}^{P}A_jY^0_{j-1;t} + \sum_{j=1}^{P}(A_j Y^0_{2;t-1})vec(B_0)$, where $Y_{2;t}$ comes from partitioning Y_t^0 in the …rst r and last $N-r$ components: $Y^0_t = (Y^0_{1;t};Y^0_{2;t})$. The corresponding …rst order conditions are given by:

$$\frac{@l}{@vec(B_0)} = \sum_{t=1}^{\rtimes}U_{t-1}A^0\Sigma^{-1}\{Y_t - (I_N \otimes Z^0_{t-1})vec(A^0)\}g=0;\tag{27}$$

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

Reinsel (1983) suggested to solve in turn equations (25), (26) and (27) until convergence is achieved, and established consistency and asymptotic normality of this estimator for …xed N. Of course these consistency and asymptotic normality results can be coupled with the standard impulse response analysis for …nite dimensional VAR models to produce standard errors for such impulse responses (see, e.g., Section 3.7 of Lütkepohl (2007)). Also, speci…c 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 …nite number of variables, and we now want to extend it to the case where N possibly diverges. This is undertaken in Appendix A. In particular, assuming that Assumption 3 of Appendix A is valid for the MAI in (2) implies Lemma 6, which in turn implies Theorems 2 and 3. In turn, these imply the following Theorem, where $k.k_F$ denotes the Frobenius norm of a matrix.

Theorem 1 Let Assumption 3 hold for the MAI in (2). De…ne $\theta_{N0} = (vec(A)^0;vec(B)^0)^0$ to be the true value of the parameters and $\hat{\theta}_N$ its MLE estimator. Then,

$$\left\|\hat{\theta}_N - \theta_{N0}\right\|_F = O^p\left(\frac{N^{5=2}}{T}\right) + O^p\left(\frac{N}{T^{1=2}}\right):\tag{28}$$

Our rate derivations require bounds that may not be as sharp as possible. 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 2.4.

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 (23), derive the conditional posterior distributions, and provide an MCMC algorithm for Bayesian estimation.

3.2.1   Priors

The model (23) has three sets of parameters, contained respectively in the matrices A^0, B_0, and . We elicit a natural conjugate Normal-Inverse Wishart prior for A^0 and   :

$$A^0_j \quad N(A_0; \quad V_0); \qquad IW(S_0; v_0): \qquad (29)$$

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 coe¢cients is set to $A_0 = 0$ (a rp N matrix of zeros). The prior variance V_0 is set to a diagonal rp rp matrix:

$$V_0 = \quad diag(V_0^1; V_0^2; \ldots; V_0^p), \qquad (30)$$

where each V_0^k for $k = 1; \ldots; p$ is a r-dimensional vector with all the entries equal to $1=k^2$. This choice for the prior mean and variance of the coe¢cients re‡ects the belief that, as the lag order increases, lagged values of the explanatory variables become increasingly less informative in explaining current values of the dependent variable, in line with the traditional Minnesota prior implementation of Doan, Litterman and Sims (1984) and Litterman (1986). The hyperparameter 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, in Section 4.2 and Section 6.1). 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 di¤use as possible while remaining proper.[4]

---

[4] Given that in the empirical application the ratio of number of observations T to number variables N is quite large (about 384 to 20) it is possible to use such a di¤use 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 coe¢cients but a slightly inferior mixing.

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 **i** and column **j** 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.[5] Then we regress each of the factors for $j = 1, \ldots, r$ onto each individual variable y_{it}, $i = 1, \ldots, N$ 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 A2 in the online Appendix contains the prior mean and standard deviations resulting from this prior elicitation strategy. These are the values used in our empirical applications.

3.2.2 Posteriors and MCMC algorithm

The joint posterior distribution $p(A^0, \tilde{B}_0, \Sigma | Y)$ has not a known form, but it can be simulated by using a Gibbs sampler drawing in turn from the conditional posterior distributions $p(A^0, \Sigma | \tilde{B}_0, Y)$ and $p(\tilde{B}_0 | A^0, \Sigma, Y)$.

Drawing from the conditional posterior $p(A^0, \Sigma | \tilde{B}_0, Y)$ is straightforward. Given knowledge of $\tilde{B}_0$ and Y, the variable Z_{t-1} is known, and (22) is a simple multivariate regression model as the one described in Zellner (1973). Then, under the natural conjugate prior described by (29), the conditional posterior distributions are:

$$A^0 | \Sigma, \tilde{B}_0, Y \sim N(\bar{A}, \Sigma \otimes \bar{V}); \quad \Sigma | \tilde{B}_0, Y \sim IW(\bar{S}, \bar{v}); \tag{31}$$

where $\bar{V} = (V_0^{-1} + Z'Z)^{-1}$, $\bar{A} = \bar{V}(V_0^{-1}A_0 + Z'Y)$, $\bar{S} = S_0 + Y'Y + A_0'V_0^{-1}A_0 - \bar{A}'\bar{V}^{-1}\bar{A}$, and $\bar{v} = v_0 + T$. Draws from $p(A^0, \Sigma | \tilde{B}_0, Y)$ can be easily obtained by generating a sequence of M draws $\{\Sigma_m\}_{m=1}^M$ from $\Sigma | \tilde{B}_0, Y \sim IW(\bar{S}, \bar{v})$ and then for each m drawing from $A^0 | \Sigma, \tilde{B}_0, Y \sim N(\bar{A}, \Sigma_m \otimes \bar{V})$, which provides the sequence $\{A^0_m, \Sigma_m\}_{m=1}^M$.

Drawing from $p(\tilde{B}_0 | A^0, \Sigma, Y)$ is less straightforward, as B_0 contains restrictions and enters the model in a nonlinear way. To draw $\tilde{B}_0$ conditional on A^0 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^0, Σ, and the remaining elements $\tilde{B}_{0ji}^-$, using a random walk proposal:

$$\tilde{B}_{0ji} = \tilde{B}_{0ji}^{m-1} + c\,\varepsilon_t, \tag{32}$$

where ε_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%.[6] The candidate draw is then accepted with

---

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

[6] To choose the scaling constant c, which is the standard deviation of the proposal density, we use the

probability

$$k = \min \left(1; \frac{p(\tilde{B}_{0ji}|\tilde{B}_{0ji}; A^0; ; Y)}{p(\tilde{B}_{0ji}^{-1}|\tilde{B}_{0ji}; A^0; ; Y)} \right).$$ (33)

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 $j = 1; \ldots; r$ and $i = 1; \ldots; N$.

Drawing in turn from $p(A^0; _j|\tilde{B}_0; Y)$ and $p(\tilde{B}_0|A^0; ; Y)$ provides a sequence of M draws $\{A^0; _j; \tilde{B}_0\}^M_{m=1}$ from the joint posterior distribution of $A^0; ; \tilde{B}_0$: 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 a¤ected by problems related to the nonlinearity. A more detailed discussion of the role of normalization 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 coe¢cient 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 …rst uses information criteria. This approach simply estimates (22) for all possible values of r and chooses the one that minimizes an information criterion (IC) that uses the …t 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

---

standard deviation of the prior density for each individual coe¢cient. 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 coe¢cients. The scaling-up is necessary as the prior is independent across coe¢cients and therefore ignores their potential correlation in the posterior distribution.

of the rank of each matrix A_i in (22). 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. Examples of tests based on this approach include Cragg and Donald (1996) and Robin and Smith (2000).[7]

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:

$$p_r(Y) = \int p(Y \mid \theta) p(\theta) d\theta , \qquad (34)$$

where $\theta = (A; \Sigma; B_0)$ contains all the coefficients of the model. The optimal rank for the system is associated with the model featuring the highest data density:

$$r = \arg\max_r p_r(Y): \qquad (35)$$

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-Blackwellization and the harmonic mean estimator proposed by Gelfand and Dey (1994), as suggested in Fuentes-Albero and Melosi (2013). In particular, given M simulated posterior draws $\{B_0\}_{m=1}^M$, we have:

$$\hat{p}_r(Y) = \left[\frac{1}{M} \sum_{m=1}^M \frac{g(B_0^m)}{p(Y \mid B_0^m) p(B_0^m)} \right]^{-1} ; \qquad (36)$$

where $g(\cdot)$ is a truncated multivariate normal distribution calibrated using the moments of the simulated posterior draws (see Geweke (1999)) and $p(B_0^m)$ is the prior distribution of B_0 evaluated at the posterior draw B_0^m. The term $p(Y \mid B_0^m)$ is the integrating constant of the conditional posterior distribution $p(A; \Sigma \mid Y; B_0)$. Since conditionally on B_0^m the model is a multivariate regression with a naturally conjugate prior, $p(Y \mid B_0^m)$ is available in closed form:

$$p(Y \mid B_0^m) = \pi^{-\frac{TN}{2}} \mid (I + Z^m V_0 Z^{m0}) \mid^{-\frac{1}{2}} \mid S_0 \mid^{\frac{v_0}{2}} \frac{\Gamma_N(\frac{v_0+T}{2})}{\Gamma_N(\frac{v_0}{2})} \qquad (37)$$
$$\mid S_0 + (Y - Z^m A_0)'(I + Z^m V_0 Z^{m0})^{-1}(Y - Z^m A_0) \mid^{-\frac{v}{2}};$$

where $\Gamma_N(\cdot)$ is denoting the N-variate gamma function and where the conditioning on B_0^m is implicit in the conditioning on $Z^m = (Z_0^m; Z_1^m; \dots; Z_{T-1}^m)'$ because $Z_t^{m0} = (Y_{t-1}' \dots; Y_{t-p}')(I_p \otimes B_0^{m0})$. The result (37) for a general multivariate regression dates back to Zellner (1973), and a

---

[7] A simple alternative to tests of rank may be 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.

straightforward derivation based on theorem A.19 in Bauwens, Lubrano and Richard (2000) can be found in Carriero, Kapetanios, and Marcellino (2011).

The marginal data density (36) can also be used to select the optimal lag length p and optimal shrinkage hyperparameter , and we do so in our empirical applications.

5 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 arti…cial data from two alternative Data Generating Processes (DGP). We recall equation (1) and rewrite it as:

$$Y_t = \sum_{u=1}^{p} {}_u Y_{t-u} + {}_t; \quad {}_t \quad i.i.d.N(0; \ ): \tag{38}$$

The …rst DGP (DGP1) is an unrestricted VAR, so it uses (38) without imposing any further restrictions. The second DGP (DGP2) is the MAI, so it imposes the rank reduction restriction:

$$ {}_u = A_u B_0; \tag{39}$$

with $u = 1; \ldots; p$. To set up the parameters and $ {}_1; \ldots; \ {}_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 (38) using a standard Bayesian approach,[8] which provides us with the estimated values $\hat{}$ and $\hat{}_1; \ldots; \hat{}_p.$[9] Similarly, for DGP2 we estimate (38) again but this time under the restriction (39), with rank set to $r = 3$, and using the estimation approach described in Section 3.2, which provides us with the estimated values and $ {}_1; \ldots; \ {}_p.$

To simulate arti…cial data from the two alternative DGPs we set and $ {}_1; \ldots; \ {}_p$ to $\hat{}$ and $\hat{}_1; \ldots; \hat{}_p$ (under DGP1) or and $ {}_1; \ldots; \ {}_p$ (under DGP2), draw 100 di¤erent disturbances vectors from ${}_t \ i.i.d.N(0; \hat{})$ and ${}_t \ i.i.d.N(0; \ )$, and project forward (38), which provides 100 di¤erent realizations of the process Y_t under DGP1 and 100 di¤erent 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, described in Section 3.1; iii) an unrestricted BVAR, estimated as in the baseline speci…cation of Carriero, Clark, and Marcellino (2015).

---

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

[9] Note 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 (39), 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 $ {}_1; \ldots; \ {}_p$ matrices.

To ascertain the properties of the di¤erent models under the di¤erent 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 coe¢cients is very large, rather than looking at individual RMSEs and MAEs for each of the coe¢cients appearing in $_1; \ldots; _p$, we focus on the average RMSE and MAEs over all the estimated coe¢cients. 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 di¤erent 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 di¤erent combinations of sample size and cross sectional size, in particular N = 5; 10; 15; 20 and T = 300; 460; 720.[10] Panel B displays results for …xed N and T (20 and 460 respectively, which are the dimensions of our empirical application) and for di¤erent values of the rank of the DGP, r = 1; 2; 3; 4; 5 in (39).[11] 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, …gures below 1 signal 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 e¤ectively handled by maximum likelihood estimation.

The considerations above are still valid when looking at results for di¤erent 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 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.

---

[10] It should be noted that while comparing results for increasing T and …xed N involves looking at the same DGP estimated with an increasing number of observations, comparing results for increasing N and …xed T is not as straightforward. Indeed, since the DGPs are calibrated using estimates obtained in a preliminary step based on actual data, qualitative di¤erences in the data used for the DGP calibration enter the picture and should be kept in mind. Another potential di¤erence lies in the fact that the overall shrinkage for the Bayesian approaches is kept …xed as N increases, 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.

[11] Recall that the rank restriction is imposed in the preliminary estimation step that provides us with the values of the DGP parameters. To obtain di¤erent ranks in the DGP it is su¢cient to impose restriction (39) with the desired alternative values for the rank.

16

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 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 e¢ciency as N increases.

Looking at results for di¤erent ranks, the Classical MAI remains systematically worse than the BVAR and the Bayesian 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 in terms of 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 -as one would expect- dependent on the true DGP. However, even with a full rank BVAR DGP, the MAI does reasonably well, in particular when N is large and/or r is small.

6   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 some alternative examples of structural shocks.

6.1   Data and selection of optimal model

We use the "medium" dataset of Banbura, Giannone, Reichlin (2010, BGR), which includes the 20 variables listed in Table 4 and Table A1 in the online Appendix. 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 e¤ects 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 di¤erent values of the shrinkage hyperparameter in the grid $\stackrel{p}{=} 2\{0.01, 0.02, 0.03, 0.04, 0.05, 0.075, 0.1\}$. This provides a total of 455 alternative speci…cations. We estimate all these speci…cations and rank them according to the marginal data density computed as shown in equation (36).

In Table 3 we provide results for the best 20 speci…cations. The …rst three columns contain the rank-lags-shrinkage combination that uniquely identi…es a speci…cation. 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 (37) and is of course insensitive to the rank). For reference, columns 6 and 7 contain the (average) Potential Scale Reduction Factors for the MAI model.[12] Columns 8 and 9 contain the Bayesian Information Criterion (BIC) computed for the MAI and the BVAR.[13] As is clear from the table, the best speci…cation selects 13 lags, a rank of 3, and a shrinkage parameter $\stackrel{p}{=} = 0{:}02$. 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 13 lags and an overall shrinkage of $0{:}1$ (a combination not shown in the table) which produces a MDD of -8956 and a BIC of $81{.}35$.[14]

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 …nancial / monetary factor by imposing restrictions on the matrix B_0, as detailed in Table 4. The resulting factors and their components are plotted in Figure A1 in the online Appendix. Once this set of restrictions is imposed, we compute again the marginal data density and …nd that its value increases from -9444 to -9380, providing support for the restrictions.[15] Table A2 in the online Appendix shows the prior and posterior mean and standard deviation of the elements of the matrix B_0 under this optimal speci…cation, which is the one we use for the structural analysis discussed in the next subsections

6.2 Structural analysis

To illustrate how to conduct empirically structural analysis using the MAI, we …rst 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 representa-

---

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

[13] We report the BIC purely as an overall indication of goodness of …t, since it is a widely used benchmark criterion. However, it is important to stress that since we are varying the prior variance, the penalty in the criterion should be given by the log determinant of the posterior covariance matrix rather than its asymptotic form of number of parameters multiplied by log sample size. The latter penalty does not account for the e¤ect of the prior variance on model complexity: as the prior variance tightens, the goodness-of-in-sample …t is reduced but the penalty is not lowered. Instead, the MDD does take into account this e¤ect (note that the BIC is an approximation of the MDD). Our model selection is based on the MDD.

[14] When comparing the MAI impulse responses with the BVAR impulse responses we use this optimal speci…cation for the BVAR.

[15] We also re-compute the optimal shrinkage and lag-length under this restricted speci…cation and the resulting optimal values for the shrinkage is $\stackrel{p}{=} = 0{:}1$ and we use this value in the empirical application.

tion of the MAI. Then, we assess the e¤ects of demand and supply shocks, modelling the same dataset but with the FAVAR-style MA representation.

6.2.1 Monetary policy shock

In line with the literature, the monetary policy shock is identi…ed with a Cholesky scheme where the federal funds rate is ordered after the slow moving variables and before the fast ones.[16] Formally, the impulse responses are based on the representation:

$$Y_t = fA(L)[I \quad B_0A(L)]^{-1}B_0 + Ig^{-1} \quad _t \tag{40}$$

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

$$_s = A_1B_0 \quad _{s-1} + \ldots + A_{\min(s;p)}B_0 \quad _{s-\min(s;p)}; \quad s > 0 \tag{41}$$

with $_0 = fA(0)[I \quad B_0A(0)]^{-1}B_0 + Ig^{-1} = ^{-1}$.

We simulate the distribution of the impulse responses using 40000 draws[17] and plot the median responses together with the 16th and 84th quantiles in Figure 1. In Figure 1 the Bayesian MAI impulse responses are overlaid with those obtained with a classical estimation of the MAI, and with the responses obtained using the unrestricted BVAR approach of Banbura, Giannone, and Reichlin (2010).

As is clear from the Figure, the impulse responses of the Bayesian MAI model (red solid lines) 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 de‡ator, and earnings. Money and reserves decrease, while the exchange rate appreciates and the reaction of the stock market is close to zero.

Comparing the Bayesian MAI responses to those obtained with maximum likelihood (black dotted lines) 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. This behaviour of the maximum likelihood estimates is in line with the Monte Carlo results presented before, and points towards the conclusion that, notwithstanding the rank reduction, classical estimation still su¤ers from overparameterization.

Comparing the Bayesian MAI responses to those obtained with an unrestricted BVAR (blue dotted lines) it emerges that the MAI model produces more reasonable responses for the real variables. Indeed the BVAR speci…cation of Banbura, Giannone and Reichlin (2010) implies a puzzling reaction for the real variables in the …rst 6 to 12 months, with variables

---

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

[17] The 40000 draws are obtained by running 2 parallel chains of 25000 draws. For each chain we retain 20000 draws and discard the …rst 5000 for burn-in.

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.

6.2.2   Demand and supply shocks

In this subsection we analyze the e¤ects of demand and supply shocks. More precisely, recall that we identi…ed an output factor, a price factor, and a …nancial / monetary factor by imposing restrictions on the matrix B_0, as detailed in Table 4.

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

$$ _s = A_1 \ _{s\ 1} + ::: + A_{min(s;p)} \ _{s\ min(s;p)}; \quad s > 0 \qquad (42)$$

with $\ _0 = f \ B^0_0 \ ^1 + A(0)[I \ B_0 A(0)] \ ^1 g P \ ^1 = \ B^0 P^0_0.$

We simulate the distribution of the impulse responses using 40000 draws and plot the median responses together with the 16th and 84th quantiles. Speci…cally, Figures 2 and 3 show the responses of the 20 macroeconomic variables to a demand and a supply shock, respectively.

The (positive) demand shock is modelled as a shock to the …rst factor. As shown in Figure 2, all the real variables react positively, and the prices also increase. As a consequence, the federal funds rate increases substantially, as well as the ten year rate, with a drop in monetary indicators and in the stock market index and an appreciation of the e¤ective exchange rate. The e¤ects are generally statistically signi…cant.

The (negative) supply shock is modelled as a shock to the second factor. As shown in Figure 3, all the real variables deteriorate, and all the price variables increase. The latter e¤ect is more marked than the former, so that there is an increase in the federal funds rate, though much smaller than in the case of the demand shock. The ten year rate also increases, and there is a drop in the monetary indicators and in the stock market index and a depreciation of the e¤ective exchange rate, followed by an appreciation that starts about one year after the shock. The e¤ects are generally statistically signi…cant, in particular at short horizons in the case of the fast variables.

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.

7   Conclusions

In this paper we addressed the issue of parameter dimensionality reduction in Vector Autoregressive models (VARs) for many variables by using the Multivariate Autoregressive Index (MAI) model of Reinsel (1983), which imposes reduced rank restrictions on the coe¢cient matrices, see also Lütkepohl (2007).

As we are particularly interested in the use of MAI models for structural analysis, we derived alternative Wold representations for them. We focused on a representation that highlights the similarities of MAI and dynamic factor models, a competing approach to model large datasets, but also discussed the di¤erences in the two methods. We then extended the analysis to general reduced rank VARs.

We reviewed classical estimation of the MAI model, extended the asymptotic results to the case of N diverging, and provided the conditional posteriors and an MCMC algorithm for Bayesian estimation of the model.

We assessed the …nite sample performance of the MAI estimation methods in a Monte Carlo experiment. The results show that Bayesian estimation of the MAI model performs much better than 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 was illustrated with empirical applications on the transmission mechanism of monetary policy, and of demand and supply shocks, in a model that includes 20 key macroeconomic variables for the US. The results we obtained are 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 Lütkepohl (2010) and Lanne, Lütkepohl and Maciejowska (2010). Hence, the MAI model is a promising tool for structural analysis using large datasets.

Appendix A: Properties of MLE Estimation when N Diverges

In this appendix we set out a framework for analyzing Maximum Likelihood estimation in the presence of a large dataset modelled through the use of a parametric model. We …rst 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 $T \times N$, $Y = (Y_1; ::::; Y_T)^0$; $Y_t = (y_{1;t}; ::::; y_{N;t})^0$, with density $F(Y; \theta_{N0})$ depending on a vector of parameters, $\theta_{N0} \in \Theta_N \subset R^{k_N N}$ for some sequence of …nite constants k_N. We assume that N is a function of T. Let θ_N be an arbitrary element of Θ_N, and let $L(Y; \theta_N) = L(\theta_N) = F(Y; \theta_N) = \prod_{t=p}^{T} F_t(Y_t | Y_{1:t-1}; \theta_N)$ denote the assumed likelihood function of Y, for some $p > 1$, where $Y_{1:t-1} = (Y_1; ::::; Y_{t-1})^0$. Then, $l(Y; \theta_N) = l(\theta_N) = \log L(\theta_N) = \sum_{t=p}^{T} f_t(Y_t | Y_{1:t-1}; \theta_N)$ denotes the log-likelihood function where $f_t(Y_t | Y_{1:t-1}; \theta_N) := f_t(Y_t; \theta_N) := \log F_t(Y_t | Y_{1:t-1}; \theta_N)$. We also de…ne the $k_N \times 1$ score vector by

$$z(Y; \theta_N) = z(\theta_N) := \partial l / \partial \theta_N := \sum_{t=p}^{T} z_t(\theta_N);$$

21

and the Hessian

$$H(Y;\theta_N) = H(\theta_N) := \frac{1}{T}\sum_{t=1}^{T} H_t(\theta_N) := \frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 f_t(Y_t;\theta_N)}{\partial\theta_N \partial\theta_N'}.$$

When $l(\theta_N)$ is di¤erentiable, the MLE $\hat{\theta}_N$ satis…es

$$z\left(Y;\hat{\theta}_N\right) = 0. \tag{43}$$

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

Assumption 1 (RC1) The support of F, $S = \{Y \in R^{T \times N} : F(Y;\theta_N) > 0\}$, is independent of θ_N; (RC2) $F(Y;\theta_N)$ is twice continuously di¤erentiable with respect to θ_N; (RC3) The matrix $J(\theta_N) := E\left[\frac{\partial^2 \log F(Y;\theta_N)}{\partial\theta_N \partial\theta_N'}\right]$ has …nite elements and is negative de…nite at $\theta_N = \theta_{N0}$.

We …rst consider consistency of MLE estimation. We make the additional assumption:

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

$$\psi(\theta_N) := El(\theta_{N0}) - El(\theta_N) > 0 \tag{44}$$

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

$$\max_{\theta_N \in \Theta_N} \frac{1}{T}\sum_{t=p}^{T}[f_t(Y_t,Y_{1:t-1};\theta_N) - Ef_t(Y_t,Y_{1:t-1};\theta_{N0})] \to_p 0 \text{ as } T \to \infty. \tag{45}$$

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

$$\hat{\theta}_N - \theta_{N0} \to_p 0 \text{ as } T \to \infty.$$

Proof. For arbitrary $\varepsilon > 0$, consider an open neighborhood around θ_{N0} of radius ε:

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

Both $N(\varepsilon)$ and its complement

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

are subsets of Θ_N which, in turn, is a subset of $R^{k_N N}$. It is easy to see that $N(\varepsilon)$ is an open set, so $\overline{N(\varepsilon)}$ is a closed set. $\overline{N(\varepsilon)}$ is also bounded, since it is a subset of the bounded set Θ_N (see (C2)). We conclude that $\overline{N(\varepsilon)}$ is a closed and bounded subset of R^{k_N}, so $\overline{N(\varepsilon)}$ is compact by the Heine-Borel theorem. Since, by (C1), $El(Y;\theta_N)$ is a continuous function

w.r.t. θ, the maximizer of $E\ell(Y;\theta)$ over $\overline{N(\epsilon)}$ belongs to $N(\epsilon)$: denoting this maximizer by θ_*, we conclude that there exists $\theta_* \in N(\epsilon)$ satisfying

$$E\ell(Y;\theta_*) \geq E\ell(Y;\theta) \quad \text{for all } \theta \in \overline{N(\epsilon)}. \tag{46}$$

We next note that (45) is equivalent to the event

$$A(\epsilon) := \left\{ \omega : \max_{\theta \in \partial N} \left| \frac{1}{T}\ell(Y(\omega);\theta) - E[\ell(Y(\omega);\theta)] \right| < \frac{\epsilon}{2} \right\}$$

occurring with probability tending to 1 for arbitrary $\epsilon > 0$, i.e., $\lim_{T \to \infty} P(A(\epsilon)) = 1$. If we can show the inequality

$$P(A(\epsilon)) \leq P\left\{ \omega : \hat\theta_N(\omega) \in N(\epsilon) \right\}; \tag{47}$$

for arbitrary $\epsilon > 0$ and some $\epsilon > 0$, consistency of $\hat\theta_N$ will follow immediately since

$$P\left\{ \omega : \hat\theta_N(\omega) \in N(\epsilon) \right\} = P\left\{ \omega : \left|\hat\theta_N(\omega) - \theta_0\right| < \epsilon \right\};$$

and the right hand side tends to 1 as $T \to \infty$ as $P(A(\epsilon)) \to 1$ for arbitrary $\epsilon > 0$. It remains to show (47). Using the identity $|x| < r \Leftrightarrow -r < x < r$, we obtain for all $\theta \in \partial N$:

$$\omega \in A(\epsilon) \Longrightarrow \begin{cases} E\left[\frac{1}{T}\ell(Y(\omega);\theta)\right] - \frac{\epsilon}{2} < \frac{1}{T}\ell(Y(\omega);\theta) \\ \frac{1}{T}\ell(Y(\omega);\theta) < E\left[\frac{1}{T}\ell(Y(\omega);\theta)\right] + \frac{\epsilon}{2}. \end{cases} \tag{48}$$

Since $\theta_0 \in \partial N$ and $\hat\theta_N \in \partial N$ (by compactness of ∂N and continuity of the log-likelihood) (48) will apply for $\theta = \theta_0$ for the top inequality and $\theta = \hat\theta_N$ for the bottom inequality. Since $\ell(Y(\omega);\theta) \leq \ell\left(Y(\omega);\hat\theta_N\right)$ by definition of the MLE, (48) implies the chain of inequalities

$$E\left[\frac{1}{T}\ell(Y(\omega);\theta_0)\right] - \frac{\epsilon}{2} < \frac{1}{T}\ell(Y(\omega);\theta_0)$$

$$\frac{1}{T}\ell\left(Y(\omega);\hat\theta_N\right) < E\left[\frac{1}{T}\ell\left(Y(\omega);\hat\theta_N\right)\right] + \frac{\epsilon}{2}.$$

We conclude that

$$\omega \in A(\epsilon) \Longrightarrow E\left[\frac{1}{T}\ell(Y(\omega);\theta_0)\right] < E\left[\frac{1}{T}\ell\left(Y(\omega);\hat\theta_N\right)\right] + \epsilon \tag{49}$$

for arbitrary $\epsilon > 0$. Since (49) holds for arbitrary $\epsilon > 0$, we may choose

$$\epsilon = \epsilon(\theta_N) = E\left[\frac{1}{T}\ell(Y(\omega);\theta_0)\right] - E\left[\frac{1}{T}\ell(Y(\omega);\theta_N)\right]$$

23

in the notation of (44). This choice is possible since $\eta_N \geq 2 \overline{N(\epsilon)} \geq \eta f_0 g$ so positivity of (η_N) is guaranteed by (C2). Imposing the choice of $\varepsilon = \varepsilon(\eta_N)$ in (49), we obtain

$$\{ ! \ge 2A(\varepsilon(\eta_N))\} \supset \left\{ E \frac{1}{T}l(Y(!); \eta_N) < E \frac{1}{T}l \ Y(!); \hat{\theta}_N \right\}$$

$$\supset \left\{ \hat{\theta}_N \ge \eta_N \setminus \overline{N(\epsilon)} \right\} \text{ by (46)}.$$

Therefore $\{ ! \ge 2A(\varepsilon(\eta_N))\} \supset \{ \hat{\theta}_N(!) \ge N(\epsilon) \}$, i.e. $P[A(\varepsilon(\eta_N))] \ge P\left[! : \hat{\theta}_N(!) \ge N(\epsilon) \right]$, establishing (47). ∎

Note that in this Theorem the convergence is in terms of T but N can also tend to in…nity as a function of T, as long as the conditions of the Theorem are satis…ed as well. We next consider the rates of convergence for the MLE estimator. We de…ne the Frobenius norm of a matrix A as $\|A\|_F$. Then we have the following Theorem.

Theorem 3 (Rates) Under (C1)-(C4), if (T1) $E[H(\theta_{N0})]$ is invertible, (T2) $\hat{\theta}_N$ is consistent, (T3) $y_{i;t}$ are strong mixing processes with mixing coe¢cients $\alpha_{i;m}$, such that

$$\sup_i \sum_{m=1}^{1} \alpha_{i;m}^{\delta=(2+\delta)} < 1;$$

for some $\delta > 0$, and

$$(T4) \quad \sup_i Var\left(\frac{1}{T} \sum_{t=1}^{T} y_{i;t} \ E(y_{i;t}) \right) = O\ T^{-1};$$

then

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

Proof. Let θ_N be a point intermediate to $\hat{\theta}_N$ and θ_{N0}, i.e.,

$$\left\| \theta_N \ \theta_{N0} \right\|_F \ \left\| \hat{\theta}_N \ \theta_{N0} \right\|_F :$$

Then, by a mean value expansion, we have

$$\left\| \hat{\theta}_N \ \theta_{N0} \right\|_F = \left\| \left[\frac{1}{T} \sum_{t=1}^{T} H_t(\theta_N) \right]^{-1} \frac{1}{T} \sum_{t=1}^{T} z_t(\theta_{N0}) \right\|_F =$$

$$\left\| \left[\frac{1}{T} \sum_{t=1}^{T} H_t(\theta_N) \right]^{-1} \frac{1}{T} \sum_{t=1}^{T} z_t(\theta_{N0}) \ \frac{1}{T} \sum_{t=1}^{T} E[H_t(\theta_{N0})]^{-1} \frac{1}{T} \sum_{t=1}^{T} z_t(\theta_{N0}) + \frac{1}{T} \sum_{t=1}^{T} E[H_t(\theta_{N0})]^{-1} \frac{1}{T} \sum_{t=1}^{T} z_t(\theta_{N0}) \right\|_F$$

24

$$\left\| \left[\frac{1}{T}\sum_{t=1}^{T} H_t(\hat\theta_N) \right]^{-1} \frac{1}{T}\sum_{t=1}^{T} z_t(\theta_{N0}) - \left[\frac{1}{T}\sum_{t=1}^{T} E[H_t(\theta_{N0})] \right]^{-1} \frac{1}{T}\sum_{t=1}^{T} z_t(\theta_{N0}) \right\|_F$$

$$+ \left\| \left[\frac{1}{T}\sum_{t=1}^{T} E[H_t(\theta_{N0})] \right]^{-1} \frac{1}{T}\sum_{t=1}^{T} z_t(\theta_{N0}) \right\|_F$$

We have

$$\left\| \left[\frac{1}{T}\sum_{t=1}^{T} H_t(\hat\theta_N) \right]^{-1} \frac{1}{T}\sum_{t=1}^{T} z_t(\theta_{N0}) - \left[\frac{1}{T}\sum_{t=1}^{T} E[H_t(\theta_{N0})] \right]^{-1} \frac{1}{T}\sum_{t=1}^{T} z_t(\theta_{N0}) \right\|_F =$$

$$\left\| \left\{ \left[\frac{1}{T}\sum_{t=1}^{T} H_t(\hat\theta_N) \right]^{-1} - \left[\frac{1}{T}\sum_{t=1}^{T} E[H_t(\theta_{N0})] \right]^{-1} \right\} \frac{1}{T}\sum_{t=1}^{T} z_t(\theta_{N0}) \right\|_F$$

$$\times \left\| \left[\frac{1}{T}\sum_{t=1}^{T} E[H_t(\theta_{N0})] \right]^{-1} \frac{1}{T}\sum_{t=1}^{T} z_t(\theta_{N0}) \right\|_F$$

$$\frac{\left\| \left[\frac{1}{T}\sum_{t=1}^{T} E[H_t(\theta_{N0})] \right]^{-1} \right\|_F^2 \left\| \frac{1}{T}\sum_{t=1}^{T} H_t(\hat\theta_N) - \frac{1}{T}\sum_{t=1}^{T} E[H_t(\theta_{N0})] \right\|_F}{1 - \left\| \left[\frac{1}{T}\sum_{t=1}^{T} E[H_t(\theta_{N0})] \right]^{-1} \right\|_F \left\| \frac{1}{T}\sum_{t=1}^{T} H_t(\hat\theta_N) - \frac{1}{T}\sum_{t=1}^{T} E[H_t(\theta_{N0})] \right\|_F} \left\| \frac{1}{T}\sum_{t=1}^{T} z_t(\theta_{N0}) \right\|_F$$

where the last inequality follows by, e.g., (2.15) of Berk (1974). We examine $\left\| \left[\frac{1}{T}\sum_{t=1}^{T} E[H_t(\theta_{N0})] \right]^{-1} \right\|_F$,
$\left\| \frac{1}{T}\sum_{t=1}^{T} H_t(\hat\theta_N) - \frac{1}{T}\sum_{t=1}^{T} E[H_t(\theta_{N0})] \right\|_F$ and $\left\| \frac{1}{T}\sum_{t=1}^{T} z_t(\theta_{N0}) \right\|_F$. By symmetry of $\frac{1}{T}\sum_{t=1}^{T} E[H_t(\theta_{N0})]$,
we have that

$$\left\| \left[\frac{1}{T}\sum_{t=1}^{T} E[H_t(\theta_{N0})] \right]^{-1} \right\|_F = \sqrt{\sum_{i=1}^{N} \lambda_{\frac{1}{T}\sum_{t=1}^{T} E[H_t(\theta_{N0})];i}^2} = O\left(\sqrt{N}\right) \tag{50}$$

where $\lambda_{A;i}$ denotes the i-th eigenvalue of A, in order of magnitude in absolute value, as long as all eigenvalues of $E[H(\theta_{N0})]$ are bounded away from zero, by the assumption of nonsingularity of $E[H(\theta_{N0})]$. By consistency of $\hat\theta_N$, and twice differentiability of f_t it follows that every element of $H(\hat\theta_N)$ converges, at rate $T^{-1/2}$, to the respective element of $E[H(\theta_{N0})]$ uniformly over all elements. Therefore,[18]

$$\left\| \frac{1}{T}\sum_{t=1}^{T} H_t(\hat\theta_N) - \frac{1}{T}\sum_{t=1}^{T} E[H_t(\theta_{N0})] \right\|_F = O_p\left(\frac{N}{T^{1/2}}\right) \tag{51}$$

---

[18] Note that in (50), which only involves a population quantity, we can use our assumptions to get a sharper rate result than the one we use in (51), which involves the convergence of a sample quantity to its population counterpart.

Further, by (T2), (T4), (C1)-(C4) and the fact that functions of mixing processes are mixing (see, e.g., Theorem 14.1 of Davidson (1994)) we have that a central limit theorem (such as, e.g., Theorem 18.5.3 of Ibragimov and Linnik (1971)) implies that all elements of $\frac{1}{T}\sum_{t=1}^{T} z_t(\theta_{N0})$ converge to zero at rate $T^{1=2}$ and so,

$$\frac{1}{T}\sum_{t=1}^{T} z_t(\theta_{N0})\bigg|_F = O_p\left(\frac{N^{1=2}}{T^{1=2}}\right) \tag{52}$$

Overall, using (51), (52) and (50) and noting that, by (51) and (T1),

$$\frac{1}{\frac{1}{T}\sum_{t=1}^{T} E[H_t(\theta_{N0})]}\bigg|_F \frac{1}{\frac{1}{T}\sum_{t=1}^{T} H_t(\theta_N)} \frac{1}{\frac{1}{T}\sum_{t=1}^{T} E[H_t(\theta_{N0})]}\bigg|_F = O_p(1);$$

we have

$$\left|\left|\frac{1}{T}\sum_{t=1}^{T} E[H_t(\theta_{N0})]\left[\frac{1}{T}\sum_{t=1}^{T} H_t(\theta_N)\right]^{-1}\frac{1}{T}\sum_{t=1}^{T} E[H_t(\theta_{N0})]\right|\right| \geq \left|\left|\frac{1}{T}\sum_{t=1}^{T} E[H_t(\theta_{N0})]\right|\right| \geq$$

$$\frac{1}{T}\sum_{t=1}^{T} z_t(\theta_{N0}) = O_p\left(\frac{N^{5=2}}{T}\right)$$

and

$$\left|\left|\frac{1}{T}\sum_{t=1}^{T} E[H_t(\theta_{N0})]\right|\right|^{-1}\frac{1}{T}\sum_{t=1}^{T} z_t(\theta_{N0}) = O_p\left(\frac{N}{T^{1=2}}\right)$$

giving

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

∎

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 (22). For ease of reference, we recall the general VAR model

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

and its MAI specialization

$$Y_t = A(L)B_0 Y_t + \epsilon_t = \sum_{u=1}^{p} A_u B_0 Y_{t-u} + \epsilon_t; \tag{53}$$

We will provide results for (53) while using the relation $\Phi(L) = A(L)B_0$ and noting that

26

straightforward extensions would make (17) 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 $_N = (vec(A)^0; vec(B)^0)^0$ where $A = (A_1; :::; A_p)$.

We make the following assumption:

Assumption 3 (i) All roots of (L) are bounded away from the unit circle uniformly over N. (ii) $_t$ is an i.i.d. sequence, which has a continuous, twice di¤erentiable and bounded probability density function and …nite $2 +$ moments for some > 0 and (iii) $E[H(_{N0})]$ 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:

$$\sup_i Var\left(\frac{1}{T}\sum_{t=1}^{T} y_{i;t} \quad E(y_{i;t})\right) = O\left(T^{-1}\right) \tag{54}$$

and

$$\max_{_N 2 _N} \frac{1}{T}\sum_{t=p}^{T} f_t(Y_t \mathbf{j} Y_{1:t-1}; _N) \quad E f_t(Y_t \mathbf{j} Y_{1:t-1}; _N) \mathbf{!}_p \; 0 \; as \; T \; \mathbf{!} \; 1: \tag{55}$$

Proof. To show (54) we use Theorem 18.5.3 of Ibragimov and Linnik (1971). Then, it is su¢cient to prove that

$$\sup_i Ey_{i;t}^{2+} < 1; \; for \; some \; > 0; \tag{56}$$

and

$$\sup_i \sum_{m=1}^{1} {}_{i;m}^{=(2+)} < 1; \tag{57}$$

where $_{i;m}$ are the strong mixing coe¢cients of $y_{i;t}$. By assumption, the eigenvalues of the companion form matrix obtained from (L) are bounded away from 1 in absolute value. This implies that $_{i;m} = {}_i^m$, $_i > 0$, where $\sup_i \; _i < 1$, which implies (57) for all > 0. Further, the above eigenvalue assumption implies that $\sup_i Ey_{i;t}^2 < 1$, which implies that $\sup_i \sum_{m=1}^{1} {}_{i;m}^2 < 1$, where $_{i;m}$ are the coe¢cients of the univariate MA representation of $y_{i;t}$. This, coupled with the Marcinkiewicz–Zygmund inequality, gives

$$\sup_i Ey_{i;t}^{2+} \quad \sup_i \left(\sum_{m=1}^{1} {}_{i;m}^2\right)^{1+ =2} E {}_{i;t}^{2+} < 1; \tag{58}$$

proving (56). Using the above we now derive the properties of $f_t(Y_i; _N)$. By the above analysis we have that every element of $Y_t \ AZ_{t\_1}$ has ...nite variance for all values of A that satisfy Assumption 3 (i), where $Z_{t\_1} = (Y_{t\_1}^{0\prime} B^0 :::; Y^{0\prime} B^0)^0$. Then, for all $_N 2 _N$,

$$E\left[(f_t(Y_t; Z_{t\_1}; _N))^2\right] \quad kAk_F^2 \, kB_0 k_F^2 \sup_i E y_{i;t}^{0;\ t\ p\ 0;}{}^2 = O(N^2)$$

Then,

$$\frac{1}{T}\sum_{t=p}^{T} f_t(Y_t j Y_{1:t\_1}; _{N0}) \quad Ef_t(Y_t j Y_{1:t\_1}; _{N0}) = O_p\left[\quad NT^{1=2}\right] = o_p(1) \tag{59}$$

To prove (55) we use (59) and note that

$$\sup_{_N2_N} E \ N^2 @I=@_N^2 \quad N^2 \sup_{_N2_N} kAk_F^2 \sup_{_N2_N} kB_0 k_F^2 \sup_i E y_{i;t}^2 = O(1)$$

and so

$$\sup_{_N2_N} N^2 @I=@_N{}_F = O_p(1)$$

Then, stochastic equicontinuity follows by Theorem (21.10) and (21.57) of Davidson (1994) proving (55).

Note that the condition of Lemma 6, in terms of the relative rates of N and T, given by $N = o T^{1=2}$, is milder than that of Theorem 1, as it relates to consistency, but the conditions of Theorem 1 also relate to the rate of convergence of the estimator.

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 ...rst 5000 draws for burn-in.

We assess convergence by looking at the Ine¢ciency Factor (IF) and the Potential Scale Reduction Factor (PSRF). The IF are related to the autocorrelation functions and measure how e¢cient 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 on the 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 A3 in the online Appendix and show that the algorithm is e¢cient and reaches convergence.

References

[1] Anderson, T. W., 1951. "Estimating Linear Restrictions on Regression Coe¢cients for Multivariate Normal Distributions," Annals of Mathematical Statistics 22, 327-351.

[2] Bai, J. and Ng, S., 2006a. "Con…dence Intervals for Di¤usion 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] Bauwens, L., Lubrano, M. and Richard, J.-F., 2000. "Bayesian Inference in Dynamic Econometric Models," Oxford: Oxford University Press,

[7] Berk, K. N., 1974, 'Consistent Autoregressive Spectral Estimates', The Annals of Statistics, 2(3), 489-502.

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

[9] 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.

[10] 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.

[11] Carriero, C., Clark, T., and Marcellino, M., 2015. "Bayesian VARs: speci…cation choices and forecast accuracy", Journal of Applied Econometrics, 30(1), 46-73, 2015.

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

[13] Carriero, C., Kapetanios, G. and Marcellino, M., 2015. "Structural Analysis with Multivariate Autoregressive Index Models", CEPR WP 10801.

[14] Centoni, M., and Cubadda, G., 2003. "Measuring the business cycle e¤ects of permanent and transitory shocks in cointegrated time series", Economics Letters 80, 45-51.

[15] 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.

[16] Davidson, J., 1994. Stochastic Limit Theory, Oxford University Press.

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

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

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

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

[21] Dufour, J.M. and Stevanovic, D., 2013. "Factor-Augmented VARMA Models With Macroeconomic Applications," Journal of Business & Economic Statistics, 31(4), 491-506,

[22] Eickmeier, S., W. Lemke, M. Marcellino, 2014. "Classical time-varying FAVAR models - estimation, forecasting and structural analysis", Journal of the Royal Statistical Society, forthcoming.

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

[24] Forni, M. and Gambetti, L. 2010. "The dynamic e¤ects of monetary policy: A structural factor model approach", Journal of Monetary Economics, 57, 203-216.

[25] Fuentes-Albero, C. and Melosi, L., 2013, "Methods for computing marginal data densities from the Gibbs output," Journal of Econometrics 175(2), 132-141.

[26] Gelfand, A. E. and Dey, D. K., 1994. "Bayesian model choice: asymptotics and exact calculations", Journal of the Royal Statistical Society, B 56, 501-514.

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

[28] George, E.I., Sun, D. and Ni, S., 2008. "Bayesian stochastic search for VAR model restrictions", Journal of Econometrics, 142, 553–580.

[29] 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).

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

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

[32] Hamilton, J.D., Waggoner, D.F. and Zha, T. 2007. "Normalization in Econometrics", Econometric Reviews, 26 pp.221–252.

[33] Kadiyala, K.R., Karlsson, S., 1997. "Numerical methods for estimation and inference in Bayesian VAR-models". Journal of Applied Econometrics 12(2), 99–132.

[34] Ibragimov, I. A. and Linnik, Y. V., 1971. "Independent and stationary sequences of random variables ", Wolters-Noordho¤.

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

[36] Kleibergen, F., van Dijk H.K., 1994. "On the Shape of the Likelihood/Posterior in Cointegration Models", Econometric Theory 10, 514-551

[37] Kleibergen, F., van Dijk H.K., 1998. "Bayesian Simultaneous Equations Analysis Using Reduced Rank Structures", Econometric Theory 14, 701-743.

[38] Koop, G.M., 2013. "Forecasting with Medium and Large Bayesian VARS," Journal of Applied Econometrics, 28(2), 177-203.

[39] 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.

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

[41] Lanne, M. and Lütkepohl, H., 2008. "Identifying Monetary Policy Shocks via Changes in Volatility", Journal of Money, Credit and Banking, 40, 1131-1149.

[42] Lanne, M. and Lütkepohl, H., 2010. "Structural Vector Autoregressions with Nonnormal Residuals", Journal of Business & Economic Statistics, 28, 159-168.

[43] Lanne, M., Lütkepohl, H. and Maciejowska K., 2010. "Structural Vector Autoregressions with Markov Switching, Journal of Economic Dynamics and Control, 34, 121-131.

[44] Leeper, E.M., Sims, C.A. and Zha, T., 1996. "What Does Monetary Policy Do?," Brookings Papers on Economic Activity, 27(2), 1-78.

[45] Litterman R. 1986. Forecasting with Bayesian Vector Autoregressions-Five Years of Experience, Journal of Business and Economic Statistics 4, 25-38.

[46] Lütkepohl, H., 1984. "Linear Transformations of Vector ARMA Processes", Journal of Econometrics, 26, 283-293.

[47] Lütkepohl, 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.

[48] Lütkepohl, H., 2007. New Introduction to Multiple Time Series Analysis, Oxford University Press.

[49] Lütkepohl, H. 2014. "Structural Vector Autoregressive Analysis in a Data Rich Environment: A Survey", SFB 649 Discussion Paper 2014-004.

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

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

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

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

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

[55] 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.

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

[57] Zellner A. 1973. An Introduction to Bayesian Inference in Econometrics. Wiley: New York.

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 average RMSE and average MAE (averages computed over all the estimated coefficients) 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 average RMSE and average MAE (averages computed over all the estimated coefficients) relative to the BVAR (i.e. ratios). The BVAR entries are the RMSE and MAE (levels).

Table 3. 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 4. 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 Echange 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.
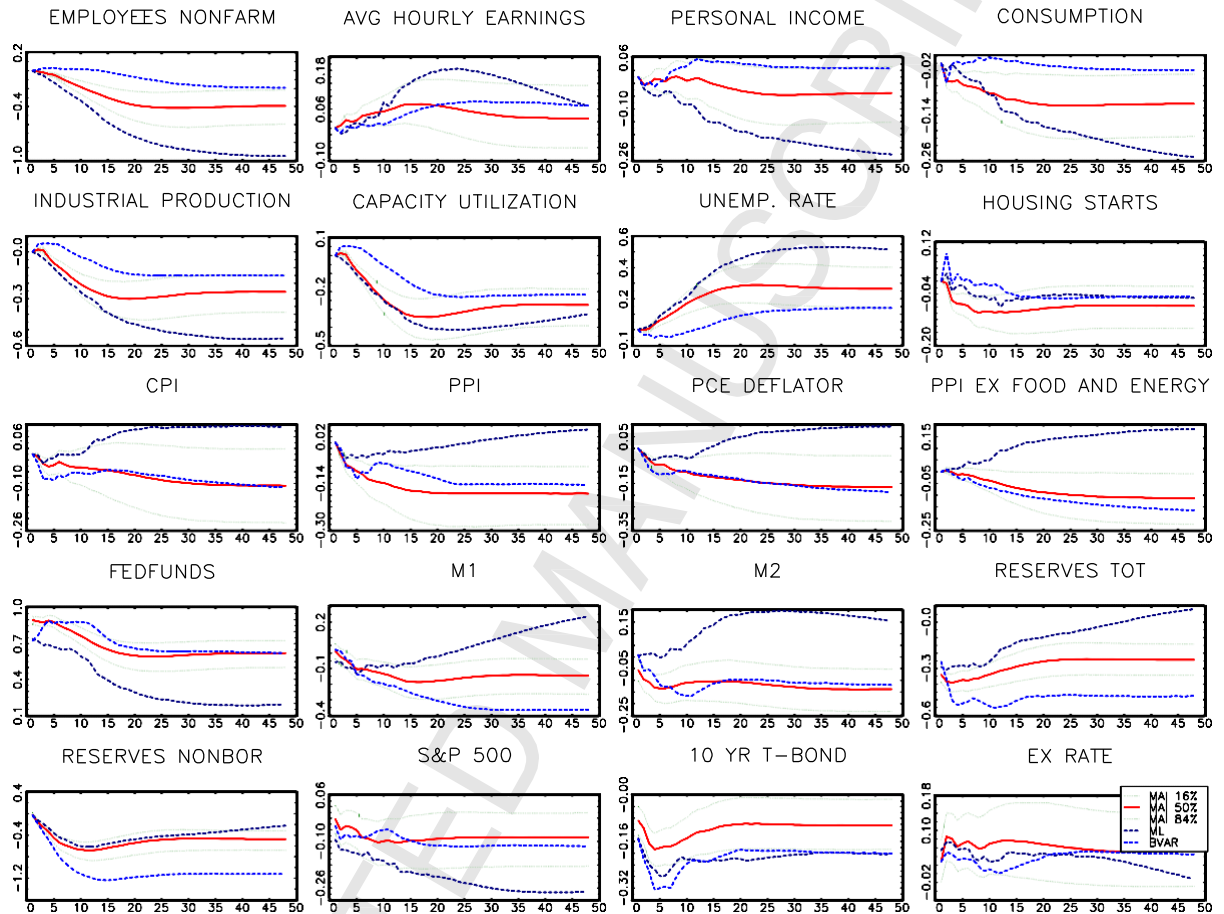
Figure 1: Impulse responses of alternative models. Responses to a permanent shock to the Federal Funds rate. Red solid line and green dashed lines are the median and 16%-84% quantiles of the Bayesian MAI impulse responses. The dotted black line represents the responses from the MAI computed using maximum likelihood estimation. The blue dotted line represents the responses computed using the unrestricted BVAR.
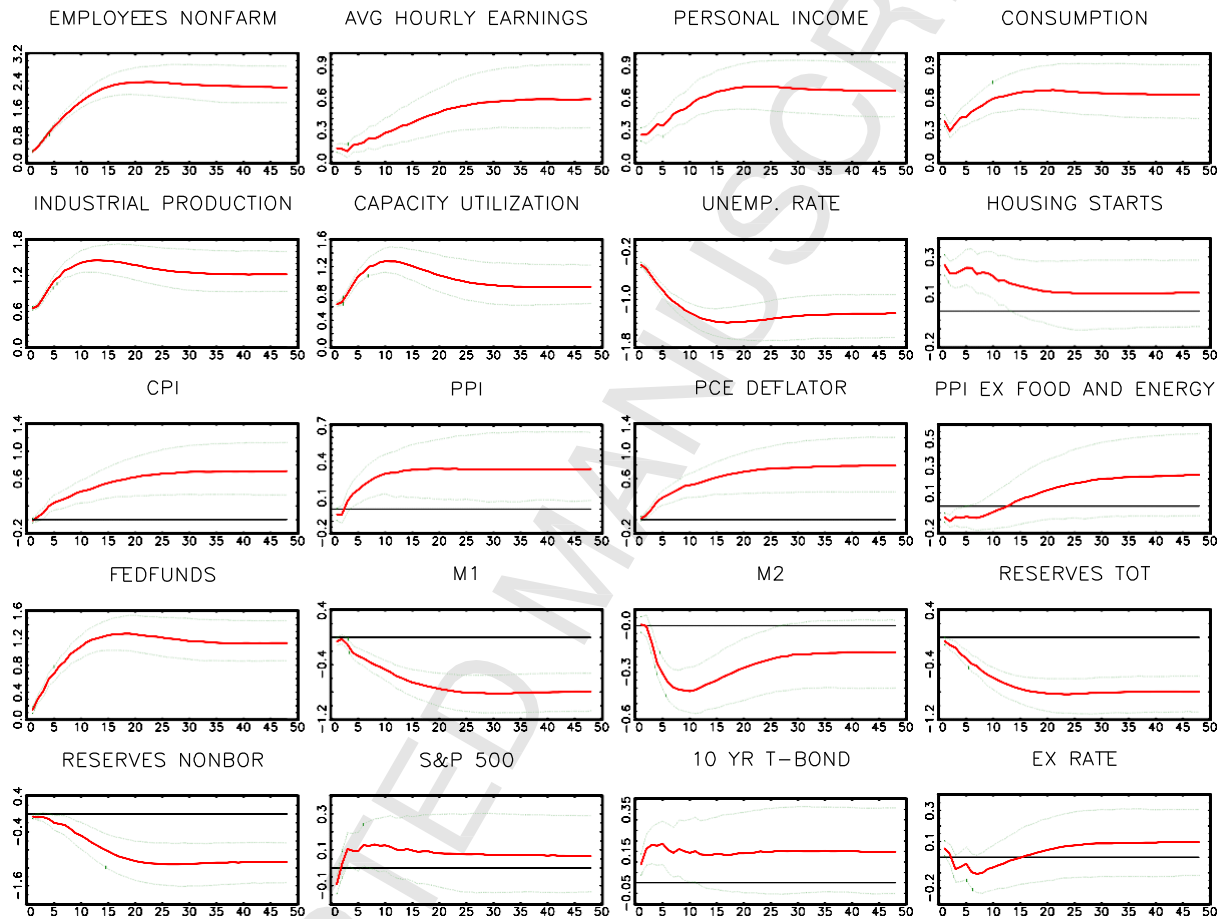
Figure 2: Demand Shock. Responses to a permanent shock to factor 1. Red solid line and green dashed lines are the median and 16%-84% quantiles of the Bayesian MAI impulse responses.
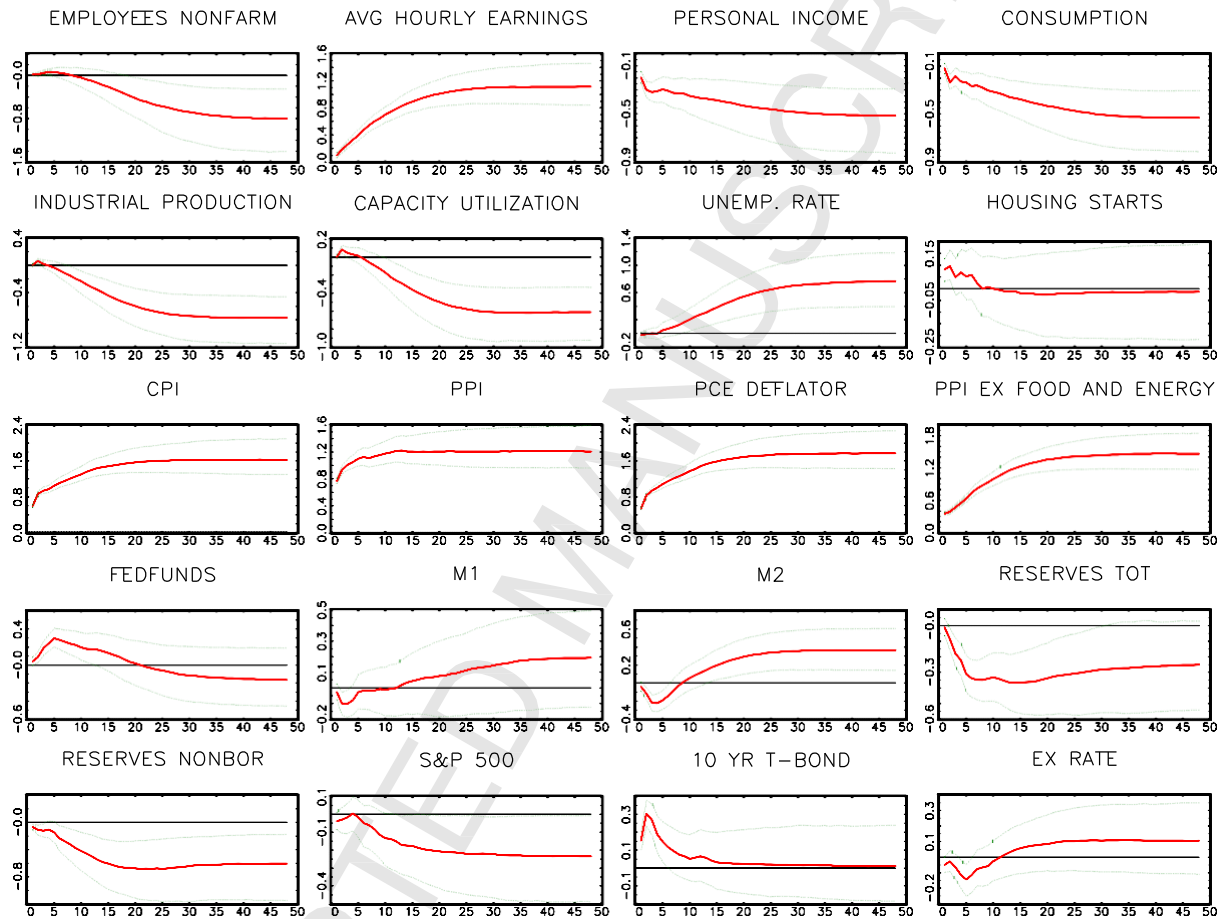
Figure 3: Supply shock. Responses to a permanent shock to factor 2. Red solid line and green dashed lines are the median and 16%-84% quantiles of the Bayesian MAI impulse responses.

CAPTIONS TO TABLES
-for the corresponding tables, open the xls file.

Table 1
For the Bayesian and Classical MAI the entries show the average RMSE and average MAE (averages computed over all the estimated coefficients) relative to the BVAR (i.e. ratios). The BVAR entries are the RMSE and MAE (levels).

Table 2
For the Bayesian and Classical MAI the entries show the average RMSE and average MAE (averages computed over all the estimated coefficients) relative to the BVAR (i.e. ratios). The BVAR entries are the RMSE and MAE (levels).

Table 3
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 4
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 $j=1,\ldots,3$ and the index i runs through different variables $i=1,\ldots,N$.