Federal Reserve Bank of Dallas
Globalization and Monetary Policy Institute
Working Paper No. 213
http://www.dallasfed.org/assets/documents/institute/wpapers/2014/0213.pdf

A Multi-Country Approach to Forecasting Output Growth Using PMIs*

Alexander Chudik
Federal Reserve Bank of Dallas

Valerie Grossman
Federal Reserve Bank of Dallas

M. Hashem Pesaran
University of Southern California and Trinity College

November 2014

Abstract

This paper derives new theoretical results for forecasting with Global VAR (GVAR) models. It is shown that the presence of a strong unobserved common factor can lead to an undetermined GVAR model. To solve this problem, we propose augmenting the GVAR with additional proxy equations for the strong factors and establish conditions under which forecasts from the augmented GVAR model (AugGVAR) uniformly converge in probability (as the panel dimensions $N,T\rightarrow\infty$ such that $N/T\rightarrow\varkappa$ for some $0<\varkappa<\infty$) to the infeasible optimal forecasts obtained from a factor-augmented high-dimensional VAR model. The small sample properties of the proposed solution are investigated by Monte Carlo experiments as well as empirically. In the empirical part, we investigate the value of the information content of Purchasing Managers Indices (PMIs) for forecasting global (48 countries) growth, and compare forecasts from AugGVAR models with a number of data-rich forecasting methods, including Lasso, Ridge, partial least squares and factor-based methods. It is found that (a) regardless of the forecasting methods considered, PMIs are useful for nowcasting, but their value added diminishes quite rapidly with the forecast horizon, and (b) AugGVAR forecasts do as well as other data-rich forecasting techniques for short horizons, and tend to do better for longer forecast horizons.

JEL codes: C53, E37

*Alexander Chudik, Research Department, Federal Reserve Bank of Dallas, 2200 N. Pearl Street, Dallas, 75201. 214-922-5769. alexander.chudik@dal.frb.org. Valerie Grossman, Research Department, Federal Reserve Bank of Dallas, 2200 N. Pearl Street, Dallas, TX 75201. 214-922-5769. valerie.grossman@dal.frb.org. M. Hashem Pesaran, Department of Economics, University of Southern California, 3620 South Vermont Avenue, Los Angeles, CA 90089. 213-740-3510. pesaran@usc.edu. We would like to thank Kundan Kishor, Evan Koenig, Kajal Lahiri, Ron Smith, Mark Wynne, participants at the 8th ECB Workshop on Forecasting Techniques, and participants at the Federal Reserve Bank of Dallas brownbag seminar for helpful comments and suggestions. The views in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Dallas or the Federal Reserve System.

1 Introduction

International datasets with relatively large cross-section (N) and time (T) dimensions are becoming increasingly available and frequently used in practice. How to work with such large datasets has been the subject of intensive research in the past decades. On the one hand, individual economies in the global system are interdependent, and a general linear dynamic framework such as high-dimensional VARs seems to be appropriate. On the other hand, estimating high-dimensional VARs is not feasible since the number of coefficients to be estimated grows at a quadratic rate with the number of variables. This problem, also known as the curse of dimensionality, has been addressed in the literature in a number of ways, but primarily in a static setting. In this paper we focus on the Global VAR modeling approach (or GVAR for short) which has been applied extensively to multi-country data sets and is designed to deal with the curse of dimensionality in dynamic contexts.

The GVAR model was proposed by Pesaran et al. (2004) and provides a feasible and coherent global reduced-form VAR representation of the data. It deals with the dimensionality problem by estimating small-scale individual country VARX* models, where domestic variables are regressed on country-specific weighted averages of foreign variables, which are treated as weakly exogenous for the purpose of estimation. The individual country VARX* models are then solved in the form of a high-dimensional VAR representation that includes all the endogenous variables of the world economy. The structure embodied in the GVAR allows for quite complex interlinkages amongst the variables (within as well as across economies), while being sufficiently compact and easy to use in forecasting, simulation and counterfactual analyses. There are numerous applications of the GVAR approach in the literature, including in the field of forecasting. Chudik and Pesaran (2014b) provide a recent survey.

In this paper we establish conditions under which forecasts from the GVAR model converge to optimal *infeasible* forecasts (as $N, T \xrightarrow{j} \infty$, such that $N/T \to \varkappa$, for some $0 < \varkappa < \infty$) when the data is generated from a high-dimensional VAR model containing unobserved common factors. It is shown that the presence of strong unobserved common factors can lead to an undetermined GVAR model with a singular contemporaneous coefficient matrix. To deal with this problem, we propose augmenting the GVAR with a sub-model for the unobserved factors that we proxy by cross-section averages. We refer to this augmented GVAR model as AugGVAR for short, and show that augmentation is effective regardless of how factors are introduced in the underlying model. Specifically, we consider two methods of augmenting VARs with factors: (i) modeling deviations from the factors as a VAR (as in Dées et al. (2007)), (ii) adding factors to the errors of the VAR model. Since factors are unobserved, we consider both specifications and provide results that are robust to the way factors are introduced in the underlying high-dimensional VAR model. We also show that only the knowledge of the maximum number of unobserved common factors ($m_{\max}$) is needed, and there is no need to identify and estimate the exact number of factors. This means that in practice it is sufficient to augment the GVAR with a sub-model in terms of $m_{\max}$ cross-section averages.

We investigate the small sample properties of the proposed approach by Monte Carlo (MC) experiments. We find that small sample performance of the AugGVAR is at least as good as the GVAR when there is no factor in the underlying model, and substantially better when factors are present. MC experiments also show that using an undetermined GVAR in the presence of factors can have serious consequences for forecasting, particularly when the time dimension is not sufficiently large. Overall, the AugGVAR is recommended irrespective of whether the underlying high-dimensional VAR contains an unobserved common factor or not.

The effectiveness of the proposed approach is also illustrated in the empirical application, where we forecast output growth across 48 countries using Purchasing Managers Indices (or PMIs for short). PMI data releases are closely watched by financial market participants for signs of improving or deteriorating economic conditions. PMIs are available across a broad range of countries in a timely manner (released monthly and with short time delay), and are considered important indicators of the current level of output growth, on which official data is often released with a considerable time delay. There is indeed a close resemblance between year-on-year economic growth and PMIs, as is evident from Figure 1, which plots data aggregated across countries at the global level. However, the usefulness of PMIs in forecasting quarterly output growth, over and above the past history of output growth rates themselves, can only be ascertained by using conditional models where forecasts are computed with and without conditioning on PMIs.

Besides the dimensionality problem, the empirical application presents us with two additional challenges.[1] The data release lags vary across countries and types of variables, and therefore forecasting of output growth, if to be carried out efficiently in real time, must be done conditional on information sets with different end dates (what we refer to as "nonsynchronous conditioning" information sets).[2] The second challenge is that PMIs are observed at a higher frequency than output growth. The literature has tackled these challenges in a number of different ways. One approach is to model all variables – PMIs and output growth – in one system written in a state-space form at the highest frequency. The problems of nonsynchronous data releases and mixed frequencies are then translated into a missing data problem, which is overcome with the use of a Kalman filter and smoother (see Evans (2005) and Giannone, Reichlin, and Small (2008)).[3] The second approach is to temporally aggregate the high-frequency PMIs into the low frequency of the output growth variable and then estimate a forecasting equation at the low frequency. Examples of this approach include Trehan (1989), Parigi and Schlitzer (1995), Kitchen and Monaco (2003), Rünstler and Sédillot (2003), Baffigi et al. (2004), Parigi and Golinelli (2007), and Diron (2008). The third

[1] The focus of our analysis is forecasting economic growth in a sample of 48 countries (with PMI manufacturing and services indices being available for a subset of the 48 countries) that comprise 92% of the world's economic output in 2013 nominal prices. We have 87 predictors for forecasting economic growth in a particular country (or region), and the use of methods that allow for a large number of predictors is therefore necessary.

[2] Nonsynchronous conditioning information sets are also called 'ragged edge' information sets in the literature, see Wallis (1986).

[3] This approach has been applied in a number of recent papers, see Rünstler et al. (2009), Angelini et al. (2010), Barhoumi et al. (2010), Camacho and Perez-Quiros (2010), Matheson (2010), Yiu and Chow (2010), Angelini et al. (2011), Arnostova et al. (2011), Bańbura and Rünstler (2011), de Winter (2011), Aastveit and Trovik (2012), D'Agostino et al. (2012), Siliverstovs (2012), Siliverstovs and Kholodilin (2012), Lahiri and Monokroussos (2013), and Bańbura and Modugno (2014).

approach is the mixed-frequency data sampling (MIDAS) regression introduced by Ghysels et al. (2004) and later extended by Ghysels et al. (2007), with applications by Clements and Galvão (2008 and 2009), Andreou et al. (2010), Marcellino and Schumacher (2010), and Kuzin et al. (2011 and 2013).

In this paper, we follow the simple approach of aggregating PMI data into a quarterly frequency, and show how to derive conditional forecasts using GVAR or AugGVAR models in the case of nonsynchronous conditioning information sets. Aggregation of PMI data into the frequency of output growth allows us to readily implement other data-rich forecasting methods as well. As an alternative to GVAR forecasts, we consider a number of commonly used methods for forecasting with a large number of predictors. In particular, we implement the Lasso, Ridge, factor models (FM), factor-augmented autoregressions (FAR), and partial least squares (PLS) methods, which are widely used in the forecasting literature (see for example reviews by Eklund and Kapetanios (2008) and Groen and Kapetanios (2008)).[4] We describe individual methods in more detail and provide references to the literature in Section 8.2.

We find that regardless of the particular forecasting method employed, the information contained in PMIs substantially improves output growth forecasts for different months within the current quarter ($h = 0$). This result is robust across the countries and methods considered. We obtain about 15-20% reduction in the cross country PPP-GDP weighted average of the mean square forecast errors over the out-of-sample forecast evaluation period of 2006Q1-2013Q2. In contrast, the contribution of PMIs to the forecasting performance of output growth is found to be rather limited beyond the current quarter. Also, in line with the theoretical and MC results, we find that the AugGVAR performs better than the non-augmented GVAR, and that AugGVAR forecasts do as well as other data-rich forecasting techniques for the months within the current quarter, but tend to do significantly better for the months in the subsequent quarters ($h \geq 1$).

The remainder of the paper is organized as follows. Section 2 sets up two alternative high-dimensional factor-augmented VAR model specifications. Section 3 discusses forecasting with factor-augmented VARs and derives a large N representation of the infeasible optimal forecasts when factors are unobserved. Section 4 discusses forecasting with GVARs, shows that the presence of a strong unobserved common factor can lead to an undetermined GVAR model, proposes the AugGVAR, and establishes uniform convergence of feasible AugGVAR forecasts to the infeasible optimal forecasts as $N, T \xrightarrow{j} \infty$ such that $N/T \to \varkappa$ for some $0 < \varkappa < \infty$. Section 5 presents an extension of the analysis to the case of multiple factors. Section 6 discusses forecasting with GVARs in the case of nonsynchronous conditioning information sets. Section 7 illustrates the theoretical findings by means of Monte Carlo experiments. Section 8 presents the empirical application to forecasting GDP growth using PMIs. This section also presents an extension of the panel Diebold and Mariano (1995) (DM) test statistic proposed by Pesaran, Schuermann, and Smith (2009) to the case where aggregation weights are unequal, and discusses the consequences of the panel DM test

[4] We note that the existing theoretical results on Lasso and Ridge do not cover the case of dynamic models. Nevertheless, these data-rich methods are commonly employed in the forecasting of economic variables and thus provide an interesting benchmark.

when the differences in forecast errors are cross-sectionally dependent. Section 9 ends with some concluding remarks. Technical proofs and further results are provided in an Appendix. Additional results are presented in an online Supplement available from the authors upon request.

A brief word on notations: $\|\mathbf{A}\|_1 \equiv \max_{1\leq j\leq n} \sum_{i=1}^{n} |a_{ij}|$, and $\|\mathbf{A}\|_\infty \equiv \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|$ denote the maximum absolute column and row sum norms of $\mathbf{A} \in \mathbb{M}^{n\times n}$, respectively, where $\mathbb{M}^{n\times n}$ is the space of real-valued $n \times n$ matrices. $\lambda_1(\mathbf{A})$ is the largest eigenvalue of $\mathbf{A}$, $\|\mathbf{A}\| = \sqrt{\varrho(\mathbf{A}'\mathbf{A})}$ is the spectral norm of $\mathbf{A}$,[5] $\varrho(\mathbf{A}) = |\lambda_1(\mathbf{A})|$ is the spectral radius of $\mathbf{A}$. Matrices are represented by bold upper case letters, and vectors are represented by bold lower case letters. All vectors are column vectors.

2 Specifications of factor-augmented VARs

We consider two alternative large dimensional factor-augmented vector autoregressive (VAR) specifications that differ in the way they are augmented with the unobserved common factor. Specifically, we consider the following two covariance stationary factor-augmented VAR models M_a and M_b

$$M_a: \qquad \mathbf{y}_t - \boldsymbol{\gamma}_a f_{at} = \boldsymbol{\Phi}_a (\mathbf{y}_{t-1} - \boldsymbol{\gamma}_a f_{a,t-1}) + \boldsymbol{\varepsilon}_{at}, \tag{1}$$

and

$$M_b: \qquad \mathbf{y}_t = \boldsymbol{\Phi}_b \mathbf{y}_{t-1} + \boldsymbol{\gamma}_b f_{bt} + \boldsymbol{\varepsilon}_{bt}, \tag{2}$$

where $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{Nt})'$, $\boldsymbol{\gamma}_s = (\gamma_{s1}, \gamma_{s2}, ..., \gamma_{sN})'$, $s = a, b$ are $N \times 1$ vectors of factor loadings, f_{st}, $s = a, b$ are common factors, which are treated as unobserved unless otherwise specified, $\boldsymbol{\Phi}_s$, $s = a, b$ are $N \times N$ matrices of unknown coefficients, and $\boldsymbol{\varepsilon}_{st} = (\varepsilon_{s1t}, \varepsilon_{s2t}, ..., \varepsilon_{sNt})'$, $s = a, b$ are $N \times 1$ vectors of idiosyncratic shocks. It is assumed that the common factors follow covariance stationary AR(1) processes

$$f_{st} = \rho_s f_{s,t-1} + v_{st}, \text{ for } s = a, b. \tag{3}$$

Equations (1)-(3) feature only one lag and one common factor for expositional convenience, and higher order lags and/or more common factors could be considered. We also abstract, without the loss of generality, from deterministic terms. Introducing these terms is relatively straightforward.

We postulate the following assumptions that, among others, restrict cross-sectional dependence of reduced form errors and for the purpose of estimation impose some suitable restrictions on the VAR coefficients as $N \to \infty$.

ASSUMPTION 1 *(Cross-sectionally weakly dependent idiosyncratic errors) Idiosyncratic errors in $\boldsymbol{\varepsilon}_{st}$, for $s = a, b$, follow the 'spatial' model*

$$\boldsymbol{\varepsilon}_{st} = \mathbf{R}_s \boldsymbol{\eta}_{st},$$

[5] Note that if $\mathbf{x}$ is a vector, then $\|\mathbf{x}\| = \sqrt{\varrho(\mathbf{x}'\mathbf{x})} = \sqrt{\mathbf{x}'\mathbf{x}}$ corresponds to the Euclidean length of vector $\mathbf{x}$.

where the $N \times N$ matrix $\mathbf{R}_s$ has bounded row and column matrix norms (in N), and $\boldsymbol{\eta}_{st} \sim IID(\mathbf{0}, \mathbf{I}_N)$.

ASSUMPTION 2 *(Unobserved common factor and its loadings)*

a. *(Model without factor) $\gamma_{si} = 0$ for $s = a, b$ and for all $i = 1, 2, ..., N$.*

b. *(Model with factor) The unobserved common factors f_{st}, for $s = a, b$, are characterized by (3) with $|\rho_s| < 1$. The macro shock v_{st} is independently distributed of idiosyncratic errors, $E(v_{st}) = 0$, $E(v_{st}^2) = \sigma_{sv}^2 = 1 - \rho_s^2$, and $E(v_{st} v_{st'}) = 0$ for $s = a, b$, and any $t \neq t'$. The factor loadings are independently and identically distributed with a nonzero mean, $\gamma_s \neq 0$, and a finite variance. In addition, the factor loadings are independently distributed of the macro and the idiosyncratic shocks.*

ASSUMPTION 3 *(Covariance stationarity and bounded variances) There exists a small positive constant ϵ such that $\|\boldsymbol{\Phi}_s\| < 1 - \epsilon$, for $s = a, b$, where $\|\boldsymbol{\Phi}_s\|$ denotes the spectral norm of $\boldsymbol{\Phi}_s$.*

ASSUMPTION 4 *(No neighbors) There exists a (finite) positive constant $K < \infty$, which does not depend on N, and such that for any $N \in \mathbb{N}$, where $\mathbb{N}$ denotes the set of natural numbers, we have*

$$|\phi_{sii}| < K, \text{ for } s = a, b \text{ and any } i = 1, 2, ..., N$$

and

$$|\phi_{sij}| < \frac{K}{N}, \text{ for } s = a, b \text{ and any } j \neq i,\ i, j = 1, 2, ..., N,$$

where ϕ_{sij} denotes the i, j-th element of the matrix $\boldsymbol{\Phi}_s$.

Remark 1 *Assumption 3 is stronger than the usual finite-N covariance stationarity assumption, which restricts the eigenvalues of $\boldsymbol{\Phi}_s$ to lie within the unit circle. Assumption 3 also ensures that the variance of y_{it} exists as $N \to \infty$. See Chudik and Pesaran (2011) for a related discussion.*

Remark 2 *Assumption 4 rules out any neighbors (with the exception of own lags). This assumption can be relaxed, at the expense of notational complexity, without any fundamental implications for the main results derived below.*

Because the common factors f_{st} are unobserved, it is unclear how one could choose between the two specifications, (1) or (2). Therefore, it is important to develop methods that are robust to the way the common factor is introduced in the VAR model. In view of this ambiguity, we proceed with both models and show that under the above assumptions the common factors can be well approximated by cross-section averages and their lags, under both specifications. The key practical difference between the two specifications turns out to be in the number of lags of cross-section averages that are required for consistent estimation and forecasting. While only contemporaneous cross-section averages are required for approximating the common factor in the case of model (1),

the consequence of the factor error structure in (2) is that a large N representation for cross-section averages features an infinite-order distributed lag function in the common factor (Chudik and Pesaran, 2014a). Under certain conditions, such infinite lag polynomials can be inverted and appropriately truncated for the purpose of consistent estimation and inference as in Chudik and Pesaran (2013a).

3 Forecasting with factor-augmented VARs

Factor-augmented VAR models considered by Bernanke, Bovian, and Eliasz (2005) and Favero, Marcellino, and Neglia (2005) are low-dimensional VARs augmented by a small set of factors that enter as additional variables. Factors are estimated from a large set of n time series, and the estimates of the factors are plugged into a VAR as if they were observed. This plug-in approach, where factors are treated as if they were observed, is justified by considering n to be sufficiently large. Factors in these models represent latent variables that summarize the behavior of a large set of time series. Our factor-augmented VAR specifications M_a or M_b differ in that we include a large number of variables in a VAR and the factor is used to capture a strong cross-sectional dependence. Model M_b is close to that of Stock and Watson (2005, equation 13), where factors are exogenous and enter the VAR in the form of a factor error structure, but $\mathbf{\Phi}_b$ is restricted to be a diagonal matrix. A version of the factor-augmented model M_a was considered by Dées, di Mauro, Pesaran, and Smith (2007), who imposed a block-diagonal structure on $\mathbf{\Phi}_b$.

Forecasting with low- or high-dimensional factor-augmented VARs is straightforward when it is assumed that the factor and coefficients are known. Consider model M_a and information set $\mathcal{I}_t \cup \mathcal{F}_{at}$, where $\mathcal{I}_t = \{\mathbf{y}_t, \mathbf{y}_{t-1}, ...\}$ is an information set containing all information on N cross-section units at time t, and $\mathcal{F}_{at} = \{f_{at}, f_{a,t-1}, ...\}$ is an information set on current and past values of the common factor. Solving (1) for $\mathbf{y}_{t+h} - \boldsymbol{\gamma}_a f_{a,t+h}$ by backward substitution yields

$$\mathbf{y}_{t+h} - \boldsymbol{\gamma}_a f_{a,t+h} = \mathbf{\Phi}_a^h \left(\mathbf{y}_t - \boldsymbol{\gamma}_a f_{at}\right) + \sum_{\ell=0}^{h-1} \mathbf{\Phi}_a^\ell \boldsymbol{\varepsilon}_{a,t+h-\ell},$$

and after repeatedly substituting equation (3) for the unobserved common factor, we obtain the following forecasting equation:

$$\mathbf{y}_{t+h} = \mathbf{\Phi}_a^h \mathbf{y}_t + \mathbf{g}_{ah} f_{at} + \boldsymbol{\xi}_{ath}, \tag{4}$$

where $\mathbf{g}_{ah} = \left(\rho_a^h \mathbf{I}_N - \mathbf{\Phi}_a^h\right) \boldsymbol{\gamma}_a$, $\mathbf{I}_N$ is an $N \times N$ identity matrix, and

$$\boldsymbol{\xi}_{ath} = \sum_{\ell=0}^{h-1} \rho_a^\ell v_{a,t+h-\ell} + \sum_{\ell=0}^{h-1} \mathbf{\Phi}_a^\ell \boldsymbol{\varepsilon}_{a,t+h-\ell}.$$

For $h > 1$, $\boldsymbol{\xi}_{ath}$ is serially correlated, but orthogonal to the information available at time t, irrespective of whether the information set includes $\mathcal{F}_{at}$ or not; namely we have $E\left(\boldsymbol{\xi}_{ath} | \mathcal{I}_t, \mathcal{F}_{at}, M_a\right) = \mathbf{0}$,

and $E\left(\boldsymbol{\xi}_{ath}|\mathcal{I}_t, M_a\right) = \mathbf{0}$, for $h = 1, 2, ...$ Assuming that the information set contains $\mathcal{F}_{at}$, the optimal h-step ahead forecasts (in mean square error sense) are given by

$$\mathbf{y}^a_{t+h|t} = E\left(\mathbf{y}_{t+h}|\mathcal{I}_t, \mathcal{F}_{at}, M_a\right) = \mathbf{\Phi}^h_a \mathbf{y}_t + \mathbf{g}_{ah} f_{at}, \text{ for } h = 1, 2, ... \tag{5}$$

Similarly, the optimal forecasts with respect to the full information set, $\mathcal{I}_t \cup \mathcal{F}_{bt}$, under model M_b are given by

$$\mathbf{y}^b_{t+h|t} = E\left(\mathbf{y}_{t+h}|\mathcal{I}_t, \mathcal{F}_{bt}, M_b\right) = \mathbf{\Phi}^h_b \mathbf{y}_t + \mathbf{g}_{bh} f_{bt}, \text{ for } h = 1, 2, ..., \tag{6}$$

where $\mathbf{g}_{bh} = \sum_{\ell=0}^{h-1} \rho_b^{h-\ell} \mathbf{\Phi}^\ell_b \boldsymbol{\gamma}_b$. Note that under both factor-augmented VAR specifications, the conditional forecasts in (5) and (6) are linear in $\mathbf{y}_t$ and the unobserved factor, f_{st}, for $s = a, b$, and neither depend on the covariance of the idiosyncratic errors.

3.1 Forecasting with high-dimensional factor-augmented VARs when factors are unobserved

The optimal forecast in (5) and (6) depends on the unobserved common factor and possibly a large number of unknown parameters. When N is small, the optimal forecasts of $\mathbf{y}_{t+h}$ based on the information set $\mathcal{I}_t$ alone can be derived using Kalman filter techniques assuming a *full* knowledge of the factor-augmented model and processes that generate the factors.[6] In practice, the requirement of having a full knowledge of the underlying model is a disadvantage, and methods that are robust to certain variations in the assumptions of the model, such as the way factors are introduced in the VAR model, are welcome. Nevertheless, application of the Kalman filter to large systems clearly deserves attention, but this is beyond the scope of the present paper, and will be left to future research. Instead here we propose an alternative large N approximation to the unobserved factor problem, and derive optimal forecasts that depend on observables and a finite number of unknown parameters, which can be consistently estimated.

We start with model M_a, and using (5) we note that the optimal forecast of $y_{i,t+h}$, the ith element of $\mathbf{y}_{t+h}$, conditional on $\mathcal{I}_t \cup \mathcal{F}_{at}$ can be written as

$$y^a_{i,t+h|t} = E\left(y_{i,t+h}|\mathcal{I}_t, \mathcal{F}_{at}, M_a\right) = \mathbf{e}'_{Ni}\mathbf{\Phi}^h_a \mathbf{y}_t + \rho^h_a \gamma_{ai} f_{at} - \mathbf{e}'_{Ni}\mathbf{\Phi}^h_a \boldsymbol{\gamma}_a f_{at} \text{ for } h = 1, 2, ..., \tag{7}$$

where $\mathbf{e}_{Ni}$ is an $N \times 1$ selection vector with its ith element unity and zeros elsewhere. The unobserved common factor can be approximated by cross-section averages along the same lines as in Pesaran (2006) and Chudik and Pesaran (2011). In particular, Chudik and Pesaran (2011) show that under Assumptions 1-3 and for any vector $\mathbf{w} = (w_1, w_2, ..., w_N)'$ such that

$$\|\mathbf{w}\|_\infty = \max_i |w_i| < \frac{K}{N}, \tag{8}$$

[6] In the Appendix we show how to derive optimal forecasts of $\mathbf{y}_{t+h}$ based on the information set $\mathcal{I}_t$ when the dependent variables are generated according to (2).

we have

$$\bar{y}_{wt} = \mathbf{w}'\mathbf{y}_t = \left(\mathbf{w}'\boldsymbol{\gamma}_a\right) f_{at} + O_p\left(N^{-1/2}\right). \tag{9}$$

Assumption 4 implies the existence of finite positive constants K_ℓ, for $\ell = 1, 2, ..., h$, such that for any $N \in \mathbb{N}$ and any $i, j \in \{1, 2, ..., N\}$ we have

$$\left|\phi_{a\ell ii} - \phi^{\ell}_{aii}\right| < \frac{K_\ell}{N}, \tag{10}$$

as well as

$$\left|\phi_{a\ell ij}\right| < \frac{K_\ell}{N}, \tag{11}$$

where $\phi_{a\ell ij}$ denotes the (i, j) element of $\mathbf{\Phi}^{\ell}_a$. Using (10)-(11) in (7), and (9) to substitute out the cross-section average $\sum_{j=1, j\neq i}^{N} \phi_{a\ell ij} y_{jt}$ yields[7]

$$y^a_{i,t+h|t} = \phi^h_{aii} y_{it} + \left(\rho^h_a - \phi^h_{aii}\right)\gamma_{ai} f_{at} + O_p\left(N^{-1/2}\right),$$

for any given fixed forecasting horizon $h > 0$. Furthermore, for any weights vector, $\mathbf{w}$, which in addition to condition (8) also satisfies

$$\sum_{i=1}^{N} w_i = 1, \tag{12}$$

we obtain

$$\bar{\gamma}_{wa} = \gamma_a + O_p\left(N^{-1/2}\right), \tag{13}$$

and hence

$$y^a_{i,t+h|t} = \phi^h_{aii} y_{it} + c_{ahi}\bar{y}_{wt} + O_p\left(N^{-1/2}\right), \tag{14}$$

for any i and a given fixed forecasting horizon h, where

$$c_{ahi} = \begin{cases} 0, & \text{under Assumption 2.a} \\ \left(\rho^h_a - \phi^h_{aii}\right)\frac{\gamma_{ai}}{\bar{\gamma}_{wa}}, & \text{under Assumption 2.b} \end{cases}.$$

Suppose now that $\mathbf{y}_t$ is generated according to (2) instead of (1). Taking cross-section averages in this case yields

$$\begin{aligned} \bar{y}_{wt} &= \mathbf{w}'\mathbf{\Phi}_b\mathbf{y}_{t-1} + \left(\mathbf{w}'\boldsymbol{\gamma}_b\right) f_{bt} + \mathbf{w}'\boldsymbol{\varepsilon}_{bt}, \\ &= \mathbf{w}'\mathbf{\Phi}_b\mathbf{y}_{t-1} + \bar{\gamma}_{wb} f_{bt} + O_p\left(N^{-1/2}\right), \end{aligned} \tag{15}$$

where $\mathbf{w}'\boldsymbol{\varepsilon}_{bt} = O_p\left(N^{-1/2}\right)$, and, for a given $\mathbf{\Phi}_b$, f_{bt} can be approximated (up to a scaling constant) by $\bar{y}_{wt} - \mathbf{w}'\mathbf{\Phi}_b\mathbf{y}_{t-1}$. But in practice $\mathbf{\Phi}_b$ is not known, and cannot be estimated consistently when N is large. Nevertheless, f_{bt} can be approximated by an infinite order distributed lag function in $\bar{y}_{wt}$. In particular, under Assumption 3 and assuming that the individual dynamic processes have

[7] We can use (9) because the vector $\left(\phi_{a\ell i1}, \phi_{a\ell i2}, ..., \phi_{a\ell i,i-1}, 0, \phi_{a\ell i,i+1}, ..., \phi_{a\ell iN}\right)'$ satisfies (8).

been in operation for some time, we have

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \mathbf{\Phi}_b^j \boldsymbol{\varepsilon}_{b,t-j} + \sum_{j=0}^{\infty} \mathbf{\Phi}_b^j \boldsymbol{\gamma}_b f_{b,t-j},$$

and hence

$$\mathbf{w}'\mathbf{\Phi}_b \mathbf{y}_{t-1} = \sum_{j=0}^{\infty} \mathbf{w}'\mathbf{\Phi}_b^{j+1} \boldsymbol{\varepsilon}_{b,t-j-1} + \sum_{j=0}^{\infty} \mathbf{w}'\mathbf{\Phi}_b^{j+1} \boldsymbol{\gamma}_b f_{b,t-j-1}.$$

Using this result in (15) now yields

$$\bar{y}_{wt} = d_b(L) f_{bt} + \sum_{\ell=0}^{\infty} \mathbf{w}'\mathbf{\Phi}_b^{\ell} \boldsymbol{\varepsilon}_{b,t-\ell}, \tag{16}$$

where the polynomial $d_b(L) = \sum_{\ell=0}^{\infty} d_{b\ell} L^{\ell} = \sum_{\ell=0}^{\infty} \mathbf{w}'\mathbf{\Phi}_b^{\ell} \boldsymbol{\gamma}_b L^{\ell}$ depends on $\boldsymbol{\gamma}_b$, $\mathbf{w}$ and *all* elements of $\mathbf{\Phi}_b$ (including the off-diagonal elements), and

$$Var\left(\sum_{\ell=0}^{\infty} \mathbf{w}'\mathbf{\Phi}_b^{\ell} \boldsymbol{\varepsilon}_{b,t-\ell}\right) = \sum_{\ell=0}^{\infty} \mathbf{w}'\mathbf{\Phi}_b^{\ell} \mathbf{R}_b \mathbf{R}_b' \mathbf{\Phi}_b^{\ell\prime} \mathbf{w}.$$

Taking the spectral matrix norm, under Assumptions 1 and 3, and condition (8) we have,

$$\left\| Var\left(\sum_{\ell=0}^{\infty} \mathbf{w}'\mathbf{\Phi}_b^{\ell} \boldsymbol{\varepsilon}_{b,t-\ell}\right) \right\| \leq \|\mathbf{w}\|^2 \|\mathbf{R}_b\|^2 \sum_{\ell=0}^{\infty} \|\mathbf{\Phi}_b\|^{2\ell} = O\left(N^{-1}\right), \tag{17}$$

where $\|\mathbf{w}\|^2 \leq \|\mathbf{w}\|_{\infty} \|\mathbf{w}\|_1 \leq N \|\mathbf{w}\|_{\infty}^2 = O\left(N^{-1}\right)$ (condition (8)), $\|\mathbf{R}_b\|^2 \leq \|\mathbf{R}_b\|_{\infty} \|\mathbf{R}_b\|_1 = O(1)$ (Assumption 1), and $\sum_{\ell=0}^{\infty} \|\mathbf{\Phi}_b\|^{2\ell} = O(1)$ (Assumption 3). Using (17) in (16) and noting that $E\sum_{\ell=0}^{\infty} \mathbf{w}'\mathbf{\Phi}_b^{\ell} \boldsymbol{\varepsilon}_{b,t-\ell} = 0$, we obtain

$$\bar{y}_{wt} = d_b(L) f_{bt} + O_p\left(N^{-1/2}\right). \tag{18}$$

Note that the coefficients in the polynomial $d_b(L)$ satisfy $|d_{b\ell}| = \left|\mathbf{w}'\mathbf{\Phi}_b^{\ell}\boldsymbol{\gamma}_b\right| \leq \|\mathbf{w}\| \left\|\mathbf{\Phi}_b^{\ell}\right\| \|\boldsymbol{\gamma}_b\| = O\left[(1-\epsilon)^{\ell}\right]$ and are thus declining at an exponential rate. Assuming that $a_b(L) = d_b^{-1}(L)$ exists and its coefficients also decline exponentially,[8] we obtain

$$f_t = a_b(L)\, \bar{y}_{wt} + O_p\left(N^{-1/2}\right), \tag{19}$$

and the error of approximating f_t with $\sum_{\ell=0}^{p} a_{b\ell} \bar{y}_{t-\ell}$ declines exponentially in the truncation lag, p. Now consider the ith element of $\mathbf{y}_{t+h|t}^b$ in (6), namely

$$y_{i,t+h|t}^b = E\left(y_{i,t+h} | \mathcal{I}_t, \mathcal{F}_{bt}, M_b\right) = \mathbf{e}_{Ni}' \mathbf{\Phi}_b^h \mathbf{y}_t + g_{bhi} f_{bt}, \text{ for } h = 1, 2, ..., \tag{20}$$

[8] See Lemma A.1 of Chudik and Pesaran (2013b) for sufficient conditions on the existence of $a(L)$ with exponentially declining coefficients.

where $g_{bhi} = \mathbf{e}'_{Ni}\mathbf{g}_{bh} = \sum_{\ell=0}^{h-1} \rho_b^{h-\ell}\mathbf{e}'_{Ni}\mathbf{\Phi}_b^{\ell}\boldsymbol{\gamma}_b$. Define $\boldsymbol{\omega}_{bhi} = \mathbf{\Phi}_b^{h\prime}\mathbf{e}_{Ni} - \phi_{bii}^h\mathbf{e}_{Ni}$ and note that (10)-(11) also holds for model M_b, and therefore $\|\boldsymbol{\omega}_{bhi}\|_\infty = O\left(N^{-1}\right)$, that is $\boldsymbol{\omega}_{bhi}$ satisfies (8). Hence, we can use the same arguments as in the derivation of (18) to obtain

$$\boldsymbol{\omega}'_{bhi}\mathbf{y}_t = \beta_{bhi}(L) f_{bt} + O_p\left(N^{-1/2}\right),$$

where $\beta_{bhi}(L) = \sum_{s=0}^{\infty} \boldsymbol{\omega}'_{bhi}\mathbf{\Phi}_b^s\boldsymbol{\gamma}_b L^s$. Using this result in $\mathbf{e}'_{Ni}\mathbf{\Phi}_b^h\mathbf{y}_t = \phi_{bii}^h y_{it} + \boldsymbol{\omega}'_{bhi}\mathbf{y}_t$ and substituting (19) in (20) yields the following large N representation of $y_{i,t+h|t}^b$:

$$y_{i,t+h|t}^b = \phi_{bii}^h y_{it} + c_{bhi}(L)\,\bar{y}_{wt} + O_p\left(N^{-1/2}\right), \tag{21}$$

where the polynomial

$$c_{bhi}(L) = \begin{cases} 0, & \text{under Assumption 2.a} \\ \left[\beta_{bhi}(L) + g_{bhi}\right] a_b(L), & \text{under Assumption 2.b} \end{cases}.$$

Note that when Assumption 2.b holds, in general, $c_{bhi}(L)$ is an infinite order polynomial in the lag operator, L.

The following proposition summarizes the main findings of this subsection.

Proposition 1 *Let $\mathbf{y}_t$ be generated by model (1) or model (2) with a factor given by (3), Assumptions 1, 2.a or 2.b, and 3-4 hold, $\mathbf{w}$ be any arbitrary vector of weights satisfying (8) and (12), and the polynomial $a_b(L) = d_b^{-1}(L)$ exists. Then for any cross-section unit $i \in \mathbb{N}$, and a given fixed forecasting horizon $0 < h < K$, the optimal forecasts of $y_{i,t+h}$, defined in (5) and (6) have a large N representation given by (14) and (21), respectively.*

Comparing (14) with (21) we see that the latter involves an infinite order lag distribution in cross-section averages that need to be truncated, whereas under the former only contemporaneous values of cross-section averages are included. In practice where the nature of factors and how they enter the VAR model are not known, the lag order selection is likely to be important when forecasting with large factor-augmented VARs. It might not be sufficient just to add factor estimates to the VAR model. The lag orders of y_{it} and $\bar{y}_{wt}$ need to be selected with care and together.

4 Forecasting with a GVAR

The GVAR approach was introduced in Pesaran et al. (2004) and has been used extensively to model cross country, regions or market interactions. Chudik and Pesaran (2014b) provide a recent survey. Consider N cross-section units (say countries) and suppose that the endogenous variables specific to unit i, denoted by the $k_i \times 1$ vector $\mathbf{y}_{it}$, are related to their own past, and current and past values of the remaining units. It is clear that without further restrictions, estimation of the full system of equations in the endogenous variables, $\mathbf{y}_t = (\mathbf{y}'_{1t}, \mathbf{y}'_{2t}, ..., \mathbf{y}'_{Nt})'$, will be subject to the

curse of dimensionality, even for moderate values of N. The GVAR approach resolves the curse of dimensionality by adopting a two-step procedure. In the first step, cross-sectionally augmented conditional models are estimated for each cross-section unit, taking the cross-section average as weakly exogenous. In the second step, the estimated conditional models are combined to form a complete system which is then used for forecasting and policy analysis. The key assumption that cross-section averages are weakly exogenous is justified under certain plausible assumptions (see Chudik and Pesaran (2011)), and are routinely tested in empirical applications of the GVAR.

More specifically, for each unit i, the following conditional model is estimated:

$$\mathbf{y}_{it} = \mathbf{\Theta}_i \mathbf{y}_{i,t-1} + \mathbf{B}_{i0}\overline{\mathbf{y}}_{wit} + \mathbf{B}_{i1}\overline{\mathbf{y}}_{wi,t-1} + \boldsymbol{\xi}_{it}, \tag{22}$$

for $i = 1, 2, ..., N$, where $\overline{\mathbf{y}}_{wit} = \mathbf{W}_i'\mathbf{y}_t$ is a $k^* \times 1$ vector of cross-section averages specific to unit i, $\mathbf{W}_i$ is a $k \times k^*$ matrix of unit-specific weights that define the k^* cross-section averages, and $k = \sum_{i=1}^{N} k_i$ is the total number of variables. We abstract from the deterministic components, observed common factors, and additional lags for the simplicity of exposition, but these additions can be readily accommodated.

In the second step, individual models in (22) are stacked and solved in one large VAR. Stacking (22) for $i = 1, 2, ..., N$ yields

$$\mathbf{y}_t = \mathbf{\Theta}\mathbf{y}_{t-1} + \mathbf{B}_0\overline{\mathbf{y}}_{wt} + \mathbf{B}_1\overline{\mathbf{y}}_{w,t-1} + \boldsymbol{\xi}_t, \tag{23}$$

where $\overline{\mathbf{y}}_{wt} = (\overline{\mathbf{y}}'_{w1t}, \overline{\mathbf{y}}'_{w2t}, ..., \overline{\mathbf{y}}'_{wNt})'$, $\xi_t = \left(\boldsymbol{\xi}'_{1t}, \boldsymbol{\xi}'_{2t}, ..., \boldsymbol{\xi}'_{Nt}\right)'$, and

$$\mathbf{\Theta} = \begin{pmatrix} \mathbf{\Theta}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Theta}_2 & & \mathbf{0} \\ \vdots & & \ddots & \\ \mathbf{0} & \mathbf{0} & & \mathbf{\Theta}_N \end{pmatrix}, \ \mathbf{B}_h = \begin{pmatrix} \mathbf{B}_{1,h} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{2,h} & & \mathbf{0} \\ \vdots & & \ddots & \\ \mathbf{0} & \mathbf{0} & & \mathbf{B}_{N,h} \end{pmatrix}, \text{ for } h = 1, 2, ...$$

Recognizing that $\overline{\mathbf{y}}_{wt} = \mathbf{W}\mathbf{y}_t$, where $\mathbf{W} = (\mathbf{W}_1, \mathbf{W}_2, ..., \mathbf{W}_N)'$, (23) can be written as

$$\mathbf{G}_0\mathbf{y}_t = \mathbf{G}_1\mathbf{x}_{t-1} + \boldsymbol{\xi}_t, \tag{24}$$

where

$$\mathbf{G}_0 = (\mathbf{I}_k - \mathbf{B}_0\mathbf{W}), \text{ and } \mathbf{G}_1 = (\mathbf{\Theta} + \mathbf{B}_1\mathbf{W}). \tag{25}$$

Finally, provided that $\mathbf{G}_0$ is invertible, we can multiply (24) by $\mathbf{G}_0^{-1}$ from the left to obtain the following GVAR model:

$$\mathbf{y}_t = \mathbf{G}\mathbf{y}_{t-1} + \mathbf{u}_t, \tag{26}$$

where $\mathbf{G} = \mathbf{G}_0^{-1}\mathbf{G}_1$, and $\mathbf{u}_t = \mathbf{G}_0^{-1}\xi_t$.

To forecast with a GVAR, one can assume that (26) is the data-generating process (DGP), or alternatively one can assume that the DGP is model M_a or M_b, and the GVAR is used as an

approximation of the model M_a or M_b. In the former case, where the DGP is (26), forecasting is straightforward. But in the latter case, where M_a or M_b is the DGP, there is no reason to believe that the inverse of $\mathbf{G}_0$ exists when the unobserved common factor is present. Even if the estimate of $\mathbf{G}_0$ is not rank deficient, the singularity of $\mathbf{G}_0$ will have adverse effects on the forecasting performance.

To show the rank deficiency of $\mathbf{G}_0$, assume that the DGP is model M_a, given by (1) and (3), that is $k_i = 1$, $k = N$, $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{Nt})'$, $k^* = 1$ and the cross-section average $\bar{y}_{wt} = \mathbf{w}'\mathbf{y}_t$, where the weights vector $\mathbf{w}$ is common across units and satisfies (8) and (12).[9] In what follows we focus on model M_a, and drop the subscript a to simplify notations. This specification (as opposed to M_b) allows us to work with a finite lag polynomial, and also allows us to use the properties of the cross-section augmented least squares (CALS) estimator developed in Chudik and Pesaran (2011). The main arguments put forward in this section apply equally to the alternative model specification, M_b, defined by (2) and (3), by relying on the CALS estimation with appropriately truncated lags as considered in Chudik and Pesaran (2013a).

Using (1), (3) and Assumptions 1-4 above, and following a similar line argument as in Chudik and Pesaran (2011), we obtain the following unit-specific equations:

$$y_{it} = \phi_{ii} y_{i,t-1} + b_{i0}\bar{y}_{wt} + b_{i1}\bar{y}_{w,t-1} + \xi_{it}, \text{ for } i \in \{1, 2, ..., N\}, \tag{27}$$

where under Assumption 2.a we have $b_{i0} = b_{i1} = 0$, and under Assumption 2.b,

$$b_{i0} = \bar{\gamma}_w^{-1}\gamma_i,\ b_{i1} = -\bar{\gamma}_w^{-1}\gamma_i\phi_{ii}, \tag{28}$$

where $\bar{\gamma}_w = \sum_{i=1}^{N} w_i\gamma_i$. Also, under both assumptions we have $\xi_{it} = \varepsilon_{it} + O_p\left(N^{-1/2}\right)$. Chudik and Pesaran (2011) established that the least squares estimates of (27) are consistent and asymptotically normally distributed. Using (27), $\mathbf{G}_0 = (\mathbf{I}_N - \mathbf{b}_0\mathbf{w}')$, which is easily seen to be rank deficient. The rank deficiency follows ($\mathbf{w}'\mathbf{b}_0 = 1$ and therefore rows of $\mathbf{G}_0$ are linearly dependent)

$$\begin{aligned} \mathbf{w}'\mathbf{G}_0 &= \mathbf{w}'\left(\mathbf{I}_N - \mathbf{b}_0\mathbf{w}'\right) \\ &= \mathbf{w}' - \bar{\gamma}_w^{-1}\left(\sum_{i=1}^{N} w_i\gamma_i\right)\mathbf{w}' = \mathbf{0}. \end{aligned} \tag{29}$$

The consequence of rank deficiency of $\mathbf{G}_0$ is that the system of N equations in (27) is undetermined, and we discuss this problem in greater detail next.

4.1 Rank deficient case

The GVAR model (26) is derived under the assumption that the contemporaneous coefficient matrix, $\mathbf{G}_0$, (defined by (25)) has full rank. To clarify the role of this assumption and to illustrate the consequences of possible rank deficiency of $\mathbf{G}_0$, abstracting from lags of $(y_{it}, \bar{y}_{wit})'$, we consider the

[9] Conditions (8) and (12) are sufficient for the usual granularity conditions (see (45) and (46)) to hold.

following illustrative GVAR model:

$$y_{it} = \lambda_i \bar{y}_{wit} + \varepsilon_{it}, \text{ for } i = 1, 2, ..., N, \tag{30}$$

where $\bar{y}_{wit} = \mathbf{w}_i' \mathbf{y}_t$. Let $\mathbf{\Lambda}$ be the $N \times N$ diagonal matrix defined by $\mathbf{\Lambda} = diag\left(\lambda_1, \lambda_2, ..., \lambda_N\right)$, and let $\mathbf{W}' = (\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_N)$. Write (30) as

$$\mathbf{y}_t = \mathbf{\Lambda} \mathbf{W} \mathbf{y}_t + \boldsymbol{\varepsilon}_t,$$

or

$$\mathbf{G}_0 \mathbf{x}_t = \boldsymbol{\varepsilon}_t, \tag{31}$$

where $\mathbf{G}_0 = \mathbf{I}_N - \mathbf{\Lambda} \mathbf{W}$. Suppose that $\mathbf{G}_0$ is rank deficient, namely $rank\left(\mathbf{G}_0\right) = N - m$, for some $m > 0$. Then the solution of (31) exists only if $\boldsymbol{\varepsilon}_t$ lies in the range of $\mathbf{G}_0$, denoted as $Col\left(\mathbf{G}_0\right)$. Assuming this is the case, system (31) does not uniquely determine $\mathbf{y}_t$, and the set of all its possible solutions can be characterized as

$$\mathbf{y}_t = \mathbf{\Gamma} \tilde{\mathbf{f}}_t + \mathbf{G}_0^+ \boldsymbol{\varepsilon}_t, \tag{32}$$

where $\tilde{\mathbf{f}}_t$ is a vector of m arbitrary stochastic processes, $\mathbf{\Gamma}$ is a $k \times m$ matrix which is a basis of the null space of $\mathbf{G}_0$, namely $\mathbf{G}_0 \mathbf{\Gamma} = \mathbf{0}$, $rank\left(\mathbf{\Gamma}' \mathbf{\Gamma}\right) = m$, and $\mathbf{G}_0^+$ is the Moore-Penrose pseudo-inverse of $\mathbf{G}_0$. To verify that (32) maps all possible solutions of (31), note that $\mathbf{G}_0^+ \boldsymbol{\varepsilon}_t$ is the particular solution of (31) and $\mathbf{\Gamma}' \widetilde{\mathbf{f}}_t$ is a general solution of the homogenous counterpart of (31), given by $\mathbf{G}_0 \mathbf{y}_t = \mathbf{0}$. To prove the former, from the property of Moore-Penrose inverses, namely $\mathbf{G}_0 \mathbf{G}_0^+ \mathbf{G}_0 = \mathbf{G}_0$, we note that $\mathbf{G}_0 \mathbf{G}_0^+ \mathbf{G}_0 \mathbf{y}_t = \mathbf{G}_0 \mathbf{y}_t$, or $\boldsymbol{\varepsilon}_t = \mathbf{G}_0 \mathbf{G}_0^+ \boldsymbol{\varepsilon}_t$, which establishes that $\mathbf{G}_0^+ \boldsymbol{\varepsilon}_t$ is indeed a solution of $\mathbf{G}_0 \mathbf{y}_t = \boldsymbol{\varepsilon}_t$. To prove the latter, we note that $\mathbf{\Gamma}$ is a basis of the null space of $\mathbf{G}_0$ and therefore $\mathbf{G}_0 \mathbf{\Gamma} \tilde{\mathbf{f}}_t = \mathbf{0}$ for any $m \times 1$ arbitrary stochastic process $\tilde{\mathbf{f}}_t$, and the set of solutions must be complete since the dimension of $Col\left(\mathbf{\Gamma}\right)$ is m.

Let $\mathbf{f}_t = \tilde{\mathbf{f}}_t - E\left(\tilde{\mathbf{f}}_t \middle| \boldsymbol{\varepsilon}_t\right) = \tilde{\mathbf{f}}_t - \mathbf{D}' \boldsymbol{\varepsilon}_t$. Then (32) can also be written as an approximate factor model, namely

$$\mathbf{y}_t = \mathbf{\Gamma} \mathbf{f}_t + \mathbf{H} \boldsymbol{\varepsilon}_t,$$

where $\mathbf{f}_t$ is uncorrelated with $\boldsymbol{\varepsilon}_t$ by construction, and

$$\mathbf{H} = \mathbf{\Gamma} \mathbf{M}' + \mathbf{G}_0^+.$$

Without any loss of generality, it is standard convention to use the normalization $Var\left(\mathbf{f}_t\right) = \mathbf{I}_m$, and to set the first non-zero element in each of the m column vectors of $\mathbf{\Gamma}$ to be positive. These normalization conditions ensure that $\mathbf{\Gamma}$ is unique, in which case $\mathbf{H}$ is unique up to the rotation matrix, $\mathbf{M}$. Therefore, the full rank condition, $rank(\mathbf{G}_0) = N$, is necessary and sufficient for $\mathbf{y}_t$, given by (30), to be uniquely determined. It also follows that $\mathbf{y}_t$ must have a factor structure in cases where $\mathbf{G}_0$ is rank deficient. Finally, note that all of the above results hold for any N, and as $N \to \infty$.

4.2 Dealing with rank deficiency by augmentation

If $\mathbf{G}_0$ is known to be rank deficient with rank $N-m$ and $m>0$, then the GVAR model (30) would need to be augmented by m equations that determine the m cross-section averages, defined by $\mathbf{\Gamma}'\mathbf{y}_t$, in order for $\mathbf{y}_t$ to be uniquely determined. In the case of system (27), $m=1$, and augmentation by one additional equation is needed in order to obtain a unique solution for $\mathbf{y}_t$. Different options could be considered for the augmentation of (27). We consider augmenting the set of conditional equations in (27) with the following marginal equation for cross-section averages:

$$\bar{y}_{wt} = \rho\bar{y}_{w,t-1} + \xi_{\bar{y}t}, \text{ where } \xi_{\bar{y}t} = \gamma v_t + O_p\left(N^{-1/2}\right), \tag{33}$$

and we treat $\bar{y}_{wt}$ as a proxy for the (scaled) unobserved common factor (see (9)). Stacking (27) and (33), we obtain the following VAR model in $\mathbf{z}_t = (\mathbf{y}_t', \bar{y}_{wt})'$:

$$\mathbf{A}_0\mathbf{z}_t = \mathbf{A}_1\mathbf{z}_{t-1} + \mathbf{e}_{zt}, \tag{34}$$

where $\mathbf{e}_{zt} = \left(\boldsymbol{\xi}_t', \xi_{\bar{y}t}\right)' = (\boldsymbol{\varepsilon}_t', \gamma v_t)' + O_p\left(N^{-1/2}\right)$,

$$\mathbf{A}_0 = \begin{pmatrix} \mathbf{I}_N & -\mathbf{b}_0 \\ \mathbf{0} & 1 \end{pmatrix}, \mathbf{A}_1 = \begin{pmatrix} \mathbf{\Theta} & \mathbf{b}_1 \\ 0 & \rho \end{pmatrix}, \tag{35}$$

and $\mathbf{\Theta}$ is an $N\times N$ diagonal matrix with elements ϕ_{ii}, for $i=1,2,...,N$, on the diagonal. The matrix $\mathbf{A}_0$ is (by construction) invertible, and let $\mathbf{A} = \mathbf{A}_0^{-1}\mathbf{A}_1$. Note that (using $\mathbf{b}_1 = -\mathbf{\Theta}\mathbf{b}_0$, see (28))

$$\mathbf{A}^\ell = \begin{pmatrix} \mathbf{\Theta}^\ell & \left(\rho^\ell\mathbf{I}_N - \mathbf{\Theta}^\ell\right)\mathbf{b}_0 \\ 0 & \rho^\ell \end{pmatrix} \text{ for } \ell = 1,2,..., \tag{36}$$

and consider the following forecast of $y_{i,t+h}$:

$$y^{aug}_{i,t+h|t} = \mathbf{e}_{N+1,i}'\mathbf{A}^h\mathbf{z}_t, \tag{37}$$

where $\mathbf{e}_{N+1,i}$ is an $N+1$ dimensional selection vector that selects the i-th element. Substituting the expression (36) for $\mathbf{A}^h$ in (37), we obtain

$$y^{aug}_{i,t+h|t} = \begin{cases} \phi^h_{ii}y_{it} & \text{under Assumption 2.a} \\ \phi^h_{ii}y_{it} + \left(\rho^h - \phi^h_{ii}\right)\frac{\gamma_i}{\bar{\gamma}_w}\bar{y}_{wt} & \text{under Assumption 2.b} \end{cases}. \tag{38}$$

It now readily follows by comparing (38) and (14) that

$$y^{aug}_{i,t+h|t} = y_{i,t+h|t} + O_p\left(N^{-1/2}\right), \tag{39}$$

which establishes the consistency of the forecast $y^{aug}_{i,t+h|t}$ defined in (37). These findings are summarized in the following proposition.

Proposition 2 *Let $\mathbf{y}_t$ be generated by (1), Assumptions 1, 2.a or 2.b, and 3-4 hold, and $\mathbf{w}$ be any arbitrary vector of weights satisfying (8) and (12). Then for any cross-section unit $i \in \mathbb{N}$, and any given fixed forecasting horizon $0 < h < K$, the forecast $y^{aug}_{i,t+h|t}$ defined in (37) is consistent, that is $\left| y^{aug}_{i,t+h|t} - y_{i,t+h|t} \right| \xrightarrow{p} 0$ as $N \to \infty$.*

$y^{aug}_{i,t+h|t}$ is still an infeasible forecast since the parameters in (37) are unknown and need to be estimated. It is therefore important to establish asymptotic results for feasible forecasts.

We consider estimation of GVAR forecasts $y_{i,t+h|t}$ and AugGVAR forecasts $y^{aug}_{i,t+h|t}$ by using least squares estimates of parameters of the conditional cross-section augmented models (27) and (in the case of the AugGVAR only) also the marginal model (33). Namely, we define

$$\hat{y}_{i,t+h|t} = \mathbf{e}'_{N,i}\widehat{\mathbf{G}}^h \mathbf{y}_t, \tag{40}$$

and

$$\hat{y}^{aug}_{i,t+h|t} = \mathbf{e}'_{N+1,i}\hat{\mathbf{A}}^h \mathbf{z}_t, \tag{41}$$

for $i = 1, 2, ..., N$ and $h = 1, 2, ...$, where we use hats on $\mathbf{G}$ and $\mathbf{A}$ to denote that these matrices are constructed based on the least squares estimates of the unknown parameters in (27) and (33).

We collect the individual forecasts in the vectors $\hat{\mathbf{y}}_{t+h|t} = \left(\hat{y}_{1,t+h|t}, \hat{y}_{2,t+h|t}, ..., \hat{y}_{N,t+h|t}\right)'$ and $\hat{\mathbf{y}}^{aug}_{t+h|t} = \left(\hat{y}^{aug}_{1,t+h|t}, \hat{y}^{aug}_{2,t+h|t}, ..., \hat{y}^{aug}_{N,t+h|t}\right)'$. We investigate the asymptotic properties of $\hat{\mathbf{y}}^{aug}_{t+h|t}$ in the case of the weakly cross-sectionally dependent model and the case of the model featuring an unobserved common factor.

Theorem 1 *Suppose $\mathbf{y}_t$ is generated by model (1), $\mathbf{w}$ is any vector satisfying conditions (8) and (12), Assumptions 1, 2.a (weakly cross-sectionally dependent model) or 2.b (model featuring unobserved common factor), and 3-4 hold, and $N, T \xrightarrow{j} \infty$ such that $N/T \to \varkappa$ for some $0 < \varkappa < \infty$. Then for any fixed $0 < h < K$, the h-step-ahead forecast $\hat{\mathbf{y}}^{aug}_{t+h|t}$ defined by (41) satisfies*

$$\left\| \mathbf{y}_{t+h|t} - \hat{\mathbf{y}}^{aug}_{t+h|t} \right\|_\infty \xrightarrow{L_1} 0. \tag{42}$$

Moreover, in the case when Assumption 2.b holds, the matrix $\mathbf{G}_0 = \mathbf{I}_N - \mathbf{b}_0\mathbf{w}'$ is singular for any $N \in \mathbb{N}$.

The proof is provided in the Appendix.

5 Extension to multiple unobserved common factors

Theorem 1 establishes that regardless of whether an unobserved common factor is included in the VAR, it is asymptotically justified to use $\hat{\mathbf{y}}^{aug}_{t+h|t}$ for forecasting individual endogenous variables in the sense that the differences between infeasible optimal forecasts and feasible forecasts are arbitrarily small as $N, T \xrightarrow{j} \infty$ such that $N/T \to \varkappa$ for some $0 < \varkappa < \infty$. This result is established under the restrictive assumption that the number of unobserved factors in the underlying VAR

model is at most equal to unity. Here we relax this assumption and consider VAR models with multiple factors. As in Chudik and Pesaran (2011), we shall assume that there are up to $m_{\max}$ factors, where $m_{\max}$ is a fixed known integer, and the data is generated by model (1) with $\boldsymbol{\gamma} f_t$ replaced by $\boldsymbol{\Gamma}\mathbf{f}_t$,

$$\mathbf{y}_t - \boldsymbol{\Gamma}\mathbf{f}_t = \boldsymbol{\Phi}\left(\mathbf{y}_{t-1} - \boldsymbol{\Gamma}\mathbf{f}_{t-1}\right) + \boldsymbol{\varepsilon}_t, \tag{43}$$

where $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, ..., \boldsymbol{\gamma}_N)'$ is the $N \times m$ matrix of factor loadings, and $\mathbf{f}_t$ is an $m \times 1$ vector of unobserved common factors, $m \leq m_{\max}$, but m is otherwise unknown. As before, we assume that $\boldsymbol{\varepsilon}_t$ is independently distributed of $\mathbf{f}_t$. Moreover, the vector of unobserved common factors is assumed to follow the covariance stationary VAR(1) process,

$$\mathbf{f}_t = \boldsymbol{\Pi}_f \mathbf{f}_{t-1} + \mathbf{v}_t. \tag{44}$$

Consider $m_{\max}$ cross-section averages $\bar{\mathbf{y}}_{wt} = \mathbf{W}'\mathbf{y}_t$, where $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_N)'$ is an $N \times m_{\max}$ matrix of predetermined granular weights satisfying the conditions

$$\|\mathbf{W}\| = O\left(N^{-\frac{1}{2}}\right), \tag{45}$$

$$\frac{\|\mathbf{w}_j\|}{\|\mathbf{W}\|} = O\left(N^{-\frac{1}{2}}\right) \text{ uniformly in } j. \tag{46}$$

Following the same steps as in the case of a single factor, we obtain the following large N representation for cross-section averages:

$$\bar{\mathbf{y}}_{wt} = \bar{\boldsymbol{\Gamma}}_w \mathbf{f}_t + O_p\left(N^{-1/2}\right), \tag{47}$$

and it is clear that the full column rank of $\bar{\boldsymbol{\Gamma}}_w$ is necessary for $\bar{\mathbf{y}}_{wt}$ to approximate the space spanned by $\mathbf{f}_t$. To this end, we postulate the following assumption instead of Assumption 2.

ASSUMPTION 5 *(Multiple unobserved common factors and their loadings) The $m \times 1$ vector of unobserved common factors is characterized by (44) with $|\lambda_1(\boldsymbol{\Pi}_f)| < 1$. The macro shocks in $\mathbf{v}_t$ are independently distributed of idiosyncratic errors, $\boldsymbol{\varepsilon}_t$, $E(\mathbf{v}_t) = 0$, $\|E(\mathbf{v}_t\mathbf{v}_t')\| < K$, and $E\left(\mathbf{v}_t\mathbf{v}_{t'}'\right) = 0$ for any $t \neq t'$. The factor loadings are bounded, $\|\boldsymbol{\gamma}_i\| < K$, and $\bar{\boldsymbol{\Gamma}}_w = \mathbf{W}'\boldsymbol{\Gamma}$ is a full column rank matrix.*

Under Assumption 5, we can multiply (47) by $\left(\bar{\boldsymbol{\Gamma}}_w'\bar{\boldsymbol{\Gamma}}_w\right)^{-1}\bar{\boldsymbol{\Gamma}}_w'$ from the left to obtain

$$\mathbf{f}_t = \left(\bar{\boldsymbol{\Gamma}}_w'\bar{\boldsymbol{\Gamma}}_w\right)^{-1}\bar{\boldsymbol{\Gamma}}_w'\bar{\mathbf{y}}_{wt} + O_p\left(N^{-1/2}\right),$$

and then using this expression in the VAR model for factors (44), we obtain the following large N VAR representation for $\bar{\mathbf{y}}_{wt}$:

$$\bar{\mathbf{y}}_{wt} = \boldsymbol{\Pi}\bar{\mathbf{y}}_{w,t-1} + \boldsymbol{\xi}_{\bar{y}t}, \tag{48}$$

where $\boldsymbol{\Pi} = \bar{\boldsymbol{\Gamma}}_w\boldsymbol{\Pi}_f\left(\bar{\boldsymbol{\Gamma}}_w'\bar{\boldsymbol{\Gamma}}_w\right)^{-1}\bar{\boldsymbol{\Gamma}}_w'$ and $\xi_{\bar{y}t} = \bar{\boldsymbol{\Gamma}}_w\mathbf{v}_t$. AugGVAR representation in the case of augmentation by $m_{\max}$ cross-section averages can be easily obtained as before, but by using marginal

model (48) instead of (33), and the conditional models (27) augmented with $m_{\max}$ cross-section averages in $\bar{\mathbf{y}}_{wt}$. In particular, we obtain the following AugGVAR representation for $\mathbf{z}_t = (\mathbf{y}_t', \bar{\mathbf{y}}_{wt}')'$:

$$\mathbf{A}_0 \mathbf{z}_t = \mathbf{A}_1 \mathbf{z}_{t-1} + \mathbf{e}_{zt}, \tag{49}$$

where $\mathbf{e}_{zt} = \left(\boldsymbol{\xi}_t', \boldsymbol{\xi}_{\bar{y}t}'\right)'$,

$$\mathbf{A}_0 = \begin{pmatrix} \mathbf{I}_N & -\mathbf{B}_0 \\ \mathbf{0} & \mathbf{I}_{m_{\max}} \end{pmatrix}, \ \mathbf{A}_1 = \begin{pmatrix} \boldsymbol{\Theta} & \mathbf{B}_1 \\ \mathbf{0} & \boldsymbol{\Pi} \end{pmatrix},$$

$\boldsymbol{\Theta}$ is the same as before and $\mathbf{B}_\ell$ for $\ell = 0, 1$ are $N \times m_{\max}$ coefficient matrices that collect the coefficients corresponding to regressors $\bar{\mathbf{y}}_{w,t-\ell}$ in the conditional models (27). Forecasts based on (49) are given by

$$y_{i,t+h|t}^{aug} = \mathbf{e}_{N+m_{\max},i}' \mathbf{A}^h \mathbf{z}_t. \tag{50}$$

As in the case when $m_{\max} = 1$, $y_{i,t+h|t}^{aug}$ can be estimated consistently using the least squares estimates of the unknown parameters on the right side of (50).

When $m < m_{\max}$, augmentation by $m_{\max}$ cross-section averages is clearly not necessary, and as can be seen from (47), $\bar{\mathbf{y}}_{wt}$ are asymptotically (as $N \to \infty$) multicollinear. Nevertheless, the asymptotic multicollinearity does not invalidate the consistency of the AugGVAR forecasts so long as $\bar{\boldsymbol{\Gamma}}_w$ has full column rank. As discussed in Chudik and Pesaran (2011), this rank condition is necessary for the consistency of estimates of individual parameters of the conditional models (27), and it is therefore also necessary for the consistency of the AugGVAR forecasts. The following theorem establishes consistency of the AugGVAR forecasts in the case of multiple factors.

Theorem 2 *Suppose $\mathbf{y}_t$ is generated by model (43), $\mathbf{W}$ is any $N \times m_{\max}$ matrix satisfying conditions (45) and (46), Assumptions 1, and 3-5 hold, and $N, T \xrightarrow{j} \infty$ such that $N/T \to \varkappa$ for some $0 < \varkappa < \infty$. Then for any fixed $0 < h < K$, the h-step-ahead forecast $\hat{\mathbf{y}}_{t+h|t}^{aug}$ satisfies*

$$\left\| \mathbf{y}_{t+h|t} - \hat{\mathbf{y}}_{t+h|t}^{aug} \right\|_\infty \xrightarrow{L_1} 0.$$

Proof of Theorem 2 is provided in a Supplement available from the authors upon request.

Instead of pre-determined cross-section averages, augmentation by principal components could be considered as well. It is analytically more convenient to work with predetermined cross-section averages as opposed to the principal components, which are essentially cross-section averages with weights that contemporaneously depend on the observations, $\mathbf{y}_t$. We leave it for future research to establish asymptotic results when $\bar{\mathbf{y}}_{wt}$ is replaced by $m_{\max}$ principal components.

6 Forecasting with nonsynchronous conditioning

Economic variables are typically released with a lag, which could widely differ across countries and variable types. As a result forecasting must often be carried out with respect to nonsynchronous information sets. An illustrative example of a nonsynchronous conditioning set arises when obser-

vations on a subset of variables are available up to time $t-1$, but for the remainder of the variables observations are available up to t. As before, let $\mathcal{I}_t = \{\mathbf{y}_t, \mathbf{y}_{t-1}, ...\}$, and $\mathcal{F}_t = \{f_t, f_{t-1}, ...\}$, and suppose that $\mathbf{y}_t$ can be partitioned as $\mathbf{y}_t = (\mathbf{y}_{1t}, \mathbf{y}_{2t})'$. Then, a simple example of a nonsynchronous information set is given by $\mathbf{y}_{2t} \cup \mathcal{I}_{t-1} \cup \mathcal{F}_t$.

6.1 Infeasible optimal forecasts with nonsynchronous information

Solving (1) from $t+h$ backward gives

$$\mathbf{y}_{t+h} - \boldsymbol{\gamma} f_{t+h} = \boldsymbol{\Phi}^{h+1} (\mathbf{y}_{t-1} - \boldsymbol{\gamma} f_{t-1}) + \boldsymbol{\Phi}^h \boldsymbol{\varepsilon}_t + \sum_{\ell=0}^{h-1} \boldsymbol{\Phi}^\ell \boldsymbol{\varepsilon}_{t+h-\ell},$$

and after substituting (3) for the factor and taking expectations conditional on $\mathbf{y}_{2t} \cup \mathcal{I}_{t-1} \cup \mathcal{F}_t$, we obtain

$$E(\mathbf{y}_{t+h} | \mathbf{y}_{2t}, \mathcal{I}_{t-1}, \mathcal{F}_t) = \rho^h \boldsymbol{\gamma} f_t + \boldsymbol{\Phi}^{h+1} (\mathbf{y}_{t-1} - \boldsymbol{\gamma} f_{t-1}) + \boldsymbol{\Phi}^h E(\boldsymbol{\varepsilon}_t | \mathbf{y}_{2t}, \mathcal{I}_{t-1}, \mathcal{F}_t). \tag{51}$$

Therefore, in the presence of nonsynchronous conditioning $E(\boldsymbol{\varepsilon}_t | \mathbf{y}_{2t}, \mathcal{I}_{t-1}, \mathcal{F}_t) \neq 0$ and must be derived. Let N_1 (N_2) denote the number of elements of $\mathbf{y}_{1t}$ ($\mathbf{y}_{2t}$), and partition $\boldsymbol{\varepsilon}_t = (\boldsymbol{\varepsilon}_{1t}', \boldsymbol{\varepsilon}_{2t}')'$ conformably, so that the dimension of $\boldsymbol{\varepsilon}_{jt} = N_j$ for $j = 1, 2$. Since $\mathbf{y}_{2t}$ is included in the conditioning set, we have

$$E(\boldsymbol{\varepsilon}_{2t} | \mathbf{y}_{2t}, \mathcal{I}_{t-1}, \mathcal{F}_t) = \boldsymbol{\varepsilon}_{2t},$$

whereas $E(\boldsymbol{\varepsilon}_{1t} | \mathbf{y}_{2t}, \mathcal{I}_{t-1}, \mathcal{F}_t)$ can differ from zero due to possible non-zero correlations between $\boldsymbol{\varepsilon}_{1t}$ and $\boldsymbol{\varepsilon}_{2t}$. Partition the covariance matrix of $\boldsymbol{\varepsilon}_t$, denoted by $\boldsymbol{\Sigma}_\varepsilon = (\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t')$, as

$$\boldsymbol{\Sigma}_\varepsilon = \begin{pmatrix} \boldsymbol{\Sigma}_{\varepsilon 11} & \boldsymbol{\Sigma}_{\varepsilon 12} \\ \boldsymbol{\Sigma}_{\varepsilon 21} & \boldsymbol{\Sigma}_{\varepsilon 22} \end{pmatrix}, \tag{52}$$

where $\boldsymbol{\Sigma}_{\varepsilon jk} = E(\boldsymbol{\varepsilon}_{jt} \boldsymbol{\varepsilon}_{kt}')$ has dimensions $N_j \times N_k$ for $j, k = 1, 2$. The conditional expectations, $E(\boldsymbol{\varepsilon}_{1t} | \mathbf{y}_{2t}, \mathcal{I}_{t-1}, \mathcal{F}_t)$, can now be readily obtained as

$$E(\boldsymbol{\varepsilon}_{1t} | \mathbf{y}_{2t}, \mathcal{I}_{t-1}, \mathcal{F}_t) = \boldsymbol{\Sigma}_{\varepsilon 12} \boldsymbol{\Sigma}_{\varepsilon 22}^{-1} \boldsymbol{\varepsilon}_{2t},$$

and hence

$$E(\boldsymbol{\varepsilon}_t | \mathbf{y}_{2t}, \mathcal{I}_{t-1}, \mathcal{F}_t) = \begin{pmatrix} \boldsymbol{\Sigma}_{\varepsilon 12} \boldsymbol{\Sigma}_{\varepsilon 22}^{-1} \\ \mathbf{I}_{N_2} \end{pmatrix} \boldsymbol{\varepsilon}_{2t}. \tag{53}$$

Substituting (53) in (51), the optimal forecasts in the presence of nonsynchronous conditioning are given by

$$E(\mathbf{y}_{t+h} | \mathbf{y}_{2t}, \mathcal{I}_{t-1}, \mathcal{F}_t) = \rho^h \boldsymbol{\gamma} f_t + \boldsymbol{\Phi}^{h+1} (\mathbf{y}_{t-1} - \boldsymbol{\gamma} f_{t-1}) + \boldsymbol{\Phi}^h \begin{pmatrix} \boldsymbol{\Sigma}_{\varepsilon 12} \boldsymbol{\Sigma}_{\varepsilon 22}^{-1} \\ \mathbf{I}_{N_2} \end{pmatrix} \boldsymbol{\varepsilon}_{2t},$$

where $\boldsymbol{\varepsilon}_{2t} = \mathbf{y}_{2t} - \boldsymbol{\gamma}_2 f_t - \mathbf{S}_2' \boldsymbol{\Phi} (\mathbf{y}_{t-1} - \boldsymbol{\gamma} f_{t-1})$, and $\mathbf{S}_2$ is a selection matrix that selects $\mathbf{y}_{2t}$, defined by $\mathbf{y}_{2t} = \mathbf{S}_2' \mathbf{y}_t$.

The covariance matrix of idiosyncratic errors, $\boldsymbol{\Sigma}_\varepsilon$, plays an important role in the case of nonsynchronous conditioning, in contrast with the case discussed in Section 3, where $\boldsymbol{\Sigma}_\varepsilon$ did not enter the forecasting equations. It is clear that a consistent estimation of $E(\boldsymbol{\varepsilon}_{1t} | \mathbf{y}_{2t}, \mathcal{I}_{t-1}, \mathcal{F}_t)$ is necessary for consistency of feasible forecasts when the conditioning information set is nonsynchronous, which adds further complexity to the forecasting exercise since when N is large, estimation of $\boldsymbol{\Sigma}_\varepsilon$ will be subject to the curse of dimensionality. Estimation of large covariance matrices is discussed in Ledoit and Wolf (2004), Bickel and Levina (2008), Cai and Liu (2011) and Bailey, Pesaran, and Smith (2014).

6.2 Forecasting with GVARs with nonsynchronous conditioning

In the case of the non-augmented GVAR specification (26), feasible forecasts based on the nonsynchronous conditioning set $\mathbf{y}_{2t} \cup \mathcal{I}_{t-1}$ can be obtained as

$$\hat{\mathbf{y}}_{t+h|\mathbf{y}_{2t},\mathcal{I}_{t-1}} = \hat{\mathbf{G}}^{h+1} \mathbf{y}_{t-1} + \hat{\mathbf{G}}^h \hat{\mathbf{u}}_t,$$

where

$$\hat{\mathbf{u}}_t = \begin{pmatrix} \hat{\boldsymbol{\Sigma}}_{u12} \hat{\boldsymbol{\Sigma}}_{u22}^{-1} \\ \mathbf{I}_n \end{pmatrix} \left(\mathbf{y}_{2t} - \mathbf{S}_2' \hat{\mathbf{G}} \mathbf{y}_{t-1} \right),$$

and $\hat{\boldsymbol{\Sigma}}_{ujk}$ for $j, k = 1, 2$ are suitably partitioned sub-matrices of $\hat{\boldsymbol{\Sigma}}_u = \hat{\mathbf{G}}_0^{-1} \hat{\boldsymbol{\Sigma}}_\xi \hat{\mathbf{G}}_0^{-1\prime}$ as in (52), and $\hat{\boldsymbol{\Sigma}}_\xi$ is an appropriate estimator of the covariance matrix of the reduced-form errors $\widehat{\boldsymbol{\xi}}_t$ defined by (24).

Consider now the following augmented GVAR specification (see (34)):

$$\mathbf{z}_t = \hat{\mathbf{A}} \mathbf{z}_{t-1} + \hat{\mathbf{u}}_{zt}, \tag{54}$$

where $\hat{\mathbf{A}} = \hat{\mathbf{A}}_0^{-1} \hat{\mathbf{A}}_1$, and $\hat{\mathbf{u}}_{zt} = \hat{\mathbf{A}}_0^{-1} \hat{\mathbf{e}}_{zt}$. In the case of AugGVAR specification (54), feasible forecasts conditional on $\mathbf{y}_{2t} \cup \mathcal{I}_{t-1}$ can be obtained in a similar way. Assuming that the country-specific and macro shocks are uncorrelated, we have

$$E\left(\widehat{\mathbf{u}_{zt} \mathbf{u}_{zt}'}\right) = \hat{\mathbf{A}}_0^{-1} \begin{pmatrix} \hat{\boldsymbol{\Sigma}}_\xi & \mathbf{0} \\ \mathbf{0} & \hat{\boldsymbol{\sigma}}_{\xi_{\bar{y}}}^2 \end{pmatrix} \hat{\mathbf{A}}_0^{-1\prime},$$

where $\hat{\boldsymbol{\Sigma}}_\xi$ is a suitable estimator of $\boldsymbol{\Sigma}_\xi$, and $\hat{\boldsymbol{\sigma}}_{\xi_{\bar{y}}}^2$ is the estimator of $Var(\xi_{\bar{y}t})$ defined by (33).

In practice, inverting large covariance matrices has proven problematic, and therefore we also consider alternative AugGVAR forecasts that avoid inverting large covariance matrices by estimating a prediction for $\bar{y}_t$ using the nonsynchronous conditioning set $\mathbf{y}_{2t} \cup \mathcal{I}_{t-1}$. Let $\overline{y}_{2t} = N_2^{-1} \sum_{i=N_1+1}^{N} y_{it}$ be the cross-section average of $\mathbf{y}_{2t}$, and note that $\bar{y}_t$ and $\overline{y}_{2t}$ are asymptotically

(as $N_2 \to \infty$) multicollinear, namely,

$$\overline{y}_t = \overline{y}_{2t} + O_p\left(N_2^{-1/2}\right).$$

Predictions for $\overline{y}_t$ based on the nonsynchronous conditioning set $\mathbf{y}_{2t} \cup \mathcal{I}_{t-1}$ can be obtained using an auxiliary regression. We consider

$$\overline{y}_t = \alpha\overline{y}_{t-1} + \beta_0\overline{y}_{2t} + \beta_1\overline{y}_{2,t-1} + \epsilon_t. \tag{55}$$

After constructing the prediction for $\overline{y}_t$, we proceed with forecasting individual elements of $\mathbf{y}_{1t}$ using the conditional models (27) and taking forecasts of $\overline{y}_t$ as given. Forecasts for $\mathbf{y}_{t+h}$, for $h = 1, 2, ...$ can subsequently be obtained recursively using formula (69) by substituting the derived forecasts for $\mathbf{y}_{1t}$.

7 Monte Carlo experiments

This section investigates the relative forecasting performance of augmented and non-augmented GVAR models denoted as before by AugGVAR and GVAR, respectively. Our main objective is to illustrate the main theoretical results of the previous sections on the need to augment GVAR models with additional equations for cross-section averages in cases where the underlying high dimensional VARs contain unobserved common factors. We consider two sets of experiments. In the first set, forecasts for the period $T + 1$ are constructed based on the observed data for time periods $t = 1, 2, ..., T$. These experiments correspond to a conventional forecasting exercise in the literature without nonsynchronous conditioning. In the second set of experiments, we consider forecasting with nonsynchronous conditioning.

7.1 MC Design

Three DGPs are considered: a high-dimensional VAR model without a common factor, and two high-dimensional VARs featuring a common factor. The latter two DGPs differ in the way the factor is introduced in the model and are used to illustrate that the GVAR and AugGVAR methods are robust to the way unobserved factors are specified to enter the underlying DGP.

DGP1: A high-dimensional VAR without a common factor. The first DGP assumes $\gamma_i = 0$ for all i, but allows for weak cross-sectional dependence of errors. $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{Nt})'$ for $t = -M + 1, ..., 0, 1, 2, ..., T$ is generated as

$$\mathbf{y}_t = \mathbf{\Phi}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{56}$$

with starting values $\mathbf{y}_{-M} = \mathbf{0}$. The first $M = 100$ observations are discarded to reduce the effects

of the initial observations on the results. Matrix $\mathbf{\Phi}$ is taken to be block-diagonal,

$$\mathbf{\Phi} = \begin{pmatrix} \mathbf{\Phi}_1 & \mathbf{0}_{2\times 2} & \cdots & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \mathbf{\Phi}_2 & & \mathbf{0}_{2\times 2} \\ \vdots & & \ddots & \vdots \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \cdots & \mathbf{\Phi}_n \end{pmatrix}, \tag{57}$$

where $n = N/2$ with N being an even integer. Matrices $\mathbf{\Phi}_s = (\phi_{sij})$, $s = 1, 2, ..., n$, are 2×2 dimensional with their elements generated randomly as

$$\begin{aligned} \phi_{sij} &\sim IIDU\left(0, 0.7\right), \text{ for } i = j, \text{ and} \\ \phi_{sij} &\sim IIDU\left(0, 0.7 - \phi_{sii}\right), \text{ for } i \neq j. \end{aligned}$$

This ensures $\|\mathbf{\Phi}_s\|_\infty \leq 0.7$, for all s, which implies that $\|\mathbf{\Phi}\|_\infty < 0.7$, and in turn ensures that the DGP is stationary for any $N \in \mathbb{N}$. Replacing the non-zero elements of $\mathbf{\Phi}$ with $O_p\left(N^{-1/2}\right)$ such that $\|\mathbf{\Phi}\|_\infty < 1$ has little effect on the MC findings reported below.[10]

The idiosyncratic errors, $\boldsymbol{\varepsilon}_t$, are generated according to the following spatial autoregressive process:

$$\boldsymbol{\varepsilon}_t = \varrho_\varepsilon \mathbf{S}_\varepsilon \boldsymbol{\varepsilon}_t + \boldsymbol{\eta}_t,\ 0 < \varrho < 1,$$

where $\boldsymbol{\eta}_t = (\eta_{1t}, \eta_{2t}, ..., \eta_{Nt})'$, $\boldsymbol{\eta}_t \sim IIDN\left(\mathbf{0}, \sigma_\eta^2 \mathbf{I}_N\right)$, and the $N \times N$ dimensional spatial weights matrix $\mathbf{S}_\varepsilon$ is given by

$$\mathbf{S}_\varepsilon = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & \cdots & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & & 0 \\ \vdots & & \ddots & \ddots & \ddots & \\ 0 & & & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \cdots & 0 & 1 & 0 \end{pmatrix}.$$

To ensure that the idiosyncratic errors are weakly correlated, the spatial autoregressive parameter, ϱ_ε, must lie in the range $[0, 1)$. We consider a low and a high value for the spatial coefficient and set $\varrho_\varepsilon = 0.2$ and 0.6. We also set σ_η^2 to ensure $N^{-1}\sum_{i=1}^N Var\left(\varepsilon_{it}\right) = 1$.[11]

DGP2: A high-dimensional VAR with an additive common factor. $\mathbf{y}_t$ and f_t, for $t = -M+1, ..., 0, 1, 2, ..., T$, are generated according to

$$\mathbf{y}_t - \boldsymbol{\gamma} f_t = \mathbf{\Phi}\left(\mathbf{y}_{t-1} - \boldsymbol{\gamma} f_{t-1}\right) + \boldsymbol{\varepsilon}_t, \tag{58}$$

[10] In particular, we have considered generating the elements outside the block-diagonal as $\phi_{ij} = \lambda_i \omega_{ij}$, where $\lambda_i \sim IIDU\left(-0.2, 0.2\right)$ and $\omega_{ij} = \varsigma_{ij} / \sum_{j=1}^N \varsigma_{ij}$,with $\varsigma_{ij} \sim IIDU\left(0, 1\right)$. These findings are available in the Supplement.

[11] More specifically, we set $\sigma_\eta^2 = N / \sum_{i=1}^N \mathbf{e}_{iN}' \mathbf{R}_{\boldsymbol{\varepsilon}} \mathbf{R}_\varepsilon' \mathbf{e}_{iN}$, where $\mathbf{e}_{iN}$ is an $N \times 1$ selection vector for the unit i, and $\mathbf{R}_\varepsilon = \left(\mathbf{I}_N - \varrho_\varepsilon \mathbf{S}_\varepsilon\right)^{-1}$.

and

$$f_t = \rho f_{t-1} + \left(1-\rho^2\right)^{1/2} v_t,$$

with the starting values $\mathbf{y}_{-M} = \mathbf{0}$, $f_{-M} = 0$. As before the first $M = 100$ observations are discarded. The coefficient matrix $\mathbf{\Phi}$ and the idiosyncratic errors in $\boldsymbol{\varepsilon}_t$ are generated in the same way as in DGP1. We set $\rho = 0.8$ and generate v_t as $N(0,1)$. Factor loadings are generated as $\gamma_i \sim IIDN\left(\gamma, \sigma_\gamma^2\right)$ with $\gamma = 1$ and $\sigma_\gamma = 0.2$.

DGP3: A high-dimensional VAR with a factor error structure. $f_t \sim IIDN\left(0,1\right)$ and $\mathbf{y}_t$, for $t = -M+1, ..., 0, 1, 2, ..., T$, are generated according to

$$\mathbf{y}_t = \mathbf{\Phi}\mathbf{y}_{t-1} + \boldsymbol{\gamma} f_t + \boldsymbol{\varepsilon}_t, \tag{59}$$

with starting values $\mathbf{y}_{-M} = \mathbf{0}$, and discarding the first $M = 100$ observations. The coefficient matrix $\mathbf{\Phi}$ and the idiosyncratic errors in $\boldsymbol{\varepsilon}_t$ are generated in the same way as in DGP1. Factor loadings are generated as $\gamma_i \sim IIDN\left(\gamma, \sigma_\gamma^2\right)$, with $\sigma_\gamma = 0.2$ and γ is set to ensure that $N^{-1}\boldsymbol{\tau}_N'\left(\mathbf{I}-\mathbf{\Phi}\right)^{-1}\boldsymbol{\tau}_N\gamma = 1$, where $\boldsymbol{\tau}_N$ is an $N \times 1$ vector of ones.

All experiments are carried out for $N, T \in \{30, 50, 100, 200, 500\}$, and replicated $R = 2,000$ times.

7.2 Individual forecasts and average MSFEs

7.2.1 Forecasting with synchronous conditioning

Our primary objective is to investigate the forecasting performance of the AugGVAR and the non-augmented GVAR for horizon $h = 1$ (one-step-ahead forecasts). We do so by comparing these forecasts with their infeasible counterparts. In particular, we compute the following average mean square forecast errors (MSFE) relative to the optimal infeasible forecasts:

$$MSFE_{RN}\left(T+1|T\right) = \frac{\sum_{r=1}^{R}\sum_{i=1}^{N}\left(\hat{y}_{i,T+1|T}^{(r)} - y_{i,T+1}^{(r)}\right)^2}{\sum_{r=1}^{R}\sum_{i=1}^{N}\left[E\left(y_{i,T+1}^{(r)}\middle|\mathcal{I}_t^{(r)}, \mathcal{F}_t^{(r)}\right) - y_{i,T+1}^{(r)}\right]^2}, \tag{60}$$

where $\mathcal{I}_t^{(r)} = \left\{\mathbf{y}_t^{(r)}, \mathbf{y}_{t-1}^{(r)}, ...\right\}$, $\mathcal{F}_t^{(r)} = \left\{f_t^{(r)}, f_{t-1}^{(r)}, ...\right\}$, and $y_{i,T+1}^{(r)}$ is the realized value for unit i, at time $T+1$, and the Monte Carlo replication, r. Similarly, we compute the MSFE for AugGVAR forecasts $\hat{y}_{i,T+1|T}^{aug}$. The optimal infeasible one-step-ahead forecasts are computed as (we are dropping the superscript (r) to simplify the notations)

$$E\left(y_{i,T+1}|\mathcal{I}_T, \mathcal{F}_T\right) = \begin{cases} \mathbf{e}_{iN}'\mathbf{\Phi}\mathbf{y}_T, & \text{in the case of DGP1} \\ \gamma_i\rho f_T + \mathbf{e}_{iN}'\mathbf{\Phi}\left(\mathbf{y}_T - \boldsymbol{\gamma} f_T\right), & \text{in the case of DGP2} \\ \gamma_i\rho f_T + \mathbf{e}_{iN}'\mathbf{\Phi}\mathbf{y}_T, & \text{in the case of DGP3} \end{cases}. \tag{61}$$

The non-augmented GVAR forecasts ($\hat{y}_{i,T+1|T}$) are based on the following regressions:

$$y_{it} = c_i + \phi_{ii} y_{i,t-1} + \phi_{i,i+1} y_{i+1,t-1} + \sum_{\ell=0}^{p} b_{i\ell} \overline{y}_{wi,t-\ell} + \xi_{it}, \text{ for } i = 1, 3, 5, ..., N-1, \tag{62}$$

$$y_{it} = c_i + \phi_{ii} y_{i,t-1} + \phi_{i,i-1} y_{i-1,t-1} + \sum_{\ell=0}^{p} b_{i\ell} \overline{y}_{wi,t-\ell} + \xi_{it}, \text{ for } i = 2, 4, 6, ..., N, \tag{63}$$

where $\overline{y}_{wit} = \sum_{j=1}^{N} w_{ij} y_{jt}$. Aggregation weights are such that $\overline{y}_{wit}$ is a simple cross-section average of units that do not directly enter individual cross-section augmented regressions in (62)-(63). In particular, when i is odd, $w_{ii} = w_{i,i+1} = 0$ and $w_{ij} = (N-2)^{-1}$ for $i \neq j, j+1$; and when i is even, $w_{ii} = w_{i,i-1} = 0$, and $w_{ij} = (N-2)^{-1}$ for $i \neq j, j-1$. Let $\mathbf{W} = [w_{ij}]$ and

$$\hat{\mathbf{B}}_\ell = diag\left(\hat{\mathbf{b}}_\ell\right) \text{ for } \ell = 0, 1, ..., p,$$

where $\hat{\mathbf{b}}_\ell$ is the least squares estimate of $\mathbf{b}_\ell = (b_{1\ell}, b_{2\ell}, ..., b_{N\ell})'$. The estimated (non-augmented) GVAR representation is

$$\mathbf{y}_t = \hat{\boldsymbol{\delta}} + \sum_{\ell=1}^{p} \hat{\boldsymbol{\Psi}}_\ell \mathbf{y}_{t-\ell} + \hat{\mathbf{u}}_t, \tag{64}$$

which yields the GVAR forecasts

$$\hat{\mathbf{y}}_{T+1|T} = \sum_{\ell=1}^{p} \hat{\boldsymbol{\Psi}}_\ell \mathbf{y}_{T+1-\ell} + \hat{\boldsymbol{\delta}}, \tag{65}$$

where $\hat{\boldsymbol{\Psi}}_\ell = \hat{\mathbf{G}}_0^{-1} \hat{\mathbf{G}}_\ell$, for $\ell = 1, 2, ..., p$, $\hat{\boldsymbol{\delta}} = \hat{\mathbf{G}}_0^{-1} \hat{\mathbf{c}}$, $\hat{\mathbf{G}}_0 = \mathbf{I}_N - \hat{\mathbf{B}}_0 \mathbf{W}$, $\hat{\mathbf{G}}_1 = \hat{\boldsymbol{\Phi}} + \hat{\mathbf{B}}_1 \mathbf{W}$, $\hat{\mathbf{G}}_\ell = \hat{\mathbf{B}}_\ell \mathbf{W}$, for $\ell = 2, 3, .., p$, $\hat{\boldsymbol{\Phi}}$ is a block-diagonal matrix constructed based on the estimates of the block-diagonal coefficients in (62)-(63), $\hat{\mathbf{c}} = (\hat{c}_1, \hat{c}_2 ..., \hat{c}_N)'$ is the vector of estimated fixed effects in (62)-(63), and $\hat{\mathbf{u}}_t = \hat{\mathbf{G}}_0^{-1} \hat{\xi}_t$.

One-step-ahead forecasts based on an augmented GVAR model ($\hat{\mathbf{y}}^{aug}_{T+1|T}$) are constructed in a similar way as described in Section 4. In particular, the following regressions are estimated instead of (62)-(63):

$$y_{it} = c_i + \phi_{ii} y_{i,t-1} + \phi_{i,i+1} y_{i+1,t-1} + \sum_{\ell=0}^{p} b_{i\ell} \overline{y}_{t-\ell} + \xi_{it}, \text{ for } i = 1, 3, 5, ..., N-1, \tag{66}$$

$$y_{it} = c_i + \phi_{ii} y_{i,t-1} + \phi_{i,i-1} y_{i-1,t-1} + \sum_{\ell=0}^{p} b_{i\ell} \overline{y}_{t-\ell} + \xi_{it}, \text{ for } i = 2, 4, 6, ..., N, \tag{67}$$

together with

$$\bar{y}_t = c_{\bar{y}} + \sum_{\ell=1}^{p} \rho_\ell \bar{y}_{t-\ell} + \xi_{\bar{y}t}, \tag{68}$$

where $\bar{y}_t = N^{-1}\sum_{i=1}^{N} y_{it}$. Individual elements of $\hat{\mathbf{y}}_{T+1|T}^{aug}$ are given by

$$\hat{y}_{i,T+1|T}^{aug} = \mathbf{e}_{i,N+1}'\hat{\mathbf{z}}_{T+1} = \mathbf{e}_{i,N+1}'\left(\hat{\boldsymbol{\delta}} + \sum_{\ell=1}^{p}\hat{\boldsymbol{\Upsilon}}_{\ell}\mathbf{z}_{T-\ell}\right), \tag{69}$$

where $\mathbf{z}_t = (\mathbf{y}_t', \bar{y}_t)'$, $\hat{\boldsymbol{\Upsilon}}_{\ell} = \hat{\mathbf{A}}_0^{-1}\hat{\mathbf{A}}_{\ell}$, for $\ell = 1, 2, ..., p$, $\hat{\boldsymbol{\delta}} = (\hat{\mathbf{c}}', \hat{c}_{\bar{y}})'$,

$$\hat{\mathbf{A}}_0 = \begin{pmatrix} \mathbf{I}_N & -\hat{\mathbf{b}}_0 \\ \mathbf{0}_{1\times N} & 1 \end{pmatrix}, \hat{\mathbf{A}}_1 = \begin{pmatrix} \hat{\boldsymbol{\Phi}} & \hat{\mathbf{b}}_1 \\ \mathbf{0}_{1\times N} & \hat{\rho}_1 \end{pmatrix}, \text{ and } \hat{\mathbf{A}}_{\ell} = \begin{pmatrix} \mathbf{0}_{N\times N} & \hat{\mathbf{b}}_{\ell} \\ \mathbf{0}_{1\times N} & \hat{\rho}_{\ell} \end{pmatrix}, \text{ for } \ell = 2, 3, ..., p,$$

in which all estimated coefficients are based on (66)-(68).

The number of lags for cross-section averages in both augmented and non-augmented GVARs is set to $p = \left[T^{1/3}\right]$, where [.] denotes the integer part.

7.2.2 Forecasting with nonsynchronous conditioning

We consider forecasting the period $T+1$ conditional on a nonsynchronous information set, which includes observations on odd cross-section units for periods $t = 1, 2, ..., T-1$, and even cross-section units for periods $t = 1, 2, ..., T$. We consider the following nonsynchronous conditioning information set:

$$\mathcal{S}_T \equiv \mathcal{S}_{1,T-1} \cup \mathcal{S}_{2T}, \tag{70}$$

where $\mathcal{S}_{1,T-1} = \{y_{it}, t = 1, 2, ..., T-1, i = 1, 3, 5, ..., N-1\}$, and $\mathcal{S}_{2T} = \{y_{it}, t = 1, 2, ..., T, i = 2, 4, 6, ..., N\}$.

As in the case of forecasting without nonsynchronous conditioning, we compute the simple cross-section average MSFE of the feasible GVAR nowcasts relative to the optimal infeasible nowcasts

$$MSFE_{RN}(T+1|T) = \frac{\sum_{r=1}^{R}\sum_{i=1}^{N}\left(\hat{y}_{i,T+1|\mathcal{S}_T}^{(r)} - y_{i,T+1}^{(r)}\right)^2}{\sum_{r=1}^{R}\sum_{i=1}^{N}\left[E\left(y_{i,T+1}^{(r)}\middle| S_T^{(r)} \cup \mathcal{F}_T^{(r)}\right) - y_{i,T+1}^{(r)}\right]^2}, \tag{71}$$

and similarly for the AugGVAR forecasts $\hat{y}_{i,T+1|\mathcal{S}_T}^{aug}$.

We compute GVAR forecasts in the presence of nonsynchronous conditioning as outlined in Section 6, and consider two options for the estimation of the covariance matrix of idiosyncratic shocks. First is Ledoit and Wolf (2004)'s estimator of $\boldsymbol{\Sigma}_{\xi}$, denoted as $\hat{\boldsymbol{\Sigma}}_{\xi,LW}$. The second option is to use the following block-diagonal specification:

$$\hat{\boldsymbol{\Sigma}}_{\xi,B} = \begin{pmatrix} \hat{\boldsymbol{\Sigma}}_1 & \mathbf{0}_{2\times 2} & \cdots & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \hat{\boldsymbol{\Sigma}}_2 & & \mathbf{0}_{2\times 2} \\ \vdots & & \ddots & \vdots \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \cdots & \hat{\boldsymbol{\Sigma}}_n \end{pmatrix}, \tag{72}$$

where $\hat{\boldsymbol{\Sigma}}_s$ is the sample estimate of the covariance matrix of the 2×1 vector $\left(\xi_{2s-1,t}\xi_{2s,t}\right)'$ for

$s = 1, 2, ..., n$.[12]

Forecasting with an AugGVAR is subject to the same problems as in the case of a non-augmented GVAR when the conditioning information set is nonsynchronous. We consider the same two options for estimating the large-dimensional covariance matrix $\mathbf{\Sigma}_\xi$, namely $\hat{\mathbf{\Sigma}}_{\xi,LW}$, and $\hat{\mathbf{\Sigma}}_{\xi,B}$. In addition we consider the AugGVAR forecasts (see Section 6.2) that avoid inverting large covariance matrices.

7.3 Monte Carlo results

7.3.1 Case of synchronous conditioning

Table 1 reports the results for the augmented and non-augmented GVAR methods in experiments with low cross-section dependence of idiosyncratic shocks ($\varrho_\varepsilon = 0.2$) and a sparse matrix $\mathbf{\Phi}$**.** The top panel of this table presents relative MSFE in the case of data generated by a high-dimensional VAR model without a common factor. We can see that both augmented and non-augmented GVAR methods converge to the infeasible forecasts as the sample size grows, and the difference between the GVAR and AugGVAR is minimal with the latter marginally better. It is interesting to observe that the augmentation with an additional equation for cross-section averages, although asymptotically redundant, does not worsen the forecasting performance. We also observe that an increase in the time dimension is crucial for the improvement in the forecasting performance, as expected, whereas increasing N (beyond 30) does not seem to make that much of a difference to the results.

In contrast, qualitatively different results are reported in the middle and bottom panels of Table 1 which report the results for the two specifications of the VAR with an unobserved common factor. AugGVAR forecasts are not affected by the inclusion of the factors, and their performance is generally similar to those reported in the top panel of the table for the VAR model without a factor. This confirms that the AugGVAR is robust to the way the unobserved factor is introduced in the analysis. However, the performance of the GVAR without augmentation deteriorates considerably with the introduction of an unobserved common factor, especially when T is small and N large. This finding is in line with our theoretical result which suggests that in the presence of a common factor the contemporaneous matrix $\mathbf{G}_0$ becomes singular as $N \to \infty$. The results clearly illustrate that the AugGVAR performs well, irrespective of whether the underlying VAR contains a factor or not. Also, when a factor is included, the results are robust to the way the factor is introduced in the VAR.

The findings for the experiments with a high spatial coefficient (namely $\varrho_\varepsilon = 0.6$) and/or a non-sparse coefficient matrix $\mathbf{\Phi}$ are qualitatively similar and are reported in a Supplement which is available upon request.

7.3.2 Case of nonsynchronous conditioning

The results for forecasting with nonsynchronous conditioning are summarized in Table 2. Similarly to Table 1, this table reports the results for experiments with $\varrho_\varepsilon = 0.2$ and a sparse matrix $\mathbf{\Phi}$.

[12] While we acknowledge that several other options for estimating large covariance matrices have been proposed in the literature, we do not consider them here. We leave this important topic for future research.

Recall that an estimate of the covariance matrix $\mathbf{\Sigma}_\xi$ is required for the computation of forecasts when the conditioning information set is nonsynchronous. Table 2 summarizes the findings when $\hat{\mathbf{\Sigma}}_{\xi,LW}$ and $\hat{\mathbf{\Sigma}}_{\xi,B}$ are used (in the case of both the GVAR and the AugGVAR) and when a nonsynchronous cross-section average is used instead of an estimate of $\mathbf{\Sigma}_\xi$ (AugGVAR).[13] Table 2 shows that with nonsynchronous conditioning, augmentation continues to be preferable. It does no harm (or marginally improves the forecasting performance) when no unobserved common factor is present, and continues to perform well when a factor is present. It is also robust to the way the unobserved common factor is introduced in the underlying VAR model. Moreover, the GVAR forecasts without augmentation perform poorly when a factor is present and T/N is small. Similar results (reported in the Supplement) are obtained in the case of experiments with a high value of the spatial AR parameter, ϱ_ε, and/or a non-sparse matrix $\mathbf{\Phi}$.

Regarding the choice of the estimator of $\mathbf{\Sigma}_\xi$, we found that no clear ordering is observed between AugGVAR$\left(\hat{\mathbf{\Sigma}}_{\xi,LW}\right)$, AugGVAR$\left(\hat{\mathbf{\Sigma}}_{\xi,B}\right)$ and the AugGVAR forecasts, where the cross-section averages are directly forecast. For experiments with a low value for the SAR parameter, $\varrho_\varepsilon = 0.2$, the forecasts from AugGVAR outperform forecasts based on $\hat{\mathbf{\Sigma}}_{\xi,LW}$, but this is not always the case when ϱ_ε is increased to 0.6 and T is relatively large.

The small sample evidence presented in this section overwhelmingly supports augmenting the GVAR with additional equations for cross-section averages when factors are present, and shows that there is no harm in augmentation when factors are absent. The results also show that under nonsynchronous conditioning, no clear conclusion regarding the choice of $\hat{\mathbf{\Sigma}}_\xi$ emerges.

8 Empirical application: forecasting GDP using PMIs

In this section we apply a number of different methods for the analysis of large data sets, including the GVAR and AugGVAR, to assess the extent to which using PMIs helps forecast GDP growth in a multi-country setting. We also provide a comparative analysis of the alternative forecasting techniques, with particular emphasis on a comparison of GVAR and AugGVAR outcomes. We begin by describing the data first, followed by a summary description of forecasting methods.

8.1 GDP and PMI data

We have compiled a panel of quarterly data on real output covering 48 countries representing 92% of world output. We chose the starting period to be 1998Q4, for which quarterly output data for all 48 countries is available, and at the same time we also have a good country coverage for PMI data. The latest available observation on output is 2013Q2. All of the output data is seasonally adjusted, most series by the source. Table A.1 describes the sources and construction of the output data in detail. We denote the first differences in the logarithm of real output in country i and quarter t by x_{it}, for $i = 1, 2, ..., N$; and $t = 1, 2, ..., T$, where $t = 1$ corresponds to 1999Q1 (due to differencing) and $T = 58$ corresponds to 2013Q2. Figure A.1 plots x_{it} for the group of advanced

[13] The variance matrix estimators $\hat{\mathbf{\Sigma}}_{\xi,LW}$ and $\hat{\mathbf{\Sigma}}_{\xi,B}$ are defined in Subsection 7.2.2.

economies (Panel A) and emerging economies (Panel B) over the period 1999Q1-2013Q2.[14]

PMIs are reported monthly as seasonally adjusted diffusion indices in which a number greater than 50 indicates an expansion, and a number below 50 indicates a contraction. We use two types of PMIs: manufacturing PMIs denoted as $\kappa_{i,\mathfrak{m},t}$, and services PMIs denoted as $s_{i,\mathfrak{m},t}$. Subscripts $\mathfrak{m}$ and t refer to month $\mathfrak{m}$ in quarter t. PMIs are not available for all countries in our dataset. We have manufacturing PMI data on 30 countries with a sufficiently long history. Country coverage on services PMIs is much less comprehensive with only 10 countries having available data with a sufficiently long history. Table A.2 provides further details on country, time coverage, and sources of the PMI data. Figures A.2 and A.3 plot the manufacturing and services PMIs, respectively.

8.1.1 Information sets

We use $\Omega_{\mathfrak{m}t}$ to denote the available information set (consisting of both quarterly and monthly data) at the end of month $\mathfrak{m} = 1, 2, 3$ of quarter t. We are interested in forecasting output growth in country i in period $t+h$ conditional on the information set available at the end of month $\mathfrak{m} = 1, 2, 3$ of quarter t. We omit reference to the information set $\Omega_{\mathfrak{m}t}$ explicitly to economize on notations, but it will be understood that all forecasts are conditional on the nonsynchronous information set $\Omega_{\mathfrak{m}t}$.

We denote the latest available observation on country i output growth in the information set $\Omega_{\mathfrak{m}t}$ as $x_{i\tau_{xi}}$, where $\tau_{xi} = \tau_x(i, \Omega_{\mathfrak{m}t})$ is a function that depends on t, the chosen month $\mathfrak{m}$, and the country i, but we abbreviate this function as τ_{xi}. We also denote the difference between t and the latest period for which an observation is available on x_{it} by $s_{xi} = t - \tau_{xi}$.

8.1.2 From monthly PMIs to quarterly PMIs

Dealing with different frequencies is not a central contribution of this paper, and we follow a simple solution of transforming monthly data into quarterly observations as opposed to developing a fully fledged mixed-frequency model (such as the MIDAS approach mentioned in the Introduction). In particular, we consider two ways of transforming monthly observations into a quarterly series.

First we employ a sequential sampling scheme where for a given month, $\mathfrak{m}$, we define

$$\bar{\kappa}^s_{it}(\mathfrak{m}) = \kappa_{i,\mathfrak{m},t}, \text{ and } \bar{s}^s_{it}(\mathfrak{m}) = s_{i,\mathfrak{m},t}, \text{ for } \mathfrak{m} = 1, 2, 3 \tag{73}$$

where superscript s stands for sequential sampling, $i = 1, 2, ..., N$ indexes individual countries and $t = 1, 2, ..., T$ indexes quarterly time periods. This gives us three sequentially sampled quarterly series. The latest available monthly observation is used in estimation of the parameters of the

[14] There are two countries with notable outliers in the group of emerging economies: Venezuela (2003Q1-Q2) and Thailand (2011Q4-2012Q2). Venezuela had a recession in 2002-03, low oil prices, a coup attempt in 2002 and a business strike. Thailand had massive flooding in late 2011 that disrupted the economy. Looking at both advanced and emerging economies, there appears to be large cross-sectional comovement across countries, especially during the 2007-08 global financial crisis.

forecasting equations. Second, we use a temporally aggregated measure, defined by

$$\bar{\kappa}_{it}^{a}(1) = (\kappa_{i,2,t-1} + \kappa_{i,3,t-1} + \kappa_{i,1,t})/3, \tag{74}$$

$$\bar{\kappa}_{it}^{a}(2) = (\kappa_{i,3,t-1} + \kappa_{i,1,t} + \kappa_{i,2,t})/3, \tag{75}$$

$$\bar{\kappa}_{it}^{a}(3) = (\kappa_{i,1,t} + \kappa_{i,2,t} + \kappa_{i,3,t})/3, \tag{76}$$

where as before $\bar{\kappa}_{it}^{a}(\mathfrak{m})$, for $\mathfrak{m} = 1, 2, 3$ denote month $\mathfrak{m}$ of quarter t. Similar temporally aggregated services PMI series can be constructed. As in the case of sequential sampling, we always select $\mathfrak{m}$ based on the latest available monthly observation in $\Omega_{\mathfrak{m}t}$.

In the case of the forecasts that make use of PMIs we compute two sets of forecasts: one based on sequentially sampled PMIs, and the other based on temporally aggregated PMIs. We report a simple average of the two forecasts. In this way we avoid the potential data mining problem that could arise due to the choice of data transformation from monthly to quarterly observations.

The timing of data releases differs across countries and by variable types. As a general rule, manufacturing PMI data is released on the first working day of the month after the reference period. Israel and New Zealand release their manufacturing PMI data in the middle of the month after the reference period. Services PMI data is released on the third working day of the month after the reference period. GDP releases vary substantially across countries—some countries adhere to a strict release schedule, while the publication date for others can be variable and/or affected by national holidays. Figure A.4 plots the GDP release lags for each of the countries in our sample ordered by the number of days after the beginning of the reference quarter for Q2 of 2013. We assume the same schedule applies to previous and subsequent releases, although release lags may vary.

8.2 Forecasting methods with a large number of predictors

We consider three basic benchmarks and a number of data-rich methods summarized below. A detailed description of individual methods is provided in the Supplement.

Let $\mathbf{y}_{it}$ be a $k_i \times 1$ vector of country-specific quarterly variables consisting of output growth (x_{it}) and, where available, manufacturing and services PMI country indices. Thus, $k_i = 3$ if all three series are available, in which case $y_{it}(\mathfrak{m}_i) = (x_{it}, \bar{\kappa}_{it}(\mathfrak{m}_i), \bar{s}_{it}(\mathfrak{m}_i))'$, whereas $k_i = 1$ or 2 if one or both PMI indices are not available.[15] We employ the GVAR model as given by (26), including an intercept term. We compute country-specific cross-section averages as $\overline{\mathbf{y}}_{wit} = (\bar{x}_{wit}, \overline{\bar{\kappa}}_{wit}, \overline{\bar{s}}_{wit})'$, where $\bar{x}_{wit} = \sum_{j=1}^{N} w_{ij} x_{it}$ is the cross-section average of output growth, $\overline{\bar{\kappa}}_{wit} = \sum_{j \in \mathcal{I}_\kappa} w_{ij}^{\kappa} \bar{\kappa}_{jt}(\mathfrak{m}_j)$ is the cross-section average of manufacturing PMI indices, and $\mathcal{I}_\kappa$ is the index set of countries with available manufacturing PMI data. Similarly, $\overline{\bar{s}}_{wit}$ is defined as $\overline{\bar{s}}_{wit} = \sum_{j \in \mathcal{I}_s} w_{ij}^{s} \bar{s}_{jt}(\mathfrak{m}_j)$, in which $\mathcal{I}_s$ is the index set of countries with available services PMI data. The weights $\{w_{ij}\}$ are based on bilateral aggregate trade flows obtained from the IMF DOTS database such that $w_{ii} = 0$ and

[15] Strictly speaking the vector of observations on country i should be defined as $y_{it}^{j}(q_i) = \left(x_{it}, \bar{\kappa}_{it}^{j}(q_i), \bar{s}_{it}^{j}(q_i)\right)'$ for $j = s, a$, which makes the choice of transformation from monthly PMIs to quarterly observations explicit. But here we have simplified the notation for ease of exposition.

$\sum_{j=1}^{N} w_{ij} = 1$ for all i.[16] Weights used for PMI indices are constructed from $\{w_{ij}\}$ as follows: $w_{ij} = 1$ if $j \in \mathcal{I}_{\varkappa}$ and 0 otherwise.

We allow for only one lag of $\mathbf{y}_{it}(\mathfrak{m}_i)$ and $\bar{\mathbf{y}}_{wit}$ in the conditional VAR models, (22), due to the short sample available. For the full sample $T = 58$, but in the out-of-sample forecasting exercise the first forecast is made for 2006Q1, which leaves us with 28 quarterly observations to estimate the conditional VAR models. We proceed with model (26) to derive conditional forecasts in the same way as outlined in Section 6.2. We denote the GVAR forecasts as GVAR-PMI and GVAR, depending on whether PMI data is included in $\mathbf{y}_{it}$, or only output growth is considered, in which case $k_i = 1$.

8.2.1 Augmented GVARs

We use the AugGVAR representation (see (49)) derived from the marginal VAR model (48) featuring arithmetic cross-section averages denoted as $\overline{\mathbf{y}}_t = (\bar{x}_t, \overline{\bar{\kappa}}_t, \overline{\bar{s}}_t)'$, and from individual conditional models (22), in which $\mathbf{y}_{it}$ is defined in the same way as in the case of the non-augmented GVAR above and the augmentation is carried out with simple cross-section averages, $\overline{\mathbf{y}}_t$. Augmented GVAR forecasts for the target variables, x_{it}, are constructed in the same way as outlined in Section 6.2, and are denoted as AugGVAR-PMI and AugGVAR, depending on whether the PMI variables are included in $\mathbf{y}_{it}$. Multi-step ahead forecasts from the GVAR and AugGVAR methods are computed iteratively. The remaining data-rich forecasts explained below are computed using the direct approach where different regressions are considered for computing forecasts at different horizons. For a discussion of iterative and direct procedures for computation of multi-step ahead forecasts see, for example, Ing (2003), Marcellino et al. (2006), and Pesaran et al. (2011).

8.2.2 Lasso regressions

Our next data-rich forecasting method is based on Lasso regressions, popularized in the literature following the seminal contribution of Tibshirani (1996). A recent textbook exposition of the Lasso regression can be found in Hastie et al. (2009).[17] The forecasts of $x_{i,t+h}$ are based on the linear penalized regressions of x_{it} on all $k = \sum_{i=1}^{N} k_i$ predictors lagged by h quarters. But before running the regressions we first standardize the predictors using the information available at time $t - h$. The estimation is carried out by minimizing the sum of squared residuals subject to the Lasso constraint, which bounds by λ_i the sum of absolute values of estimated coefficients. We denote the Lasso forecasts by LASSO-PMI or LASSO depending on whether PMI data is included in the set of predictors.

[16] Weights are constructed as a ratio of total exports (from i to j) and imports (from j to i) over total foreign trade (of country i) using 2000-2010 trade data.

[17] However, it is important to note that the use of Lasso (and Ridge below) is theoretically justified in the case of exogenous predictors and does not necessarily apply to the dynamic case where the predictors are lagged values of the dependent variables from a large dimensional VAR.

8.2.3 Ridge regressions

Ridge forecasts are constructed similarly to Lasso forecasts, with the difference that instead of constraining the sum of absolute values of coefficients, it is the sum of squared coefficients which is restricted to not exceed λ_i. The consequence of the Ridge constraint is that it does not penalize small coefficients as much as the Lasso constraint. The main difference between the Ridge and Lasso is therefore the tendency of the Ridge regression to favor many small coefficients as opposed to the Lasso which tends to select a small number of nonzero coefficients. Ridge regression can also be interpreted as a Bayesian normal regression with Gaussian priors. For further details and applications of the Ridge approach in economics see De Mol et al. (2008), Lin and Tsay (2006), Groen and Kapetanios (2008), and Eickmeier and Ng (2011). We denote the Ridge forecasts as RIDGE-PMI (when both output and PMI data are included in the set of predictors), and RIDGE (when only output data is included in the set of predictors).

8.2.4 Factor models

Instead of estimating a linear relationship between the target variable, x_{it}, and k predictors, an alternative strategy considered in the literature is to shrink the large number of available predictors first into a small $m \times 1$ dimensional vector of factors (pooled predictors) and then forecast the target variable in terms of these m factors. To this end both static (principal components) and dynamic factors are used. Dynamic factor models were introduced by Geweke (1977) and Sargent and Sims (1977), and later generalized to allow for weak cross-sectional dependence by Forni and Lippi (2001), Forni et al. (2000) and Forni et al. (2004). In a typical macroeconomic dataset, empirical evidence suggests that few factors are needed to explain a significant portion of the co-variations of the predictors under consideration (see Stock and Watson (1999), Stock and Watson (2002), Giannone, Reichlin, and Sala (2005), Bai and Ng (2007) and Stock and Watson (2005)).

We use the method of principal components and extract the first m principal components of the k predictors available, after standardization. A key choice is the number of factors to use in the subsequent analysis. We estimate separate models for $m = 1, 2, ..., 5$ factors and then average the corresponding forecasts. We denote the corresponding forecasts as FM-PMI and FM, depending on whether the set of predictors contains PMIs. This procedure is followed as a diversification device to avoid the difficult choice of determining the optimal number of factors.

8.2.5 Factor-augmented AR models

Factor-augmented autoregressive (FAR) forecasts are computed in the same way as the FM forecasts, but the model is augmented with lagged values of x_{it}. We consider again up to 5 factors ($m = 1, 2, ..., 5$) and average the forecasts that result for each value of m. Depending on whether or not PMIs are included when extracting the factors, the corresponding forecasts are denoted by FAR-PMI and FAR, respectively.

8.2.6 Partial least squares regressions

Partial least squares (PLS) regressions are due to Wold (1982), who proposed constructing factors based on the covariance of the predictors with the target variable(s). We estimate PLS factors from the set of standardized predictors in the same way as in Groen and Kapetanios (2008). As with the FM method, we consider up to 5 factors and then average across the corresponding forecasts. We use PLS-PMI and PLS to denote the forecasts based on PLS regressions with and without PMI data.

8.3 Choice of the penalty parameter

The selection of the shrinkage parameter, λ_i, has important consequences for the forecasting performance in a data-rich environment and the choice of λ_i should therefore be made with care. This problem has been addressed in different ways in the literature. Perhaps the most common solution is to choose λ_i by cross-validation. Although, *a priori* fixed values for λ_i have also been used in the literature. See, for example, Groen and Kapetanios (2008). In our forecasting exercise we consider a number of different options for the selection of λ_i. For the Lasso and Ridge methods, we employ the following 7 options.

Option 1 : λ_i is set to 0.25 for all i, as in Groen and Kapetanios (2008).

Option 2 : λ_i is chosen based on an 80%-20% split of the available observations, with the first 80% of the observations used as the training sub-sample and the last 20% as the evaluation sub-sample. We compute forecasts using a fine grid of $\lambda_i \in \{0.01, 0.02, ..., 2\}$ and choose λ_i with the smallest MSFE computed based on the evaluation sub-sample.

Option 3 : λ_i is set to a simple average of the penalty parameters estimated under Option 2.

Option 4 : λ_i is restricted to be the same across all i, but unlike Option 3, we choose the value of λ for which the average of the RMSFEs from Option 2 above is minimized.

Option 5 : λ_i is chosen by standard 10-fold cross-validation for all i.

Option 6 : λ_i is set equal to a simple average of the penalty parameters estimated under Option 5.

Option 7 : λ_i is restricted to be the same across all i, but unlike Option 6, we choose the value of $\lambda_i = \lambda$ for which the average of the RMSFEs from Option 5 above is minimized.

In the case of the AugGVAR model, we choose the shrinkage parameter, λ_{Aug}, based on an 80%-20% split of the available sample with the first 80% of the observations used as the training sub-sample and the last 20% as the evaluation sub-sample (as in Option 2). We compute forecasts using a fine grid of $\lambda_A \in \{0.01, 0.02, ..., 2\}$ and choose λ_{Aug} with the smallest MSFE based on the evaluation sub-sample.

8.4 Benchmark forecasts

We consider three benchmarks. A random walk (RW) benchmark where the forecasts (at all horizons) are set to the latest available observation on output growth. A first-order autoregression, AR(1), benchmark where output growth forecasts at different horizons are computed using the direct approach where $x_{i,t+h}$ is regressed on an intercept and x_{it}. The third benchmark is an extension of the AR(1) benchmark where the AR(1) model is augmented with domestic PMIs. As in the case of data-rich methods that use PMI data, we compute two benchmark forecasts, one using sequentially sampled PMIs and a second one using temporally aggregated PMIs. The PMI-augmented AR benchmark is then given by a simple average of the two forecasts which we denote by AR-PMI.

8.5 Empirical results

Using the alternative forecasting schemes set out above we generated recursive quarterly forecasts of GDP growth for all 48 countries over the period $2006Q1-2013Q2$ using an expanding estimation window starting in $1999Q1$. To compare the average forecasting performance of the different schemes we first computed MSFEs for each country over the evaluation sample, $2006Q1-2013Q2$, for different PMI release months within a quarter, $\mathfrak{m}=1,2,3$, and the forecast horizons, $h=0,1,2$ quarters ahead. We then computed a GDP-weighted average of these MSFEs using 2013 GDP measures in PPP terms which we report in the tables below.

First we consider how AugGVAR forecasts perform as compared to the GVAR forecasts without augmentation. Table 3 reports the average GDP-weighted MSFEs for the AugGVAR-PMI relative to the non-augmented GVAR-PMI, when Ledoit and Wolf (2004)'s estimator of the error covariance matrix, $\hat{\mathbf{\Sigma}}_{\xi,LW}$, is used to take account of the nonsynchronous nature of the GDP and PMI release dates (see Section 6.2). As can be seen from this table the average MSFE of the augmented GVAR at horizon $h=0$ for the different PMI release months, $\mathfrak{m}=1,2,3$, range between 13 and 30 percent of the MSFE of the non-augmented GVAR, which means that the augmented GVAR has about 3 to 7 times smaller MSFE than the benchmark. The differences in the forecasting performance of the augmented and non-augmented procedures are even more pronounced at longer horizons. Similar results are also obtained when other estimators of the covariance matrix of errors are used (reported in the Supplement). Therefore, augmentation of the GVAR model with an additional equation for cross-section averages improves the forecasting performance for all choices of $\hat{\mathbf{\Sigma}}_{\xi}$ and horizons considered.

Table 4 investigates how the choice of $\hat{\mathbf{\Sigma}}_{\xi}$ and the shrinkage estimation of individual country models affect the forecasting performance of the AugGVAR-PMI method. In this table, we choose the AugGVAR-PMI with $\hat{\mathbf{\Sigma}}_{\xi,LW}$ as a benchmark and report the GDP-weighted cross-section average MSFE of individual AugGVAR-PMI methods relative to this benchmark. There are some important differences in forecasting performance for different choices of $\hat{\mathbf{\Sigma}}_{\xi}$. The block-diagonal estimate $\hat{\mathbf{\Sigma}}_{\xi,B}$ and the AugGVAR-PMI, which makes use of a nonsynchronous cross-section average, perform better at $h=0$, but not at longer horizons. But the AugGVAR-PMI forecasts that are

based on shrinkage estimators of the individual country models perform marginally better than the AugGVAR-PMI forecasts without shrinkage, with the former performing about 15 percent better than the latter.

Table 5 gives the GDP-weighted average MSFEs of the other data-rich forecasting techniques as well as the AR benchmark forecasts. The results in this table show how the different forecasts compare with the random walk (RW) benchmark. The top panel (a) of the table gives the results when PMI data are not used in forecasting whilst the bottom panel (b) gives the results when PMI data are used.

In the case where PMI data are not used, depending on the choice of the forecast horizon, h, and data release month, $\mathfrak{m}$, the AR forecasts show between 22 to 47 percent improvement over the RW benchmark, which is quite substantial. Adding the PMI data does not improve the AR forecasts much and seems to help only in the case of nowcasting ($h = 0$). A similar picture also emerges when we consider the data-rich techniques. It is clear that regardless of the forecasting method considered, the inclusion of PMIs always decreases the MSFE at horizon $h = 0$, by about 19 percent on average for $\mathfrak{m} = 1$, 14 percent for $\mathfrak{m} = 2$, and 20 percent for $\mathfrak{m} = 3$. The information contained in PMIs is still useful at horizon $h = 1$, but the average improvement is smaller, about 8 to 13 percent. At the longer forecast horizon, $h = 2$, the use of PMI data does not seem to help. In fact, for $h = 2$ the simple AR forecasts do slightly better than the AR-PMI forecasts for all release months $\mathfrak{m}$.

Consider now the performance of the forecasts based on the data-rich methods. The results are mixed and depend on the choice of the forecasting scheme, forecast horizon, h, data release date, $\mathfrak{m}$, and whether PMI data are used in forecasting. But on average data-rich methods tend to outperform AR forecasts when $h = 0$ and PMI data are used in forecasting. But for longer forecast horizons neither PMI nor data-rich techniques seem to help, with the possible exception of the AugGVAR-PMI forecasts which outperform or perform as well as AR forecasts for all forecast horizons and release months.

Overall, perhaps not surprisingly, the use of PMIs helps for the nowcasting of GDP growth and its added value diminishes quite rapidly with the forecast horizon.

8.6 Panel DM test statistics

The forecast comparisons in Table 5 provide clear-cut evidence of improvements when AR and data-rich forecasts are compared to the RW benchmark, but the evidence is much less clear-cut when one considers the relative performance of simple AR and data-rich forecasting techniques. To check the statistical significance of the relative performance of forecasting schemes we use an extension of the panel Diebold and Mariano (1995) (DM) test statistic proposed in Pesaran, Schuermann, and Smith (2009) that allows for unequal weights in the pooling of the country specific MSFEs, and also discuss the robustness of the panel DM test to possible cross-sectional dependence of the differences in squared forecast errors.

Let $z_{it} = e_{itA}^2 - e_{itB}^2$ be the difference in the squared forecasting errors of models A and B, and

consider the following pooled test statistic:

$$\bar{z}_\omega = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\omega_i z_{it},$$

where the weights $\{\omega_i\}_{i=1}^N$ are given and are not necessarily granular. Initially, suppose that z_{it} is serially uncorrelated, but could be correlated over the cross-section units. Decompose z_{it} as $z_{it} = \alpha_i + \eta_{it}$, where α_i represents the systematic difference between the two forecasts, and η_{it} the idiosyncratic component. Let $\boldsymbol{\eta}_t = (\eta_{1t}, \eta_{2t}, ..., \eta_{Nt})'$ and suppose that $\boldsymbol{\eta}_t \sim IID\left(\mathbf{0}_{N\times 1}, \boldsymbol{\Sigma}_\eta\right)$. The implicit null and alternative hypotheses of interest are now given by $H_0 : \bar{\alpha}_\omega = \sum_{i=1}^N \omega_i\alpha_i = 0$ and $H_1 : \bar{\alpha}_\omega < 0$, respectively. Under the null hypothesis $E\left(\bar{z}_\omega\right) = 0$, whereas under the alternative $E\left(\bar{z}_\omega\right) = \bar{\alpha}_\omega \neq 0$, with forecast A preferred to forecast B if $\bar{\alpha}_\omega < 0$, and the reverse if $\bar{\alpha}_\omega > 0$.

To derive a test based on $\bar{z}_\omega$ we first note that under H_0

$$V\left(\bar{z}_\omega\right) = E\left(\bar{z}_\omega^2\right) = E\left[\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\omega_i\eta_{it}\right)^2\right].$$

Under the assumption that $\boldsymbol{\eta}_t$ are serially uncorrelated we have

$$V\left(\bar{z}_\omega\right) = \frac{1}{T^2}\sum_{t=1}^{T}E\left(\sum_{i=1}^{N}\omega_i\eta_{it}\right)^2 = \frac{1}{T}\boldsymbol{\omega}'\boldsymbol{\Sigma}_\eta\boldsymbol{\omega},$$

where $\boldsymbol{\omega} = (\omega_1, \omega_2, ..., \omega_N)'$. Denoting the elements of $\boldsymbol{\Sigma}_\eta$ by $\sigma_{\eta,ij}$, then $V\left(\bar{z}_\omega\right)$ can be written equivalently as

$$V\left(\bar{z}_\omega\right) = \frac{\sum_{i=1}^N \omega_i^2}{T}\left(\vartheta_1 + \vartheta_2\right),$$

where

$$\vartheta_1 = \left(\sum_{i=1}^{N}\omega_i^2\right)^{-1}\cdot\sum_{i=1}^{N}\omega_i^2\sigma_{\eta,ii},$$

and

$$\vartheta_2 = \left(\sum_{i=1}^{N}\omega_i^2\right)^{-1}\cdot\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\omega_i\omega_j\sigma_{ij}.$$

In the special case when $\boldsymbol{\Sigma}_\eta$ is a diagonal matrix, $\vartheta_2 = 0$, and $V\left(\bar{z}_\omega\right)$ converges towards zero at the rate of $T^{-1/2}\left(\sum_{i=1}^N \omega_i^2\right)^{1/2}$, which yields the standard rate of $(NT)^{-1/2}$ when the weights are granular. In the non-diagonal case the limiting behavior of $V\left(\bar{z}_\omega\right)$ depends on the degree of cross-sectional dependence of z_{it}. A distinction can be made depending on whether the row (column) norm of $\boldsymbol{\Sigma}_\eta$ is bounded in N. In the bounded case the cross-sectional dependence is weak and the rate at which $V\left(\bar{z}_\omega\right)$ converges towards zero is the same as in the diagonal case. In contrast, when the row (column) norm of $\boldsymbol{\Sigma}_\eta$ is not bounded in N then the rate of convergence of $V\left(\bar{z}_\omega\right)$ towards zero is slower than $\sqrt{T}\cdot\left(\sum_{i=1}^N w_i^2\right)^{-1/2}$ and inference based on $\bar{z}_\omega$ will depend on the

off-diagonal elements of $\mathbf{\Sigma}_\eta$, and in general require T to be much larger than N. In the current pair-wise comparisons where the forecast errors are obtained conditional on a common set of factors, it is reasonable to expect that the dependence of z_{it} across i is reasonably weak and when making inference the off-diagonal elements of $\mathbf{\Sigma}_\eta$ can be ignored. Accordingly, we base the panel DM tests on the following weighted pooled DM test statistic:

$$WPDM = \sqrt{T} \cdot \left(\sum_{i=1}^{N} w_i^2 \right)^{-1/2} \frac{\bar{z}_w}{\sqrt{\hat{\vartheta}_1}}, \tag{77}$$

where

$$\hat{\vartheta}_1 = \left(\sum_{i=1}^{N} w_i^2 \right)^{-1} \cdot \sum_{i=1}^{N} w_i^2 \hat{\sigma}_{LRi},$$

in which $\hat{\sigma}_{LRi}$ is the Newey and West (1987) estimator of the long-run variance of z_{it} to take into account possible serial correlations of z_{it}. We set the truncation lag in the Newey-West estimator to 2. Under the null hypothesis the $WPDM$ is asymptotically normally distributed with mean zero and a unit variance as $N, T \xrightarrow{j} \infty$ but only if $\vartheta_2 \to 0$. Hence, $WPDM$ is valid when the weighted sum of off-diagonal elements of $\mathbf{\Sigma}$ is sufficiently small. We leave the further development of the panel DM test statistics under more general form of cross-sectional dependence to future research and present test results based on $WPDM$ as defined by (77).

All methods that use PMIs are significantly better than the RW at the 1% level for all the three months in the current quarter, $h = 0$, and the vast majority at the 1% level for $h > 0$.[18] These findings are not surprising given the differences in MSFE reported in Table 5. We consider next testing whether adding PMIs significantly improves the MSFEs. The top panel of Table 6 presents pair-wise GDP-weighted panel DM test statistics comparing the performance of individual forecasting techniques with and without the use of PMIs. We see that using PMIs significantly improves the forecasting performance at the 1% level for the vast majority of tests when $h = 0$, but this is no longer the case for $h > 0$. We also provide panel DM test statistics for all the forecasting methods against the AugGVAR-PMI forecasts at the bottom panel of Table 6. These results show that, for $h = 0$, AugGVAR-PMI is not significantly better (or worse) than the other methods that use PMIs. In contrast, statistically significant differences at the 1% level can be observed for longer horizons ($h > 0$), where AugGVAR performs significantly better in the majority of cases.

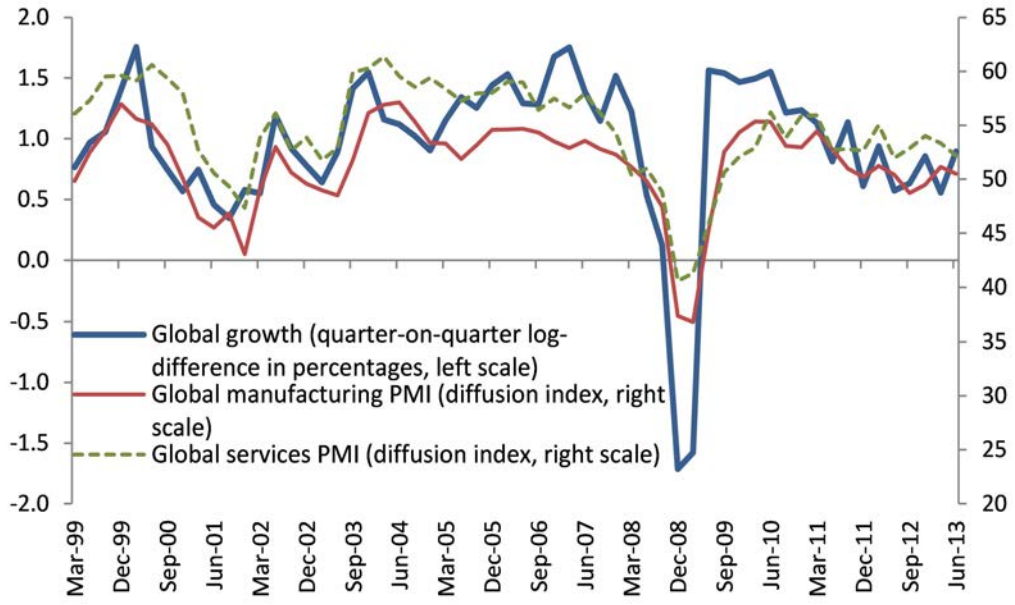
9 Conclusion

In this paper we have shown that the GVAR model can be undetermined when strong unobserved common factors are present, and propose augmenting the GVAR model with additional equations in cross-sectional averages that proxy the common factors. The validity of the augmentation procedure is established theoretically for N and $T \to \infty$, jointly such that $N/T \to \varkappa$ for some $0 < \varkappa < \infty$. The theoretical results are illustrated by MC experiments, and extended to the case of forecasting

[18] WPDM tests using the RW benchmark are reported in the Supplement.

with GVARs in the presence of nonsynchronous conditioning sets. Empirical application to the forecasting of output growth with PMIs using a sample of 48 countries also confirms the superior forecasting performance of the AugGVARs relative to the non-augmented GVARs. A number of other data-rich methods were also implemented. It was found that, regardless of the forecasting method considered, PMIs are useful in nowcasting ($h = 0$), but their value added is rather limited for forecasting when $h > 0$. It is also found that AugGVAR forecasts do as well as other data-rich forecasting techniques for $H = 0$, and tend to do better for longer forecast horizons. Furthermore, the AugGVAR approach has the added advantage that it can be used for impulse response and other forms of counterfactual analyses whilst the single equation data-rich techniques are limited in this respect.

Figure 1: Global growth (thick blue line, left scale, quarter-on-quarter log-difference in percentages), global manufacturing PMI (thin red line, right scale, diffusion index) and global services PMI (dashed green line, right scale, diffusion index), 1999Q1-2013Q2.



Notes: See Section 8.1 for more information on diffusion indices. Global manufacturing PMI and global services PMI are series reported by JP Morgan (see www.markiteconomics.com), and global output growth is calculated using PPP-weighted GDP from 48 countries.

Table 1: Cross-section average MSFE of one-step-ahead GVAR forecasts relative to infeasible optimal forecasts in Monte Carlo experiments without mixed conditioning, SAR parameter set equal to 0.6 and sparse coefficient matrix.

	GVAR					AugGVAR				
(N,T)	**30**	**50**	**100**	**200**	**500**	**30**	**50**	**100**	**200**	**500**
	DGP1: High-dimensional VAR without common factor									
30	1.35	1.17	1.09	1.05	1.02	1.31	1.15	1.08	1.04	1.02
50	1.35	1.17	1.08	1.05	1.02	1.32	1.15	1.08	1.05	1.02
100	1.36	1.16	1.08	1.04	1.02	1.34	1.15	1.08	1.04	1.02
200	1.35	1.16	1.08	1.04	1.02	1.33	1.15	1.08	1.04	1.02
500	1.34	1.16	1.08	1.04	1.02	1.32	1.16	1.08	1.04	1.02
	DGP2: High-dimensional VAR with an additive common factor									
30	2.08	1.49	1.22	1.17	1.11	1.31	1.16	1.08	1.05	1.03
50	17.46	1.67	1.35	1.22	1.18	1.29	1.16	1.08	1.04	1.03
100	483.07	2.21	1.54	1.36	1.26	1.30	1.15	1.08	1.05	1.02
200	477.42	5.67	1.97	1.60	1.51	1.30	1.15	1.07	1.04	1.02
500	$>10^3$	$>10^3$	3.70	2.58	2.12	1.32	1.15	1.08	1.04	1.02
	DGP3: High-dimensional VAR with a factor error structure									
30	1.67	1.29	1.14	1.09	1.05	1.29	1.14	1.07	1.04	1.02
50	2.46	1.35	1.18	1.11	1.07	1.29	1.14	1.08	1.04	1.02
100	2.45	1.48	1.24	1.15	1.10	1.29	1.15	1.07	1.04	1.02
200	129.18	1.96	1.42	1.25	1.17	1.30	1.14	1.08	1.04	1.02
500	$>10^3$	31.92	1.90	1.58	1.44	1.31	1.14	1.08	1.04	1.02

Notes: This table reports the simple cross-section average mean square forecast error of GVAR and AugGVAR forecasts relative to infeasible optimal forecasts, see (60). DGPs 1-3 are given by models (56), (58) and (59), respectively. Infeasible forecasts are defined as $E\left(y_{i,T+1}|\mathcal{I}_t,\mathcal{F}_t\right)$. See (61). Computations of GVAR and AugGVAR forecasts are explained in Subsection 7.2.1. In particular, see (65) and (69).

Table 2: Cross-section average MSFE of one-step ahead GVAR forecasts relative to infeasible optimal forecasts in Monte Carlo experiments with additive common factor, mixed conditioning and sparse coefficient matrix.

	Experiments with SAR coef. $\varrho = 0.2$					Experiments with SAR coef. $\varrho = 0.6$				
(N,T)	**30**	**50**	**100**	**200**	**500**	**30**	**50**	**100**	**200**	**500**
	GVAR without augmentation									
	GVAR($\hat{\mathbf{\Sigma}}_{\xi,LW}$)									
30	2.67	1.47	1.24	1.16	1.11	3.65	1.50	1.20	1.14	1.09
50	12.47	1.66	1.31	1.21	1.16	3.65	1.66	1.30	1.18	1.13
100	$>10^3$	2.55	1.58	1.37	1.27	$>10^3$	222.05	1.45	1.27	1.18
200	$>10^3$	436.35	2.21	1.73	1.55	$>10^3$	4.11	1.78	1.43	1.35
500	$>10^3$	$>10^3$	210.17	2.96	2.41	$>10^3$	$>10^3$	3.19	2.15	1.78
	GVAR($\hat{\mathbf{\Sigma}}_{\xi,B}$)									
30	2.52	1.46	1.23	1.15	1.11	3.32	1.46	1.19	1.14	1.10
50	11.61	1.62	1.30	1.19	1.16	3.43	1.59	1.27	1.17	1.13
100	$>10^3$	2.41	1.54	1.34	1.25	$>10^3$	226.85	1.41	1.24	1.17
200	$>10^3$	377.57	2.09	1.65	1.49	$>10^3$	3.36	1.66	1.38	1.29
500	$>10^3$	$>10^3$	221.81	2.73	2.22	$>10^3$	$>10^3$	2.68	1.91	1.68
	Augmented GVAR									
	AugGVAR($\hat{\mathbf{\Sigma}}_{\xi,LW}$)									
30	1.50	1.22	1.12	1.06	1.04	1.51	1.24	1.10	1.06	1.03
50	1.49	1.23	1.11	1.06	1.03	1.51	1.25	1.11	1.06	1.03
100	1.53	1.23	1.12	1.06	1.03	1.55	1.25	1.13	1.06	1.03
200	1.53	1.22	1.11	1.06	1.03	1.58	1.28	1.13	1.07	1.03
500	1.52	1.22	1.11	1.06	1.03	1.61	1.28	1.14	1.08	1.04
	AugGVAR($\hat{\mathbf{\Sigma}}_{\xi,B}$)									
30	1.49	1.21	1.11	1.06	1.04	1.47	1.23	1.11	1.07	1.05
50	1.48	1.23	1.11	1.06	1.03	1.46	1.23	1.11	1.07	1.04
100	1.51	1.22	1.11	1.06	1.03	1.47	1.22	1.12	1.07	1.04
200	1.51	1.21	1.10	1.06	1.03	1.49	1.23	1.11	1.07	1.04
500	1.50	1.21	1.11	1.06	1.03	1.52	1.23	1.11	1.07	1.04
	AugGVAR									
30	1.38	1.17	1.10	1.05	1.04	1.46	1.24	1.13	1.10	1.07
50	1.37	1.18	1.09	1.06	1.03	1.44	1.24	1.14	1.09	1.07
100	1.39	1.18	1.10	1.05	1.03	1.46	1.23	1.14	1.10	1.07
200	1.39	1.17	1.09	1.05	1.03	1.47	1.25	1.13	1.09	1.06
500	1.39	1.17	1.09	1.05	1.03	1.49	1.24	1.13	1.09	1.07

Notes: This table reports the simple cross-section average mean square forecast error of GVAR forecasts relative to the infeasible optimal forecasts. See (71). DGPs 1-3 are given by models (56), (58) and (59), respectively. Infeasible forecasts are given by $E\left(y_{i,T+1}|\,\Omega_T\right)$. $\hat{\mathbf{\Sigma}}_{\xi,LW}$ is Ledoit and Wolf (2004)'s estimator of $\mathbf{\Sigma}_{\xi}$, $\hat{\mathbf{\Sigma}}_{\xi,B}$ is the block-diagonal estimator defined by (72), and AugGVAR uses nonsynchronous cross-section averages and auxiliary regression (55). See Subsection 7.2.2 for a detailed description of forecasting methods.

Table 3: GDP-weighted cross-section average MSFE of AugGVAR-PMI relative to non-augmented GVAR-PMI

forecasting horizon (quarters):	$h=0$			$h=1$			$h=2$		
month:	$\mathfrak{m}=1$	$\mathfrak{m}=2$	$\mathfrak{m}=3$	$\mathfrak{m}=1$	$\mathfrak{m}=2$	$\mathfrak{m}=3$	$\mathfrak{m}=1$	$\mathfrak{m}=2$	$\mathfrak{m}=3$
GVAR-PMI($\hat{\mathbf{\Sigma}}_{\xi,LW}$)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	(11.7)	(7.4)	(6.4)	(59.3)	(29.3)	(19.5)	(98.9)	(79.6)	(365.2)
AugGVAR-PMI($\hat{\mathbf{\Sigma}}_{\xi,LW}$)	0.135	0.223	0.296	0.027	0.054	0.085	0.017	0.022	0.005

Notes: MSFE is computed based on the evaluation sample 2006Q1-2013Q2. The GDP-weighted cross-section average MSFE of the non-augmented GVAR-PMI with $\hat{\mathbf{\Sigma}}_{\xi,LW}$ is reported in parentheses.

Table 4: GDP-weighted cross-section average MSFE of AugGVAR-PMI methods relative to the benchmark AugGVAR-PMI with $\hat{\mathbf{\Sigma}}_{LW}$ and without shrinkage

	forecasting horizon (quarters):	$h=0$			$h=1$			$h=2$		
	month:	$\mathfrak{m}=1$	$\mathfrak{m}=2$	$\mathfrak{m}=3$	$\mathfrak{m}=1$	$\mathfrak{m}=2$	$\mathfrak{m}=3$	$\mathfrak{m}=1$	$\mathfrak{m}=2$	$\mathfrak{m}=3$
		Individual country models estimated without shrinkage								
1	AugGVAR-PMI($\hat{\mathbf{\Sigma}}_{\xi,LW}$)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		(1.58)	(1.66)	(1.88)	(1.60)	(1.59)	(1.65)	(1.72)	(1.74)	(1.77)
2	AugGVAR-PMI($\hat{\mathbf{\Sigma}}_{\xi,B}$)	0.84	0.86	0.79	1.04	1.04	1.02	1.01	1.02	0.99
3	AugGVAR-PMI	0.89	0.82	0.71	0.96	0.95	1.00	0.94	0.93	0.96
		Individual country models estimated with shrinkage								
4	AugGVAR-PMI($\hat{\mathbf{\Sigma}}_{\xi,LW}$)	0.82	0.79	0.66	0.88	0.87	0.83	0.91	0.88	0.87
5	AugGVAR-PMI($\hat{\mathbf{\Sigma}}_{\xi,B}$)	0.74	0.71	0.60	0.86	0.85	0.80	0.90	0.87	0.86
6	AugGVAR-PMI	0.75	0.70	0.58	0.90	0.91	0.84	0.89	0.88	0.84

Notes: $\hat{\mathbf{\Sigma}}_{\xi,LW}$ is Ledoit and Wolf (2004)'s covariance matrix estimator and $\hat{\mathbf{\Sigma}}_{\xi,B}$ is the block-diagonal covariance matrix estimator of $\mathbf{\Sigma}_{\xi}$. AugGVAR-PMI does not make use of $\hat{\mathbf{\Sigma}}_{\xi}$, but augment the GVAR with an additional equation for forecasting cross-section averages computed using the nonsynchronous conditioning information set. MSFE is computed based on the evaluation sample 2006Q1-2013Q2. The GDP-weighted average MSFE of the AugGVAR-PMI with $\hat{\mathbf{\Sigma}}_{\xi,LW}$ and without shrinkage is reported in parentheses.

Table 5: GDP-weighted cross-section average MSFE of individual methods relative to RW

forecasting horizon (quarters):		$h=0$			$h=1$			$h=2$		
	month:	$\mathfrak{m}=1$	$\mathfrak{m}=2$	$\mathfrak{m}=3$	$\mathfrak{m}=1$	$\mathfrak{m}=2$	$\mathfrak{m}=3$	$\mathfrak{m}=1$	$\mathfrak{m}=2$	$\mathfrak{m}=3$
1	RW (benchmark)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		(2.01)	(1.77)	(1.73)	(2.39)	(2.31)	(2.28)	(2.83)	(2.59)	(2.54)
	(a) Models without PMI									
2.a	AR	0.71	0.77	0.78	0.61	0.64	0.66	0.53	0.59	0.60
3.a	Lasso	0.66	0.74	0.76	0.67	0.70	0.72	0.52	0.60	0.62
4.a	Ridge	0.68	0.79	0.77	0.71	0.84	0.80	0.63	0.77	0.77
5.a	FM	0.75	0.92	0.81	0.72	0.97	0.96	0.65	0.75	0.77
6.a	FM-AR	0.77	0.93	0.83	0.72	0.98	0.97	0.67	0.78	0.80
7.a	PLS	0.81	0.95	0.91	0.92	1.10	1.00	0.85	1.10	1.11
8.a	AugGVAR	0.79	0.76	0.75	0.62	0.66	0.68	0.55	0.61	0.62
	(b) Models with PMI									
2.b	AR-PMI	0.63	0.66	0.64	0.66	0.68	0.62	0.59	0.64	0.65
3.b	Lasso-PMI	0.61	0.69	0.66	0.62	0.69	0.69	0.52	0.59	0.62
4.b	Ridge-PMI	0.57	0.70	0.62	0.65	0.77	0.70	0.66	0.74	0.78
5.b	FM-PMI	0.59	0.79	0.62	0.72	0.88	0.82	0.76	0.81	0.84
6.b	FM-AR-PMI	0.61	0.81	0.64	0.75	0.91	0.85	0.79	0.83	0.88
7.b	PLS-PMI	0.61	0.74	0.65	0.70	0.85	0.77	0.76	0.87	0.88
8.b	AugGVAR-PMI	0.58	0.66	0.62	0.58	0.59	0.58	0.54	0.59	0.59

Notes: The GDP-weighted cross-section average MSFE of RW forecasts is reported in parentheses. MSFE is computed based on the evaluation sample 2006Q1-2013Q2. The AugGVAR-PMI is the simple average of AugGVAR-PMI models with shrinkage (models 4-6 in Table 4). Similarly, AugGVAR is the simple average of AugGVAR models with shrinkage. All methods are described in Subsection 8.2.

Table 6: GDP-weighted pair-wise panel DM test statistics

forecasting horizon (quarters):		$h = 0$			$h = 1$			$h = 2$		
	month:	$\mathfrak{m} = 1$	$\mathfrak{m} = 2$	$\mathfrak{m} = 3$	$\mathfrak{m} = 1$	$\mathfrak{m} = 2$	$\mathfrak{m} = 3$	$\mathfrak{m} = 1$	$\mathfrak{m} = 2$	$\mathfrak{m} = 3$
	(a) Benchmark is the same method without PMI									
2.a	AR-PMI	-2.11	-3.06	-3.43	2.04	1.34	-1.38	2.58	2.30	2.41
3.a	Lasso-PMI	-3.39	-2.78	-3.95	-3.89	-0.80	-2.48	-0.03	-0.88	-0.50
4.a	Ridge-PMI	-4.14	-3.94	-4.45	-2.85	-2.66	-3.53	2.62	-1.20	0.75
5.a	FM-PMI	-4.41	-3.20	-4.40	0.07	-2.26	-3.19	3.92	2.28	2.50
6.a	FM-AR-PMI	-4.37	-2.96	-4.09	0.83	-2.01	-2.90	3.97	1.51	2.55
7.a	PLS-PMI	-5.05	-4.19	-5.15	-5.12	-4.45	-4.46	-2.74	-3.38	-3.93
8.a	AugGVAR-PMI	-3.58	-1.92	-2.07	-3.43	-2.62	-3.19	-1.43	-2.97	-2.78
	(b) Benchmark is AugGVAR-PMI									
2.b	AR-PMI	1.38	0.20	0.55	3.53	3.49	1.19	1.84	1.95	2.25
3.b	Lasso-PMI	1.22	0.76	1.00	2.14	3.91	3.87	-1.71	0.02	1.34
4.b	Ridge-PMI	-0.10	0.73	-0.07	2.39	4.05	3.47	4.76	4.15	3.97
5.b	FM-PMI	0.27	1.92	-0.06	3.94	4.69	4.07	4.78	4.12	4.19
6.b	FM-AR-PMI	0.94	2.04	0.38	4.29	4.98	4.43	5.25	4.32	4.29
7.b	PLS-PMI	0.74	1.32	0.57	3.42	5.00	4.30	6.62	5.57	5.03
8.b	AugGVAR-PMI	-	-	-	-	-	-	-	-	-

Notes: Panel DM test statistics are computed based on the evaluating sample 2006Q1-2013Q2. The panel DM test is a one-sided test and asymptotically normal, so the relevant 1% and 5% critical values for a given method to outperform the benchmark are -2.326 and -1.645, respectively. The AugGVAR-PMI is the simple average of AugGVAR-PMI models with shrinkage (models 4-6 in Table 4).

A Appendix

A.1 Derivation of optimal forecasts when factors are unobserved

Consider the problem of optimal forecasts of y_{t+h} generated by (2) based on the information set $\mathcal{I}_t$ alone. To derive optimal forecasts in this case we also assume that $\boldsymbol{\varepsilon}_{bt}$ and v_{bt} are normally distributed. In particular, let $\boldsymbol{\varepsilon}_{bt} \sim IIDN\left(\mathbf{0}, \boldsymbol{\Sigma}_{b\varepsilon}\right)$, $v_t \sim IIDN\left(0, \sigma_{bv}^2\right)$, and assume that $\boldsymbol{\varepsilon}_{bt}$ and $v_{bt'}$ are independently distributed for all t and t'. We have

$$E\left(\mathbf{y}_{t+h} | \mathcal{I}_t, M_b\right) = \boldsymbol{\Phi}_b^h \mathbf{y}_t + \mathbf{g}_{bh} E\left(f_{bt} | \mathcal{I}_t, M_b\right).$$

Optimal prediction of the common factor, $E\left(f_{bt} | \mathcal{I}_t, M_b\right)$, can be obtained (under the above assumptions) using a Kalman filter, noting that

$$\mathbf{u}_{bt} \equiv \mathbf{y}_t - \boldsymbol{\Theta}_b \mathbf{y}_{t-1} = \boldsymbol{\gamma}_b f_{bt} + \boldsymbol{\varepsilon}_{bt}.$$

$\mathcal{I}_t$ contains information on the infinite past of y_t, and under stationarity requirements $|\rho_b| < 1$ and $|\lambda_1\left(\boldsymbol{\Theta}_b\right)| < 1$, the steady-state Kalman filter gives

$$E\left(f_{bt} | \mathcal{I}_t, M_b\right) = \hat{f}_{b,\,t|t-1} + \mathbf{q}_b' \left(\mathbf{u}_{bt} - \boldsymbol{\gamma}_b \hat{f}_{b,\,t|t-1}\right),$$

where $\mathbf{q}_b' = p_b \boldsymbol{\gamma}_b' \left(p_b \boldsymbol{\gamma}_b \boldsymbol{\gamma}_b' + \boldsymbol{\Sigma}_{b\varepsilon}\right)^{-1}$, p_b is the unique solution of

$$p_b = 1 - \rho_b^2 + \left(1 - \rho_b^2\right)^{-1} p_b \mathbf{q}_b' \boldsymbol{\gamma}_b,$$

and $\hat{f}_{b,\,t|t-1} = E\left(f_{bt} | \mathcal{I}_{t-1}, M_b\right)$ is a stationary process given by

$$\hat{f}_{b,\,t|t-1} = \rho_b \left(1 - \mathbf{q}_b' \boldsymbol{\gamma}_b\right) \hat{f}_{b,\,t-1|t-2} + \rho_b \mathbf{q}_b' \mathbf{u}_{b,t-1}.$$

A.2 Proofs

Proof of Theorem 1. We provide proof for the weakly cross-sectionally dependent model first, namely under Assumptions 1, 2.a, and 3-4. For $h = 1$ we have

$$\hat{\mathbf{y}}_{t+1|t}^{aug} = \left[\widehat{\boldsymbol{\Theta}} + \left(\widehat{\rho}\widehat{\mathbf{b}}_0 + \widehat{\mathbf{b}}_1\right) \mathbf{w}'\right] \mathbf{y}_t.$$

Consider

$$\begin{aligned} E\left(\mathbf{y}_{t+1} \mid \mathcal{I}_t, \mathcal{F}_t\right) - \hat{\mathbf{y}}_{t+1|t}^{aug} &= \boldsymbol{\Phi}\mathbf{y}_t - \left[\widehat{\boldsymbol{\Theta}} + \left(\widehat{\rho}\widehat{\mathbf{b}}_0 + \widehat{\mathbf{b}}_1\right) \mathbf{w}'\right] \mathbf{y}_t \\ &= \left(\boldsymbol{\Phi} - \boldsymbol{\Theta}\right) \mathbf{y}_t - \left(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\right) \mathbf{y}_t - \left(\widehat{\rho}\widehat{\mathbf{b}}_0 + \widehat{\mathbf{b}}_1\right) \mathbf{w}'\mathbf{y}_t. \end{aligned} \tag{A.1}$$

Consider the individual elements on the right side of (A.1) below. Note that the row i of matrix $\left(\boldsymbol{\Phi} - \boldsymbol{\Theta}\right)$, namely

$$\mathbf{e}_{N,i}' \left(\boldsymbol{\Phi} - \boldsymbol{\Theta}\right) = \boldsymbol{\phi}_{-i}',$$

satisfies condition (8). By Liapunov's inequality

$$\begin{aligned}\left(E\left|\boldsymbol{\phi}_{-i}'\mathbf{y}_t\right|\right)^2 &\leq E\left(\boldsymbol{\phi}_{-i}'\mathbf{y}_t\right)^2 \\ &\leq \varrho\left(E\left(\mathbf{y}_t\mathbf{y}_t'\right)\right)\boldsymbol{\phi}_{-i}'\boldsymbol{\phi}_{-i} \\ &\leq \frac{K}{N},\end{aligned}$$

where constant $K < \infty$ does not depend on N. We have used the Rayleigh-Ritz theorem[19] to obtain the second inequality and

$$\begin{aligned}\varrho\left(E\left(\mathbf{y}_t\mathbf{y}_t'\right)\right) &\leq \left\|E\left(\mathbf{y}_t\mathbf{y}_t'\right)\right\|, \\ &\leq \left\|\mathbf{R}\mathbf{R}'\right\|\sum_{\ell=0}^{\infty}\left\|\boldsymbol{\Phi}\right\|^{2\ell}, \\ &= O\left(1\right), \end{aligned} \tag{A.2}$$

follows from Assumptions 1 and 3. Therefore

$$\max_{i\in\{1,..,N\}} E\left|\boldsymbol{\phi}_{-i}'\mathbf{y}_t\right| \to 0, \tag{A.3}$$

and similarly it can be shown that (replacing $\boldsymbol{\phi}_{-i}$ with $\mathbf{e}_{N,i}$)

$$\max_{i\in\{1,..,N\}} E\left|\mathbf{e}_{N,i}'\mathbf{y}_t\right| < K. \tag{A.4}$$

Equation (A.3) implies

$$E\left\|\left(\boldsymbol{\Phi}-\boldsymbol{\Theta}\right)\mathbf{y}_t\right\|_\infty \to 0. \tag{A.5}$$

Now consider the second term on the right side of (A.1), namely $\left(\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}\right)\mathbf{y}_t$. Equation (A.4) implies that the elements of $\mathbf{y}_t$ are uniformly bounded in L_1 norm, namely

$$E\left\|\mathbf{y}_t\right\|_\infty < K. \tag{A.6}$$

Chudik and Pesaran (2011, Theorem 2) established asymptotic distribution of the diagonal elements of $\widehat{\boldsymbol{\Theta}}$ in the special case when there is no common factor (rank deficient case with $m = 0$), and we have

$$\sqrt{T}\left(\widehat{\phi}_{ii}-\phi_{ii}\right) = O_p\left(1\right), \tag{A.7}$$

uniformly in i. It follows that $\left\|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}\right\|_\infty = O_p\left(T^{-1/2}\right)$, and together with (A.6) we obtain

$$E\left\|\left(\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}\right)\mathbf{y}_t\right\|_\infty \to 0. \tag{A.8}$$

Now consider the last term on the right side of (A.1). Let us define $\tilde{y}_t = \sqrt{N}\mathbf{w}'\mathbf{y}_t$, $\tilde{b}_{si} = N^{-1/2}b_{si}$, for $s = 0, 1$ and for all i, and consider the least squares regression:

$$\begin{aligned} y_{it} &= \phi_{ii}y_{i,t-1} + b_{0i}\mathbf{w}'\mathbf{y}_t + b_{0i}\mathbf{w}'\mathbf{y}_{t-1} + e_{it} \\ &= \phi_{ii}y_{i,t-1} + \tilde{b}_{0i}\tilde{y}_{wt} + \tilde{b}_{1i}\tilde{y}_{w,t-1} + e_{it}. \end{aligned} \tag{A.9}$$

[19] See Horn and Johnson (1985, p. 176).

Under the assumptions of Theorem 1, which rule out strong cross-sectional dependence in y_{it}, and using (A.2), we obtain

$$\mathbf{w}'\mathbf{y}_t = O_p\left(N^{-1/2}\right), \tag{A.10}$$

$$\tilde{y}_{wt} = \sqrt{N}\mathbf{w}'\mathbf{y}_t = O_p(1),$$

and note that all of the regressors in (A.9) are $O_p(1)$. Using similar arguments as in Chudik and Pesaran (2011), it can be established that $p\lim \widehat{\tilde{b}}_{si} = 0$, for $s = 0, 1$, uniformly in i, which in turn implies $\widehat{b}_{si} = \sqrt{N}\widehat{\tilde{b}}_{si} = o\left(N^{1/2}\right)$, for $s = 0, 1$, uniformly in i. This result together with $\widehat{\rho} = O_p(1)$, and (A.10) establish

$$E\left\|\left(\widehat{\rho}\widehat{\mathbf{b}}_0 + \widehat{\mathbf{b}}_1\right)\mathbf{w}'\mathbf{y}_t\right\|_\infty \to 0. \tag{A.11}$$

Using (A.5), (A.8), and (A.11) in (A.1) establish $E\left\|E\left(\mathbf{y}_{t+1} \mid \mathcal{I}_t, \mathcal{F}_t\right) - \hat{\mathbf{y}}^{aug}_{t+1|t}\right\|_\infty \to 0$. This completes the proof of result (42) for $h = 1$ in the weakly cross-sectionally dependent model. The proof of (42) for $h > 1$ in the weakly cross-sectionally dependent model can be constructed in a similar way.

Next, we provide proof for the model featuring an unobserved common factor, namely under Assumptions 1, 2.b, and 3-4. For $h = 1$ we have

$$E\left(\mathbf{y}_{t+1} \mid \mathcal{I}_t, \mathcal{F}_t\right) = \mathbf{\Phi}\left(\mathbf{y}_t - \boldsymbol{\gamma} f_t\right) + \rho\boldsymbol{\gamma} f_t = \mathbf{\Theta}\mathbf{y}_t - \mathbf{\Theta}\boldsymbol{\gamma} f_t + \rho\boldsymbol{\gamma} f_t + \left(\mathbf{\Phi} - \mathbf{\Theta}\right)\left(\mathbf{y}_t - \boldsymbol{\gamma} f_t\right), \tag{A.12}$$

and as before

$$\hat{\mathbf{y}}^{aug}_{t+1|t} = \widehat{\mathbf{\Theta}}\mathbf{y}_t - \widehat{\mathbf{\Theta}}\widehat{\mathbf{b}}_0\mathbf{w}'\mathbf{y}_t + \widehat{\rho}\widehat{\mathbf{b}}_0\mathbf{w}'\mathbf{y}_t. \tag{A.13}$$

Subtracting (A.13) from (A.12) yields

$$\begin{aligned} E\left(\mathbf{y}_{t+1} \mid \mathcal{I}_t, \mathcal{F}_t\right) - \hat{\mathbf{y}}^{aug}_{t+1|t} &= \left(\mathbf{\Theta} - \widehat{\mathbf{\Theta}}\right)\mathbf{y}_t - \left(\rho\boldsymbol{\gamma} f_t - \widehat{\rho}\widehat{\mathbf{b}}_0\mathbf{w}'\mathbf{y}_t\right) \\ &\quad + \left(\mathbf{\Theta}\boldsymbol{\gamma} f_t - \widehat{\mathbf{\Theta}}\widehat{\mathbf{b}}_0\mathbf{w}'\mathbf{y}_t\right) + \left(\mathbf{\Phi} - \mathbf{\Theta}\right)\left(\mathbf{y}_t - \boldsymbol{\gamma} f_t\right). \end{aligned} \tag{A.14}$$

We now investigate the properties of the individual elements on the right side of (A.14). First, consider $\left(\mathbf{\Theta} - \widehat{\mathbf{\Theta}}\right)\mathbf{y}_t$ and note that

$$E\left(y_{it}^2\right) = \sum_{\ell=0}^{\infty}\mathbf{e}'_{N,i}\mathbf{\Phi}^\ell\mathbf{R}\mathbf{R}'\mathbf{\Phi}^{\ell\prime}\mathbf{e}_{N,i} + E\left(\gamma_i^2 f_t^2\right),$$

where $E\left(\gamma_i^2 f_t^2\right) < K$ under Assumption 2.b and

$$\sum_{\ell=0}^{\infty}\mathbf{e}'_{N,i}\mathbf{\Phi}^\ell\mathbf{R}\mathbf{R}'\mathbf{\Phi}^{\ell\prime}\mathbf{e}_{N,i} \leq \left\|\mathbf{e}_{N,i}\right\|^2\left\|\mathbf{R}\right\|^2\sum_{\ell=0}^{\infty}\left\|\mathbf{\Phi}\right\|^{2\ell} < K,$$

in which $\left\|\mathbf{e}_{N,i}\right\| = 1$, $\left\|\mathbf{R}\right\|^2 \leq \left\|\mathbf{R}\right\|_1\left\|\mathbf{R}\right\|_\infty < K$ by Assumption 1, and $\sum_{\ell=0}^{\infty}\left\|\mathbf{\Phi}\right\|^{2\ell} < K$ by Assumption 3. It therefore follows that $E\left(y_{it}^2\right) < K$, where K does not depend on i nor on N, and similar to the weakly dependent case, we obtain

$$E\left\|\mathbf{y}_t\right\|_\infty < K. \tag{A.15}$$

Chudik and Pesaran (2011, Theorem 1) establishes that

$$\sqrt{T}\left(\hat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i\right) = O_p(1) \tag{A.16}$$

uniformly in i as $N, T \xrightarrow{j} \infty$ such that $N/T \to \varkappa$ for some $0 < \varkappa < \infty$, where $\hat{\boldsymbol{\pi}}_i = \left(\widehat{\phi}_{ii}, \widehat{b}_{i0}, \widehat{b}_{i1}\right)'$ is the vector of least squares estimates of $\boldsymbol{\pi}_i = (\phi_{ii}, b_{i0}, b_{i1})'$. This implies $\left\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\right\|_\infty = o_p(1)$ and together with (A.15) we obtain

$$E\left\|\left(\boldsymbol{\Theta} - \widehat{\boldsymbol{\Theta}}\right)\mathbf{y}_t\right\|_\infty \to 0. \tag{A.17}$$

Consider next

$$\begin{aligned}
\rho\boldsymbol{\gamma} f_t - \widehat{\rho}\widehat{\mathbf{b}}_0\mathbf{w}'\mathbf{y}_t &= \rho\left(\boldsymbol{\gamma} f_t - \widehat{\mathbf{b}}_0\mathbf{w}'\mathbf{y}_t\right) - \left(\widehat{\rho} - \rho\right)\widehat{\mathbf{b}}_0\mathbf{w}'\mathbf{y}_t \\
&= \rho\left(\boldsymbol{\gamma} f_t - \mathbf{b}_0\mathbf{w}'\mathbf{y}_t\right) - \rho\left(\widehat{\mathbf{b}}_0 - \mathbf{b}_0\right)\mathbf{w}'\mathbf{y}_t - \left(\widehat{\rho} - \rho\right)\widehat{\mathbf{b}}_0\mathbf{w}'\mathbf{y}_t.
\end{aligned}$$

Since $\mathbf{w}'\mathbf{y}_t - \bar{\gamma}_w f_t = O_p\left(N^{-1/2}\right)$, it can be shown that $\widehat{\rho}$ is a consistent estimator of ρ and therefore $E\left|\widehat{\rho} - \rho\right| \to 0$. Furthermore, (A.15) and (8) imply $E\left\|\mathbf{w}'\mathbf{y}_t\right\|_\infty < K$ and (A.16) implies $E\left\|\widehat{\mathbf{b}}_0 - \mathbf{b}_0\right\|_\infty \to 0$. It therefore follows that

$$E\left\|\rho\boldsymbol{\gamma} f_t - \widehat{\rho}\widehat{\mathbf{b}}_0\mathbf{w}'\mathbf{y}_t\right\|_\infty \to 0. \tag{A.18}$$

Similarly, (A.16) also implies that $E\left\|\widehat{\mathbf{b}}_0 - \mathbf{b}_0\right\|_\infty \to 0$ and $E\left\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\right\|_\infty \to 0$, and it follows that

$$E\left\|\boldsymbol{\Theta}\boldsymbol{\gamma} f_t - \widehat{\boldsymbol{\Theta}}\widehat{\mathbf{b}}_0\mathbf{w}'\mathbf{y}_t\right\|_\infty \to 0. \tag{A.19}$$

Consider now the i-th element of $(\boldsymbol{\Phi} - \boldsymbol{\Theta})(\mathbf{y}_t - \boldsymbol{\gamma} f_t)$, denoted as

$$\begin{aligned}
\vartheta_{it} &\equiv \mathbf{e}_{N,i}'(\boldsymbol{\Phi} - \boldsymbol{\Theta})(\mathbf{y}_t - \boldsymbol{\gamma} f_t) = \boldsymbol{\phi}_{-i}'(\mathbf{y}_t - \boldsymbol{\gamma} f_t) \\
&= \boldsymbol{\phi}_{-i}'\sum_{\ell=0}^{\infty}\boldsymbol{\Phi}^\ell\mathbf{R}\boldsymbol{\eta}_{t-\ell},
\end{aligned}$$

where $\boldsymbol{\phi}_{-i} = (\boldsymbol{\Phi} - \boldsymbol{\Theta})'\mathbf{e}_{N,i}$ satisfies condition (8) under Assumption 4, and $\mathbf{y}_t - \boldsymbol{\gamma} f_t = \sum_{\ell=0}^{\infty}\boldsymbol{\Phi}^\ell\mathbf{R}\boldsymbol{\eta}_{t-\ell}$. The second moment of ϑ_{it} is uniformly bounded by KN^{-1}, using similar arguments as before:

$$\begin{aligned}
E\left(\vartheta_{it}^2\right) &= \sum_{\ell=0}^{\infty}\boldsymbol{\phi}_{-i}'\boldsymbol{\Phi}^\ell\mathbf{R}\mathbf{R}'\boldsymbol{\Phi}^{\ell\prime}\boldsymbol{\phi}_{-i} \\
&\leq \left\|\boldsymbol{\phi}_{-i}\right\|^2\left\|\mathbf{R}\right\|^2\sum_{\ell=0}^{\infty}\left\|\boldsymbol{\Phi}\right\|^{2\ell} \\
&\leq \frac{K}{N},
\end{aligned} \tag{A.20}$$

where $\left\|\boldsymbol{\phi}_{-i}\right\|^2 < KN^{-1}$ by Assumption 4, $\left\|\mathbf{R}\right\|^2 \leq \left\|\mathbf{R}\right\|_1\left\|\mathbf{R}\right\|_\infty < K$ by Assumption 1, and $\sum_{\ell=0}^{\infty}\left\|\boldsymbol{\Phi}\right\|^{2\ell} < K$ by Assumption 3. (A.20) implies

$$E\left\|(\boldsymbol{\Phi} - \boldsymbol{\Theta})(\mathbf{y}_t - \boldsymbol{\gamma} f_t)\right\|_\infty \to 0. \tag{A.21}$$

Using (A.17)-(A.19), and (A.21) in (A.14), we obtain

$$E\left\|E\left(\mathbf{y}_{t+1} \mid \mathcal{I}_t, \mathcal{F}_t\right) - \hat{\mathbf{y}}_{t+1|t}^{aug}\right\|_\infty \to 0,$$

as desired. Result (42) in the case of the model featuring a common factor for $h > 1$ can be established in a similar way. Singularity of $\mathbf{G}_0 = \mathbf{I}_N - \mathbf{b}_0\mathbf{w}'$ is implied by (29). This completes the proof. ■

A.3 Additional Tables and Figures

Table A.1: Classification of countries and description of output data

Country	Class.	Data Description	Source	Start Date	SA and Notes
U.S.	Adv.	GDP in Bil. Chained 2009 Dollars	Bureau of Economic Analysis	Q1-1947	by source
U.K.	Adv.	GDP in Mil. Chained 2010 Pounds	Office for National Statistics	Q1-1955	by source
Australia	Adv.	GDP in Mil. Chained Q3:10-Q2:11 AUD	Australian Bureau of Statistics	Q3-1959	by source
Sweden	Adv.	GDP in Mil. Chained 2012 SEK	Statistiska Centralbyran	Q1-1993	by source
Denmark	Adv.	GDP in Mil. Chained 2005 Kroner	Danmarks Statistik	Q1-1991	by source
Switzerland	Adv.	GDP in Mil. Chained 2005 Francs	State Secretariat for Economic Affairs	Q1-1980	by source
Hungary	Eme.	GDP in Bil. Chained 2005 Forints	Central Statistical Office	Q1-1995	by source
Israel	Adv.	GDP in Mil. Chained 2010 NIS	Central Bureau of Statistics	Q1-1995	by source
Germany	Adv.	GDP in Bil. Chained 2005 Euros	Deutsche Bundesbank	Q1-1970	by source
Italy	Adv.	GDP in Mil. Chained 2005 Euros	Istituto Nazionale di Statistica	Q1-1981	by source
Russia	Eme.	GDP in Bil. Chained 2008 Rubles	State Committee of the Russian Federation	Q1-1995	by source
Austria	Adv.	GDP in Bil. Chained 2005 Euros	Austrian Institute of Economic Research/Statistik Austria/Haver	Q1-1988	by Haver
Spain	Adv.	GDP in Mil. Chained 2008 Euros	Instituto Nacional de Estadística	Q1-1995	by source
France	Adv.	GDP in Mil. Chained 2005 Euros	Institut National de la Statistique/Economique	Q1-1949	by source
Ireland	Adv.	GDP in Mil. Chained 2011 Euros	Central Statistics Office Ireland	Q1-1997	by source
Poland	Eme.	GDP in Mil. Chained 2005 Zloty	Central Statistical Office/Haver Analytics	Q2-1995	by source
Hong Kong	Adv.	GDP in Mil. Chained 2011 Hong Kong Dollars	Hong Kong Census and Statistics Department/Haver Analytics	Q1-1973	by Haver
Singapore	Adv.	GDP in Mil. 2005 Singapore Dollars	Department of Statistics	Q1-1975	by source
Argentina	Eme.	GDP in Mil. 1993 Pesos	Min de Economía y Obras y Servicios Públicos	Q1-1993	by source
Brazil	Eme.	GDP at Market Prices in Mil. Chained 1995 Reais	Instituto Brasileiro de Geografia e Estatística/Haver	Q1-1990	by source
Canada	Adv.	GDP in Mil. Chained 2007 Canadian Dollars	Statistics Canada	Q1-1981	by source
Chile	Eme.	GDP in Mil. Chained 2008 Chilean Pesos	Banco Central de Chile	Q1-1986	by Haver code
China	Eme.	GDP in Bil. 2000 Yuan	China National Bureau of Statistics/Haver Analytics	Q1-1992	by Haver
Colombia	Eme.	GDP in Bil. Chained 2005 Pesos	Departamento Administrativo Nacional de Estadísticas	Q1-2000	by source
		Spliced with Annual GDP Data	World Bank	1960	(*)
India	Eme.	GDP at Market Prices in Bil. FY04 Rupees	Central Statistical Organization, India/Haver Analytics	Q2-1996	by Haver
Indonesia	Eme.	GDP in Bil. 2000 Rupiahs	Biro Pusat Statistik/Haver Analytics	Q1-1983	by Haver
Japan	Adv.	GDP in Bil. Chained 2005 Yen	Cabinet Office of Japan	Q1-1980	by source
Korea	Adv.	GDP in Bil. Chained 2005 Won	Bank of Korea	Q1-1970	by source
Malaysia	Eme.	GDP in Mil. 2005 Ringgit	Department of Statistics, Malaysia/Haver Analytics	Q1-1991	by Haver
Mexico	Eme.	GDP at Market Prices in Mil. 2008 New Pesos	Instituto Nacional de Estadística Geografía e Informática	Q1-1980	by source
Netherlands	Adv.	GDP in Mil. Chained 2005 Euros	Centraal Bureau voor de Statistiek	Q1-1977	by source
Philippines	Eme.	GDP in Mil. 2000 Pesos	National Statistical Coordination Board/Haver Analytics	Q1-1981	by Haver
Saudi Arabia	Eme.	GDP in Mil. 1999 Riyals	Central Department of Statistics & Information	Q1-2010	by Haver code
		Spliced with Annual GDP Data incl Import Duties	Central Department of Statistics & Information	1968	(*)
Thailand	Eme.	GDP in Mil. 1988 Baht	National Economic and Social Development Board	Q1-1993	by source
Venezuela	Eme.	GDP in Thous. 1997 Bolivares	Banco Central de Venezuela/Haver Analytics	Q1-1997	by Haver
Czech Republic	Adv.	GDP in Mil. Chained 2005 CZK	Czech Statistical Office	Q1-1996	by source
Greece	Adv.	GDP in Mil. Chained 2005 Euros	Statistical Office of the European Communities	Q1-1995	by Haver code
New Zealand	Adv.	GDP in Mil. Chained Q3/1995-Q2/1996 NZD	Statistics New Zealand	Q2-1987	by source
Norway	Adv.	GDP in Mil. Chained 2010 NOK	Statistisk Sentralbyra	Q1-1978	by source
Taiwan	Adv.	GDP in Mil. 2006 New Taiwan Dollars	Directorate-General of Budget Accounting & Stats	Q1-1981	by source
Turkey	Eme.	GDP in Thous. 1998 TL	Turkish Statistical Institute/Haver Analytics	Q1-1987	by Haver
South Africa	Eme.	GDP in Mil. 2005 Rand	South African Reserve Bank	Q1-1960	by source
Belgium	Adv.	GDP in Mil. Chained 2010 Euros	Banque Nationale de Belgique	Q1-1995	by source
Portugal	Adv.	GDP in Mil. Chained 2006 Euros	Instituto Nacional de Estatistica	Q1-1995	by source
Costa Rica	Eme.	GDP in Mil. 1991 Col	Banco Central de Costa Rica/Haver Analytics	Q1-1991	by Haver
Peru	Eme.	GDP in Mil. 1994 New Soles	Banco Central de Reserva del Perú/Haver Analytics	Q1-1980	by Haver
Nigeria	Eme.	GDP in Mil. 1990 Naira	National Bureau of Statistics, Nigeria	Q1-2007	by Haver code
		Spliced with Annual GDP Data	World Bank	1960	(*)
Bulgaria	Eme.	GDP in Thous. Chained 2005 Leva	National Statistical Institute	Q1-1997	by source

Notes: (*) Interpolated to quarterly frequency using cubic splines. We have interpolated partial sums of annual series and then first differenced the interpolated series to obtain quarterly output data. Interpolating partial sums and then differencing ensures that the sum of quarters matches the annual data for each year.

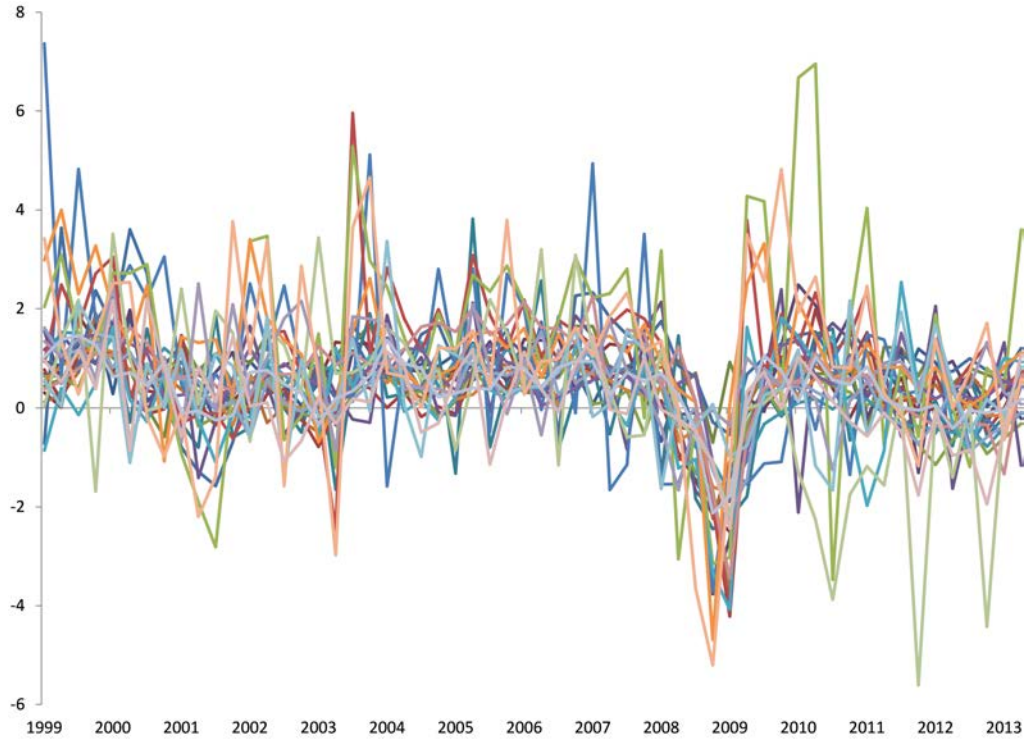
Table A.2: Description of PMI data

Country	Data Description	Source	Start Date	SA and Notes
		A. *Manufacturing PMIs*		
U.S.	Manufacturing PMI	Institute for Supply Management	Jan-1948	by source
U.K.	Manufacturing PMI	CIPS/Markit	Jan-1992	by source
Australia	Manufacturing PMI	Australian Industry Group-PricewaterhouseCoopers	Sep-1992	by source
Sweden	Manufacturing PMI	Swedbank	Nov-1994	by source
Denmark	Manufacturing PMI	Danish Purchasing and Logistics Forum	Jan-1994	by source
Switzerland	Manufacturing PMI	SVME/Credit Suisse/Markit	Jan-1995	by source
Hungary	Manufacturing PMI	Hungary Assoc Logistics, Purchasing, Inventory	Sep-1995	by source
Israel	Manufacturing PMI	Bank Hapoalim/Israeli Purchasing and Logistics Managers Association	Jan-1996	by source
Germany	Manufacturing PMI	Markit	Apr-1996	by source
Italy	Manufacturing PMI	Markit/Associazione Italiana Acquisti e Supply Management	Jun-1997	by source
Russia	Manufacturing PMI	HSBC/Markit	Sep-1997	by source
Austria	Manufacturing PMI	Creditanstalt	Oct-1998	by source
Spain	Manufacturing PMI	Markit	Feb-1998	by source
France	Manufacturing PMI	Markit/Comp des Dirigeants et Acheteurs France	Apr-1998	by source
Ireland	Manufacturing PMI	NCB Stockbrokers	May-1998	by source
Poland	Manufacturing PMI	HSBC/Markit	Jun-1998	by source
Hong Kong	Total Economy PMI	HSBC/Markit	Jul-1998	by source, ([b])
Singapore	Manufacturing PMI	Singapore Institute of Purchasing and Materials Management/Markit	Jan-1999	by source
Brazil	Manufacturing PMI	HSBC/Markit	Feb-2006	by source, ([d])
Canada	Total Economy PMI	Richard Ivey School of Business/Univ W Ontario	Jan-2001	by Haver, ([a,b])
China	Manufacturing PMI	HSBC/Markit	Apr-2004	by source, ([a])
India	Manufacturing PMI	HSBC/Markit	Mar-2005	by source, ([d])
Japan	Manufacturing PMI	Markit/Japan Materials Management Association	Oct-2001	by source, ([a])
Korea	Manufacturing PMI	HSBC/Markit	Apr-2004	by source, ([a])
Mexico	Manufacturing PMI	Instituto Nacional de Estadística Geografía e Informática	Jan-2004	by source, ([a])
Netherlands	Manufacturing PMI	NEVI	Mar-2000	by source, ([a])
Saudi Arabia	Total Economy PMI	SABB/HSBC/Markit	Aug-2009	by source, ([b,d])
Czech Republic	Manufacturing PMI	HSBC/Markit	Jun-2001	by source, ([a])
Greece	Manufacturing PMI	Markit	May-1999	by source, ([a])
New Zealand	Manufacturing PMI	Business NZ/Markit	Aug-2002	by source, ([a])
Norway	Manufacturing PMI	Norsk Forbund for Innkjop og Logistikk	Feb-2004	by source, ([a])
Taiwan	Manufacturing PMI	HSBC/Markit	Apr-2004	by source, ([a])
Turkey	Manufacturing PMI	HSBC/Markit	Jun-2005	by source, ([d])
South Africa	Manufacturing PMI	Investec/IPSA/Markit	Sep-1999	by source, ([a])
		B. *Services PMIs*		
U.S.	Services PMI: Business Activity Index	Institute for Supply Management	Jul-1997	by source
U.K.	Services PMI: Business Activity Index	CIPS/Markit	Jul-1996	by source
Australia	Services PMI Composite	Australian Industry Group-Commonwealth Bank	Feb-2003	by source, ([a,c])
Germany	Services PMI: Business Activity Index	Markit	Jun-1997	by source
Italy	Services PMI: Business Activity Index	Markit/Associazione Italiana Acquisti e Supply Management	Jan-1998	by source
Russia	Services PMI: Business Activity Index	HSBC/Markit	Oct-2001	by source, ([a])
Spain	Services PMI: Business Activity Index	Markit	Aug-1999	by source, ([a])
France	Services PMI: Business Activity Index	Markit/Comp des Dirigeants et Acheteurs France	May-1998	by source
Ireland	Services PMI: Business Activity Index	NCB Stockbrokers	May-2000	by source, ([a])
Brazil	Services PMI: Business Activity Index	HSBC/Markit	Mar-2007	by source, ([d])
China	Services PMI: Business Activity Index	HSBC/Markit	Nov-2005	by source, ([d])
India	Services PMI: Business Activity Index	HSBC/Markit	Dec-2005	by source, ([d])
Japan	Services PMI: Business Activity Index	Markit	Sep-2007	by source, ([d])

Notes: (a) Missing historical data going back to 1999M1 are backcasted with the global manufacturing (or service) PMI index using a regression estimated from the beginning of the series until 2005M12. (b) PMI data for the manufacturing sector for Hong Kong and Saudi Arabia are not available and total economy PMI is used instead. The Canadian manufacturing PMI series starts in October 2010, so we use total economy PMI also for Canada. (c) The Australian Performance of Services Index (PSI) is a composite index based on the diffusion indices for sales, new orders, employment, inventories, and deliveries all with varying weights. (d) Series not used in the forecasting exercise since they are available only over relatively short periods.

Figure A.1: Output (1st differences of logs)

A. Advanced Economies



B. Emerging Economies

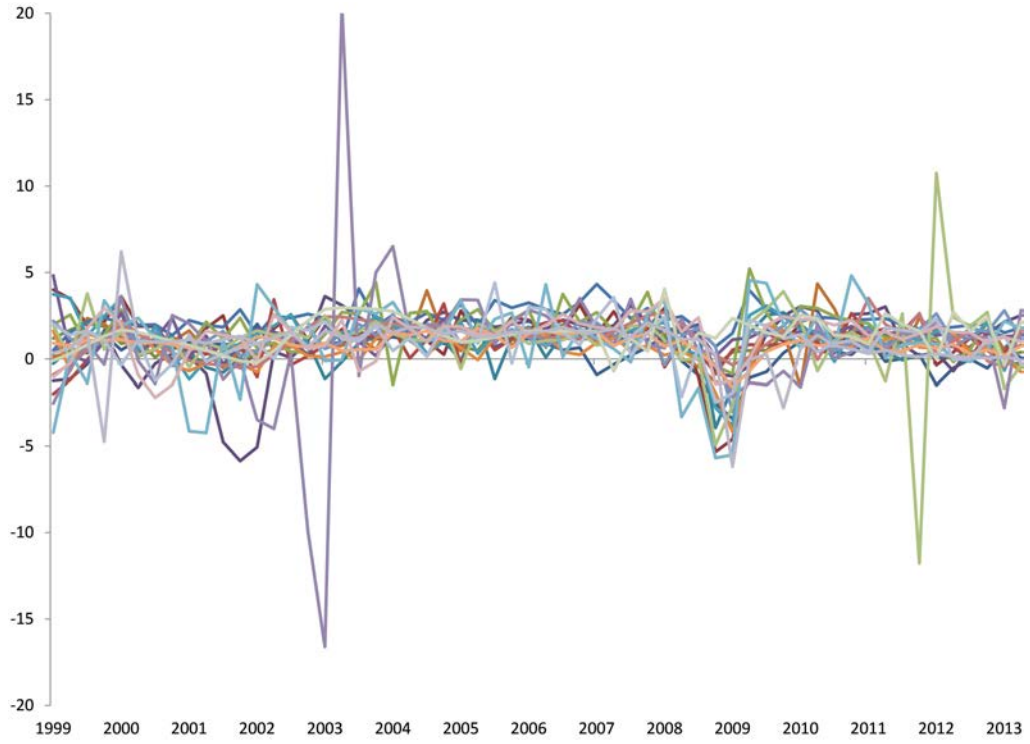
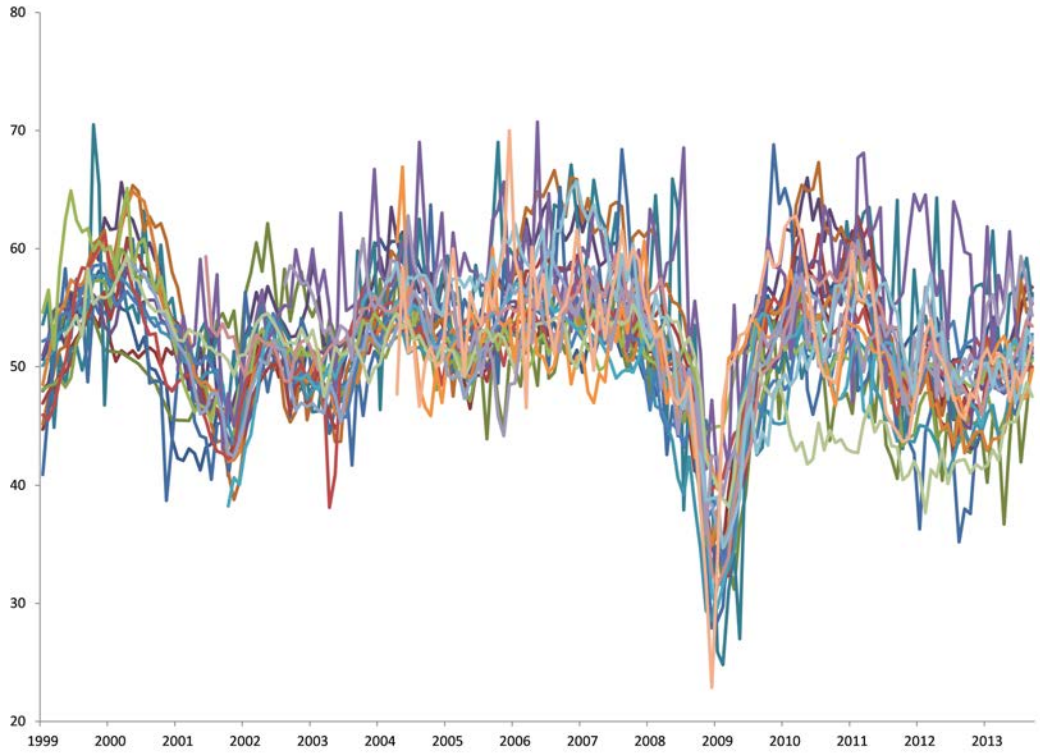


Figure A.2: Manufacturing Purchasing Managers Indices

A. Advanced Economies



B. Emerging Economies

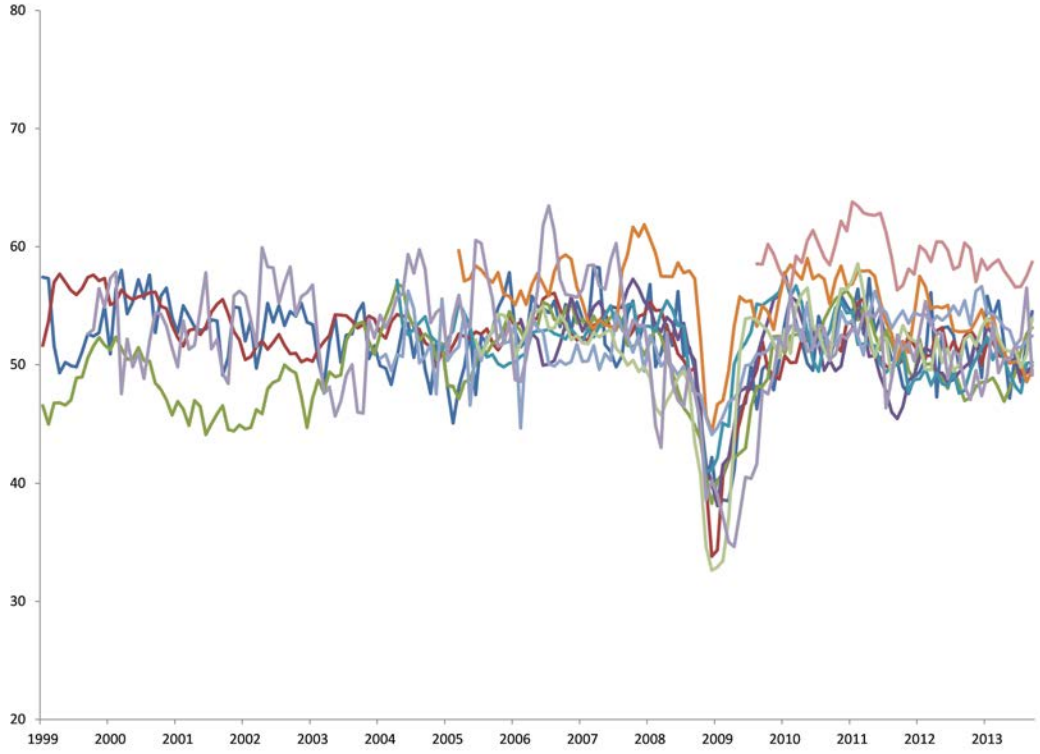


Figure A.3: All Economies' Services Purchasing Managers Indices

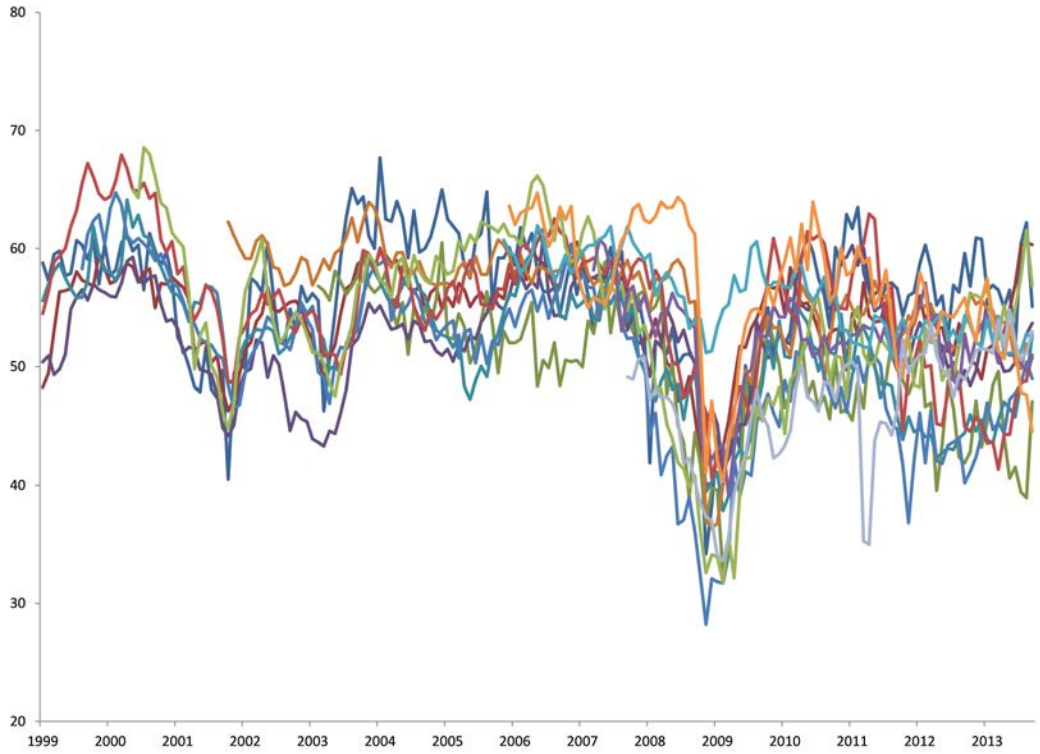
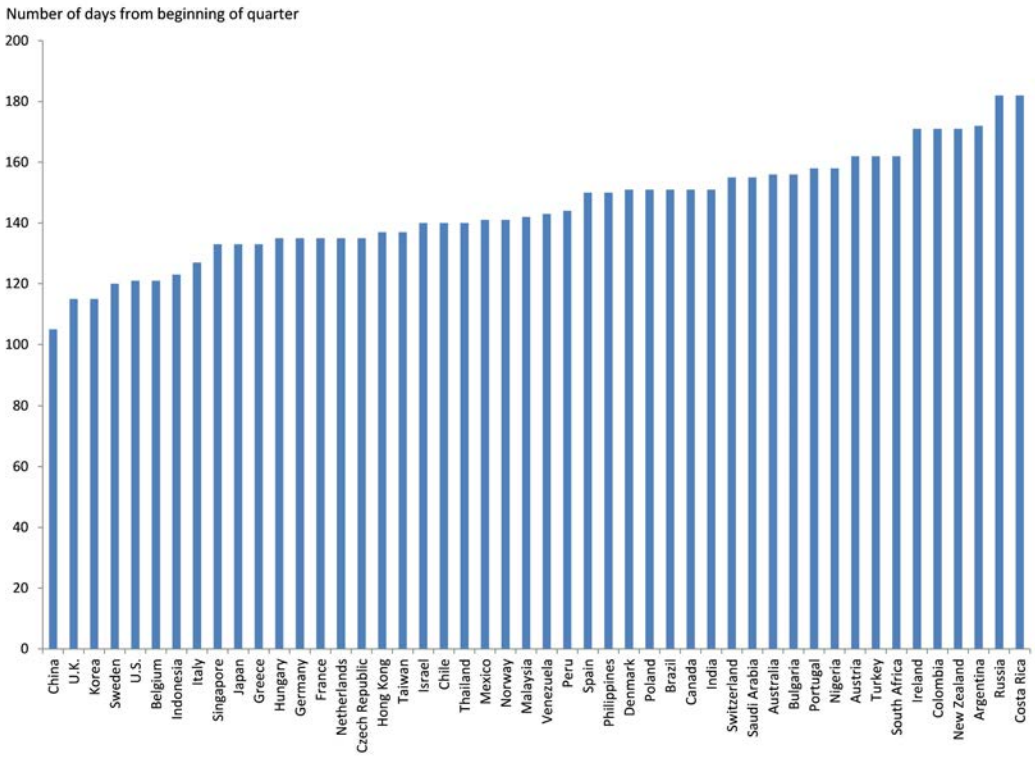


Figure A.4: GDP Release Lags



References

Aastveit, K. and T. Trovik (2012, February). Nowcasting Norwegian GDP: The role of asset prices in a small open economy. *Empirical Economics 42*(1), 95–119.

Andreou, E., E. Ghysels, and A. Kourtellos (2010). Should macroeconomic forecasters use daily financial data and how? Working Paper Series 42-10, The Rimini Centre for Economic Analysis.

Angelini, E., M. Banbura, and G. Rünstler (2010). Estimating and forecasting the euro area monthly national accounts from a dynamic factor model. *OECD Journal: Journal of Business Cycle Measurement and Analysis 2010*(1), 1–22.

Angelini, E., G. Camba-Mendez, D. Giannone, L. Reichlin, and G. Rünstler (2011). Short-term forecasts of euro area GDP growth. *Econometrics Journal 14*(1), C25–C44.

Arnostova, K., D. Havrlant, L. Rùđièka, and P. Tóth (2011). Short-term forecasting of Czech quarterly GDP using monthly indicators. *Czech Journal of Economics and Finance (Finance a uver) 61*(6), 566–583.

Baffigi, A., R. Golinelli, and G. Parigi (2004). Bridge models to forecast the euro area GDP. *International Journal of Forecasting 20*(3), 447–460.

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

Bailey, N., M. H. Pesaran, and L. V. Smith (2014). A multiple testing approach to the regularisation of large sample correlation matrices. CESifo Working Paper No. 4834.

Bańbura, M. and M. Modugno (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics 29*(1), 133–160.

Bańbura, M. and G. Rünstler (2011). A look into the factor model black box: publication lags and the role of hard and soft data in forecasting GDP. *International Journal of Forecasting 27*(2), 333–346.

Barhoumi, K., O. Darné, and L. Ferrara (2010). Are disaggregate data useful for factor analysis in forecasting French GDP? *Journal of Forecasting 29*(1-2), 132–144.

Bernanke, B. S., J. Bovian, and P. Eliasz (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics 120*, 387–422.

Bickel, P. J. and E. Levina (2008). Covariance regularization by thresholding. *The Annals of Statistics 36*, 2577–2604.

Cai, T. and W. Liu (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of American Statistical Association 106*, 672–684.

Camacho, M. and G. Perez-Quiros (2010). Introducing the euro-sting: Short-term indicator of euro area growth. *Journal of Applied Econometrics 25*(4), 663–694.

Chudik, A. and M. H. Pesaran (2011). Infinite dimensional VARs and factor models. *Journal of Econometrics 163*, 4–22.

Chudik, A. and M. H. Pesaran (2013a). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. CESifo Working Paper No. 4232.

Chudik, A. and M. H. Pesaran (2013b). Econometric analysis of high dimensional VARs featuring a dominant unit. *Econometric Reviews 32*, 592–649.

Chudik, A. and M. H. Pesaran (2014a). Aggregation in large dynamic panels. *Journal of Econometrics 178*, 273–285.

Chudik, A. and M. H. Pesaran (2014b). Theory and practice of GVAR modeling. *forthcoming in Journal of Economic Surveys*.

Clements, M. P. and A. B. Galvão (2008). Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States. *Journal of Business & Economic Statistics 26*(4), 546–554.

Clements, M. P. and A. B. Galvão (2009). Forecasting US output growth using leading indicators: an appraisal using MIDAS models. *Journal of Applied Econometrics 24*(7), 1187–1206.

D'Agostino, A., K. McQuinn, and D. O'Brien (2012). Nowcasting Irish GDP. *OECD Journal: Journal of Business Cycle Measurement and Analysis 2012*(2), 21–31.

De Mol, C., D. Giannone, and L. Reichlin (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics 146*, 318–328.

de Winter, J. (2011). Forecasting GDP growth in times of crisis: private sector forecasts versus statistical models. DNB Working Papers 320, Netherlands Central Bank, Research Department.

Diebold, F. X. and R. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics 13*, 253–263.

Diron, M. (2008). Short-term forecasts of euro area real GDP growth: an assessment of real-time performance based on vintage data. *Journal of Forecasting 27*(5), 371–390.

Dées, S., F. di Mauro, M. H. Pesaran, and L. V. Smith (2007). Exploring the international linkages of the euro area: A global VAR analysis. *Journal of Applied Econometrics 22*, 1–38.

Eickmeier, S. and T. Ng (2011). Forecasting national activity using lots of international predictors: An application to New Zealand. *International Journal of Forecasting 27*, 496–511.

Eklund, J. and G. Kapetanios (2008). A review of forecasting techniques for large data sets. *National Institute Economic Review 203*(1), 109–115.

Evans, M. D. D. (2005, September). Where are we now? Real-time estimates of the macroeconomy. *International Journal of Central Banking 1*(2).

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

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The generalized dynamic factor model: Identification and estimation. *Review of Economics and Statistic 82*, 540–554.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2004). The generalized dynamic factor model: Consistency and rates. *Journal of Econometrics 119*, 231–235.

Forni, M. and M. Lippi (2001). The generalized factor model: Representation theory. *Econometric Theory 17*, 1113–1141.

Geweke, J. (1977). The dynamic factor analysis of economic time series. In D. Aigner and A. Goldberger (Eds.), *Latent variables in socio-economic models*. Amsterdam: North-Holland.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2004, May). The MIDAS touch: Mixed data sampling regression models. CIRANO Working Papers 2004s-20, CIRANO.

Ghysels, E., A. Sinko, and R. Valkanov (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews 26*(1), 53–90.

Giannone, D., L. Reichlin, and L. Sala (2005). Monetary policy in real time. In M. Gertler and K. Rogoff (Eds.), *NBER Macroeconomics Annual 2004*, Volume 19, pp. 161–200. MIT Press.

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

Groen, J. J. J. and G. Kapetanios (2008). Revisiting useful approaches to data-rich macroeconomic forecasting. Federal Reserve Bank of New York, Staff Report No. 327, revised September 2009.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning* (2 ed.). Springer.

Horn, R. A. and C. A. Johnson (1985). *Matrix Analysis*. Cambridge University Press.

Ing, C.-K. (2003). Multistep prediction in autoregressive processes. *Econometric Theory 19*, 254–279.

Kitchen, J. and R. Monaco (2003, October). Real-time forecasting in practice: The U.S. treasury staff's real-time GDP forecast system. MPRA Paper 21068, University Library of Munich, Germany.

Kuzin, V., M. Marcellino, and C. Schumacher (2011). MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area. *International Journal of Forecasting 27*(2), 529–542.

Kuzin, V., M. Marcellino, and C. Schumacher (2013). Pooling versus model selection for nowcasting GDP with many predictors: Empirical evidence for six industrialized countries. *Journal of Applied Econometrics 28*(3), 392–411.

Lahiri, K. and G. Monokroussos (2013). Nowcasting US GDP: The role of ISM business surveys. *International Journal of Forecasting 29*(4), 644–658.

Ledoit, O. and M. Wolf (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis 88*, 365–411.

Lin, J.-L. and R. S. Tsay (2006). Comparisons of forecasting methods with many predictors. Mimeo.

Marcellino, M. and C. Schumacher (2010). Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP. *Oxford Bulletin of Economics and Statistics 72*(4), 518–550.

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

Matheson, T. D. (2010, January). An analysis of the informational content of New Zealand data releases: The importance of business opinion surveys. *Economic Modelling 27*(1), 304–314.

Newey, W. K. and K. D. West (1987, May). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–08.

Parigi, G. and R. Golinelli (2007). The use of monthly indicators to forecast quarterly GDP in the short run: an application to the G7 countries. *Journal of Forecasting 26*(2), 77–94.

Parigi, G. and G. Schlitzer (1995). Quarterly forecasts of the Italian business cycle by means of monthly economic indicators. *Journal of Forecasting 14*(2), 117–141.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with multifactor error structure. *Econometrica 74*, 967–1012.

Pesaran, M. H., A. Pick, and A. Timmermann (2011). Variable selection, estimation and inference for multi-period forecasting problems. *Journal of Econometrics 164*, 173–187.

Pesaran, M. H., T. Schuermann, and L. V. Smith (2009). Forecasting economic and financial variables with global VARs. *International Journal of Forecasting 25*, 642–675.

Pesaran, M. H., T. Schuermann, and S. M. Weiner (2004). Modelling regional interdependencies using a global error-correcting macroeconometric model. *Journal of Business and Economics Statistics 22*, 129–162.

Rünstler, G., K. Barhoumi, S. Benk, R. Cristadoro, A. Den Reijer, A. Jakaitiene, P. Jelonek, A. Rua, K. Ruth, and C. Van Nieuwenhuyze (2009). Short-term forecasting of GDP using large datasets: a pseudo real-time forecast evaluation exercise. *Journal of forecasting 28*(7), 595–611.

Rünstler, G. and F. Sédillot (2003, September). Short-term estimates of euro area real GDP by means of monthly data. Working Paper Series 0276, European Central Bank.

Sargent, T. J. and C. A. Sims (1977). Business cycle modeling without pretending to have too much a-priori economic theory. In C. Sims (Ed.), *New methods in business cycle research*. Minneapolis: Federal Reserve Bank of Minneapolis.

Siliverstovs, B. (2012). Keeping a finger on the pulse of the economy: Nowcasting Swiss GDP in real-time squared. KOF Working paper No. 12-302.

Siliverstovs, B. and K. A. Kholodilin (2012, July). Assessing the real-time informational content of macroeconomic data releases for now-/forecasting GDP: Evidence for Switzerland. *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik) 232*(4), 429–444.

Stock, J. H. and M. W. Watson (1999). Forecasting inflation. *Journal of Monetary Economics 44*, 293–335.

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

Stock, J. H. and M. W. Watson (2005). Implications of dynamic factor models for VAR analysis. NBER Working Paper No. 11467.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society Series B 58*, 267–288.

Trehan, B. (1989). Forecasting growth in current quarter real GNP. *Economic Review* (Win), 39–52.

Wallis, K. F. (1986). Forecasting with an econometric model: The Śragged edgeŠ problem. *Journal of Forecasting 5*, 1–13.

Wold, H. (1982). Soft modeling: The basic design and some extensions. In K. G. Joreskog and H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction: Vol. 2*, pp. 589–591. Amsterdam: North-Holland.

Yiu, M. S. and K. K. Chow (2010). Nowcasting Chinese GDP: Information content of economic and financial data. *China Economic Journal 3*(3), 223–240.