Journal of Business & Economic Statistics
Publication details, including instructions for authors and subscription information:
http://amstat.tandfonline.com/loi/ubes20

Factor-Augmented VARMA Models with Macroeconomic Applications

Jean-Marie Dufour [a] & Dalibor Stevanović [b]

[a] McGill University

[b] Université du Québec à Montréal

Accepted author version posted online: 25 Aug 2013.

Factor-Augmented VARMA Models with Macroeconomic Applications *

Jean-Marie Dufour †
McGill University

Dalibor Stevanović ‡
Université du Québec à Montréal

First version: October 2008
Revised: December 2008, April 2013, May 2013
This version: May 2013

ABSTRACT

We study the relationship between VARMA and factor representations of a vector stochastic process. We observe that, in general, vector time series and factors cannot both follow finite-order VAR models. Instead, a VAR factor dynamics induces a VARMA process, while a VAR process entails VARMA factors. We propose to combine factor and VARMA modeling by using factor-augmented VARMA (FAVARMA) models. This approach is applied to forecasting key macroeconomic aggregates using large U.S. and Canadian monthly panels. The results show that FAVARMA models yield substantial improvements over standard factor models, including precise representations of the effect and transmission of monetary policy.

Key words: factor analysis, VARMA process, forecasting, structural analysis.
Journal of Economic Literature classification: C32, C51, C52, C53.

---

*The authors thank Jean Boivin, Carsten Trenkler, two anonymous referees, an Associate Editor, and the Editor Jonathan Wright, for several useful comments. Special thanks go to Denis Pelletier for sharing Matlab codes. This work was supported by the William Dow Chair in Political Economy (McGill University), the Bank of Canada (Research Fellowship), the Toulouse School of Economics (Pierre-de-Fermat Chair of excellence), the Universitad Carlos III de Madrid (Banco Santander de Madrid Chair of excellence), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds de recherche sur la société et la culture (Québec).

† William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 4400 ext. 09156; FAX: (1) 514 398 4800; e-mail: jean-marie.dufour@mcgill.ca . Web page: http://www.jeanmariedufour.com

‡Département des sciences économiques, Université du Québec à Montréal. Mailing address: Université du Québec à Montréal, Département des sciences économiques, 315, rue Ste-Catherine Est, Montréal, QC, H2X 3X2. e-mail: dstevanovic.econ@gmail.com

SUMMARY

We study the relationship between VARMA and factor representations of a vector stochastic process, and we propose to use factor-augmented VARMA (FAVARMA) models as an alternative to usual VAR models. We start by observing that vector time series and the associated factors do not both follow a finite-order VAR process, except in very special cases. When factors are defined as linear combinations of observable series, the observable series follows a VARMA process, not a finite-order VAR as typically assumed. Second, even if the factors follow a finite-order VAR model, this entails a VARMA representation for the observable series. In view of these observations, we propose to use a FAVARMA framework which combines two dimension reduction techniques in order to represent the dynamic interactions between a large number of time series: factor analysis and VARMA modeling. We apply this approach in two out-of-sample forecasting exercises using large U.S. and Canadian monthly panels. The results show that VARMA factors provide better forecasts for several key macroeconomic aggregates relative to standard factor models. Finally, we estimate the effect of monetary policy using the data and the identification scheme of Bernanke, Boivin and Eliasz (2005). We find that impulse responses from a parsimonious 6-factor FAVARMA(2,1) model give an accurate and plausible picture of the effect and transmission of monetary policy in the U.S. To get similar responses from a standard FAVAR model, the Akaike information criterion leads to a lag order of 14. The FAVARMA model requires the estimation of 84 coefficients in order to represent the system dynamics, while the corresponding FAVAR model includes 510 VAR parameters.

1. INTRODUCTION

As information technology improves, the availability of economic and financial time series grows in terms of both time and cross-section size. However, a large amount of information can lead to a dimensionality problem when standard time series tools are used. Since most of these series are correlated, at least within some categories, their co-variability and information content can be approximated by a smaller number of variables. A popular way to address this issue is to use "large dimensional approximate factor analysis", an extension of classical factor analysis which allows for limited cross-section and time correlations among idiosyncratic components.

While factor models were introduced in macroeconomics and finance by Sargent and Sims (1977), Geweke (1977), and Chamberlain and Rothschild (1983), the literature on the large factor models starts with Forni, Hallin, Lippi and Reichlin (2000) and Stock and Watson (2002). Further theoretical advances were made, among others, by Bai and Ng (2002), Bai (2003), and Forni, Hallin, Lippi and Reichlin (2004). These models can be used to forecast macroeconomic aggregates [Stock and Watson (2002b), Forni, Hallin, Lippi and Reichlin (2005), Banerjee, Marcellino and Masten (2006)], structural macroeconomic analysis [Bernanke et al. (2005), Favero, Marcellino and Neglia (2005)], for now-casting and economic monitoring [Giannone, Reichlin and Small (2008), Aruoba, Diebold and Scotti (2009)], to deal with weak instruments [Bai and Ng (2010), Kapetanios and Marcellino (2010)], and the estimation of dynamic stochastic general equilibrium models [Boivin and Giannoni (2006)].

Vector autoregressive moving-average (VARMA) models provide another way to obtain a parsimonious representation of a vector stochastic process. VARMA models are especially appropriate in forecasting, since they can represent the dynamic relations between time series while keeping the number of parameters low; see Lütkepohl (1987) and

ACCEPTED MANUSCRIPT

Boudjellaba, Dufour and Roy (1992). Further, VARMA structures emerge as reduced-form representations of structural models in macroeconomics. For instance, the linear solution of a standard dynamic stochastic general equilibrium model generally implies a VARMA representation on the observable endogenous variables [Ravenna (2006), Komunjer and Ng (2011), and Poskitt (2011)].

In this paper, we study the relationship between VARMA and factor representations of a vector stochastic process, and we propose a new class of factor-augmented VARMA models. We start by observing that, in general, multivariate time series and the associated factors do not typically both follow finite-order VAR processes. When the factors are obtained as linear combinations of observable series, the dynamic process obeys a VARMA model, not a finite-order VAR as usually assumed in the literature. Further, if the latent factors follow a finite-order VAR process, this implies a VARMA representation for the observable series. Consequently, we propose to combine two techniques for representing in a parsimonious way the dynamic interactions between a huge number of time series: dynamic factor reduction and VARMA modelling. Thus lead us to consider factor-augmented VARMA (FAVARMA) models.Besides parsimony, the class of VARMA models is closed under marginalization and linear transformations (in contrast with VAR processes). This represents an additional advantage if the number of factors is underestimated.

The importance of the factor process specification depends on the technique used to estimate the factor model and the research goal. In the two-step method developed by Stock and Watson (2002), the factor process does not matter for the approximation of factors, but this might be an issue if we use a likelihood-based technique which relies on a completely specified process. Moreover, if predicting observable variables depends on factor forecasting, a reliable and parsimonious approximation of the factor dynamic process is important. In Deistler, Anderson, Filler, Zinner and Chen (2010), the authors study identification of the generalized dynamic factor model where the common component has a singular rational spectral density. Under the assumption that transfer functions are tall and zeroless (*i.e.,* the number of common shocks is less than the number of static factors), they argue that static factors have a finite-order AR singular representation which can be estimated by generalized Yule-Walker equations. Note that Yule-Walker equations are not unique for such systems, but Deistler, Filler and Funovits (2011) propose a particular canonical form for estimation purposes.

After showing that FAVARMA models yield a theoretically consistent specification, we study whether VARMA factors can help in forecasting time series. We compare the forecasting performance (in terms of MSE) of four FAVARMA specifications, with standard AR(p), ARMA(p, q) and factor models where the factor dynamics is approximated by a finite-order VAR. An out-of-sample forecasting exercise is performed using a U.S. monthly panel from Boivin, Giannoni and Stevanović (2009).

The results show that VARMA factors help in predicting several key macroeconomic aggregates, relative to standard factor models, and across different forecasting horizons. We find important gains, up to a reduction of 42% in MSE, when forecasting the growth rates of industrial production, employment and consumer price index inflation. In particular, the FAVARMA specifications generally outperform the VAR-factor forecasting models. We also report simulation results which show that VARMA factor modelling noticeably improves forecasting in finite samples.

Finally, we perform a structural factor analysis exercise. We estimate the effect of a monetary policy shock using the data and identification scheme of Bernanke et al. (2005). We find that impulse responses from a parsimonious 6-factor FAVARMA$(2, 1)$ model give a precise and plausible picture of the effect and transmission of monetary policy in the U.S. To get similar responses from a standard FAVAR model, the Akaike information criterion leads to a lag order of 14. So we need to estimate 84 coefficients governing the factor dynamics in the FAVARMA framework, while the FAVAR model requires 510 VAR parameters.

In Section 2, we summarize some important results on linear transformations of vector stochastic processes and present four identified VARMA forms. In Section 3, we study the link between VARMA and factor representations. The FAVARMA model is proposed in Section 4, and estimation is discussed in Section 5. Monte Carlo simulations are discussed in Section 6. The empirical forecasting exercise is presented in Section 7, and the structural analysis in Section 8. Proofs and simulation results are reported in Appendix.

2. FRAMEWORK

In this section, we summarize a number of important results on linear transformations of vector stochastic processes, and we present four identified VARMA forms we will use in forecasting applications.

2.1. Linear transformations of vector stochastic processes

Exploring the features of transformed processes is important since data are often obtained by temporal and spatial aggregation, and/or transformed through linear filtering techniques, before they are used to estimate models and evaluate theories. In macroeconomics, researchers model dynamic interactions by specifying a multivariate stochastic process on a small number of economic indicators. Hence, they work on marginalized processes, which can be seen as linear transformations of the original series. Finally, dimension-reduction methods, such as principal components, lead one to consider linear transformations of the observed series. Early contributions on these issues include Zellner and Palm (1974), Rose (1977), Wei (1978), Abraham (1982), and Lütkepohl (1984).

The central result we shall use focuses on linear transformations of a N-dimensional, stationary, strictly indeterministic stochastic process. Suppose X_t satisfies the model

$$X_t = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j} = \Psi(L)\varepsilon_t\,, \quad \Psi_0 = I_K\,, \tag{2.1}$$

where ε_t is a weak white noise, with $E(\varepsilon_t) = 0$, $E(\varepsilon_t \varepsilon_t') = \Sigma_\varepsilon$, $\det[\Sigma_\varepsilon] > 0$, $E(X_t X_t') = \Sigma_X$, $E(X_t X_{t+h}') = \Gamma_X(h)$, $\Psi(L) = \sum_{i=0}^{\infty} \Psi_i L^i$ and $\det[\Psi(z)] \neq 0$ for $|z| < 1$. (2.1) can be interpreted as the Wold representation of X_t, in which case $\varepsilon_t = X_t - P_L[X_t | X_{t-1}, X_{t-2}, \ldots]$ and $P_L[X_t | X_{t-1}, X_{t-2}, \ldots]$ is the best linear forecast of X_t based on its own past (*i.e.*, ε_t is the *innovation process* of X_t). Consider the following linear transformation of X_t:

$$F_t = C X_t \tag{2.2}$$

where C is a $K \times N$ matrix of rank K. Then F_t is also stationary, indeterministic and has zero mean, so it has an MA representation of the form:

$$F_t = \sum_{j=0}^{\infty} \Phi_j v_{t-j} = \Phi(L) v_t, \quad \Phi_0 = I_K\,, \tag{2.3}$$

where v_t is K-dimensional white noise with $E(v_t v_t') = \Sigma_v$. These properties hold whenever X_t is a vector stochastic process with an MA representation. If it is invertible, finite and infinite-order VAR processes are covered.

In practice, only a finite number of parameters can be estimated. Consider the MA(q) process

$$X_t = \varepsilon_t + M_1 \varepsilon_{t-1} + \cdots + M_q \varepsilon_{t-q} = M(L)\varepsilon_t \tag{2.4}$$

with $\det[M(z)] \neq 0$ for $|z| < 1$ and nonsingular white noise noise covariance matrix Σ_ε, and a $K \times N$ matrix C with rank K. Then, the transformed process $F_t = C X_t$ has an invertible MA(q_*) representation

$$F_t = v_t + N_1 v_{t-1} + \cdots + N_{q_*} v_{t-q_*} = N(L) v_t \tag{2.5}$$

with $\det[N(z)] \neq 0$ for $|z| < 1$, where v_t is a K-dimensional white noise with nonsingular matrix Σ_v, each N_i is a $K \times K$ coefficient matrix, and $q_* \leq q$.

Some conditions in the previous results can be relaxed. The nonsingularity of the covariance matrix Σ_ε and the full rank of C are not necessary so there may be exact linear dependencies among the components of X_t and F_t [see Lütkepohl (1984b)]. It is also possible that $q_* < q$.

It is well known that weak VARMA models are closed under linear transformations. Let X_t be an N-dimensional, stable, invertible VARMA(p, q) process

$$\Phi(L)X_t = \Theta(L)\varepsilon_t \tag{2.6}$$

and let C be a $K \times N$ matrix of rank $K < N$. Then $F_t = CX_t$ has a VARMA(p_*, q_*) representation with $p_* \leq (N-K+1)p$ and $q_* \leq (N-K)p+q$; see Lütkepohl (2005, Corollary 11.1.2). A linear transformation of a finite-order VARMA process still has a finite-order VARMA representation, but with possibly higher autoregressive and moving-average orders.

When modeling economic time series, the most common specification is a finite-order VAR. Therefore, it is important to notice that this class of models is not closed with respect to linear transformations reducing the dimensions of the original process.

2.2. Identified VARMA models

An identification problem arises since the VARMA representation of X_t is not unique. There are several ways to identify the process in (2.6). In the following, we state four unique VARMA representations: the well-known final-equation form and three representations proposed in Dufour and Pelletier (2013).

Definition 2.1 Final AR equation form (FAR). *The VARMA representation in* (2.6) *is said to be in final AR equation form if $\Phi(L) = \phi(L)I_N$, where $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$ is a scalar polynomial with $\phi_p \neq 0$.*

Definition 2.2 Final MA equation form (FMA). *The VARMA representation in* (2.6) *is said to be in final MA equation form if $\Theta(L) = \theta(L)I_N$, where $\theta(L) = 1 - \theta_1 L - \cdots - \theta_q L^q$ is a scalar polynomial with $\theta_q \neq 0$.*

Definition 2.3 Diagonal MA equation form (DMA). *The VARMA representation in* (2.6) *is said to be in diagonal MA equation form if* $\Theta(L) = \mathrm{diag}[\theta_{ii}(L)] = I_N - \Theta_1 L - \cdots - \Theta_q L^q$*, where* $\theta_{ii}(L) = 1 - \theta_{ii,1} L - \cdots - \theta_{ii,q_i} L^{q_i}$, $\theta_{ii,q_i} \neq 0$, *and* $q = \max_{1 \le i \le N}(q_i)$.

Definition 2.4 Diagonal AR equation form (DAR). *The VARMA representation in* (2.6) *is said to be in diagonal AR equation form if* $\Phi(L) = \mathrm{diag}[\phi_{ii}(L)] = I_N - \Phi_1 L - \cdots - \Phi_p L^p$*, where* $\phi_{ii}(L) = 1 - \phi_{ii,1} L - \cdots - \phi_{ii,p_i} L^{p_i}$, $\phi_{ii,p_i} \neq 0$, *and* $p = \max_{1 \le i \le N}(p_i)$.

The identification of these VARMA representations is discussed in Dufour and Pelletier (2013, Section 3). In particular, the identification of diagonal MA form is established under the simple assumption of no common root.

From standard results on the linear aggregation of VARMA processes [see, e.g., Zellner and Palm (1974), Rose (1977), Wei (1978), Abraham (1982), and Lütkepohl (1984)], it is easy to see that an aggregated process such as F_t also has an identified VARMA representation in final AR or MA equation form. But this type of representation may not be attractive for several reasons. First, it is far from the usual VAR model, because it excludes lagged values of other variables in each equation. Moreover, the AR coefficients are the same in all equations, which typically leads to a high-order AR polynomial. Second, the interaction between different variables is modeled through the MA part of the model, and may be difficult to assess in empirical and structural analysis.

The diagonal MA form is especially appealing. In contrast with the echelon form [Deistler and Hannan (1981), Hannan and Deistler (1988), and Lütkepohl (1991, Chapter 7)], it is relatively simple and intuitive. In particular, there is no complex structure of zero off-diagonal elements in the AR and MA operators. For practitioners, this is quite appealing since adding lags of ε_{it} to the i^{th} equation is a simple natural extension of the VAR model. The MA operator has a simple diagonal form, so model nonlinearity is reduced and estimation becomes numerically simpler.

3. VARMA AND FACTOR REPRESENTATIONS

In this section, we study the link between VARMA and factor representations of a vector stochastic process X_t, and the dynamic process of the factors. In the theorems below, we suppose that X_t is a N-dimensional regular (strictly indeterministic) discrete-time process in $\mathbb{R}^N : X = \{X_t : t \in \mathbb{R}^N,\ t \in \mathbb{Z}\}$ with Wold representation (2.1). In Theorem **3.1**, we postulate a factor model for X_t where factors follow a finite-order VAR process:

$$X_t = \Lambda F_t + u_t \tag{3.1}$$

where Λ is an $N \times K$ matrix of factor loadings with rank K, and u_t is a (weak) white noise process with covariance matrix Σ_u such that

$$\mathsf{E}[F_t u_t'] = 0 \text{ for all } t. \tag{3.2}$$

We now show that finite-order VAR factors induce a finite-order VARMA process for the observable series. Proofs are supplied in the Appendix.

Theorem 3.1 Observable process induced by finite-order VAR factors. *Suppose X_t satisfies the assumptions* (3.1) - (3.2) *and F_t follows the VAR*(p) *process*

$$F_t = \Phi(L)F_{t-1} + a_t \tag{3.3}$$

such that $e_t = [u_t \vdots a_t]'$ is a (*weak*) *white noise process with*

$$\mathsf{E}[F_{t-j}e_t'] = 0 \textit{ for } j \geq 1,\ \forall t, \tag{3.4}$$

$\Phi(L) = \Phi_1 L - \cdots - \Phi_p L^p$, and the equation $\det[I_K - \Phi(z)] = 0$ *has all its roots outside the unit circle. Then, for all t, $\mathsf{E}[X_{t-j}e_t'] = 0$ for $j \geq 1$, and X_t has the following representations:*

$$A(L)X_t = B(L)e_t\,, \tag{3.5}$$

$$A(L)X_t = \bar{\Psi}(L)\varepsilon_t\,, \tag{3.6}$$

where $A(L) = [I - \Lambda\Phi(L)(\Lambda'\Lambda)^{-1}\Lambda' L]$, $B(L) = [A(L) \vdots \Lambda]$, $\bar{\Psi}(L) = \sum_{j=0}^{p+1} \bar{\Psi}_j L^j$ with $\bar{\Psi}_j = \sum_{i=0}^{p+1} A_i \Psi_{j-i}$, the matrices Ψ_j are the coefficients of the Wold representation (2.1), *and ε_t is the innovation process of X_t.*

ACCEPTED MANUSCRIPT

This result can be extended to the case where the factors have VARMA representations. It is not surprising that the induced process for X_t is again a finite-order VARMA, though possibly with a different MA order. This is summarized in the following theorem.

Theorem 3.2 Observable process induced by VARMA factors. *Suppose X_t satisfies the assumptions* (3.1) - (3.2) *and F_t follows the VARMA(p, q) process*

$$F_t = \Phi(L)F_{t-1} + \Theta(L)a_t \tag{3.7}$$

where $e_t = [u_t \vdots a_t]'$ is a (weak) white noise process which satisfies the orthogonality condition (3.4), *$\Phi(L) = \Phi_1 L - \cdots - \Phi_p L^p$, $\Theta(L) = I_K - \Theta_1 L - \cdots - \Theta_q L^q$, and the equation* $\det[I_K - \Phi(z)] = 0$ *has all its roots outside the unit circle. Then X_t has representations of the form* (3.5) *and* (3.6), *with $B(L) = [A(L) \vdots \Lambda\Theta(L)]$, $\bar{\Psi}(L) = \sum_{j=0}^{p_*} \bar{\Psi}_j L^j$, $\bar{\Psi}_j = \sum_{i=0}^{p_*} A_i \Psi_{j-i}$, and* $p_* = \max(p+1, q)$.

Note that the usual invertibility assumption on the factor VARMA process (3.7) is not required. The next issue we consider concerns the factor representation of X_t. What are the implications of the underlying structure of X_t on the representation of latent factors when the latter are calculated as linear transformations of X_t? This is summarized in the following theorem.

Theorem 3.3 Dynamic factor models associated with VARMA processes. *Suppose $F_t = CX_t$, where C is a $K \times N$ full row rank matrix. Then the following properties hold:*

(*i*) *if X_t has a VARMA(p, q) representation as in* (2.6), *then F_t has VARMA(p_*, q_*) representation with $p_* \leq (N - K + 1)p$ and $q_* \leq q + (N - K)p$;*

(*ii*) *if X_t has a VAR(p) representation, then F_t has VARMA(p_*, q_*) representation with $p_* \leq Np$ and $q_* \leq (N-1)p$;*

(*iii*) *if X_t has an MA representation as in* (2.4), *then F_t has an MA(q_*) representation with $q \leq q_*$.*

From the Wold decomposition of common components, Deistler et al. (2010) argue that latent variables can have ARMA or state-space representations, but given the singularity and zero-free nature of transfer functions, they can also be modeled as finite-order singular AR processes. Theorem **3.3** does not assume the existence of a dynamic factor structure, so it holds for any linear aggregation of X_t.

Arguments in favor of using a FAVARMA specification can be summarized as follows.

(i) Whenever X_t follows a VAR or a VARMA process, the factors defined through a linear cross-sectional transformation (such as principal components) follow a VARMA process. Moreover, a VAR or VARMA-factor structure on X_t entails a VARMA structure for X_t.

(ii) VARMA representations are more parsimonious, so they easily lead to more efficient estimation and tests. As shown in Dufour and Pelletier (2013), the introduction of the MA operator allows for a reduction of the required AR order so we can get more precise estimates. Moreover, in terms of forecasting accuracy, VARMA models have theoretical advantages over the VAR representation [see Lütkepohl (1987)].

(iii) The use of VARMA factors can be viewed from two different perspectives. First, if we use factor analysis as a dimension-reduction method, a VARMA specification is a natural process for factors (Theorem **3.3**). Second, if factors are given a deep ("structural") interpretation, the factor process has intrinsic interest, and a VARMA specification on factors – rather than a finite-order VAR – is an interesting generalization motivated by usual arguments of theoretical coherence, parsimony, and marginalization. In particular, even if F_t has a finite-order VAR representation, subvectors of F_t typically follow a VARMA process.

4. FACTOR-AUGMENTED VARMA MODELS

We have shown that the observable VARMA process generally induces a VARMA representation for factors, not a finite-order VAR. Following these results, we propose to consider factor-augmented VARMA (FAVARMA) models. Following the notation of Stock and Watson (2005), the dynamic factor model (DFM) where factors have a finite-order

ACCEPTED MANUSCRIPT

VARMA(p_f, q_f) representation can be written as

$$X_{it} = \tilde{\lambda}_i(L) f_t + u_{it}, \tag{4.1}$$

$$u_{it} = \delta_i(L) u_{i,t-1} + \nu_{it}, \tag{4.2}$$

$$f_t = \Gamma(L) f_{t-1} + \Theta(L)\eta_t, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{4.3}$$

where f_t is $q \times 1$ factor vector, $\tilde{\lambda}_i(L)$ is a $1 \times q$ vector of lag polynomials, $\tilde{\lambda}_i(L) = (\tilde{\lambda}_{i1}(L), \ldots, \tilde{\lambda}_{iq}(L))$, $\tilde{\lambda}_{ij}(L) = \sum_{k=0}^{p_{i,j}} \tilde{\lambda}_{i,j,k} L^k$, $\delta_i(L)$ is a $p_{x,i}$-degree lag polynomial, $\Gamma(L) = \Gamma_1 L + \cdots + \Gamma_{p_f} L^{p_f}$, $\Theta(L) = I - \Theta_1 L - \cdots - \Theta_{q_f} L^{q_f}$, and ν_{it} is a N-dimensional white noise uncorrelated with the q-dimensional white noise process η_t. The exact DFM is obtained if the following assumption is satisfied:

$$\mathsf{E}(u_{it} u_{js}) = 0, \ \forall i, j, t, s, \quad i \neq j. \tag{4.4}$$

We obtain the approximate DFM by allowing for cross-section correlations among the idiosyncratic components as in Stock and Watson (2005). We assume the idiosyncratic errors ν_{it} are uncorrelated with the factors f_t at all leads and lags.

On premultiplying both sides of (4.1) by $1 - \delta_i(L)$, we get the DFM with serially uncorrelated idiosyncratic errors:

$$X_{it} = \lambda_i(L) f_t + \delta_i(L) X_{it-1} + \nu_{it} \tag{4.5}$$

where $\lambda_i(L) = [1 - \delta_i(L)L]\tilde{\lambda}_i(L)$. Then, we can rewrite the DFM in the following form:

$$X_t = \lambda(L) f_t + D(L) X_{t-1} + \nu_t, \tag{4.6}$$

$$f_t = \Gamma(L) f_{t-1} + \Theta(L)\eta_t, \tag{4.7}$$

where

$$\lambda(L) = \begin{bmatrix} \lambda_1(L) \\ \vdots \\ \lambda_n(L) \end{bmatrix}, \quad D(L) = \begin{bmatrix} \delta_1(L) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \delta_n(L) \end{bmatrix}, \quad \nu_t = \begin{bmatrix} \nu_{1t} \\ \vdots \\ \nu_{nt} \end{bmatrix}.$$

To obtain the static version, we suppose that $\lambda(L)$ has degree $p-1$, and let $F_t = [f_t', f_{t-1}', \ldots, f_{t-p+1}']'$, where the dimension of F_t is K, with $q \le K \le qp$. Then,

$$
\begin{aligned}
X_t &= \Lambda F_t + u_t, && (4.8)\\
u_t &= D(L)u_{t-1} + \nu_t, && (4.9)\\
F_t &= \Phi(L)F_{t-1} + G\Theta(L)\eta_t, && (4.10)
\end{aligned}
$$

where Λ is a $N \times K$ matrix where the i-th row consists of coefficients of $\tilde{\lambda}_i(L)$, $\Phi(L)$ contains coefficients of $\Gamma(L)$ and zeros, and G is a $K \times q$ matrix which loads (structural) shocks η_t to static factors (it consists of 1's and 0's). Note that if $\Theta(L) = I$ we obtain the static factor model which has been used to forecast time series [Stock and Watson (2002b), Stock and Watson (2006), Boivin and Ng (2005)] and study the impact of monetary policy shocks in a FAVAR model [Bernanke et al. (2005), Boivin et al. (2009)].

5. ESTIMATION

Several estimation methods have been proposed for factor models and VARMA processes (separately). One possibility is to estimate the system (4.8)-(4.10) simultaneously after making distributional assumptions on the error terms. This method is already computationally difficult when the factors have a simple VAR structure. Adding the MA part to the factor process makes this task even more difficult, for estimating VARMA models is typically not easy.

We use here the two-step Principal Component Analysis (PCA) estimation method; see Stock and Watson (2002) and Bai and Ng (2008) for theoretical results concerning the PCA estimator. In the first step, $\hat{F}_t$ are computed as K principal components of X_t. In the second step, we estimate the VARMA representation (4.10) on $\hat{F}_t$. The number of factors can be estimated through different procedures proposed by Amengual and Watson (2007), Bai and Ng (2002), Bai and Ng (2007), Hallin and Liska (2007), and Onatski (2009). In forecasting we estimate the number of factors using the Bayesian information criterion as in Stock and Watson (2002b), while the number of factors in the structural FAVARMA model is the same as in Bernanke et al. (2005).

The standard estimation methods for VARMA models are maximum likelihood and nonlinear least squares. Unfortunately, these methods require nonlinear optimization, which may not be feasible when the number of parameters is large. Here, we use the GLS method proposed in Dufour and Pelletier (2013), which generalizes the regression-based estimation method introduced by Hannan and Rissanen (1982). Consider a K-dimensional zero mean process Y_t generated by the VARMA(p, q) model:

$$A(L)Y_t = B(L)U_t \tag{5.1}$$

where $A(L) = I_K - A_1 L - \cdots - A_p L^p$, $B(L) = I_K - B_1 L - \cdots - B_q L^q$, and U_t is a weak white noise. Assume $\det[A(z)] \neq 0$ for $|z| \leq 1$ and $\det[B(z)] \neq 0$ for $|z| \leq 1$ so the process Y_t is stable and invertible. Set $A_k = [a'_{1\bullet,k}, \ldots, a'_{K\bullet,k}]'$, $k = 1, \ldots, K$, where $a_{j\bullet,k}$ is the j-th row of A_k, and $B(L) = \text{diag}[b_{11}(L), \ldots, b_{KK}(L)]$, $b_{jj}(L) = 1 - b_{jj,1}L - \cdots - b_{jj,q_j}L^{q_j}$, when $B(L)$ is in MA diagonal form. Then, when the model is in diagonal MA form, we can write the parameters of the VARMA model as a vector $\gamma = [\gamma_1, \gamma_2]'$ where γ_1 contains the AR parameters and γ_2 the MA parameters, as follows:

$$\gamma_1 = [a_{1\bullet,1}, \ldots, a_{1\bullet,p}, \ldots, a_{K\bullet,1}, \ldots, a_{K\bullet,p}], \tag{5.2}$$

$$\gamma_2 = [b_{11,1}, \ldots, b_{11,q_1}, \ldots, b_{KK,1}, \ldots, b_{KK,q_K}]. \tag{5.3}$$

The estimation method involves three steps.

Step 1. Estimate a VAR(n_T) model by least squares, where $n_T < T/(2K)$, and compute the residuals:

$$\hat{U}_t = Y_t - \sum_{l=1}^{n_T} \hat{\Pi}_l(n_T) Y_{t-l}. \tag{5.4}$$

Step 2. From the residuals of step 1, compute $\hat{\Sigma}_U = \frac{1}{T}\sum_{t=n_T+1}^{T}\hat{U}_t\hat{U}_t'$, *i.e.* the corresponding estimate of the covariance matrix of U_t, and apply GLS to the multivariate regression

$$A(L)Y_t = [B(L) - I_K]\hat{U}_t + e_t \tag{5.5}$$

to get estimates $\tilde{A}(L)$ and $\tilde{B}(L)$. The estimator is

$$\hat{\gamma} = [\sum_{t=l}^{T}\hat{Z}_{t-1}'\hat{\Sigma}_U^{-1}\hat{Z}_{t-1}]^{-1}\sum_{t=l}^{T}\hat{Z}_{t-1}'\hat{\Sigma}_U^{-1}Y_t \tag{5.6}$$

with $l = n_T + \max(p, q) + 1$. Setting

$$\mathbf{Y}_{t-1}(p) = \left[y_{1,t-1}, \ldots, y_{K,t-1}, \ldots, y_{1,t-p}, \ldots, y_{K,t-p}\right], \tag{5.7}$$

$$\hat{\mathbf{U}}_{t-1} = \left[\hat{u}_{1,t-1}, \ldots, \hat{u}_{K,t-1}, \ldots, \hat{u}_{1,t-q}, \ldots, \hat{u}_{K,t-q}\right], \ \hat{\mathbf{u}}_{k,t-1} = \left[\hat{u}_{k,t-1}, \ldots, \hat{u}_{k,t-q_k}\right], \tag{5.8}$$

the matrix $\hat{Z}_{t-1}$ is defined as

$$\hat{Z}_{t-1} = \begin{bmatrix} \mathbf{Y}_{t-1}(p) & \cdots & 0 & \hat{\mathbf{u}}_{1,t-1} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{Y}_{t-1}(p) & 0 & \cdots & \hat{\mathbf{u}}_{K,t-1} \end{bmatrix}.$$

Step 3. Using the second step estimates, form new residuals

$$\tilde{U}_t = Y_t - \sum_{i=1}^{p}\tilde{A}_iY_{t-i} + \sum_{j=1}^{q}\tilde{B}_j\tilde{U}_{t-j}$$

with $\tilde{U}_t = 0$ for $t \le \max(p, q)$, and define

$$X_t = \sum_{j=1}^{q}\tilde{B}_jX_{t-j} + Y_t,\ W_t = \sum_{j=1}^{q}\tilde{B}_jW_{t-j} + \tilde{U}_t,\ \tilde{V}_t = \sum_{j=1}^{q}\tilde{B}_j\tilde{V}_{t-j} + \tilde{Z}_t,$$

where $X_t = W_t = 0$ for $t \le \max(p, q)$, and $\tilde{Z}_t$ is defined like $\hat{Z}_t$ in step 2, with $\hat{U}_t$ replaced by $\tilde{U}_t$. Then, compute a new estimate of Σ_U, $\tilde{\Sigma}_U = \frac{1}{T}\sum_{t=\max(p,q)+1}^{T}\tilde{U}_t\tilde{U}_t'$,

ACCEPTED MANUSCRIPT

and regress by GLS $\tilde{U}_t + X_t - W_t$ on $\tilde{V}_{t-1}$ to obtain the following estimate of γ:

$$\hat{\gamma} = \left[\sum_{t=\max(p,q)+1}^{T} \tilde{V}'_{t-1} \tilde{\Sigma}_U^{-1} \tilde{V}_{t-1} \right]^{-1} \left[\sum_{t=\max(p,q)+1}^{T} \tilde{V}'_{t-1} \tilde{\Sigma}_U^{-1} \left[\tilde{U}_t + X_t - W_t \right] \right]. \tag{5.9}$$

The consistency and asymptotic normality of the above estimators are established in Dufour and Pelletier (2013). In the previous steps, the orders of the AR and MA operators are taken as known. In practice, they are usually estimated by statistical methods or suggested by theory. Dufour and Pelletier (2013) propose an information criterion to be applied in the second step of the estimation procedure. For all $p_i \le P$ and $q_i \le Q$ compute

$$\log[\det(\tilde{\Sigma}_U)] + \dim(\gamma) \frac{(\log T)^{1+\delta}}{T}, \quad \delta > 0\,. \tag{5.10}$$

Choose $\hat{p}_i$ and $\hat{q}_i$ as the set which minimizes the information criteria (5.10). The properties of estimators $\hat{p}_i$ and $\hat{q}_i$ are given in the paper.

6. FORECASTING

In this section, we study whether the introduction of VARMA factors can improve forecasting. We consider a simplified version of the static model (4.8) - (4.10) where F_t is scalar:

$$X_{it} = \lambda_i F_t + u_{it}\,, \tag{6.1}$$

$$u_{it} = \delta_i u_{it-1} + \nu_{it}\,,\ i, \ldots, N, \tag{6.2}$$

$$F_t = \phi F_{t-1} + \eta_t - \theta \eta_{t-1}. \tag{6.3}$$

On replacing F_t and u_{it} in the observation equation (6.1) with the expressions in (6.2) - (6.3), we get the following forecast equation for $X_{i,T+1}$ based on the information available at time T:

$$X_{i,T+1|T} = \delta_i X_{iT} + \lambda_i(\phi - \delta_i) F_T - \lambda_i \theta \eta_T\,.$$

Suppose $\lambda_i \neq 0$, *i.e.* there is indeed a factor structure applicable to X_{it}, and $\phi \neq \delta_i$, *i.e.* the common and specific components do not have the same dynamics. With the additional assumption that F_t follows an AR(1) process, Boivin and Ng (2005) show that taking into account the factor F_t allows one to obtain better forecasts of X_{it} [in terms of the mean squared error (MSE)]. We allow here for an MA component in the dynamic process of F_t, which provides a parsimonious way of representing an infinite-order AR structure for the factor.

Forecast performance depends on the way factors are estimated as well as the choice of forecasting model. Boivin and Ng (2005) consider static and dynamic factor estimation along with three types of forecast equations: (1) *unrestricted*, where $X_{i,T+h}$ is predicted using X_{iT}, F_T and their lags; (2) *direct*, where F_{T+h} is first predicted using its dynamic process, a forecast then used to predict $X_{i,T+h}$ with the factor equation (6.3); (3) *nonparametric*, where no parametric assumption is made on factor dynamics and its relationship with observables. The simulation and empirical results of Boivin and Ng (2005) show that the unrestricted forecast equation with static factors generally yields the best performance in terms of MSE.

6.1. Forecasting models

A popular way to evaluate the predictive power of a model is to conduct an out-of-sample forecasting exercise. Here, we compare the FAVARMA approach with common factor-based methods. The forecast equations are divided in two categories. First, we consider methods where no explicit dynamic factor model is used, such as diffusion-index (DI) and diffusion autoregressive (DI-AR) models [Stock and Watson (2002b)]:

$$X_{i,T+h|T} = \alpha^{(h)} + \sum_{j=1}^{m} \beta_{ij}^{(h)} F_{T-j+1} + \sum_{j=1}^{p} \rho_{ij}^{(h)} X_{i,T-j+1} .$$

ACCEPTED MANUSCRIPT

In this case, three variants are studied: (1) "unrestricted" (with $m \geq 1$ and $p \geq 0$); (2) "DI" (with $m = 1$ and $p = 0$); (3) "DI-AR" (with $m = 1$). Second, we consider two-step methods where common and specific components are first predicted from their estimated dynamic processes, and then combined to forecast the variables of interest using the estimated observation equation. Moreover, we distinguish between sequential (or iterative) and direct methods to calculate forecasts [see Marcellino, Stock and Watson (2006) for details]:

$$X_{i,T+h|T} = \lambda_i' F_{T+h|T} + u_{i,T+h|T}$$

where $u_{i,T+h|T}$ is obtained after fitting an AR(p) process on u_{it}, while the factor forecasts are obtained using "sequential" [$F_{T+h|T} = \hat{\Phi}_{T+h-1}(L) F_{T+h-1|T}$] or "direct" methods [$F_{T+h|T} = \hat{\Phi}_T^{(h)}(L) F_T$].

In this exercise, the factors are defined as principal components of X_t. Thus, only the second type of forecast method is affected by allowing for VARMA factors. We consider four identified VARMA forms labeled: "Diag MA", "Diag AR", "Final MA" and "Final AR". The FAVARMA forecasting equations have the form:

$$X_{i,T+h|T} = \lambda_i' F_{T+h|T} + u_{i,T+h|T} , \quad F_{T+h|T} = \hat{\Phi}_{T+h-1}(L) F_{T+h-1|T} + \hat{\Theta}_{T+h-1}(L) \eta_{T+h-1|T} .$$

Our benchmark forecasting model is an AR(p) model, as in Stock and Watson (2002b) and Boivin and Ng (2005). However, given the postulated factor structure, a finite-order autoregressive model is only an approximation of the process of X_{it}. From Theorem **3.1**, the marginal process for each element of X_t typically has an ARMA form. If the MA polynomial has roots close to the non-invertibility region, a long autoregressive model may be needed to approximate the process. For this reason, we also consider ARMA models as benchmarks, to see how they fare with respect to AR and factor-based models.

ACCEPTED MANUSCRIPT

6.2. Monte-Carlo simulations

To assess the performance of our approach, we performed a Monte Carlo simulation comparing the forecasts of FAVARMA models (in four identified forms) with those of FAVAR models. The data were simulated using a static factor model with MA(1) factors and idiosyncratic components similar to the ones considered by Boivin and Ng (2005) and Onatski (2009b):

$$X_{it} = \lambda_i F_t + u_{it},\ F_t = \eta_t - B\eta_{t-1}\,,$$

$$u_{it} = \rho_N u_{i-1,t} + \xi_{it}\,,\ \xi_{it} = \rho_T \xi_{i,t-1} + \epsilon_{it}, \quad \epsilon_{it} \sim N(0,1)\,,\ i = 1,\ldots,N,\ t = 1,\ldots,T,$$

where $\eta_t \overset{iid}{\sim} N(0,1)$, $\rho_N \in \{0.1, 0.5, 0.9\}$ determines the cross-sectional dependence, $\rho_T \in \{0.1, 0.9\}$ the time dependence, the number of factors is 2, $B = \text{diag}[0.5,\ 0.3]$, $N = \{50, 100, 130\}$, and $T \in \{50, 100, 600\}$. VARMA orders are estimated as in Dufour and Pelletier (2013), the AR order for idiosyncratic component is 1, and the lag order in VAR approximation of factors dynamics is set to 6.

The results from this simulation exercise are presented in Appendix (Table 1). The numbers represent the MSE of four FAVARMA identified forms over the MSE of FAVAR direct forecasting models. When the number of time periods is small ($T = 50$), FAVARMA models strongly outperform FAVAR models, especially at long horizons. The huge improvement at horizons 24 and 36 is due to the small sample size. When compared to the iterative FAVAR model (not reported), FAVARMA models still produce better forecasts in terms of MSE, but the improvement is smaller relative to the multi-step-ahead VAR-based forecasts. When the number of time periods increases ($T = 100$, 600), the improvement of VARMA-based models is moderate, but the latter still yield better forecasts, especially at longer horizons. Another observation of interest is that FAVARMA models perform better when the factor structure is weak, *i.e.* in cases where the cross-section size is relatively small ($N = 50$ compared to $N = 100$) and idiosyncratic components are correlated.

ACCEPTED MANUSCRIPT

We performed additional simulation exercises (not reported), which also demonstrate a better performance of FAVARMA-based forecasts when the number of factors increases. The description and results are available in the online appendix Dufour and Stevanović (2013*b*)and in the working paper version Dufour and Stevanović (2013*a*).

7. APPLICATION: FORECASTING U.S. MACROECONOMIC AGGREGATES

In this section, we present an out-of-sample forecasting exercise using a balanced monthly panel from Boivin et al. (2009) which contains 128 monthly U.S. economic and financial indicators observed from 1959M01 to 2008M12. The series were initially transformed to induce stationarity.

The MSE results relative to benchmark AR(p) models are presented in Table 1. The out-of-sample evaluation period is 1988M01-2008M12. In the forecasting models "unrestricted", "DI", and "DI-AR", the number of factors, the number of lags for both factors and X_{it} are estimated with BIC, and are allowed to vary over the whole evaluation period. For "unrestricted" model the number of factors is 3, $m = 1$ and $p = 0$. In the case of "DI-AR" and "DI", 6 factors are used, plus 5 lags of X_{it} within "DI-AR" representation.

In the FAVAR and FAVARMA models, the number of factors is set to 4. For all evaluation periods and forecasting horizons the estimated VARMA orders (AR and MA respectively) are low: 1 and $[1,1,1,1]$ for DMA form, $[1,2,1,1]$ and 1 for DAR, 1 and 2 for FMA, and $[2-4]$ and 1 for FAR form. The estimated VAR order is most of the time equal to 2, while the lag order of each idiosyncratic AR(p) process is between 1 and 3. In robustness analysis, the VAR order has been set to 4, 6 and 12, but the results did not change substantially. Both univariate ARMA orders are estimated to 1, while the number of lags in the benchmark AR model fluctuates between 1 and 2.

The results in Table 1 show that VARMA factors improve the forecasts of key macroeconomic indicators across several horizons. For industrial production growth, the diffusion-index model exhibits the best performance at the one-month horizon, while diagonal MA and final MA FAVARMA models outperform the other methods for horizons of 2, 4 and 6 months. Finally, univariate ARMA models yield the smallest RMSE for the long-term forecasts. When forecasting employment growth, three FAVARMA forms outperform all other factor-based models for short and mid-term horizons. ARMA models still produce the smallest RMSE for most of the long-term horizons.

For CPI inflation, the DI model provides the smallest MSE at horizon 1, while the final AR FAVARMA models do a better job at horizons 2, 4 and 6. Several VARMA-based models perform the best for longer horizons (18, 24 and 48 months), while sequential and DI approaches dominate in forecasting 12 and 36 months ahead.

From Theorem **3.1**, it is easy to see that each component of X_t follows a univariate ARMA process. The forecasts based on factor and univariate ARMA models are not in general equivalent, because different information sets are used. Even though multivariate models (such as factor models) use more variables, univariate ARMA models tend to be more parsimonious in practice, which may reduce estimation uncertainty. So these two modelling strategies can produce quite different forecasts. In Table 2 we present MSE of all factor model predictions relative to ARMA forecasts. Boldface numbers highlight cases where the ARMA model outperforms the factor-based alternatives in terms of MSE.

For industrial production, ARMA specifications do better than all diffusion-index and FAVAR models (except at the one-month horizon). For employment, the conclusion is quite similar relative to FAVARMA, while diffusion-index models perform better than ARMA at horizons 1, 2, 4, and 48. Finally, in the case of CPI inflation, ARMA model seem to be a better choice for most of the horizons relatively to diffusion-index and FAVAR alternatives. On the other hand, FAVARMA models do much better, *e.g.* the final MA form beats the ARMA models at all horizons.

ACCEPTED MANUSCRIPT

Based on these results, ARMA models appears to be a very good alternative to standard factor-based models at long horizons. This is not surprising since ARMA models are very parsimonious. However, FAVARMA models outperform ARMA models in most cases.

It is also of interest to see more directly how FAVARMA forecasts compare to those from FAVAR models. In Table 3, we present MSE of FAVARMA forecasting models relative to Direct and Sequential FAVAR specifications. The numbers in bold character present cases where the FAVARMA model performs better than the FAVAR.

Most numbers in Table 3 are boldfaced, *i.e.* FAVARMA models outperform standard FAVAR specifications at most horizons. This is especially the case for industrial production, where both MA VARMA forms produce smaller MSE at all horizons. At best, the FAVARMA model improves the forecasting accuracy by 32% at horizon 12. In the case of Civilian labor force, VARMA factors do improve the predicting power, but the Direct FAVAR model performs better for longer horizons. Finally, both diagonal and final MA FAVARMA specifications provide smaller MSEs over all horizons in predicting CPI inflation. The improvement increases with the forecast horizons, and reaches a maximum of 15%.

We performed a similar exercise with a Canadian data set from Boivin, Giannoni and Stevanović (2009b). We found that VARMA factors help in predicting several key Canadian macroeconomic aggregates, relative to standard factor models, and at many forecasting horizons. The description and results are available in the online appendix Dufour and Stevanović (2013*b*) and in the working paper version Dufour and Stevanović (2013*a*).

ACCEPTED MANUSCRIPT

8. APPLICATION: EFFECTS OF MONETARY POLICY SHOCKS

In the recent empirical macroeconomic literature, structural factor analysis has become popular: using hundreds of observable economic indicators appears to overcome several difficulties associated with standard structural VAR modelling. In particular, bringing more information, while keeping the model parsimonious, may provide corrections for omitted and measurement errors; see Bernanke et al. (2005) and Forni, Giannone, Lippi and Reichlin (2009).

We reconsider the empirical study of Bernanke et al. (2005) with the same data, the same method to extract factors (principal components) and the same observed factor (Federal Funds Rate). So we set $D(L) = 0$ and $G = I$ in equations (4.9)-(4.10). The difference is that we estimate VARMA dynamics on static factors instead of imposing a finite-order VAR representation. The monetary policy shock is identified from the Cholesky decomposition of the residual covariance matrix in (4.10), where the observed factor is ordered last. We consider all four identified VARMA forms, but retain only the diagonal MA representation. The number of latent factors is set to five, and we estimate a VARMA (2.1) model [these orders were estimated using the information criterion in Dufour and Pelletier (2013)].

In Figure 1, we present FAVARMA$(2, 1)$-based impulse responses, with 90% confidence intervals (computed from 5000 bootstrap replications). A contractionary monetary policy shock generates a significant and very persistent economic downturn. The confidence intervals are more informative than those from FAVAR models. We conclude that impulse responses from a parsimonious 6-factor FAVARMA(2, 1) model provide a precise and plausible picture of the effect and transmission of monetary policy in the U.S.

In Figure 2, we compare the impulse responses to a monetary policy shock estimated from FAVAR and FAVARMA-DMA models. The FAVAR impulse coefficients were computed for several VAR orders. To get similar responses from a standard FAVAR model, the Akaike information criterion leads to a lag order of 14. So we need to estimate 84 coefficients governing the factors dynamics in the FAVARMA framework, while the FAVAR model requires 510 VAR parameters.

The approximation of the true factor process could be important when choosing the parametric bootstrap procedure to obtain statistical inference on objects of interest. The confidence intervals are produced as follows [see Yamamoto (2011) for theoretical justification of this bootstrap procedure].

Step 1 Shuffle the time periods, with replacement, of the residuals in (4.10) to get the bootstrap sample $\tilde{\eta}_t$. Then, resample static factors using estimated VARMA coefficients:

$$\tilde{F}_t = \hat{\Phi}(L)\tilde{F}_{t-1} + \hat{\Theta}\tilde{\eta}_t \,.$$

Step 2 Shuffle the time periods, with replacement, of the residuals in (4.8) to get the bootstrap sample $\tilde{u}_t$. Then resample the observable series using $\tilde{F}_t$ and the estimated loadings:

$$\tilde{X}_t = \hat{\Lambda}\tilde{F}_t + \tilde{u}_t \,.$$

Step 3 Estimate FAVARMA model on $\tilde{X}_t$, identify structural shocks and produce impulse responses.

9. CONCLUSION

In this paper, we have studied the relationship between VARMA and factor representations of a vector stochastic process and proposed the FAVARMA model. We started by observing that multivariate time series and their associated factors cannot in general both follow a finite-order VAR process. When the factors are obtained as linear combinations of observable series, the dynamic process of the latter has a VARMA structure, not a finite-order VAR form. In addition, even if the factors follow a finite-order VAR process, this implies a VARMA representation for the observable series. As a result, we proposed the FAVARMA framework, which combines two parsimonious methods to represent the dynamic interactions between a large number of time series: factor analysis and VARMA modeling.

To illustrate the performance of the proposed approach, we performed Monte Carlo simulations and found that VARMA modelling is quite helpful, especially in small samples cases – where the best improvement occurred at long horizons – but also in cases where the sample size is comparable to the one in our empirical data.

We applied our approach in an out-of-sample forecasting exercises based on a large U.S. monthly panel. The results show that VARMA factors help predict several key macroeconomic aggregates relative to standard factor models. In particular, FAVARMA models generally outperform FAVAR forecasting models, especially if we use MA VARMA-factor specifications.

Finally, we estimated the effect of monetary policy using the data and the identification scheme of Bernanke et al. (2005). We found that impulse responses from a parsimonious 6-factor FAVARMA (2.1) factor model yields a precise and plausible picture of the effect and the transmission of monetary policy in the U.S. To get similar responses from a standard FAVAR model, the Akaike information criterion leads to a lag order of 14. So we need to estimate 84 coefficients governing the factors dynamics in the FAVARMA framework, while the FAVAR model requires 510 parameters

References

Abraham, B. (1982), 'Temporal aggregation and time series', *International Statistical Review* **50**, 285–291.

Amengual, D. and Watson, M. (2007), 'Consistent estimation of the number of dynamic factors in large N and T panel', *Journal of Business and Economic Statistics* **25**(1), 91–96.

Aruoba, S., Diebold, F. and Scotti, C. (2009), 'Real-time measurement of business conditions', *Journal of Business and Economic Statistics* **27**, 417–427.

Bai, J. (2003), 'Inferential theory for factor models of large dimensions', *Econometrica* **71**(1), 135–172.

Bai, J. and Ng, S. (2002), 'Determining the number of factors in approximate factor models', *Econometrica* **70**(1), 191–221.

Bai, J. and Ng, S. (2007), 'Determining the number of primitive shocks', *Journal of Business and Economic Statistics* **25**(1), 52–60.

Bai, J. and Ng, S. (2008), 'Large dimensional factor analysis', *Foundations and Trends in Econometrics* **3**(2), 89–163.

Bai, J. and Ng, S. (2010), 'Instrumental variable estimation in a data-rich environment', *Econometric Theory* **26**(6), 1577–1606.

Banerjee, A., Marcellino, M. and Masten, I. (2006), Forecasting macroeconomic variables using diffusion indexes in short samples with structural change, *in* D. Rapach and M. Wohar, eds, 'Presence of Structural Breaks and Model Uncertainty'.

Bernanke, B., Boivin, J. and Eliasz, P. (2005), 'Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach', *Quarterly Journal of Economics* **120**, 387422.

Boivin, J. and Giannoni, M. (2006), DSGE models in a data-rich environment, Technical report, Columbia Business School, Columbia University.

Boivin, J., Giannoni, M. and Stevanović, D. (2009), Dynamic effects of credit shocks in a data-rich environment, Technical report, Columbia Business School, Columbia University.

Boivin, J., Giannoni, M. and Stevanović, D. (2009b), Monetary transmission in a small open economy: more data, fewer puzzles, Technical report, Columbia Business School, Columbia University.

Boivin, J. and Ng, S. (2005), 'Understanding and comparing factor-based forecasts', *International Journal of Central Banking* **1**, 117–151.

Boudjellaba, H., Dufour, J.-M. and Roy, R. (1992), 'Testing causality between two vectors in multivariate ARMA models', *Journal of the American Statistical Association* **87**(420), 1082–1090.

Chamberlain, G. and Rothschild, M. (1983), 'Arbitrage, factor structure and mean-variance analysis in large asset markets', *Econometrica* **51**(5), 1281–1304.

Deistler, M., Anderson, B., Filler, A., Zinner, C. and Chen, W. (2010), 'Generalized linear dynamic factor models - an approach via singular autoregressions', *European Journal of Control* **16:3**, 211–224.

Deistler, M., Filler, A. and Funovits, B. (2011), 'AR systems and AR processes: The singular case', *Communications in Information and Systems* **11:3**, 225–236.

Deistler, M. and Hannan, E. J. (1981), 'Some properties of the parameterization of ARMA systems with unknown order', *Journal of Multivariate Analysis* **11**, 474–484.

Dufour, J.-M. and Pelletier, D. (2013), Practical methods for modelling weak VARMA processes: Identification, estimation and specification with a macroeconomic application, Technical report, Department of Economics, McGill University, CIREQ and CIRANO, Montréal, Canada.

Dufour, J.-M. and Stevanović, D. (2013*a*), Factor-augmented VARMA models with macroeconomic applications, Technical report, Department of Economics, McGill University, Montréal, Canada.

Dufour, J.-M. and Stevanović, D. (2013*b*), Factor-augmented VARMA models with macroeconomic applications: Online appendix, Technical report, Department of Economics, McGill University, Montréal, Canada.

Favero, C., Marcellino, M. and Neglia, F. (2005), 'Principal components at work: The empirical analysis of monetary policy with large datasets', *Journal of Applied Econometrics* **20**, 603–620.

Forni, M., Giannone, D., Lippi, M. and Reichlin, L. (2009), 'Opening the black box: identifying shocks and propagation mechanisms in VAR and factor models', *Econometric Theory* **25**, 13191347.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000), 'The generalized factor model: identification and estimation', *The Review of Economics and Statistics* **82**, 540554.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2004), 'The generalized factor model: consistency and rates', *Journal of Econometrics* **119**, 231–255.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2005), 'The generalized dynamic factor model: one-sided estimation and forecasting', *Journal of the American Statistical Association* **100**, 830–839.

Geweke, J. (1977), The dynamic factor analysis of economic time series, *in* D. Aigner and A. Goldberger, eds, 'Latent Variables in Socio Economic Models', Vol. 77, North Holland, Amsterdam, pp. 304–313.

Giannone, D., Reichlin, L. and Small, D. (2008), 'Nowcasting: the real-time informational content of macroeconomic data', *Journal of Monetary Economics* **55**, 665–676.

Hallin, M. and Liska, R. (2007), 'Determining the number of factors in the general dynamic factor model', *Journal of the American Statistical Association* **102**, 603–617.

Hannan, E. J. and Deistler, M. (1988), *The Statistical Theory of Linear Systems*, John Wiley & Sons, New York.

Hannan, E. J. and Rissanen, J. (1982), 'Recursive estimation of mixed autoregressive-moving-average order', *Biometrika* **69**(1), 81–94. Errata 70 (1983), 303.

Kapetanios, G. and Marcellino, M. (2010), 'Factor-GMM estimation with large sets of possibly weak instruments', *Computational Statistics and Data Analysis* **54**, 2655–2675.

Komunjer, I. and Ng, S. (2011), 'Dynamic identification of dynamic stochastic general equilibrium models', *Econometrica* **79**, 1995–2032.

Lütkepohl, H. (1984), 'Forecasting contemporaneously aggregated vector ARMA processes', *Journal of Business and Economic Statistics* **2**, 201–214.

Lütkepohl, H. (1984b), 'Linear transformations of vector ARMA processes', *Journal of Econometrics* **26**, 283–293.

Lütkepohl, H. (1987), *Forecasting Aggregated Vector ARMA Processes*, Springer-Verlag, Berlin.

Lütkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.

Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Springer, Springer-Verlag, Berlin Heidelberg, Germany.

Marcellino, M., Stock, J. H. and Watson, M. W. (2006), 'A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series', *Journal of Econometrics* **135**, 499–526.

Onatski, A. (2009), 'A formal statistical test for the number of factors in the approximate factor models', *Econometrica* **77**(5), 1447–1480.

Onatski, A. (2009b), Asymptotics of the principal components estimator of large factor models with weak factors, Technical report, Department of Economics, Columbia University.

Poskitt, D. S. (2011), Vector autoregressive moving average identification for macroeconomic modeling: A new methodology, Technical report, Monash University.

Ravenna, F. (2006), 'Vector autoregressions and reduced form representation of dynamic stochastic general equilibrium models', *Journal of Monetary Economics* **54**(7), 715–719.

Rose, D. (1977), 'Forecasting aggregates of independent arima processes', *Journal of Econometrics* **5**, 323–345.

Sargent, T. and Sims, C. (1977), Business cycle modeling without pretending to have too much a priori economic theory, *in* C. Sims, ed., 'New Methods in Business Cycle Research', Federal Reserve Bank of Minneapolis, Minneapolis.

Stock, J. H. and Watson, M. W. (2002), 'Forecasting using principal components from a large number of predictors', *Journal of the American Statistical Association* **97**, 11671179.

Stock, J. H. and Watson, M. W. (2002b), 'Macroeconomic forecasting using diffusion indexes', *Journal of Business and Economic Statistics* **20**(2), 147–162.

Stock, J. H. and Watson, M. W. (2005), Implications of dynamic factor models for var analysis, Technical report, NBER WP 11467.

Stock, J. H. and Watson, M. W. (2006), An empirical comparison of methods for forecasting using many predictors, Technical report, Department of Economics, Harvard University.

Wei, W. S. (1978), *Some consequences of temporal aggregation in seasonal time series models*, U.S. Government Printing Office (U.S. Department of Commerce, Bureau of the Census), Washington, D.C.

Yamamoto, Y. (2011), Bootstrap inference for the impulse response functions in factor-augmented vector autoregressions, Technical report, Faculty of Business, University of Alberta, Edmonton, Alberta.

Zellner, A. and Palm, F. (1974), 'Time series analysis and simultaneous equation econometric model', *Journal of Econometrics* **2**(1), 17–54.

ACCEPTED MANUSCRIPT

APPENDIX

A. PROOFS

Proof of Theorem 3.1 Since Λ has full rank, we can multiply (3.1) by $(\Lambda'\Lambda)^{-1}\Lambda'$ to get

$$F_{t-1} = (\Lambda'\Lambda)^{-1}\Lambda' X_{t-1} - (\Lambda'\Lambda)^{-1}\Lambda' u_{t-1}. \tag{A.1}$$

If we now substitute F_{t-1} in (3.3), we see that

$$F_t = \Phi(L)(\Lambda'\Lambda)^{-1}\Lambda' X_{t-1} - \Phi(L)(\Lambda'\Lambda)^{-1}\Lambda' u_{t-1} + a_t,$$

hence, on substituting the latter expression for F_t in (3.1), and defining $A_1(L) = \Lambda\Phi(L)(\Lambda'\Lambda)^{-1}\Lambda'$,

$$X_t = \Lambda F_t + u_t = A_1(L)X_{t-1} + u_t - A_1(L)u_{t-1} + \Lambda a_t = A_1(L)X_{t-1} + A(L)u_t + \Lambda a_t = A_1(L)X_{t-1} + B(L)e_t$$

where $A(L) = I - A_1(L)L$ and $e_t = [u_t \vdots a_t]'$. This yields the representation (3.5).

We will now show that X_t can can be written as a VARMA process where the noise is the innovation process of X_t. Since X_t is regular strictly indeterministic weakly stationary process, it has a moving-average representation of the form (2.1) where $\varepsilon_t = X_t - P_L[X_t | X_{t-1}, X_{t-2}, \ldots]$ and $P_L[X_t | X_{t-1}, X_{t-2}, \ldots]$ is the best linear forecast of X_t based on its own past, $\Sigma_\varepsilon = E[\varepsilon_t \varepsilon_t']$ and $\det[\Sigma_\varepsilon] > 0$. Using the assumptions (3.2) and (3.4), it is easy to see that

$$\mathsf{E}[X_{t-j}u_t'] = \mathsf{E}[X_{t-j}a_t'] = \mathsf{E}[u_t\varepsilon_{t-j}'] = \mathsf{E}[a_t\varepsilon_{t-j}'] = 0 \text{ for } j \geq 1. \tag{A.2}$$

Then

$$A(L)X_t = A(L)\Psi(L)\varepsilon_t = \bar{\Psi}(L)\varepsilon_t = \sum_{j=0}^{\infty} \bar{\Psi}_j \varepsilon_{t-j} \tag{A.3}$$

where $\bar{\Psi}_j = \sum_{i=0}^{p+1} A_i \Psi_{j-i}$ and $\Psi_s = 0$ for $s < 0$, $s = j - i$. Let us now multiply $A(L)X_t$ by ε'_{t-k} and take the expected value: using (A.3) and (3.5), we get

$$\mathsf{E}[A(L)X_t\varepsilon'_{t-k}] = \sum_{j=0}^{\infty} \bar{\Psi}_j \mathsf{E}[\varepsilon_{t-j}\varepsilon'_{t-k}] = B_j\Sigma_\varepsilon \tag{A.4}$$

$$= \mathsf{E}[(A(L)u_t + \Lambda a_t)\varepsilon'_{t-k}] = 0 \text{ for } k > p+1, \tag{A.5}$$

hence $\bar{\Psi}_j = 0$ for $k > p+1$, so that X_t has the following VARMA$(p+1,\, p+1)$ representation:

$$A(L)X_t = \bar{\Psi}(L)\varepsilon_t \tag{A.6}$$

where $\bar{\Psi}(L) = \sum_{j=0}^{p+1} \bar{\Psi}_j L^j$.

Proof of Theorem 3.2 To obtain the representations of X_t, we follow the same steps as in the previous proof except we substitute (A.1) for F_{t-1} in (3.7), which yields

$$X_t = \Lambda\Phi(L)(\Lambda'\Lambda)^{-1}\Lambda' X_{t-1} + u_t - \Lambda\Phi(L)(\Lambda'\Lambda)^{-1}\Lambda' u_{t-1} + \Lambda\Theta(L)a_t.$$

Defining $A(L)$ and e_t as above, with $B(L) = [A(L) \vdots \Lambda\Theta(L)]$, gives the representation as in (3.5). Then, remark that (A.5) becomes

$$\mathsf{E}[(A(L)u_t + \Lambda\Theta(L)a_t)\varepsilon'_{t-k}] = 0 \text{ for } k > \max(p+1,\, q)\,, \tag{A.7}$$

so X_t has a VARMA$(p+1, \max(p+1, q))$.

Proof of Theorem 3.3 $F_t = CX_t$, where C is a $K \times N$ full row rank matrix. Properties (i) and (ii) are easily proved using Lütkepohl (2005, Corollaries 11.1.1 and 11.1.2). For (iii), if X_t has an MA representation as in (2.1) or (2.4), the result is obtained using Lütkepohl (1987, Propositions 4.1 and 4.2).

B. SIMULATION RESULTS: FAVARMA AND FAVAR FORECASTS

Table 1 contains the results of the Monte Carlo simulation exercise presented in Section 6.2. The numbers represent the MSE of four FAVARMA identified forms over the MSE of FAVAR direct forecasting models.

Table B.1: Comparison between FAVARMA and FAVAR forecasts: Monte Carlo simulations

	$\rho_T = 0.9, \rho_N = 0.5$							
	$T = 50$, $N = 50$				$T = 50$, $N = 100$			
Horizon	Diag MA	Diag AR	Final MA	Final AR	Diag MA	Diag AR	Final MA	Final AR
1	1.0078	1.1405	0.9235	1.3858	1.0061	1.0945	0.9084	1.4722
2	1.0199	1.0852	0.9483	1.3189	1.0302	1.0762	0.9383	1.3660
4	0.8872	0.9459	0.8350	1.0746	0.9338	1.0242	0.8745	1.1542
6	0.8122	0.9181	0.7635	0.9536	0.8514	0.9375	0.7954	1.0010
12	0.6311	0.8392	0.6072	0.7198	0.6857	0.9278	0.6533	0.8036
18	0.4913	0.7186	0.4754	0.5339	0.5181	0.8285	0.4955	0.5744
24	0.3762	0.6192	0.3706	0.4237	0.3846	0.7215	0.3788	0.4291
36	0.1394	0.2429	0.1369	0.1480	0.1445	0.3006	0.1422	0.1560
	$T = 100$, $N = 50$				$T = 600$, $N = 130$			
1	1.0761	1.1170	1.0004	1.6656	1.0130	1.0126	1.0093	1.0070
2	1.0865	1.1495	1.0193	1.5676	0.9962	0.9956	0.9952	0.9951
4	1.0537	1.0890	1.0038	1.4432	0.9945	0.9950	0.9947	0.9947
6	1.0168	1.0392	0.9686	1.3060	0.9945	0.9954	0.9946	0.9946
12	0.9183	0.9915	0.8960	1.2573	0.9871	0.9883	0.9873	0.9873
18	0.8886	0.9848	0.8552	1.1123	0.9831	0.9880	0.9832	0.9832
24	0.8643	0.9706	0.8198	1.1203	0.9831	0.9830	0.9828	0.9828
36	0.8078	0.9754	0.7956	1.0742	0.9863	0.9846	0.9847	0.9847

	$\rho_T = 0.9, \rho_N = 0.1$							
	$T = 50$, $N = 50$				$T = 50$, $N = 100$			
Horizon	Diag MA	Diag AR	Final MA	Final AR	Diag MA	Diag AR	Final MA	Final AR
1	1.0203	1.0656	0.8897	1.2688	0.9977	1.0303	0.9026	1.3464
2	0.9689	1.0113	0.8982	1.1708	1.0013	1.0406	0.9038	1.1735
4	0.9142	0.9508	0.8616	1.0391	0.9032	0.9166	0.8461	1.0029
6	0.8420	0.8656	0.7851	0.9213	0.8841	0.8798	0.8054	0.9182
12	0.6401	0.7487	0.6235	0.7038	0.7042	0.8089	0.6850	0.7582
18	0.5208	0.6774	0.5133	0.5609	0.5469	0.6970	0.5296	0.5742
24	0.4095	0.5979	0.4124	0.4322	0.4380	0.5724	0.4282	0.4499
36	0.1417	0.2169	0.1402	0.1447	0.1453	0.2152	0.1424	0.1535
	$T = 100$, $N = 50$				$T = 600$, $N = 130$			
1	1.0622	1.0751	0.9990	1.3846	0.9978	0.9980	0.9984	0.9927
2	1.0578	1.0368	0.9913	1.2818	0.9935	0.9951	0.9935	0.9933
4	1.0254	1.0088	0.9729	1.2141	0.9890	0.9894	0.9891	0.9891
6	1.0058	0.9720	0.9477	1.1812	0.9892	0.9892	0.9892	0.9892
12	0.9480	0.9163	0.8819	1.0303	0.9919	0.9918	0.9919	0.9919
18	0.9371	0.9068	0.8823	1.0173	0.9784	0.9784	0.9784	0.9784
24	0.9441	0.8755	0.8626	1.0214	0.9807	0.9807	0.9807	0.9807
36	0.8591	0.8376	0.8013	0.9264	0.9796	0.9796	0.9796	0.9796

ACCEPTED MANUSCRIPT

Table B.1: Monte Carlo simulation results (continued)

	$\rho_T = 0.1, \rho_N = 0.9$							
	$T = 50$, $N = 50$				$T = 50$, $N = 100$			
Horizon	Diag MA	Diag AR	Final MA	Final AR	Diag MA	Diag AR	Final MA	Final AR
1	0.8978	0.9108	0.8924	0.9362	0.9329	0.8880	0.8585	0.9208
2	0.8522	0.8716	0.8606	0.9168	0.8289	0.8194	0.8228	0.8642
4	0.8381	0.8420	0.8524	0.8601	0.8195	0.8213	0.8187	0.8238
6	0.8213	0.8227	0.8225	0.8210	0.7852	0.7806	0.7799	0.7795
12	0.7923	0.7906	0.7905	0.7907	0.7630	0.7569	0.7567	0.7568
18	0.6803	0.6770	0.6771	0.6772	0.6582	0.6576	0.6577	0.6577
24	0.5367	0.5363	0.5364	0.5364	0.4865	0.4864	0.4863	0.4862
36	0.0946	0.0956	0.0944	0.0944	0.0801	0.0799	0.0799	0.0800
	$T = 100$, $N = 50$				$T = 600$, $N = 130$			
1	0.9680	0.9676	0.9560	0.9515	0.9931	0.9995	0.9926	0.9921
2	0.9332	0.9304	0.9306	0.9310	0.9881	0.9929	0.9882	0.9878
4	0.9338	0.9261	0.9257	0.9257	0.9882	0.9895	0.9893	0.9894
6	0.9467	0.9350	0.9351	0.9351	0.9831	0.9831	0.9830	0.9830
12	0.9358	0.9359	0.9359	0.9359	0.9825	0.9825	0.9825	0.9825
18	0.9297	0.9298	0.9297	0.9297	0.9874	0.9873	0.9873	0.9873
24	0.9140	0.9142	0.9143	0.9143	0.9887	0.9886	0.9886	0.9886
36	0.9044	0.9047	0.9043	0.9043	0.9929	0.9930	0.9930	0.9930

	$\rho_T = 0.1, \rho_N = 0.1$							
	$T = 50$, $N = 50$				$T = 50$, $N = 100$			
Horizon	Diag MA	Diag AR	Final MA	Final AR	Diag MA	Diag AR	Final MA	Final AR
1	0.9439	0.8761	0.7969	0.9618	0.9289	0.8919	0.8155	0.9675
2	0.8029	0.7863	0.7764	0.8459	0.7888	0.7900	0.7736	0.8569
4	0.7894	0.7542	0.7533	0.7742	0.7513	0.7533	0.7525	0.7687
6	0.7580	0.7420	0.7409	0.7438	0.7477	0.7488	0.7446	0.7458
12	0.6773	0.6751	0.6751	0.6754	0.6575	0.6604	0.6560	0.6613
18	0.5757	0.5700	0.5701	0.5761	0.5741	0.5753	0.5704	0.5728
24	0.4106	0.4074	0.4073	0.4084	0.4329	0.4303	0.4304	0.4317
36	0.0726	0.0721	0.0721	0.0721	0.0719	0.0722	0.0721	0.0721
	$T = 100$, $N = 50$				$T = 600$, $N = 130$			
1	0.9702	0.9672	0.9290	0.9316	0.9838	0.9874	0.9868	0.9840
2	0.8998	0.9053	0.8985	0.8993	0.9816	0.9904	0.9811	0.9811
4	0.9095	0.9003	0.9000	0.8997	0.9891	0.9894	0.9891	0.9891
6	0.8806	0.8771	0.8767	0.8767	0.9821	0.9822	0.9821	0.9821
12	0.8855	0.8841	0.8839	0.8839	0.9778	0.9778	0.9778	0.9778
18	0.8725	0.8704	0.8702	0.8702	0.9852	0.9852	0.9852	0.9852
24	0.8711	0.8707	0.8709	0.8709	0.9815	0.9815	0.9815	0.9815
36	0.8183	0.8185	0.8183	0.8183	0.9790	0.9790	0.9790	0.9790

Table 1: RMSE relative to direct AR(p) forecasts

Industrial production growth rate: total										
Horizon	Unrestricted	DI	DI AR	Direct	Sequential	Diag MA	Diag AR	Final MA	Final AR	ARMA
1	0.8706	**0.8457**	0.8958	0.9443	0.9443	0.8971	0.9019	0.9132	0.8985	0.9700
2	1.0490	0.9938	1.0106	1.0157	1.0665	**0.9074**	0.9202	0.9112	0.9123	1.0026
4	1.1934	1.0411	1.0527	1.0711	1.2214	**0.8947**	0.9906	0.8970	0.9481	0.9710
6	1.1496	1.0238	1.0245	1.1743	1.3528	0.9248	1.0494	**0.9202**	0.9847	0.9918
12	1.2486	1.0445	1.0389	1.0933	1.3682	1.0008	1.2215	1.0075	1.0371	**0.9713**
18	1.0507	1.0048	1.0207	1.0662	1.2508	1.0511	1.5098	1.0615	1.1206	**0.9910**
24	1.0393	1.0628	1.0748	1.0128	1.0863	0.9858	1.7920	0.9959	1.1061	**0.9604**
36	1.0092	1.0906	1.1437	1.2364	1.0421	0.9855	3.0304	0.9883	1.1795	**0.9826**
48	1.0147	1.1110	1.1212	1.1063	1.0355	0.9921	5.5321	0.9922	1.1681	**0.9856**
Civilian labor force growth rate: employed. total										
Horizon	Unrestricted	DI	DI AR	Direct	Sequential	Diag MA	Diag AR	Final MA	Final AR	ARMA
1	0.8264	0.8832	0.8451	0.8202	0.8202	**0.8004**	0.8075	0.8027	0.8008	1.0496
2	0.9407	0.9391	0.9381	0.9477	0.9591	0.8931	**0.8805**	0.8961	0.8852	1.0422
4	0.9766	0.9739	0.9937	1.0204	1.0551	0.9213	0.8997	0.9200	**0.8991**	0.9993
6	1.0776	1.0799	1.0937	1.0714	1.1550	0.9667	0.9526	0.9636	**0.9455**	1.0032
12	1.0741	1.0742	1.0722	1.0137	1.1654	0.9718	0.9912	0.9704	0.9558	**0.9507**
18	1.0471	1.0488	1.0472	0.9735	1.1391	1.0073	1.1386	1.0096	1.0391	**0.9721**
24	1.0237	1.0580	1.0268	**0.9641**	1.1002	1.0154	1.2806	1.0177	1.0856	0.9893
36	0.9573	0.9099	0.9703	0.9507	0.9477	0.9070	1.5452	0.9043	1.0098	**0.8957**
48	**0.9227**	0.9236	0.9250	0.9576	0.9989	0.9652	2.4022	0.9624	1.0482	0.9550
Consumer price index growth rate: all items										
Horizon	Unrestricted	DI	DI AR	Direct	Sequential	Diag MA	Diag AR	Final MA	Final AR	ARMA
1	0.8806	**0.8700**	**0.8700**	0.9228	0.9228	0.9144	0.9432	0.8856	0.9072	1.0143
2	0.9866	0.9942	0.9942	0.9612	0.9730	0.9309	0.9427	0.9274	**0.9170**	0.9856
4	1.0656	1.0732	1.0732	1.0398	1.0170	1.0007	1.0665	0.9895	**0.9792**	1.0129
6	1.1343	1.1334	1.1334	1.0349	1.0101	0.9946	1.0752	0.9939	**0.9928**	1.0364
12	1.1173	1.1279	1.1279	1.0821	**0.9513**	0.9572	1.1958	0.9553	1.0408	1.0297
18	1.0311	1.0379	1.0379	1.0430	0.9654	**0.8894**	1.1021	0.8909	0.9673	0.9391
24	0.9644	1.0712	1.0712	0.9510	0.9980	0.8819	1.1851	**0.8791**	0.9713	0.8805
36	0.7645	**0.7627**	**0.7627**	0.9870	0.9470	0.8329	1.4591	0.8385	0.9126	0.8619
48	0.8663	0.8488	0.8488	0.9361	0.9536	**0.8292**	2.2640	0.8335	0.8864	0.8511

Note – The numbers in bold character present the model producing the best forecasts in terms of MSE.

Table 2: RMSE relative to ARMA(p, q) forecasts

Industrial production growth rate: total									
Horizon	Unrestricted	DI	DI AR	Direct	Sequential	Diag MA	Diag AR	Final MA	Final AR
1	0.8975	0.8719	0.9235	0.9735	0.9735	0.9248	0.9298	0.9414	0.9263
2	**1.0463**	0.9912	**1.0080**	**1.0131**	**1.0637**	0.9050	0.9178	0.9088	0.9099
4	**1.2290**	**1.0722**	**1.0841**	**1.1031**	**1.2579**	0.9214	**1.0202**	0.9238	0.9764
6	**1.1591**	**1.0323**	**1.0330**	**1.1840**	**1.3640**	0.9324	**1.0581**	0.9278	0.9928
12	**1.2855**	**1.0754**	**1.0696**	**1.1256**	**1.4086**	**1.0304**	**1.2576**	**1.0373**	**1.0677**
18	**1.0602**	**1.0139**	**1.0300**	**1.0759**	**1.2622**	**1.0606**	**1.5235**	**1.0711**	**1.1308**
24	**1.0822**	**1.1066**	**1.1191**	**1.0546**	**1.1311**	**1.0264**	**1.8659**	**1.0370**	**1.1517**
36	**1.0271**	**1.1099**	**1.1640**	**1.2583**	**1.0606**	**1.0030**	**3.0841**	**1.0058**	**1.2004**
48	**1.0295**	**1.1272**	**1.1376**	**1.1225**	**1.0506**	**1.0066**	**5.6129**	**1.0067**	**1.1852**
Civilian labor force growth rate: employed. total									
Horizon	Unrestricted	DI	DI AR	Direct	Sequential	Diag MA	Diag AR	Final MA	Final AR
1	0.7873	0.8415	0.8052	0.7814	0.7814	0.7626	0.7693	0.7648	0.7630
2	0.9026	0.9011	0.9001	0.9093	0.9203	0.8569	0.8448	0.8598	0.8494
4	0.9773	0.9746	0.9944	**1.0211**	**1.0558**	0.9219	0.9003	0.9206	0.8997
6	**1.0742**	**1.0765**	**1.0902**	**1.0680**	**1.1513**	0.9636	0.9496	0.9605	0.9425
12	**1.1298**	**1.1299**	**1.1278**	**1.0663**	**1.2258**	**1.0222**	**1.0426**	**1.0207**	**1.0054**
18	**1.0772**	**1.0789**	**1.0773**	**1.0014**	**1.1718**	**1.0362**	**1.1713**	**1.0386**	**1.0689**
24	**1.0348**	**1.0694**	**1.0379**	0.9745	**1.1121**	**1.0264**	**1.2945**	**1.0287**	**1.0973**
36	**1.0688**	**1.0159**	**1.0833**	**1.0614**	**1.0581**	**1.0126**	**1.7251**	**1.0096**	**1.1274**
48	0.9662	0.9671	0.9686	**1.0027**	**1.0460**	**1.0107**	**2.5154**	**1.0077**	**1.0976**
Consumer price index growth rate: all items									
Horizon	Unrestricted	DI	DI AR	Direct	Sequential	Diag MA	Diag AR	Final MA	Final AR
1	0.8682	0.8577	0.8577	0.9098	0.9098	0.9015	0.9299	0.8731	0.8944
2	**1.0010**	**1.0087**	**1.0087**	0.9752	0.9872	0.9445	0.9565	0.9409	0.9304
4	**1.0520**	**1.0595**	**1.0595**	**1.0266**	**1.0040**	0.9880	**1.0529**	0.9769	0.9667
6	**1.0945**	**1.0936**	**1.0936**	0.9986	0.9746	0.9597	**1.0374**	0.9590	0.9579
12	**1.0851**	**1.0954**	**1.0954**	**1.0509**	0.9239	0.9296	**1.1613**	0.9277	**1.0108**
18	**1.0980**	**1.1052**	**1.1052**	**1.1106**	**1.0280**	0.9471	**1.1736**	0.9487	**1.0300**
24	**1.0953**	**1.2166**	**1.2166**	**1.0801**	**1.1334**	**1.0016**	**1.3459**	0.9984	**1.1031**
36	0.8870	0.8849	0.8849	**1.1451**	**1.0987**	0.9664	**1.6929**	0.9729	**1.0588**
48	**1.0179**	0.9973	0.9973	**1.0999**	**1.1204**	0.9743	**2.6601**	0.9793	**1.0415**

Note – The numbers in bold character present cases where the ARMA model outperforms the factor-based alternatives in terms of MSE.

ACCEPTED MANUSCRIPT

Table 3: MSE of FAVARMA relative to FAVAR forecasting models

	Industrial production growth rate: total							
	VARMA/Direct				VARMA/Sequential			
Horizon	Diag MA	Diag AR	Final MA	Final AR	Diag MA	Diag AR	Final MA	Final AR
1	**0.9500**	**0.9551**	**0.9671**	**0.9515**	**0.9500**	**0.9551**	**0.9671**	**0.9515**
2	**0.8934**	**0.9060**	**0.8971**	**0.8982**	**0.8508**	**0.8628**	**0.8544**	**0.8554**
4	**0.8353**	**0.9248**	**0.8375**	**0.8852**	**0.7325**	**0.8110**	**0.7344**	**0.7762**
6	**0.7875**	**0.8936**	**0.7836**	**0.8385**	**0.6836**	**0.7757**	**0.6802**	**0.7279**
12	**0.9154**	1.1173	**0.9215**	**0.9486**	**0.7315**	**0.8928**	**0.7364**	**0.7580**
18	**0.9858**	1.4161	**0.9956**	1.0510	**0.8403**	1.2071	**0.8487**	**0.8959**
24	**0.9733**	1.7694	**0.9833**	1.0921	**0.9075**	1.6496	**0.9168**	1.0182
36	**0.7971**	2.4510	**0.7993**	**0.9540**	**0.9457**	2.9080	**0.9484**	1.1318
48	**0.8968**	5.0005	**0.8969**	1.0559	**0.9581**	5.3424	**0.9582**	1.1281
	Civilian labor force growth rate: employed. total							
	VARMA/Direct				VARMA/Sequential			
Horizon	Diag MA	Diag AR	Final MA	Final AR	Diag MA	Diag AR	Final MA	Final AR
1	**0.9759**	**0.9845**	**0.9787**	**0.9763**	**0.9759**	**0.9845**	**0.9787**	**0.9763**
2	**0.9424**	**0.9291**	**0.9456**	**0.9341**	**0.9312**	**0.9180**	**0.9343**	**0.9229**
4	**0.9029**	**0.8817**	**0.9016**	**0.8811**	**0.8732**	**0.8527**	**0.8720**	**0.8521**
6	**0.9023**	**0.8891**	**0.8994**	**0.8825**	**0.8370**	**0.8248**	**0.8343**	**0.8186**
12	**0.9587**	**0.9778**	**0.9573**	**0.9429**	**0.8339**	**0.8505**	**0.8327**	**0.8201**
18	1.0347	1.1696	1.0371	1.0674	**0.8843**	**0.9996**	**0.8863**	**0.9122**
24	1.0532	1.3283	1.0556	1.1260	**0.9229**	1.1640	**0.9250**	**0.9867**
36	**0.9540**	1.6253	**0.9512**	1.0622	**0.9571**	1.6305	**0.9542**	1.0655
48	1.0079	2.5086	1.0050	1.0946	**0.9663**	2.4048	**0.9635**	1.0494
	Consumer price index growth rate: all items							
	VARMA/Direct				VARMA/Sequential			
Horizon	Diag MA	Diag AR	Final MA	Final AR	Diag MA	Diag AR	Final MA	Final AR
1	**0.9909**	1.0221	**0.9597**	0.9831	**0.9909**	1.0221	**0.9597**	**0.9831**
2	**0.9685**	**0.9808**	**0.9648**	0.9540	**0.9567**	**0.9689**	**0.9531**	**0.9424**
4	**0.9624**	1.0257	**0.9516**	0.9417	**0.9840**	1.0487	**0.9730**	**0.9628**
6	**0.9611**	1.0389	**0.9604**	0.9593	**0.9847**	1.0644	**0.9840**	**0.9829**
12	**0.8846**	1.1051	**0.8828**	0.9618	1.0062	1.2570	1.0042	1.0941
18	**0.8527**	1.0567	**0.8542**	0.9274	**0.9213**	1.1416	**0.9228**	1.0020
24	**0.9273**	1.2462	**0.9244**	1.0213	**0.8837**	1.1875	**0.8809**	**0.9732**
36	**0.8439**	1.4783	**0.8495**	0.9246	**0.8795**	1.5408	**0.8854**	**0.9637**
48	**0.8858**	2.4185	**0.8904**	0.9469	**0.8695**	2.3742	**0.8741**	**0.9295**

Note – The numbers in bold character present cases where the FAVARMA model performs better than the FAVAR.

ACCEPTED MANUSCRIPT

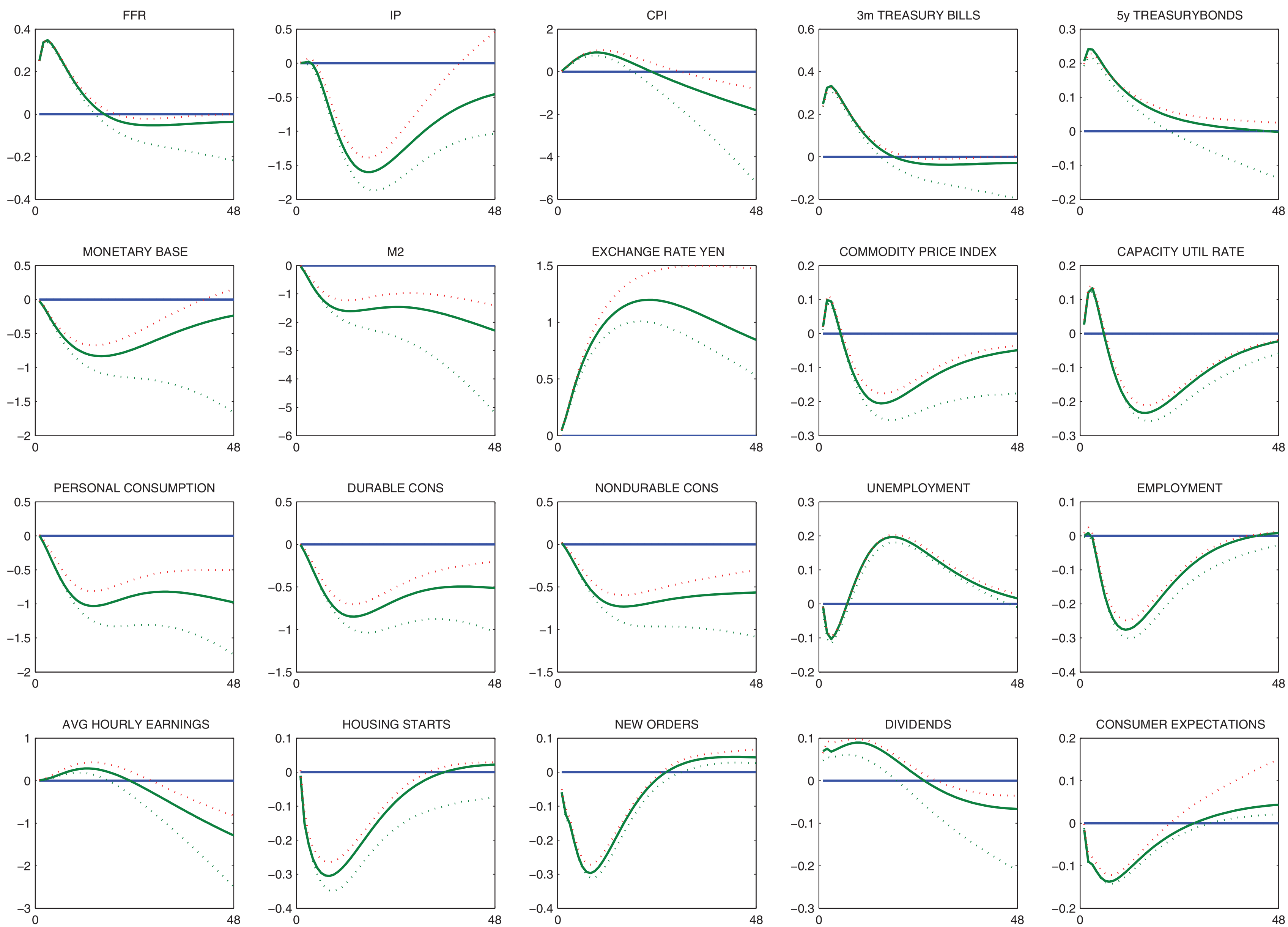

Figure 1: FAVARMA-DMA impulse responses to monetary policy shock

ACCEPTED MANUSCRIPT

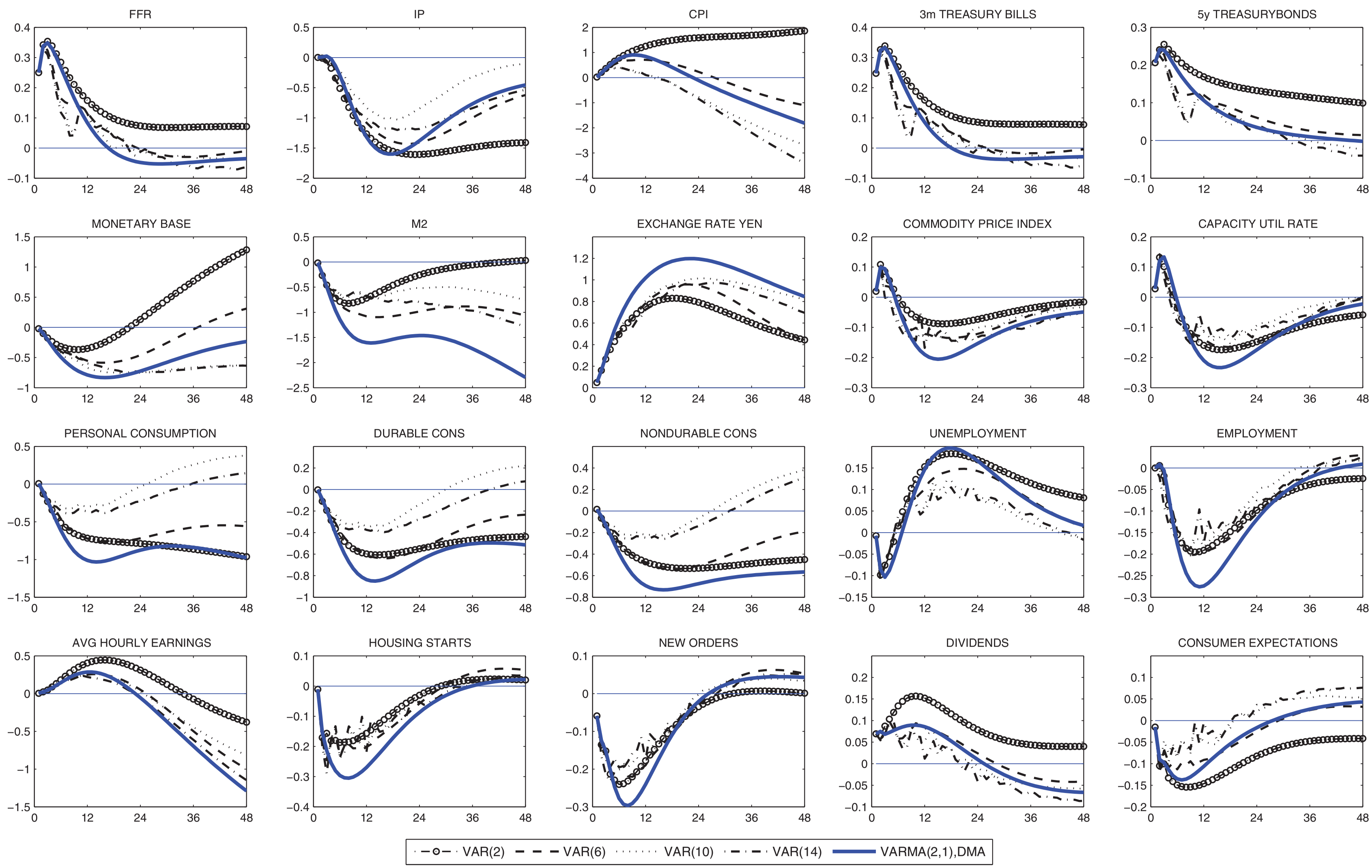

Figure 2: Comparison between FAVAR and FAVARMA impulse responses to a monetary policy shock

ACCEPTED MANUSCRIPT