A multi-country approach to forecasting output growth using PMIs*

Alexander Chudik†
Federal Reserve Bank of Dallas, CAFE and CIMF

Valerie Grossman‡
Federal Reserve Bank of Dallas

M. Hashem Pesaran§
USC Dornsife INET and Trinity College, Cambridge

January 4, 2016

Abstract

This paper derives new theoretical results for forecasting with Global VAR (GVAR) models. It is shown that the presence of strong unobserved common factors 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 $N, T \xrightarrow{j} \infty$) 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) output growth, and compare forecasts from AugGVAR models with a number of data-rich forecasting methods, including Lasso, Ridge, partial least squares, and factor-based techniques. 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.

Keywords: Global VAR, High-dimensional VAR, Augmented GVAR, Forecasting, Nowcasting, Data-rich methods, GDP and PMIs

JEL Classification: C53, E37

*We would like to thank Ana Beatriz Galvão, Kundan Kishor, Evan Koenig, Kajal Lahiri, Ron Smith, Mark Wynne, participants at the 8^{th} ECB workshop on forecasting techniques, Frankfurt, Germany, 2014, 32nd Biannual CIRET conference, Hangzhou, China, 2014, EABCN conference on econometric methods for business cycle analysis, forecasting and policy simulations, Oslo, Norway, 2015, and participants at the Federal Reserve Bank of Dallas brownbag seminar for helpful comments and suggestions. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Dallas.

†Federal Reserve Bank of Dallas, 2200 N. Pearl Street, Dallas, Texas; e-mail: alexander.chudik@dal.frb.org.

‡Federal Reserve Bank of Dallas, 2200 N. Pearl Street, Dallas, Texas; e-mail: valerie.grossman@dal.frb.org.

§USC Dornsife INET, Department of Economics, University of Southern California, 3620 South Vermont Avenue, Kaprielian Hall 300, Los Angeles, CA 90089-0253; e-mail: pesaran@usc.edu; http://pesaran.com/.

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 vector autoregressive (VAR) models 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 datasets 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. A handbook introduction to the GVAR approach, covering a number of early applications, is provided in di Mauro and Pesaran (2013). Further developments of the GVAR approach together with more recent applications, including in the field of forecasting, are reviewed in Chudik and Pesaran (2015b) and Pesaran (2015).

In this paper we establish conditions under which forecasts from the GVAR model converge to optimal *infeasible* forecasts (as $N, T \xrightarrow{j} \infty$) 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 popular methods of augmenting VARs with factors as special cases of a more general model: (i) modeling deviations from the factors as a VAR (as in Dées et al. (2007)), and (ii) a VAR model with a multi-factor error structure. Since factors are unobserved, we 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 the 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 unobserved common factors or not.

The effectiveness of the proposed approach is also illustrated in an empirical application, where forecasts of output growth across 48 countries are obtained 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.

In this paper, we follow the simple approach of aggregating PMI data into a quarterly frequency, and 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)). We describe individual methods in more detail and provide references to the literature in Section 6.2.

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 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 strong unobserved common factors can lead to an undetermined GVAR model, proposes the AugGVAR, and establishes uniform convergence of feasible AugGVAR forecasts to the infeasible optimal forecasts as $N, T \xrightarrow{j} \infty$. Section 5 illustrates the theoretical findings by means of MC experiments. Section 6 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 when the differences in forecast errors are cross-sectionally dependent. Section 7 ends with some concluding remarks. Technical proofs and further results are provided in the Appendix and in Chudik et al. (2014).

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 $\mathbf{A}$, and $\|\mathbf{A}\| = \sqrt{|\lambda_1(\mathbf{A}'\mathbf{A})|}$ is the spectral norm 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

Let $\mathbf{y}_{it}$ be an $n_i \times 1$ vector of variables for the i-th cross-section unit (e.g., country), for $i = 1, 2, ..., N$. We assume n_i is bounded, $|n_i| < K$, and generally small, whereas N is large. Let $\mathbf{y}_t = (\mathbf{y}'_{1t}, \mathbf{y}'_{2t}, ..., \mathbf{y}'_{Nt})'$ be a vector of all $n = \sum_{i=1}^{N} n_i$ variables. We model $\mathbf{y}_t$ using a VAR model augmented with unobserved common factors. The augmentation can be done in different ways. One possible approach would be to model deviations of $\mathbf{y}_t$ from common factors, $\mathbf{f}_t$, as a VAR, namely

$$M_a: \qquad \tilde{\mathbf{y}}_t = \mathbf{\Phi}\tilde{\mathbf{y}}_t + \boldsymbol{\varepsilon}_t, \tag{1}$$

where $\tilde{\mathbf{y}}_t = \mathbf{y}_t - \mathbf{\Gamma}_0\mathbf{f}_t$, $\mathbf{\Gamma}_0 = \left(\mathbf{\Gamma}'_{0,1}, \mathbf{\Gamma}'_{0,2}, ..., \mathbf{\Gamma}'_{0,N}\right)'$ is an $n \times m$ matrix of factor loadings, with $\mathbf{\Gamma}_{0,i}$ being $n_i \times m$, for $i = 1, 2, ..., N$, $\mathbf{f}_t$ is an $m \times 1$ vector of unobserved common factors, $\mathbf{\Phi}$ is an $n \times n$ matrix of unknown coefficients, and $\boldsymbol{\varepsilon}_t = (\boldsymbol{\varepsilon}'_{1t}, \boldsymbol{\varepsilon}'_{2t}, ..., \boldsymbol{\varepsilon}'_{Nt})'$ is an $n \times 1$ vector of idiosyncratic shocks with $\boldsymbol{\varepsilon}_{it}$ denoting an $n_i \times 1$ vector of unit-specific reduced-form innovations. Alternatively, one could consider a standard VAR model with a multi-factor error structure, namely

$$M_b: \quad \mathbf{y}_t = \mathbf{\Phi}\mathbf{y}_{t-1} + \mathbf{\Gamma}_0\mathbf{f}_t + \boldsymbol{\varepsilon}_t. \tag{2}$$

The two specifications can be combined to yield the following general specification:

$$\mathbf{y}_t = \mathbf{\Phi}\mathbf{y}_{t-1} + \mathbf{\Gamma}_0\mathbf{f}_t + \mathbf{\Gamma}_1\mathbf{f}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{3}$$

where $\mathbf{\Gamma}_1 = \left(\mathbf{\Gamma}'_{1,1}, \mathbf{\Gamma}'_{1,2}, ..., \mathbf{\Gamma}'_{1,N}\right)'$, and $\mathbf{\Gamma}_{1,i}$ are $n_i \times m$ coefficient matrices, for $i = 1, 2, ..., N$. It is clear that (1) and (2) can be obtained as special cases of (3). Specification M_a results by setting

$\mathbf{\Gamma}_1 = -\mathbf{\Phi}\mathbf{\Gamma}_0$, and specification M_b follows if we set $\mathbf{\Gamma}_1 = \mathbf{0}$. The factor augmented VAR model, (3), can also be extended to allow for deterministic terms, observed common factors, and higher order lags, but such extensions do not impact the main message of the paper and will be abstracted from for expositional convenience.

ASSUMPTION 1 *(Cross-sectionally weakly dependent idiosyncratic errors) Idiosyncratic errors in $\boldsymbol{\varepsilon}_t$ follow the 'spatial' model*

$$\boldsymbol{\varepsilon}_t = \mathbf{R}\boldsymbol{\eta}_t, \tag{4}$$

where the $n \times n$ matrix $\mathbf{R}$ has bounded row and column matrix norms (in n), $\boldsymbol{\eta}_t = (\eta_{1t}, \eta_{2t}, ..., \eta_{nt})'$, η_{it}, $i = 1, 2, ..., n$ are independently (across i and t) and identically distributed with zero mean, unit variance, and finite fourth moments.

ASSUMPTION 2 *(Unobserved common factors and their loadings)*

a. *(Model without factors) $\mathbf{\Gamma}_0 = \mathbf{\Gamma}_1 = \mathbf{0}$.*

b. *(Model with factors) The $m \times 1$ vector of unobserved common factors, $\mathbf{f}_t$, follows the stationary VAR(1) process*

$$\mathbf{f}_t = \mathbf{\Pi}\mathbf{f}_{t-1} + \mathbf{v}_t, \tag{5}$$

where $\mathbf{\Pi}$ is an $m \times m$ matrix of unknown coefficients, $\mathbf{v}_t$ is an $m \times 1$ vector of factor innovations, and $|\lambda_1(\mathbf{\Pi})| < 1$. The vector of factor innovations $\mathbf{v}_t$ is independently distributed of idiosyncratic errors, $E(\mathbf{v}_t) = 0$, $\|E(\mathbf{v}_t\mathbf{v}_t')\| < K$, $E\left(\mathbf{v}_t\mathbf{v}_{t'}'\right) = \mathbf{0}$ for any $t \neq t'$, and fourth moments of innovations in $\mathbf{v}_t$ are finite. Factor loadings in $\mathbf{\Gamma}_0$ and $\mathbf{\Gamma}_1$ are bounded, namely $\|\mathbf{\Gamma}_0\|_\infty < K$ and $\|\mathbf{\Gamma}_1\|_\infty \leq K$.

ASSUMPTION 3 *(Covariance stationarity and bounded variances) There exists a positive constant ϱ such that $\|\mathbf{\Phi}\| \leq \varrho < 1$, where $\|\mathbf{\Phi}\|$ denotes the spectral norm of $\mathbf{\Phi}$.*

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*

$$\|\mathbf{\Phi}_{ii}\| < K, \text{ for any } i = 1, 2, ..., N$$

and

$$\|\mathbf{\Phi}_{ij}\| < \frac{K}{N}, \text{ for any } j \neq i,\ i, j = 1, 2, ..., N,$$

where the $n \times n$ matrix $\mathbf{\Phi}$ is partitioned conformably to the partitioning of $\mathbf{y}_t = (\mathbf{y}_{1t}', \mathbf{y}_{2t}', ..., \mathbf{y}_{Nt}')'$ into $n_i \times n_j$ sub-matrices denoted by $\mathbf{\Phi}_{ij}$.

Remark 1 *Assumption 3 is stronger than the usual finite-N covariance stationarity assumption, which restricts the eigenvalues of $\mathbf{\Phi}$ to lie within the unit circle. Assumption 3 also ensures that the variance of $\mathbf{y}_{it}$ 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, $\mathbf{f}_t$, are unobserved, in general it is not possible to know which of the many possible factor-augmented VARs should be considered, and it is therefore desirable to develop forecasting techniques that are robust to the way unobserved factors are introduced in the VAR model. In view of this ambiguity, we proceed with the encompassing model (3) and show that under the above assumptions, the common factors can be well approximated by cross-section averages and their lags. The key practical difference between the specifications (1) and (2) turns out to be in the number of lags of cross-section averages that is 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, 2014). 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 (2015a).

3 Forecasting with factor-augmented VARs

Factor-augmented VAR models considered by Bernanke, Boivin, 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 specification (3) and its two special cases M_a and M_b differ in that we include a large number of variables in a VAR and the factors are 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 $\mathbf{\Phi}$ 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 $\mathbf{\Phi}$.

Forecasting with low- or high-dimensional factor-augmented VARs is straightforward when it is assumed that the factors and coefficients are known. Consider the general model (3) and information set $\mathcal{I}_t \cup \mathcal{F}_t$, 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 $\mathcal{F}_t = \{\mathbf{f}_t, \mathbf{f}_{t-1}, ...\}$ is an information set on current and past values of the common factors. Solving (1) for $\mathbf{y}_{t+h}$ by backward substitution yields

$$\mathbf{y}_{t+h} = \mathbf{\Phi}^h \mathbf{y}_t + \sum_{\ell=0}^{h-1} \mathbf{\Phi}^\ell \mathbf{\Gamma}_0 \mathbf{f}_{t+h-\ell} + \sum_{\ell=0}^{h-1} \mathbf{\Phi}^\ell \mathbf{\Gamma}_1 \mathbf{f}_{t+h-\ell-1} + \sum_{\ell=0}^{h-1} \mathbf{\Phi}^\ell \boldsymbol{\varepsilon}_{t+h-\ell},$$

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

$$\mathbf{y}_{t+h} = \mathbf{\Phi}^h \mathbf{y}_t + \mathbf{G}_h \mathbf{f}_t + \boldsymbol{\xi}_{th}, \tag{6}$$

where $\mathbf{G}_h = \sum_{\ell=0}^{h-1} \mathbf{\Phi}^\ell \left(\mathbf{\Gamma}_0 \mathbf{\Pi} + \mathbf{\Gamma}_1\right) \mathbf{\Pi}^{h-\ell-1}$, $\mathbf{I}_N$ is an $N \times N$ identity matrix,

$$\boldsymbol{\xi}_{th} = \boldsymbol{\kappa}_h \left(L\right) v_{t+h} + \sum_{\ell=0}^{h-1} \mathbf{\Phi}^\ell \boldsymbol{\varepsilon}_{t+h-\ell},$$

and

$$\kappa_h \left(L\right) = \begin{cases} \mathbf{\Gamma}_0, & h = 1, \\ \sum_{\ell=0}^{h-1} \mathbf{\Phi}^\ell \mathbf{\Gamma}_0 \sum_{s=0}^{h-\ell-1} \mathbf{\Pi}^s L^{\ell+s} + \sum_{\ell=0}^{h-2} \mathbf{\Phi}^\ell \mathbf{\Gamma}_1 \sum_{s=0}^{h-\ell-1} \mathbf{\Pi}^s L^{\ell+s+1}, & h > 1. \end{cases}$$

For $h > 1$, $\boldsymbol{\xi}_{th}$ is serially correlated, but orthogonal to the information available at time t, irrespective of whether the information set includes $\mathcal{F}_t$ or not; namely we have $E\left(\boldsymbol{\xi}_{th} | \mathcal{I}_t, \mathcal{F}_t\right) = \mathbf{0}$, and $E\left(\boldsymbol{\xi}_{th} | \mathcal{I}_t\right) = \mathbf{0}$, for $h = 1, 2, ...$ Assuming that the information set contains $\mathcal{F}_t$, the optimal h-step-ahead forecasts (in mean square error sense) are given by

$$\mathbf{y}_{\,t+h|t} = E\left(\mathbf{y}_{t+h} | \mathcal{I}_t, \mathcal{F}_t\right) = \mathbf{\Phi}^h \mathbf{y}_t + \mathbf{G}_h \mathbf{f}_t, \text{ for } h = 1, 2, ... \tag{7}$$

In the special case of model M_a and M_b, $\mathbf{G}_h$ reduces to

$$\mathbf{G}_h = \begin{cases} \mathbf{\Gamma}_0 \mathbf{\Pi}^h - \mathbf{\Phi}^h \mathbf{\Gamma}_0, & \text{if } \mathbf{\Gamma}_1 = -\mathbf{\Phi}\mathbf{\Gamma}_0 \ (M_a), \\ \sum_{\ell=0}^{h-1} \mathbf{\Phi}^\ell \mathbf{\Gamma}_0 \mathbf{\Pi}^{h-\ell}, & \text{if } \mathbf{\Gamma}_1 = \mathbf{0} \ (M_b). \end{cases}$$

Note that, regardless of whether $\mathbf{\Gamma}_1 = \mathbf{0}$ or $\mathbf{\Gamma}_1 = -\mathbf{\Phi}\mathbf{\Gamma}_0$, the conditional forecasts in (7) are linear in $\mathbf{y}_t$ and $\mathbf{f}_t$, and do not 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 (7) depends on the unobserved common factors 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 the processes generating the factors.[1] 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.

[1] In the Appendix of Chudik et al. (2014), 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).

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.

Suppose that $\mathbf{y}_t$ is generated according to (3) and let $\mathbf{W}$ be an $n \times n^*$ weights matrix such that

$$\|\mathbf{W}\|_\infty = \max_{s=1,2,\ldots,n} \sum_{j=1}^{n^*} |w_{sj}| < \frac{K}{N}, \tag{8}$$

and the individual columns of $\mathbf{W}$ sum to one, namely

$$\mathbf{W}'\boldsymbol{\tau}_n = \boldsymbol{\tau}_{n^*}, \tag{9}$$

where $\boldsymbol{\tau}_n = (1, 1, ..., 1)'$ is an $n \times 1$ vector of ones. n^* is finite and, as we shall see, is related to m, the number of unobserved factors. Taking weighted cross-section averages of $\mathbf{y}_t$ yields

$$\begin{aligned} \bar{\mathbf{y}}_{wt} &= \mathbf{W}'\boldsymbol{\Phi}\mathbf{y}_{t-1} + \mathbf{W}'\boldsymbol{\Gamma}_0\mathbf{f}_t + \mathbf{W}'\boldsymbol{\Gamma}_1\mathbf{f}_{t-1} + \mathbf{W}'\boldsymbol{\varepsilon}_t, \\ &= \mathbf{W}'\boldsymbol{\Phi}\mathbf{y}_{t-1} + \bar{\boldsymbol{\Gamma}}_{W,0}\mathbf{f}_t + \bar{\boldsymbol{\Gamma}}_{W,1}\mathbf{f}_{t-1} + O_p\left(N^{-1/2}\right), \end{aligned} \tag{10}$$

where $\bar{\mathbf{y}}_{wt}$ is an $n^* \times 1$ vector of cross-section averages, $\bar{\boldsymbol{\Gamma}}_{W,0} = \mathbf{W}'\boldsymbol{\Gamma}_0$, and $\bar{\boldsymbol{\Gamma}}_{W,1} = \mathbf{W}'\boldsymbol{\Gamma}_1$. We have used the result, $\|\mathbf{W}'\boldsymbol{\varepsilon}_t\| = O_p\left(N^{-1/2}\right)$, which follows since $\boldsymbol{\varepsilon}_t$ in (4) represents a weakly cross-sectionally correlated process, and $\mathbf{W}$ is granular. More formally, we have

$$\begin{aligned} \left\| E\left[\left(\mathbf{W}'\boldsymbol{\varepsilon}_t\right)^2\right]\right\| &= \left\|\mathbf{W}'\mathbf{R}\mathbf{R}'\mathbf{W}\right\|, \\ &\leq \|\mathbf{W}\|^2 \|\mathbf{R}\|^2 = O\left(N^{-1}\right), \end{aligned}$$

in which $\|\mathbf{W}\|^2 \leq \|\mathbf{W}\|_\infty \|\mathbf{W}\|_1 \leq n \|\mathbf{W}\|_\infty^2 = O\left(N^{-1}\right)$ (see condition (8) and assumption $|n_i| < K$), and $\|\mathbf{R}\|^2 \leq \|\mathbf{R}\|_\infty \|\mathbf{R}\|_1 = O(1)$ (Assumption 1). For a given $\boldsymbol{\Phi}$, and provided that the left inverse of $\bar{\boldsymbol{\Gamma}}_{W,0} + \bar{\boldsymbol{\Gamma}}_{W,1}L$ exists, denoted as $\left(\bar{\boldsymbol{\Gamma}}_{W,0} + \bar{\boldsymbol{\Gamma}}_{W,1}L\right)^-$, $\mathbf{f}_t$ can be approximated (up to a rotation matrix) by $\left(\bar{\boldsymbol{\Gamma}}_{W,0} + \bar{\boldsymbol{\Gamma}}_{W,1}L\right)^-\left(\bar{\mathbf{y}}_{wt} - \mathbf{W}'\boldsymbol{\Phi}\mathbf{y}_{t-1}\right)$. But $\boldsymbol{\Phi}$ cannot be estimated consistently when N is large, and this approach to estimation of the factors does not work. However, it is possible to approximate $\mathbf{f}_t$ by $\bar{\mathbf{y}}_{wt}$ and its lagged values, under certain assumptions. In particular, under Assumption 3 and assuming that the individual dynamic processes have been in operation for some time, we have

$$\mathbf{y}_t = \mathbf{Q}(L)\,\mathbf{f}_t + \sum_{j=0}^{\infty} \boldsymbol{\Phi}^j \boldsymbol{\varepsilon}_{t-j}, \tag{11}$$

where $\mathbf{Q}(L) = \boldsymbol{\Gamma}_0 + \sum_{j=1}^{\infty} \boldsymbol{\Phi}^{j-1}\left(\boldsymbol{\Phi}\boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_1\right) L^j$. Forming cross-sectional averages by pre-multiplying both sides of the above by $\mathbf{W}$, we have

$$\bar{\mathbf{y}}_{wt} = \mathbf{D}(L)\,\mathbf{f}_t + \sum_{j=0}^{\infty} \mathbf{W}'\boldsymbol{\Phi}^j \boldsymbol{\varepsilon}_{t-j}, \tag{12}$$

where

$$\mathbf{D}(L)=\sum_{j=0}^{\infty}\mathbf{D}_j L^j=\mathbf{W}'\mathbf{Q}(L)=\bar{\mathbf{\Gamma}}_{W,0}+\sum_{j=1}^{\infty}\mathbf{W}'\mathbf{\Phi}^{j-1}\left(\mathbf{\Phi}\mathbf{\Gamma}_0+\mathbf{\Gamma}_1\right)L^j$$

depends on $\mathbf{\Gamma}_0$, $\mathbf{\Gamma}_1$, $\mathbf{W}$, and *all* elements of $\mathbf{\Phi}$ (including the off-diagonal elements). But

$$Var\left(\sum_{j=0}^{\infty}\mathbf{W}'\mathbf{\Phi}^j\boldsymbol{\varepsilon}_{t-j}\right)=\sum_{j=0}^{\infty}\mathbf{W}'\mathbf{\Phi}^j\mathbf{R}\mathbf{R}'\mathbf{\Phi}^{j\prime}\mathbf{W}.$$

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

$$\left\|Var\left(\sum_{j=0}^{\infty}\mathbf{W}'\mathbf{\Phi}^j\boldsymbol{\varepsilon}_{t-j}\right)\right\|\leq\|\mathbf{W}\|^2\|\mathbf{R}\|^2\sum_{j=0}^{\infty}\|\mathbf{\Phi}\|^{2j}=O\left(N^{-1}\right), \tag{13}$$

where $\sum_{j=0}^{\infty}\|\mathbf{\Phi}\|^{2j}=O(1)$, under Assumption 3, and, as before, $\|\mathbf{W}\|^2=O\left(N^{-1}\right)$ and $\|\mathbf{R}\|^2=O(1)$. Using (13) in (12) and noting that $E\left(\sum_{j=0}^{\infty}\mathbf{W}'\mathbf{\Phi}^j\boldsymbol{\varepsilon}_{t-j}\right)=\mathbf{0}$, we obtain

$$\bar{\mathbf{y}}_{wt}=\mathbf{D}(L)\mathbf{f}_t+O_p\left(N^{-1/2}\right). \tag{14}$$

Note that the coefficient matrices, $\mathbf{D}_j$, in the polynomial $\mathbf{D}(L)$ satisfy $\|\mathbf{D}_j\|=\left\|\mathbf{W}'\mathbf{\Phi}^{j-1}\left(\mathbf{\Phi}\mathbf{\Gamma}_0+\mathbf{\Gamma}_1\right)\right\|\leq\|\mathbf{W}\|\left\|\mathbf{\Phi}^{j-1}\right\|\|\mathbf{\Phi}\mathbf{\Gamma}_0+\mathbf{\Gamma}_1\|=O\left(\varrho^j\right)$, for $j>0$, and are thus declining at an exponential rate. Assuming that the left inverse of $\mathbf{D}(L)$, namely the polynomial $\mathbf{H}(L)=\mathbf{D}^-(L)$ satisfying $\mathbf{H}(L)\mathbf{D}(L)=\mathbf{I}_m$, exists and its coefficients also decline exponentially,[2] we obtain

$$\mathbf{f}_t=\mathbf{H}(L)\bar{\mathbf{y}}_{wt}+O_p\left(N^{-1/2}\right), \tag{15}$$

and the error of approximating $\mathbf{f}_t$ with $\sum_{j=0}^{p}\mathbf{H}_j\bar{y}_{t-j}$ declines exponentially in the truncation lag, p.

Consider now the $n_i\times 1$ vector in the i^{th} partition of $\mathbf{y}_{t+h|t}$ defined by (7), namely

$$\mathbf{y}_{i,t+h|t}=E\left(\mathbf{y}_{i,t+h}|\mathcal{I}_t,\mathcal{F}_t\right)=\mathbf{E}_{i,n}'\mathbf{\Phi}^h\mathbf{y}_t+\mathbf{G}_{hi}\mathbf{f}_t,\text{ for } h=1,2,..., \tag{16}$$

where $\mathbf{E}_{i,n}'$ is an $n_i\times n$ selection matrix defined by $\mathbf{y}_{it}=\mathbf{E}_{i,n}'\mathbf{y}_t$, and $\mathbf{G}_{hi}=\mathbf{E}_{i,n}'\mathbf{G}_h$. Denote the $n_i\times n_j$ sub-matrices of $\mathbf{\Phi}^h$ by $\mathbf{\Phi}_{h,ij}$, for $i,j=1,2,...,N$, and write the first term on the right side

[2] See Lemma A.1 of Chudik and Pesaran (2013) for sufficient conditions on the existence of $\mathbf{H}(L)$ with exponentially declining coefficients when $n^*=m=1$.

of (16) as

$$
\begin{aligned}
\mathbf{E}'_{i,n}\mathbf{\Phi}^h\mathbf{y}_t &= \sum_{j=1}^{N}\mathbf{\Phi}_{h,ij}\mathbf{y}_{jt} = \sum_{j=1,j\neq i}^{N}\mathbf{\Phi}_{h,ij}\mathbf{y}_{jt} + \mathbf{\Phi}_{h,ii}\mathbf{y}_{it}, \\
&= \sum_{j=1,j\neq i}^{N}\mathbf{\Phi}_{h,ij}\mathbf{y}_{jt} + \left(\mathbf{\Phi}_{h,ii} - \mathbf{\Phi}_{ii}^h\right)\mathbf{y}_{it} + \mathbf{\Phi}_{ii}^h\mathbf{y}_{it}, \\
&= \mathbf{\Psi}'_{h,i}\mathbf{y}_t + \mathbf{\Phi}_{ii}^h\mathbf{y}_{it}, \qquad (17)
\end{aligned}
$$

where $\mathbf{\Psi}'_{h,i} = \mathbf{E}'_{i,n}\mathbf{\Phi}^h - \mathbf{\Phi}_{ii}^h\mathbf{E}'_{i,n}$, $\mathbf{\Psi}'_{h,i}\mathbf{y}_t = \sum_{j=1}^{N}\mathbf{\Psi}_{h,ij}\mathbf{y}_{jt}$, and

$$
\mathbf{\Psi}_{h,ij} = \begin{cases} \mathbf{\Phi}_{h,ii} - \mathbf{\Phi}_{ii}^h & \text{for } i = j, \\ \mathbf{\Phi}_{h,ij} & \text{for } i \neq j. \end{cases}
$$

Using (11), which decomposes $\mathbf{y}_t$ into its weakly and strongly cross-sectionally dependent components, we obtain

$$
\mathbf{\Psi}'_{h,i}\mathbf{y}_t = \mathbf{\Psi}'_{h,i}\sum_{j=0}^{\infty}\mathbf{\Phi}^j\boldsymbol{\varepsilon}_{t-j} + \mathbf{\Psi}'_{h,i}\mathbf{Q}(L)\mathbf{f}_t.
$$

Lemma A.1 in the Appendix establishes that there exists a finite constant K_h such that for any $N \in \mathbb{N}$ and any $i = 1, 2, ..., N$ we have $\|\mathbf{\Psi}_{h,ij}\| < K_h/N$, where $|K_h| < \infty$ does not depend on i, j, or N. Since the dimension of $\mathbf{\Psi}_{h,ij}$ is $n_i \times n_j$ and n_i is finite for all $i = 1, 2, ..., N$, it follows from Lemma A.1 that $\left\|\mathbf{\Psi}'_{h,ij}\right\|_\infty = O\left(N^{-1}\right)$ uniformly in i and j, and

$$
\|\mathbf{\Psi}_{h,i}\|_\infty < \frac{K_h}{N}.
$$

We can now use the same arguments as in the derivation of (14) to establish the result $\mathbf{\Psi}'_{h,i}\sum_{j=0}^{\infty}\mathbf{\Phi}^j\boldsymbol{\varepsilon}_{t-j} = O_p\left(N^{-1/2}\right)$, and hence

$$
\mathbf{\Psi}'_{h,i}\mathbf{y}_t = \mathbf{\Xi}_{h,i}(L)\mathbf{f}_t + O_p\left(N^{-1/2}\right), \qquad (18)
$$

where $\mathbf{\Xi}_{h,i}(L) = \mathbf{\Psi}'_{h,i}\mathbf{Q}(L) = \mathbf{\Psi}'_{h,i}\mathbf{\Gamma}_0 + \sum_{j=1}^{\infty}\mathbf{\Psi}'_{h,i}\mathbf{\Phi}^{j-1}(\mathbf{\Phi}\mathbf{\Gamma}_0 + \mathbf{\Gamma}_1)L^j$. Using (17) and (18) in (16) and then substituting (15) yields the following large N representation of $\mathbf{y}_{i,t+h|t}$:

$$
\mathbf{y}_{i,t+h|t} = \mathbf{\Phi}_{ii}^h\mathbf{y}_{it} + \mathbf{C}_{h,i}(L)\bar{\mathbf{y}}_{wt} + O_p\left(N^{-1/2}\right), \qquad (19)
$$

where

$$
\mathbf{C}_{h,i}(L) = \begin{cases} \mathbf{0}, & \text{under Assumption 2.a,} \\ [\mathbf{\Xi}_{h,i}(L) + \mathbf{G}_{h,i}]\mathbf{H}(L), & \text{under Assumption 2.b.} \end{cases} \qquad (20)
$$

In general, $\mathbf{C}_{h,i}(L)$ will be an infinite order polynomial in the lag operator, L. But there are also some special cases under which $\mathbf{C}_{h,i}(L)$ is finite order. In particular, consider the specification M_a (see (1)). In this special case, $\mathbf{\Gamma}_1 = -\mathbf{\Phi}\mathbf{\Gamma}_0$ and the matrix polynomial $\mathbf{C}_{h,i}(L)$ in (19) reduces to

$$\mathbf{C}_{h,i}=\left\{\begin{array}{cc}\mathbf{0}, & \text{under Assumption 2.a}\\ \left(\mathbf{\Gamma}_{0,i}\mathbf{\Pi}^{h}-\mathbf{\Phi}_{ii}^{h}\mathbf{\Gamma}_{0,i}\right)\left(\bar{\mathbf{\Gamma}}_{W,0}^{\prime}\bar{\mathbf{\Gamma}}_{W,0}\right)^{+}\bar{\mathbf{\Gamma}}_{W,0}^{\prime}, & \text{under Assumption 2.b}\end{array}\right. . \tag{21}$$

The following proposition summarizes the main findings of this subsection.

Proposition 1 *Let $\mathbf{y}_t$ and $\mathbf{f}_t$ be generated by (3) with (5), suppose that Assumptions 1, 2.a or 2.b, and 3-4 hold, $\mathbf{W}$ is any arbitrary vector of weights satisfying (8) and (9), and the left inverse of $\mathbf{D}(L)$, denoted by $\mathbf{H}(L)=\mathbf{D}^{-}(L)$ exists. Then for any cross-section unit $i\in\mathbb{N}$, and a given forecasting $h>0$, the optimal forecast of $\mathbf{y}_{i,t+h}$, defined in (7) has a large N representation given by (19).*

Comparing (20) with (21), we see that the former involves an infinite order lag distribution in cross-section averages that need to be truncated, whereas under the latter special case, only contemporaneous values of cross-section averages are included. In practice, where the nature of factors and how they enter the VAR model are unknown, 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 $\mathbf{y}_{it}$ and $\bar{\mathbf{y}}_{wt}$ 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, cross-region or market interactions. Suppose that the endogenous variables specific to unit i, the $n_i\times 1$ 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 n endogenous variables, $\mathbf{y}_t=(\mathbf{y}_{1t}^{\prime},\mathbf{y}_{2t}^{\prime},...,\mathbf{y}_{Nt}^{\prime})^{\prime}$, 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 (say country), taking the cross-section averages (say country-specific weighted averages of foreign variables, also commonly referred to as 'star' variables) 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 is routinely tested and generally accepted in empirical applications of the GVAR.

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

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

for $i=1,2,...,N$, where $\overline{\mathbf{y}}_{wit}=\mathbf{W}_i^{\prime}\mathbf{y}_t$ is an $n_i^*\times 1$ vector of cross-section averages specific to unit i, and $\mathbf{W}_i$ is an $n\times n_i^*$ matrix of unit-specific weights that define the n_i^* cross-section averages. As before, 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

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

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

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

Recognizing that $\overline{\mathbf{y}}_{\mathcal{W}t} = \mathcal{W}\mathbf{y}_t$, where $\mathcal{W} = (\mathbf{W}_1, \mathbf{W}_2, ..., \mathbf{W}_N)'$, (23) can be written as

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

where

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

Finally, provided that $\mathbf{G}_0$ is invertible, we can multiply (24) by $\mathbf{G}_0^{-1}$ from the left to obtain the following high-dimensional VAR 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 $\mathbf{u}_t = \mathbf{G}_0^{-1}\boldsymbol{\xi}_t$.

In forecasting with a GVAR, it is important to distinguish between the case where (26) is the data generating process (DGP) from the case where (3) is the DGP, and the GVAR is used as an approximating model. In the former case, where the DGP is (26), forecasting is straightforward. But in the latter case, where (3) is the DGP, there is no reason to believe that the inverse of $\mathbf{G}_0$ exists when the unit-specific variables are affected by unobserved common factors. Even if the estimate of $\mathbf{G}_0$ is not rank deficient, the singularity of $\mathbf{G}_0$ can have adverse effects on the forecasting performance and could deteriorate with N.

4.1 Rank deficient case

The rank deficiency of $\mathbf{G}_0$ arises due to the external nature of the factors under specification (3). This can be seen readily in the following simple example which is reminiscent of the capital asset pricing model (CAPM).

Example 1 *Suppose that variable y_{it}, for $i = 1, 2, ..., N$, follows the static one-factor model*

$$y_{it} = \gamma_i f_t + \varepsilon_{it},\ f_t = \rho f_{t-1} + v_t. \tag{27}$$

The infeasible forecasts are $y_{i,t+h|t} = E\left(y_{i,t+h} | \mathcal{I}_t, \mathcal{F}_t\right) = \gamma_i \rho^h f_t$, for $i = 1, 2, ..., N$ and $h = 1, 2, ...$ Applying the GVAR approach to the above, using equal weights, we have

$$y_{it} = b_i \bar{y}_t + \xi_{it},\ \text{for } i = 1, 2, ..., N, \tag{28}$$

where $\bar{y}_t = N^{-1} \sum_{i=1}^{N} y_{it}$, $b_i = \gamma_i / \bar{\gamma}$, $\bar{\gamma} = N^{-1} \sum_{i=1}^{N} \gamma_i$, $\xi_{it} = \varepsilon_{it} - b_i \bar{\varepsilon}_t$, and $\bar{\varepsilon}_t = N^{-1} \sum_{i=1}^{N} \varepsilon_{it}$. If y_{it} was the return for asset i and $\bar{y}_t$ was the market return, then (28) represents the standard CAPM. Writing (28) in vector notations, we obtain the corresponding GVAR model

$$\mathbf{G}_0 \mathbf{y}_t = \boldsymbol{\xi}_t, \tag{29}$$

where $\boldsymbol{\xi}_t = (\xi_{1t}, \xi_{2t}, ..., \xi_{Nt})'$, $\mathbf{G}_0 = \mathbf{I}_N - \mathbf{b}\boldsymbol{\tau}'_N / N$, $\mathbf{b} = (b_1, b_2, ..., b_N)'$ and $\boldsymbol{\tau}_N$ is an $N \times 1$ vector of ones. $\mathbf{G}_0$ is rank deficient since $\boldsymbol{\tau}'_N \mathbf{G}_0 = \mathbf{0}$, noting that $N^{-1} \sum_{i=1}^{N} b_i = N^{-1} \sum_{i=1}^{N} \gamma_i / \bar{\gamma} = 1$.

Returning to the dynamic setting of model (1), we also assume that

ASSUMPTION 5 *$\bar{\mathbf{\Gamma}}_{W,0}$ is full column rank for all N, and $\lim_{N\to\infty} \bar{\mathbf{\Gamma}}_{W,0}$ is full column rank.*

Following a similar line of argument as in Chudik and Pesaran (2011), using (1) and (5), and under Assumptions 1, 2.b, 3-5 we obtain the following unit-specific equations:

$$\mathbf{y}_{it} = \mathbf{\Phi}_{ii} \mathbf{y}_{i,t-1} + \mathbf{B}_{i0} \bar{\mathbf{y}}_{wt} + \mathbf{B}_{i1} \bar{\mathbf{y}}_{w,t-1} + \boldsymbol{\xi}_{it}, \text{ for } i = 1, 2, ..., N, \tag{30}$$

where we have used the same weights in the computation of cross-section averages for all units ($\mathbf{W}_i = \mathbf{W}$ for all i). It is also easily seen that

$$\mathbf{B}_{i0} = \mathbf{\Gamma}_{0,i} \left(\bar{\mathbf{\Gamma}}'_{W,0} \bar{\mathbf{\Gamma}}_{W,0}\right)^{-1} \bar{\mathbf{\Gamma}}'_{W,0},\ \mathbf{B}_{i1} = -\mathbf{\Phi}_{ii} \mathbf{B}_{i0}, \tag{31}$$

and

$$\boldsymbol{\xi}_{it} = \boldsymbol{\varepsilon}_{it} + O_p\left(N^{-1/2}\right).$$

Chudik and Pesaran (2011) establish that the least squares estimates of (30) are consistent and asymptotically normally distributed. Stacking (30) yields the system (24) with the matrix $\mathbf{G}_0$ given by $\mathbf{I}_n - \mathbf{B}_0 \mathcal{W}$ which is rank deficient. The rank deficiency follows by observing that the weighted cross-section averages of $\mathbf{B}_{i0}$ sum to an identity matrix, and therefore the rows of $\mathbf{G}_0$ are linearly dependent, namely

$$\begin{aligned}
\mathbf{W}' \mathbf{G}_0 &= \mathbf{W}' \left(\mathbf{I}_n - \mathbf{B}_0 \mathcal{W}\right), \\
&= \mathbf{W}' - \left(\mathbf{W}' \mathbf{\Gamma}_0\right) \left(\bar{\mathbf{\Gamma}}'_{W,0} \bar{\mathbf{\Gamma}}_{W,0}\right)^{-1} \bar{\mathbf{\Gamma}}'_{W,0} \mathbf{W}', \\
&= \left[\mathbf{I}_m - \bar{\mathbf{\Gamma}}'_{W,0} \left(\bar{\mathbf{\Gamma}}'_{W,0} \bar{\mathbf{\Gamma}}_{W,0}\right)^{-1} \bar{\mathbf{\Gamma}}'_{W,0}\right] \mathbf{W}'.
\end{aligned}$$

But in the case where $\bar{\mathbf{\Gamma}}_{W,0}$ is full column rank, we have $\bar{\mathbf{\Gamma}}'_{W,0}\left(\bar{\mathbf{\Gamma}}'_{W,0}\bar{\mathbf{\Gamma}}_{W,0}\right)^{-1}\bar{\mathbf{\Gamma}}'_{W,0}=\mathbf{I}_m$, and

$$\mathbf{W}'\mathbf{G}_0=\left(\mathbf{I}_m-\mathbf{I}_m\right)\mathbf{W}'=\mathbf{0}. \tag{32}$$

Rank deficiency can also be illustrated in the case of the more general model, (3), by basing the analysis on CALS regressions with appropriately truncated lags as considered in Chudik and Pesaran (2015a). This is not pursued here for the sake of expositional brevity.

4.1.1 Solution sets in the 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, we consider the following simple GVAR model:

$$\mathbf{y}_{it}=\mathbf{B}_{i0}\bar{\mathbf{y}}_{w,it}+\boldsymbol{\varepsilon}_{it},\ \text{for } i=1,2,...,N, \tag{33}$$

where $\bar{\mathbf{y}}_{w,it}=\mathbf{W}_i'\mathbf{y}_t$. As before, $\mathbf{B}_0$ is the $n\times n$ block-diagonal matrix defined by $\mathbf{B}_0=diag\left(\mathbf{B}_{1,0},\mathbf{B}_{2,0},...,\mathbf{B}_{N,0}\right)$, and $\mathcal{W}=\left(\mathbf{W}_1,\mathbf{W}_2,...,\mathbf{W}_N\right)'$. Write (33) as

$$\mathbf{y}_t=\mathbf{B}_0\mathcal{W}\mathbf{y}_t+\boldsymbol{\varepsilon}_t,$$

or

$$\mathbf{G}_0\mathbf{y}_t=\boldsymbol{\varepsilon}_t, \tag{34}$$

where $\mathbf{G}_0=\mathbf{I}_n-\mathbf{B}_0\mathcal{W}$. Suppose that $\mathbf{G}_0$ is rank deficient, namely $rank\left(\mathbf{G}_0\right)=n-m$, for some $m>0$. Then solutions to (34) exist 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, (34) admits multiple solutions with all its possible solutions characterized by

$$\mathbf{y}_t=\mathbf{\Gamma}\tilde{\mathbf{f}}_t+\mathbf{G}_0^+\boldsymbol{\varepsilon}_t, \tag{35}$$

where $\tilde{\mathbf{f}}_t$ is a vector of m arbitrary stochastic processes, $\mathbf{\Gamma}$ is an $n\times m$ 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 (35) maps all possible solutions of (34), note that $\mathbf{G}_0^+\boldsymbol{\varepsilon}_t$ is the particular solution of (34) and $\mathbf{\Gamma}'\tilde{\mathbf{f}}_t$ is a general solution of the homogenous counterpart of (34), 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 (35) 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 $Var(\mathbf{f}_t) = \mathbf{I}_m$, 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 (33), 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

Suppose now $\mathbf{G}_0$ is known to be rank deficient with rank $N-m$, $m > 0$. Then the GVAR model (33) would need to be augmented by at least 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. Different options could be considered for the augmentation of (30). We consider augmenting the set of conditional equations in (30) with the following VAR model in the n^* ($\geq m$) cross-section averages:

$$\bar{\mathbf{y}}_{wt} = \mathbf{\Pi}_{\bar{y}}\bar{y}_{w,t-1} + \boldsymbol{\xi}_{\bar{y}t}. \tag{36}$$

As before, $\bar{\mathbf{y}}_{wt}$ is viewed as a proxy for the unobserved common factors (up to a rotation matrix). See (14), and note that $\mathbf{D}(L) = \bar{\mathbf{\Gamma}}_{W,0}$ under M_a, namely

$$\bar{\mathbf{y}}_{wt} = \bar{\mathbf{\Gamma}}_{W,0}\mathbf{f}_t + O_p\left(N^{-1/2}\right). \tag{37}$$

Using (5) in (37), we have

$$\bar{\mathbf{y}}_{wt} = \bar{\mathbf{\Gamma}}_{W,0}\left(\mathbf{\Pi}\mathbf{f}_{t-1} + \mathbf{v}_t\right) + O_p\left(N^{-1/2}\right). \tag{38}$$

Under the full column rank of $\bar{\mathbf{\Gamma}}_{W,0}$, (37) for period $t-1$ can be multiplied by $\left(\bar{\mathbf{\Gamma}}'_{W,0}\bar{\mathbf{\Gamma}}_{W,0}\right)^+\bar{\mathbf{\Gamma}}'_{W,0}$ from the left to obtain

$$\mathbf{f}_{t-1} = \left(\bar{\mathbf{\Gamma}}'_{W,0}\bar{\mathbf{\Gamma}}_{W,0}\right)^+\bar{\mathbf{\Gamma}}'_{W,0}\bar{\mathbf{y}}_{w,t-1} + O_p\left(N^{-1/2}\right). \tag{39}$$

Substituting (39) in (38), we obtain (36) with

$$\mathbf{\Pi}_{\bar{y}} = \bar{\mathbf{\Gamma}}_{W,0}\mathbf{\Pi}\left(\bar{\mathbf{\Gamma}}'_{W,0}\bar{\mathbf{\Gamma}}_{W,0}\right)^+\bar{\mathbf{\Gamma}}'_{W,0}, \tag{40}$$

and

$$\boldsymbol{\xi}_{\bar{y}t} = \bar{\mathbf{\Gamma}}_{W,0}\mathbf{v}_t + O_p\left(N^{-1/2}\right).$$

In practice, higher order VARs in $\bar{\mathbf{y}}_{wt}$ can also be considered. Stacking (30) and (36), we obtain the following augmented VAR model in $\mathbf{z}_t = (\mathbf{y}_t', \bar{\mathbf{y}}_{wt}')'$:

$$\mathbf{A}_0\mathbf{z}_t = \mathbf{A}_1\mathbf{z}_{t-1} + \mathbf{e}_{zt}, \tag{41}$$

where $\mathbf{e}_{zt} = \left(\boldsymbol{\xi}_t', \boldsymbol{\xi}_{\bar{y}t}\right)' = \left(\boldsymbol{\varepsilon}_t', \mathbf{v}_t'\bar{\boldsymbol{\Gamma}}_{W,0}'\right)' + O_p\left(N^{-1/2}\right)$,

$$\mathbf{A}_0 = \begin{pmatrix} \mathbf{I}_N & -\mathring{\mathbf{B}}_0 \\ \mathbf{0} & \mathbf{I}_{n^*} \end{pmatrix}, \ \mathbf{A}_1 = \begin{pmatrix} \boldsymbol{\Theta} & \mathring{\mathbf{B}}_1 \\ \mathbf{0} & \boldsymbol{\Pi}_{\bar{y}} \end{pmatrix}, \tag{42}$$

$\mathring{\mathbf{B}}_s = \left(\mathbf{B}_{s,1}', \mathbf{B}_{s,2}', ..., \mathbf{B}_{s,N}'\right)'$ for $s = 0, 1$, and $\boldsymbol{\Theta}$ is an $n \times n$ block-diagonal matrix with matrices $\boldsymbol{\Phi}_{ii}$, for $i = 1, 2, ..., N$, on its diagonal. Let $\mathbf{A} = \mathbf{A}_0^{-1}\mathbf{A}_1$, noting that $\mathbf{A}_0$ is invertible, and using $\mathring{\mathbf{B}}_1 = -\boldsymbol{\Theta}\mathring{\mathbf{B}}_0$ (see (31)) we have

$$\mathbf{A}^h = \begin{pmatrix} \boldsymbol{\Theta}^h & \mathring{\mathbf{B}}_0\boldsymbol{\Pi}_{\bar{y}}^h - \boldsymbol{\Theta}^h\mathring{\mathbf{B}}_0 \\ 0 & \boldsymbol{\Pi}_{\bar{y}}^h \end{pmatrix} \text{ for } h = 1, 2, ... \tag{43}$$

The AugGVAR forecast $\mathbf{y}_{i,t+h|t}^{aug}$ is then given by

$$\mathbf{y}_{i,t+h|t}^{aug} = \mathbf{E}_{i,n+n^*}'\mathbf{A}^h\mathbf{z}_t, \tag{44}$$

where $\mathbf{E}_{i,n+n^*}'$ is an $n_i \times (n + n_i)$ selection matrix that selects variables for the cross-section unit i, namely $\mathbf{E}_{i,n+n^*}'\mathbf{z}_t = \mathbf{y}_{it}$.

The following proposition relates the AugGVAR forecast $\mathbf{y}_{i,t+h|t}^{aug}$ given by (44) and the infeasible optimal forecast $\mathbf{y}_{i,t+h|t}$ given by (7).

Proposition 2 *Let $\mathbf{y}_t$ be generated by (1), Assumptions 1, 2.a or 2.b, and 3-5 hold, and $\mathbf{W}$ be any arbitrary $n \times n^*$ matrix of weights satisfying (8) and (9). Then for any cross-section unit $i \in \mathbb{N}$, and any given forecasting horizon $h > 0$, $\mathbf{y}_{i,t+h|t}^{aug}$ defined in (44) is consistent, in the sense that $\left\|\mathbf{y}_{i,t+h|t}^{aug} - \mathbf{y}_{i,t+h|t}\right\|_\infty \xrightarrow{p} 0$, as $N \to \infty$.*

$\mathbf{y}_{i,t+h|t}^{aug}$ is still an infeasible forecast since the parameters in (44) are unknown and need to be estimated. To this end we consider estimation of GVAR forecast $\mathbf{y}_{i,T+h|T}$ and AugGVAR forecast $\mathbf{y}_{i,T+h|T}^{aug}$ by using least squares estimates of parameters of the conditional cross-section augmented models (30) and (in the case of the AugGVAR only) also the marginal model (36). Namely, we define

$$\hat{\mathbf{y}}_{i,T+h|T} = \mathbf{E}_{i,n}'\widehat{\mathbf{G}}^h\mathbf{y}_T, \tag{45}$$

and

$$\hat{\mathbf{y}}_{i,T+h|T}^{aug} = \mathbf{E}_{i,n+n^*}'\hat{\mathbf{A}}^h\mathbf{z}_T, \tag{46}$$

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 (30) and (36).

We collect the individual forecasts in the vectors $\hat{\mathbf{y}}_{T+h|T} = \left(\hat{\mathbf{y}}'_{1,T+h|T}, \hat{\mathbf{y}}'_{2,T+h|T}, ..., \hat{\mathbf{y}}'_{N,T+h|T} \right)'$ and $\hat{\mathbf{y}}^{aug}_{T+h|T} = \left(\hat{\mathbf{y}}^{aug\prime}_{1,T+h|T}, \hat{\mathbf{y}}^{aug\prime}_{2,T+h|T}, ..., \hat{\mathbf{y}}^{aug\prime}_{N,T+h|T} \right)'$. We investigate the asymptotic properties of $\hat{\mathbf{y}}^{aug}_{T+h|T}$ in the case where the underlying VAR model contains unobserved common factors.

Theorem 1 *Suppose $\mathbf{y}_t$ is generated by model (1), $\mathbf{W}$ is any arbitrary $n \times n^*$ matrix of weights satisfying conditions (8) and (9), Assumptions 1, 2.b (model featuring unobserved common factors), and 3-5 hold, and $N, T \overset{j}{\to} \infty$. Then for any given $h > 0$, the h-step-ahead forecast $\hat{\mathbf{y}}^{aug}_{T+h|T}$ defined by (46) satisfies*

$$\left\| \mathbf{y}_{T+h|T} - \hat{\mathbf{y}}^{aug}_{T+h|T} \right\|_{\infty} \overset{L_1}{\to} 0. \tag{47}$$

Moreover, matrix $\mathbf{G}_0$ defined in (25) is singular for any $N \in \mathbb{N}$.

The proof is provided in the Appendix.

Remark 3 *We do not require any restrictions on the relative expansion rates of N and T for consistency of the feasible AugGVAR forecasts in Theorem 1. Restrictions on the ratio N/T will be required for conducting inference on the individual parameters of the country-specific regressions (30). See Chudik and Pesaran (2011) for further details.*

Remark 4 *The consistency of the feasible AugGVAR forecasts can be also established in the case of model (3) without imposing $\mathbf{\Gamma}_1 = -\mathbf{\Phi}\mathbf{\Gamma}_0$. Additional conditions will be required in this case. Similar to Proposition 1, we will need to assume that the left inverse of $\mathbf{D}(L)$, denoted by $\mathbf{H}(L) = \mathbf{D}^{-}(L)$, exists, and its coefficients decay at an exponential rate. To approximate distributed lag relations of infinite order, $p = p(T)$ lags of cross-section averages will be required in the conditional unit-specific regressions in (30) and in the marginal model (36), where p will need to increase in T so that the error from omitting higher order lags will become asymptotically negligible, but not too fast so that there are still sufficient degrees of freedom for consistent estimation. As shown in Chudik and Pesaran (2015a), a sufficient rate for p_T is $p_T/T^3 \to \kappa$ for some $0 < \kappa < \infty$. In the next section, we will investigate the small sample performance of GVAR and AugGVAR forecasts based on p lags for cross-section averages, with p chosen to be the integer part of $T^{1/3}$.*

Instead of predetermined 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.

It is also worth bearing in mind that the GVAR has its own logic and structure and need not be related to a particular factor-augmented high-dimensional VAR. The use of country-specific weights in construction of foreign-specific (or star) variables in the GVAR approach also allows for inclusion of local or spill-over effects which is not typically addressed in standard factor-augmented VAR models. Finally, using cross-sectional averages computed using different variables in $\mathbf{y}_{it}$, is not necessarily the same as using different weighted averages of the same set of variables, which is the approach followed when one uses principle components.

5 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.

5.1 MC design

Three DGPs are considered: a high-dimensional VAR model without a common factor, and two high-dimensional VARs featuring $m = 1$ and $m = 2$ common factors. 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 no factors, but allows for weak cross-sectional dependence of errors. We consider two variables per cross-section unit i, $\mathbf{y}_{it} = (y_{i1}, y_{i2})'$ for $i = 1, 2, ..., N$. The vector of all $n = 2N$ variables $\mathbf{y}_t = (\mathbf{y}'_{1t}, \mathbf{y}'_{2t}, ..., \mathbf{y}'_{Nt})'$ for $t = -M + 1, ..., 0, 1, 2, ..., T$ is generated as

$$\mathbf{y}_t = \mathbf{\Phi y}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{48}$$

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. The $2N \times 2N$ matrix $\mathbf{\Phi}$ is partitioned into N^2 2×2 dimensional sub-matrices denoted by $\mathbf{\Phi}_{ij}$, for $i, j = 1, 2, ..., N$. The diagonal sub-matrices, $\mathbf{\Phi}_{ii}$, are generated randomly as

$$\mathbf{\Phi}_{ii} = \begin{pmatrix} \kappa_{\phi i1} & \delta_{\phi i1} \\ \delta_{\phi i2} & \kappa_{\phi i2} \end{pmatrix}, \text{ for } i = 1, 2, ..., N,$$

where $\kappa_{\phi is} \sim IIDU\,(0, 0.7)$ and $\delta_{\phi is} \sim IIDU\,(0, 0.7 - \kappa_{\phi is})$ for $s = 1, 2$. For the off-diagonal sub-matrices, $\mathbf{\Phi}_{ij}$, for $i \neq j$, we consider two options. Under the first option, we set $\mathbf{\Phi}_{ij} = \mathbf{0}$ for all $i \neq j$. Under the second option, we generate the elements of $\mathbf{\Phi}_{ij}$ sub-matrices ($i \neq j$) as $\phi_{rs} = \lambda_r \omega_{rs}$, where $\lambda_r \sim IIDU\,(-0.2, 0.2)$ and $\omega_{rs} = \varsigma_{rs} / \sum_{s=1}^{n} \varsigma_{rs}$, with $\varsigma_{rs} \sim IIDU\,(0, 1)$. Under both options the underlying DGP is stationary (for any $N \in \mathbb{N}$) and the spectral norm of $\mathbf{\Phi}$ is less than one, on average.

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

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

where $\boldsymbol{\eta}_t = \left(\eta_{1t}, \eta_{2t}, ..., \eta_{n,t}\right)'$, $\boldsymbol{\eta}_t \sim IIDN\left(\mathbf{0}, \sigma_\eta^2 \mathbf{I}_n\right)$, 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 $n^{-1}\sum_{r=1}^{n} Var\left(\varepsilon_{rt}\right) = 1$.

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

$$\mathbf{y}_t - \mathbf{\Gamma}\mathbf{f}_t = \mathbf{\Phi}\left(\mathbf{y}_{t-1} - \mathbf{\Gamma}\mathbf{f}_{t-1}\right) + \boldsymbol{\varepsilon}_t, \tag{49}$$

where $\mathbf{f}_t = (f_{1t}, f_{2t}, ..., f_{mt})'$, and

$$f_{jt} = \rho_j f_{j,t-1} + \left(1-\rho_j^2\right)^{1/2} v_{jt}, \text{ for } j = 1, 2,$$

with the starting values $\mathbf{y}_{-M} = \mathbf{0}$, $\mathbf{f}_{-M} = \mathbf{0}$. 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 consider two options for the number of common factors, $m = 1$ and 2. We set $\rho_1 = 0.8$, $\rho_2 = 0.7$ and generate v_{jt} as $N(0,1)$. Factor loadings are generated as $\gamma_{ij} \sim IIDN\left(\gamma_j, \sigma_{j\gamma}^2\right)$ with $\gamma_j = \gamma = 1$ and $\sigma_{\gamma j} = \sigma_\gamma = 0.2$.

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

$$\mathbf{y}_t = \mathbf{\Phi}\mathbf{y}_{t-1} + \mathbf{\Gamma}\mathbf{f}_t + \boldsymbol{\varepsilon}_t, \tag{50}$$

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 $\gamma_{ij} \sim IIDN\left(\gamma_j, \sigma_{\gamma j}^2\right)$, with $\sigma_{\gamma j} = \sigma_\gamma = 0.2$ and γ_j is set to ensure that $N^{-1}\boldsymbol{\tau}_j'\left(\mathbf{I} - \mathbf{\Phi}\right)^{-1}\boldsymbol{\tau}_j\gamma_j = 1$, for $j = 1, 2$, where $\boldsymbol{\tau}_j$ is an $n \times 1$ vector with odd (for $j = 1$) or even (for $j = 2$) elements set equal to one and the remaining elements are set to zero. In this notation, $N^{-1}\boldsymbol{\tau}_j'\mathbf{y}_t = N^{-1}\sum_{i=1}^{N} y_{ijt} = \bar{y}_{jt}$, for $j = 1, 2$, where y_{ijt} is the j-th element of $\mathbf{y}_{it}$.

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

5.2 Individual forecasts and average MSFEs

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:

$$MSFE_{RN}\left(T+1|\,T\right) = \frac{\sum_{r=1}^{R}\sum_{i=1}^{N}\left\|\hat{\mathbf{y}}_{i,T+1|T}^{(r)} - \mathbf{y}_{i,T+1}^{(r)}\right\|^2}{\sum_{r=1}^{R}\sum_{i=1}^{N}\left\|E\left(\mathbf{y}_{i,T+1}^{(r)}\middle|\,\mathcal{I}_t^{(r)},\mathcal{F}_t^{(r)}\right) - \mathbf{y}_{i,T+1}^{(r)}\right\|^2}, \tag{51}$$

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 $\mathbf{y}_{i,T+1}^{(r)}$ is the realized vector of observations for cross-section unit i, at time $T+1$, and the Monte Carlo replication, r. Similarly, we compute the MSFE for AugGVAR forecasts $\hat{\mathbf{y}}_{i,T+1|T}^{aug}$. The optimal infeasible one-step-ahead forecasts are computed as (we are dropping the superscript (r) to simplify the notations)

$$E\left(\mathbf{y}_{i,T+1}|\,\mathcal{I}_T,\mathcal{F}_T\right) = \begin{cases} \mathbf{E}_{i,n}'\mathbf{\Phi}\mathbf{y}_T, & \text{in the case of DGP1} \\ \mathbf{\Gamma}_i\mathbf{\Pi}\mathbf{f}_T + \mathbf{E}_{i,n}'\mathbf{\Phi}\left(\mathbf{y}_T - \mathbf{\Gamma}\mathbf{f}_T\right), & \text{in the case of DGP2} \\ \mathbf{\Gamma}_i\mathbf{\Pi}\mathbf{f}_T + \mathbf{E}_{i,n}'\mathbf{\Phi}\mathbf{y}_T, & \text{in the case of DGP3} \end{cases}, \tag{52}$$

where $\mathbf{\Gamma}$ is partitioned into a set of N sub-matrices of dimension $2\times m$ as $\mathbf{\Gamma} = (\mathbf{\Gamma}_1', \mathbf{\Gamma}_2', ..., \mathbf{\Gamma}_N')'$, $\mathbf{E}_{i,n}'$ is a $2\times n$ selection matrix that selects the 2×1 variable vector for the i^{th} unit, namely $\mathbf{y}_{it} = \mathbf{E}_{i,n}'\,\mathbf{y}_t$, and $\mathbf{\Pi}$ is an $m\times m$ diagonal matrix with elements ρ_j, $j = 1,2,...,m$, on the diagonal. The non-augmented GVAR forecasts ($\hat{\mathbf{y}}_{\,i,T+1|T}$) are based on the stacked version of the following regressions:

$$\mathbf{y}_{it} = \mathbf{c}_i + \mathbf{\Phi}_{ii}\mathbf{y}_{i,t-1} + \sum_{\ell=0}^{p}\mathbf{B}_{i\ell}\overline{\mathbf{y}}_{wi,t-\ell} + \boldsymbol{\xi}_{it}, \text{ for } i = 1,2,...,N, \tag{53}$$

where $\overline{\mathbf{y}}_{wit} = \sum_{j=1}^{N} w_{ij}\mathbf{y}_{jt}$. Aggregation weights are such that $\overline{\mathbf{y}}_{wit}$ is a simple cross-section average of units that do not directly enter individual cross-section augmented regressions in (53). In particular, $w_{ii} = 0$ and $w_{ij} = (N-1)^{-1}$ for $i \neq j$. Let $\mathcal{W} = (\mathbf{W}_1, \mathbf{W}_2, ..., \mathbf{W}_N)'$, where $\mathbf{W}_i$ is an $n\times 2$ weight matrix defined by $\overline{\mathbf{y}}_{wit} = \mathbf{W}_i'\mathbf{y}_t$. Moreover, let

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

where $\hat{\mathbf{B}}_{i\ell}$ is the least squares estimate of $\mathbf{B}_{i\ell}$. The estimated (non-augmented) GVAR representation is

$$\mathbf{y}_t = \hat{\boldsymbol{\delta}} + \sum_{\ell=1}^{p}\hat{\mathbf{\Psi}}_\ell\mathbf{y}_{t-\ell} + \hat{\mathbf{u}}_t, \tag{54}$$

which yields the GVAR forecasts

$$\hat{\mathbf{y}}_{T+1|T} = \sum_{\ell=1}^{p} \hat{\mathbf{\Psi}}_\ell \mathbf{y}_{T+1-\ell} + \hat{\boldsymbol{\delta}}, \tag{55}$$

where $\hat{\mathbf{\Psi}}_\ell = \hat{\mathbf{G}}_0^{-1}\hat{\mathbf{G}}_\ell$, for $\ell = 1, 2, ..., p$, $\hat{\boldsymbol{\delta}} = \hat{\mathbf{G}}_0^{-1}\hat{\mathbf{c}}$, $\hat{\mathbf{G}}_0 = \mathbf{I}_n - \hat{\mathbf{B}}_0\mathcal{W}$, $\hat{\mathbf{G}}_1 = \hat{\mathbf{\Theta}} + \hat{\mathbf{B}}_1\mathcal{W}$, $\hat{\mathbf{G}}_\ell = \hat{\mathbf{B}}_\ell\mathcal{W}$, for $\ell = 2, 3, ..., p$, $\hat{\mathbf{\Theta}}$ is a block-diagonal matrix constructed based on the estimates of $\mathbf{\Phi}_{ii}$ in (53), $\hat{\mathbf{c}} = (\hat{\mathbf{c}}_1', \hat{\mathbf{c}}_2', ..., \hat{\mathbf{c}}_N')'$ is the vector of estimated fixed effects in (53), and $\hat{\mathbf{u}}_t = \hat{\mathbf{G}}_0^{-1}\hat{\boldsymbol{\xi}}_t$.

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 unit-specific conditional models are estimated by least squares:

$$\mathbf{y}_{it} = \mathbf{c}_i + \mathbf{\Phi}_{ii}\mathbf{y}_{i,t-1} + \sum_{\ell=0}^{p} \mathbf{B}_{i\ell}\bar{\mathbf{y}}_{t-\ell} + \boldsymbol{\xi}_{it}, \tag{56}$$

and augmented with the estimates of the VAR(p) model in $\bar{\mathbf{y}}_t$

$$\bar{\mathbf{y}}_t = \mathbf{c}_{\bar{y}} + \sum_{\ell=1}^{p} \mathbf{\Pi}_{\bar{y}\ell}\bar{\mathbf{y}}_{t-\ell} + \boldsymbol{\xi}_{\bar{y}t}, \tag{57}$$

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

$$\hat{\mathbf{y}}^{aug}_{i,T+1|T} = \mathbf{E}'_{i,n+2}\hat{\mathbf{z}}_{T+1} = \mathbf{E}'_{i,n+2}\left(\hat{\boldsymbol{\delta}} + \sum_{\ell=1}^{p} \hat{\mathbf{\Upsilon}}_\ell \mathbf{z}_{T-\ell}\right), \tag{58}$$

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

$$\hat{\mathbf{A}}_0 = \begin{pmatrix} \mathbf{I}_N & -\hat{\mathbf{B}}_0 \\ \mathbf{0}_{2\times k} & \mathbf{I}_2 \end{pmatrix}, \ \hat{\mathbf{A}}_1 = \begin{pmatrix} \hat{\mathbf{\Theta}} & \hat{\mathbf{B}}_1 \\ \mathbf{0}_{2\times k} & \hat{\mathbf{\Pi}}_{\bar{y}1} \end{pmatrix}, \text{ and } \hat{\mathbf{A}}_\ell = \begin{pmatrix} \mathbf{0}_{k\times k} & \hat{\mathbf{B}}_\ell \\ \mathbf{0}_{2\times k} & \hat{\mathbf{\Pi}}_{\bar{y}\ell} \end{pmatrix}, \text{ for } \ell = 2, 3, ..., p,$$

in which all estimated coefficients are based on (56)-(57).

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.

5.3 Monte Carlo results

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 for different specifications DGP1-DGP3 set out above. 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 performing marginally better. It is interesting to observe that the augmentation with additional equations for cross-section averages, although asymptotically redundant in the case of this DGP, does not unduly 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 $m = 1$ and $m = 2$ unobserved common factors. AugGVAR forecasts are not affected by the inclusion of the factors, and their performance is generally similar to those reported at 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 factors are introduced in the analysis. However, the performance of the GVAR without augmentation deteriorates considerably with the introduction of unobserved common factors, especially when T is small and N large. This finding is in line with our theoretical result which suggests that in the presence of common factors, the contemporaneous matrix $\mathbf{G}_0$ becomes singular. The results clearly illustrate that the AugGVAR performs well, irrespective of whether the underlying VAR contains unobserved common factors or not. Also, when factors are included, the results are robust to the way the factors are introduced in the underlying VAR model.[3]

Table 2 reports the findings for the experiments with a high spatial coefficient for the idiosyncratic errors (namely $\varrho_\varepsilon = 0.6$) and the non-sparse coefficient matrix $\mathbf{\Phi}$. These findings are qualitatively similar to Table 1, confirming the robustness of the AugGVAR approach and the singularity problem of the GVAR forecasts when at least one strong unobserved common factor is present.

6 Empirical application: forecasting GDP using PMIs

In this section we apply a number of different methods for the analysis of large datasets, 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.

6.1 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 we also have a good country coverage for the PMI data. The latest available observation on output growth is 2013Q2. All of the output data is seasonally adjusted, most series by the source. 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.

[3] The bias component of the mean square forecast errors of the GVAR and AugGVAR forecasts is very small (in most cases less than 7%) in all the experiments.

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.

A detailed description of the sources and construction of the output and PMI data and plots of the data are presented in the Appendix of Chudik et al. (2014).

6.1.1 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 notation, but it will be understood that all forecasts are conditional on the nonsynchronous information set $\Omega_{\mathfrak{m}t}$.

6.2 Forecasting methods

We consider three basic benchmark forecasts: 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}; and an extension of the AR(1) benchmark where the AR(1) model is augmented with domestic PMIs.

These benchmarks are not subject to the dimensionality problem since they do not utilize the foreign variables. In addition to the benchmark methods, we consider seven data-rich methods: GVAR; AugGVAR; Lasso; Ridge; factor model (FM), factor-augmented AR model (FAR), and partial least squares (PLS). In all data-rich methods, we consider forecasting with output data alone, and with en extended information set that includes the output and PMI data. Forecasts using the latter extended information set are denoted with the suffix PMI. We refer the reader to Chudik et al. (2014) for a detailed description of individual methods and related references in order to economize on space.

In the implementation of the individual forecasting methods, we have to overcome two challenges: the discrepancy in frequency of (quarterly) output data and (monthly) PMI data; and the timing of data releases that differ 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.[4] We refer to the second challenge as "nonsynchronous conditioning" information sets. We follow a simple solution of transforming monthly data into quarterly observations as opposed to developing a full-fledged mixed-frequency model (such as the 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. (2013), Marcellino and Schumacher (2010), and Kuzin et al. (2011 and 2013)). In particular, we consider two ways of transforming monthly observations into a quarterly series: temporal aggregation (we use three-month non-overlapping averages) and sequential sampling (we sample every third observation). 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. Dealing with different frequencies and the nonsynchronous conditioning sets is not a central contribution of this paper, and we provide a detailed discussion of these issues in Chudik et al. (2014).

6.3 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 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.[5] As can be seen from this table, the average MSFE of the augmented GVAR at horizon $h=0$ for the different PMI release months, $\mathfrak{m}=1,2,3$, ranges 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. 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 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

[4] See Chudik et al. (2014) for more details on the GDP release lags for each of the countries in our sample.

[5] See Chudik et al. (2014) for details.

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.

6.4 Panel DM test statistics

The forecast comparisons in Table 4 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

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

and

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

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 $\boldsymbol{\Sigma}_\eta$, 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 $\boldsymbol{\Sigma}_\eta$ can be ignored. Accordingly, we base the panel DM tests on the following weighted pooled DM test statistic:

$$WPDM = \sqrt{T}\left(\sum_{i=1}^{n} w_i^2\right)^{-1/2} \frac{\bar{z}_w}{\sqrt{\hat{\vartheta}_1}}, \tag{59}$$

where

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

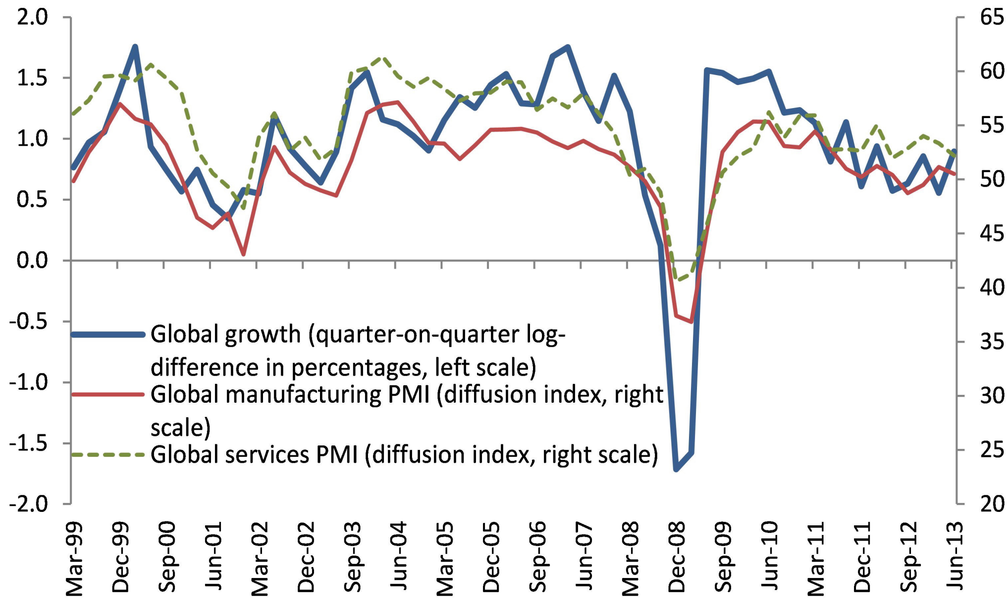
in which $\hat{\sigma}_{LRi}$ is the Newey and West (1987) estimator of the long-run variance of z_{it} to take into account possible serial correlations of z_{it}. We set the truncation lag in the Newey-West estimator to 2. Under the null hypothesis, the $WPDM$ is asymptotically normally distributed 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 $\boldsymbol{\Sigma}$ is sufficiently small. We leave the further development of the panel DM test statistics under a more general form of cross-sectional dependence to future research and present test results based on $WPDM$ as defined by (59).

All methods that use PMIs are significantly better than the RW at the 1% level for all three months in the current quarter, $h = 0$, and the vast majority at the 1% level for $h > 0$. These findings are not surprising given the differences in MSFE reported in Table 4. We consider next testing whether adding PMIs significantly improves the MSFEs. The top panel of Table 5 presents pairwise 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 5. 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.

7 Conclusion

In this paper we have shown that the GVAR model can be under-determined when strong unobserved common factors are present, and propose augmenting the GVAR model with additional equations for cross-sectional averages that proxy the common factors. The validity of the augmentation procedure is established theoretically for N and $T \to \infty$, jointly without any restrictions on N/T. The theoretical results are illustrated by MC experiments, and extended to the case of forecasting with GVARs in the presence of nonsynchronous conditioning sets (reported in Chudik et al. (2014)). 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 output 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 6.1 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, SAR parameter ϱ_ε for idiosyncratic errors set equal to 0.4 and block-diagonal coefficient matrix $\mathbf{\Phi}$.

	GVAR					**AugGVAR**				
(N,T)	**30**	**50**	**100**	**200**	**500**	**30**	**50**	**100**	**200**	**500**
DGP1: High-dimensional VAR without common factor										
30	1.63	1.27	1.13	1.07	1.04	1.57	1.26	1.13	1.07	1.04
50	1.64	1.26	1.14	1.08	1.03	1.59	1.25	1.13	1.07	1.03
100	1.63	1.27	1.13	1.07	1.04	1.58	1.26	1.13	1.07	1.04
200	1.63	1.26	1.14	1.07	1.04	1.59	1.25	1.14	1.07	1.04
500	1.63	1.27	1.14	1.07	1.04	1.59	1.26	1.13	1.07	1.04
DGP2: High-dimensional VAR with additive common factors										
	$m = 1$ unobserved common factor									
30	37.89	2.29	1.61	1.44	1.31	1.54	1.24	1.14	1.07	1.04
50	>100	3.32	1.96	1.63	1.51	1.57	1.25	1.13	1.08	1.04
100	>100	>100	2.97	2.13	2.05	1.56	1.23	1.13	1.08	1.03
200	>100	>100	11.83	3.79	3.04	1.54	1.25	1.13	1.07	1.04
500	>100	>100	>100	16.60	6.32	1.53	1.24	1.13	1.07	1.04
	$m = 2$ unobserved common factors									
30	>100	87.98	3.05	2.55	2.09	1.47	1.22	1.13	1.08	1.04
50	>100	92.48	4.64	3.51	2.78	1.51	1.24	1.13	1.07	1.04
100	>100	>100	>100	6.27	4.68	1.53	1.23	1.13	1.07	1.04
200	>100	>100	>100	24.97	9.66	1.51	1.24	1.13	1.07	1.04
500	>100	>100	>100	>100	27.50	1.51	1.22	1.14	1.08	1.04
DGP3: High-dimensional VAR with a multi-factor error structure										
	$m = 1$ unobserved common factor									
30	>100	>100	4.74	3.41	2.86	1.46	1.23	1.12	1.06	1.03
50	>100	>100	8.84	5.16	4.24	1.46	1.20	1.12	1.06	1.03
100	>100	>100	65.14	10.40	7.84	1.49	1.22	1.13	1.07	1.04
200	>100	>100	>100	>100	15.04	1.50	1.21	1.12	1.07	1.03
500	>100	>100	>100	>100	96.68	1.51	1.24	1.12	1.07	1.04
	$m = 2$ unobserved common factors									
30	>100	>100	26.59	7.01	5.38	1.44	1.20	1.12	1.06	1.04
50	>100	>100	>100	14.22	8.45	1.43	1.21	1.12	1.07	1.04
100	>100	>100	>100	>100	16.70	1.40	1.21	1.12	1.06	1.03
200	>100	>100	>100	>100	55.68	1.43	1.21	1.12	1.06	1.03
500	>100	>100	>100	>100	>100	1.46	1.20	1.14	1.05	1.03

Notes: This table reports the simple cross-section average mean square forecast error of GVAR and AugGVAR forecasts relative to infeasible optimal forecasts. See (51). DGPs 1-3 are given by models (48), (49), and (50), respectively. Infeasible forecasts are defined as $E\left(\mathbf{y}_{i,T+1}|\mathcal{I}_t, \mathcal{F}_t\right)$. See (52). Computations of GVAR and AugGVAR forecasts are explained in Subsection 5.2. In particular, see (55) and (58).

Table 2: Cross-section average MSFE of one-step-ahead GVAR forecasts relative to infeasible optimal forecasts in Monte Carlo experiments, SAR parameter ϱ_ε for idiosyncratic errors set equal to 0.6 and non-sparse coefficient matrix $\mathbf{\Phi}$.

	GVAR					**AugGVAR**				
(N,T)	**30**	**50**	**100**	**200**	**500**	**30**	**50**	**100**	**200**	**500**
DGP1: High-dimensional VAR without common factor										
30	1.66	1.27	1.13	1.08	1.04	1.58	1.25	1.13	1.08	1.03
50	1.64	1.27	1.14	1.08	1.04	1.57	1.26	1.14	1.07	1.04
100	1.63	1.27	1.14	1.07	1.04	1.58	1.25	1.13	1.07	1.04
200	1.63	1.27	1.14	1.07	1.04	1.58	1.26	1.13	1.07	1.03
500	1.63	1.27	1.14	1.07	1.04	1.58	1.26	1.14	1.07	1.04
DGP2: High-dimensional VAR with additive common factors										
	$m = 1$ unobserved common factor									
30	47.92	2.05	1.45	1.31	1.21	1.56	1.26	1.13	1.08	1.04
50	>100	2.62	1.62	1.40	1.29	1.55	1.25	1.13	1.07	1.04
100	>100	54.26	2.20	1.72	1.49	1.55	1.25	1.13	1.08	1.04
200	>100	72.98	3.32	2.30	1.94	1.54	1.25	1.12	1.07	1.04
500	>100	>100	>100	5.04	3.57	1.56	1.24	1.13	1.07	1.04
	$m = 2$ unobserved common factors									
30	>100	4.71	2.39	1.86	1.58	1.52	1.24	1.13	1.09	1.05
50	>100	>100	3.09	2.31	1.95	1.48	1.23	1.14	1.07	1.04
100	>100	>100	7.52	3.72	3.03	1.49	1.24	1.12	1.08	1.04
200	>100	>100	39.97	6.39	4.74	1.52	1.24	1.13	1.08	1.03
500	>100	>100	>100	51.10	13.15	1.53	1.23	1.12	1.07	1.04
DGP3: High-dimensional VAR with a multi-factor error structure										
	$m = 1$ unobserved common factor									
30	>100	59.04	3.33	2.43	1.92	1.45	1.21	1.13	1.07	1.03
50	>100	95.73	5.04	3.33	2.52	1.43	1.22	1.12	1.07	1.03
100	>100	>100	14.97	5.52	3.96	1.48	1.21	1.15	1.05	1.03
200	>100	>100	>100	11.51	7.45	1.48	1.20	1.12	1.07	1.03
500	>100	>100	>100	>100	18.49	1.48	1.22	1.12	1.07	1.04
	$m = 2$ unobserved common factors									
30	>100	>100	6.62	3.99	3.26	1.45	1.20	1.13	1.06	1.04
50	>100	>100	27.86	6.65	4.61	1.50	1.19	1.11	1.07	1.02
100	>100	>100	>100	15.61	8.72	1.42	1.20	1.10	1.06	1.03
200	>100	>100	>100	>100	17.07	1.40	1.20	1.10	1.07	1.02
500	>100	>100	>100	>100	>100	1.47	1.19	1.11	1.07	1.03

Notes: This table reports the simple cross-section average mean square forecast error of GVAR and AugGVAR forecasts relative to infeasible optimal forecasts. See (51). DGPs 1-3 are given by models (48), (49), and (50), respectively. Infeasible forecasts are defined as $E\left(\mathbf{y}_{i,T+1}|\mathcal{I}_t, \mathcal{F}_t\right)$. See (52). Computations of GVAR and AugGVAR forecasts are explained in Subsection 5.2. In particular, see (55) and (58).

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:	$\mathtt{m} = 1$	$\mathtt{m} = 2$	$\mathtt{m} = 3$	$\mathtt{m} = 1$	$\mathtt{m} = 2$	$\mathtt{m} = 3$	$\mathtt{m} = 1$	$\mathtt{m} = 2$	$\mathtt{m} = 3$
GVAR-PMI	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	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 is reported in parentheses. GVAR and AugGVAR forecasts reported in this table use Ledoit and Wolf (2004)'s estimator of the error covariance matrix to take account of the nonsynchronous nature of the GDP and PMI release dates. See Chudik et al. (2014) for details.

Table 4: GDP-weighted cross-section average MSFE of individual methods relative to RW

	forecasting horizon (quarters):	$h = 0$			$h = 1$			$h = 2$		
	month:	$\mathtt{m} = 1$	$\mathtt{m} = 2$	$\mathtt{m} = 3$	$\mathtt{m} = 1$	$\mathtt{m} = 2$	$\mathtt{m} = 3$	$\mathtt{m} = 1$	$\mathtt{m} = 2$	$\mathtt{m} = 3$
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 the AugGVAR-PMI models with shrinkage (models 4-6 in Table 4 in Chudik et al. (2014)). Similarly, the AugGVAR is the simple average of the AugGVAR models with shrinkage. See Chudik et al. (2014) for details.

Table 5: GDP-weighted pair-wise panel DM test statistics

forecasting horizon (quarters):		$h = 0$			$h = 1$			$h = 2$		
	month:	$\mathtt{m} = 1$	$\mathtt{m} = 2$	$\mathtt{m} = 3$	$\mathtt{m} = 1$	$\mathtt{m} = 2$	$\mathtt{m} = 3$	$\mathtt{m} = 1$	$\mathtt{m} = 2$	$\mathtt{m} = 3$
	(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 the AugGVAR-PMI models with shrinkage (models 4-6 in Table 4 in Chudik et al. (2014)). See Chudik et al. (2014) for details.

A Appendix

Lemma A.1 *Let $h > 0$ be a given (finite) integer,*

$$\mathbf{\Psi}_{h,ij} = \begin{cases} \mathbf{\Phi}_{h,ii} - \mathbf{\Phi}_{ii}^{h}, & \text{for } i = 1,2,...,N, \\ \mathbf{\Phi}_{h,ij}, & \text{for } i \neq j,\ i,j = 1,2,...,N, \end{cases}$$

where $\mathbf{\Phi}_{h,ij}$, denotes the $n_i \times n_j$ sub-matrix associated with the (i,j) block of $\mathbf{\Phi}^h$, and $\mathbf{\Phi}$ satisfies Assumption 4. Then there exists a finite constant $K_h = K(h)$, such that for any $N \in \mathbb{N}$ and any $i,j = 1,2,...,N$, we have

$$\|\mathbf{\Psi}_{h,ij}\| < \frac{K_h}{N}. \tag{A.1}$$

Proof. We prove (A.1) by induction.

1. Suppose $h = 1$. Then $\mathbf{\Phi}_{1,ij} = \mathbf{\Phi}_{ij}$, and

$$\mathbf{\Psi}_{1,ij} = \begin{cases} \mathbf{0}, & \text{for } i = j = 1,2,...,N, \\ \mathbf{\Phi}_{ij}, & \text{for } i \neq j,\ i,j \in \{1,2,...,N\} \end{cases}.$$

   But under Assumption 4, we have $\|\mathbf{\Phi}_{ij}\| < \frac{K}{N}$, for any $i \neq j$, $i,j = 1,2,...,N$, and therefore there exists a finite constant $K_1 = K$ such that $\|\mathbf{\Psi}_{1,ij}\| < \frac{K_1}{N}$.

2. Suppose (A.1) holds for some $h > 0$. We show next that (A.1) will also hold for $h+1$. Assuming (A.1) holds for some $h > 0$, then

$$\|\mathbf{\Psi}_{h,ij}\| < \frac{K_h}{N} \text{ for any } N \in \mathbb{N} \text{ and any } i,j = 1,2,...,N. \tag{A.2}$$

   Define an $n \times n$ dimensional matrix $\mathbf{\Psi}_h = (\mathbf{\Psi}_{h,ij}, i,j = 1,2,...,N)$, and note that

$$\mathbf{\Psi}_h = \mathbf{\Phi}^h - \mathbf{\Theta}^h,$$

   where $\mathbf{\Theta}$ is an $n \times n$ dimensional block-diagonal matrix with matrices $\mathbf{\Phi}_{ii}$, for $i = 1,2,...,N$, on its diagonal. Note that

$$\mathbf{\Psi}_{h+1} = \mathbf{\Phi}\mathbf{\Psi}_h + \mathbf{\Phi}\mathbf{\Theta}^h - \mathbf{\Theta}^{h+1}.$$

   Consider $\mathbf{\Psi}'_{h+1,i}$ and note that

$$\begin{aligned} \mathbf{\Psi}'_{h+1,i} &= \mathbf{E}'_{i,n}\mathbf{\Psi}_{h+1}, \\ &= \mathbf{E}'_{i,n}\mathbf{\Phi}\mathbf{\Psi}_h + \left(\mathbf{E}'_{i,n}\mathbf{\Phi}\mathbf{\Theta}^h - \mathbf{E}'_{i,n}\mathbf{\Theta}^{h+1}\right), \\ &= \mathbf{A}'_{h,i} + \mathbf{B}_{h,i}, \end{aligned} \tag{A.3}$$

   where $\mathbf{A}_{h,i} = \mathbf{E}'_{i,n}\mathbf{\Phi}\mathbf{\Psi}_h$ and $\mathbf{B}_{h,i} = \mathbf{E}'_{i,n}\mathbf{\Phi}\mathbf{\Theta}^h - \mathbf{E}'_{i,n}\mathbf{\Theta}^{h+1}$. Let $\mathbf{A}'_{h,i} = (\mathbf{A}_{h,i1}, \mathbf{A}_{h,i2}, ..., \mathbf{A}_{h,iN})$, where $\mathbf{A}_{h,ij}$ is $n_i \times n_j$. We have

$$\mathbf{A}_{h,ij} = \sum_{s=1}^{N} \mathbf{\Phi}_{is}\mathbf{\Psi}_{h,sj} = \sum_{s=1, j\neq 1}^{N} \mathbf{\Phi}_{is}\mathbf{\Psi}_{h,sj} + \mathbf{\Phi}_{ii}\mathbf{\Psi}_{h,ij},$$

and taking the spectral norm, we obtain

$$\|\mathbf{A}_{h,ij}\| \leq \sum_{s=1,j\neq 1}^{N} \|\mathbf{\Phi}_{is}\| \|\mathbf{\Psi}_{h,sj}\| + \|\mathbf{\Phi}_{ii}\| \|\mathbf{\Psi}_{h,ij}\| \leq \frac{K_{a,h+1}}{N}, \tag{A.4}$$

for $i,j = 1,2,...,N$, where $\|\mathbf{\Phi}_{ij}\| < \frac{K}{N}$ for $i \neq j$ and $\|\mathbf{\Phi}_{ii}\| < K$ are both implied by Assumption 4, and $\|\mathbf{\Psi}_{h,js}\| < \frac{K_h}{N}$ (see (A.2)). Consider next the term $\mathbf{B}_{h,i}$ on the right side of (A.3). We partition $\mathbf{B}_{h,i}$ similarly to $\mathbf{A}_{h,i}$, namely $\mathbf{B}'_{h,i} = (\mathbf{B}_{h,i1}, \mathbf{B}_{h,i2}, ..., \mathbf{B}_{h,iN})$, where the dimensions of $\mathbf{B}_{h,ij}$ are $n_i \times n_j$. We have

$$\mathbf{B}_{h,ij} = \begin{cases} \mathbf{0}, & \text{for } i = j = 1,2,...,N, \\ \mathbf{\Phi}_{ij}\mathbf{\Phi}_{jj}^{h}, & \text{for } i \neq j,\ i,j \in \{1,2,...,N\} \end{cases}.$$

Taking the spectral norm and noting that $\|\mathbf{\Phi}_{ij}\| < \frac{K}{N}$ for $i \neq j$ and that $\left\|\mathbf{\Phi}_{jj}^{h}\right\| \leq \|\mathbf{\Phi}_{jj}\|^{h} < K$ (both hold by Assumption 4), we obtain

$$\|\mathbf{B}_{h,ij}\| < \frac{K_{b,h+1}}{N}, \tag{A.5}$$

where $K_{b,h+1}$ is finite for a finite h. Setting $K_{h+1} = K_{a,h+1} + K_{b,h+1}$ and using (A.4)-(A.5) in (A.3) yields $\|\mathbf{\Psi}_{h+1,ij}\|_{\infty} < \frac{K_{h+1}}{N}$ for any $N \in \mathbb{N}$ and any $i,j = 1,2,...,N$, as desired.

■

Proof of Proposition 2. We assume that model M_a, defined by (1) holds, and the augmented forecasts are based on (44). Note that $\mathbf{E}'_{i,n+n^*}\mathbf{\Theta}^h = \mathbf{\Phi}_{ii}^h$, and

$$\mathbf{E}'_{i,n+n^*}\mathring{\mathbf{B}}_0 = \begin{cases} \mathbf{0}, & \text{under Assumption 2.a,} \\ \mathbf{\Gamma}_{0,i}\left(\bar{\mathbf{\Gamma}}'_{W,0}\bar{\mathbf{\Gamma}}_{W,0}\right)^{+}\bar{\mathbf{\Gamma}}'_{W,0}, & \text{under Assumption 2.b.} \end{cases}$$

Thus, using (44) with $\mathbf{A}^h$ given by (43), $\mathbf{y}^{aug}_{i,t+h|t}$ can be written as

$$\mathbf{y}^{aug}_{i,t+h|t} = \begin{cases} \mathbf{\Phi}_{ii}^{h}\mathbf{y}_{it}, & \text{under Assumption 2.a} \\ \mathbf{\Phi}_{ii}^{h}\mathbf{y}_{it} + \mathbf{\Gamma}_{0,i}\left(\bar{\mathbf{\Gamma}}'_{W,0}\bar{\mathbf{\Gamma}}_{W,0}\right)^{+}\bar{\mathbf{\Gamma}}'_{W,0}\mathbf{\Pi}_{\bar{y}}^{h}\bar{\mathbf{y}}_{wt} - \mathbf{\Phi}_{ii}^{h}\mathbf{\Gamma}_{0,i}\left(\bar{\mathbf{\Gamma}}'_{W,0}\bar{\mathbf{\Gamma}}_{W,0}\right)^{+}\bar{\mathbf{\Gamma}}'_{W,0}\bar{\mathbf{y}}_{wt}, & \text{under Assumption 2.b} \end{cases}. \tag{A.6}$$

Using (40) we first note that

$$\mathbf{\Pi}_{\bar{y}}^{h} = \bar{\mathbf{\Gamma}}_{W,0}\mathbf{\Pi}^{h}\left(\bar{\mathbf{\Gamma}}'_{W,0}\bar{\mathbf{\Gamma}}_{W,0}\right)^{+}\bar{\mathbf{\Gamma}}'_{W,0}, \text{ for } h = 1,2,....$$

Substituting this result in (A.6), we obtain

$$\mathbf{y}^{aug}_{i,t+h|t} = \begin{cases} \mathbf{\Phi}_{ii}^{h}\mathbf{y}_{it}, & \text{under Assumption 2.a} \\ \mathbf{\Phi}_{ii}^{h}\mathbf{y}_{it} + \left(\mathbf{\Gamma}_{0,i}\mathbf{\Pi}^{h} - \mathbf{\Phi}_{ii}^{h}\mathbf{\Gamma}_{0,i}\right)\left(\bar{\mathbf{\Gamma}}'_{W,0}\bar{\mathbf{\Gamma}}_{W,0}\right)^{+}\bar{\mathbf{\Gamma}}'_{W,0}\bar{\mathbf{y}}_{wt}, & \text{under Assumption 2.b} \end{cases}. \tag{A.7}$$

Comparing (A.7) to (19) with $\mathbf{C}_{h,i}(L) = \mathbf{C}_{h,i}$ given by (21), it now readily follows that

$$\mathbf{y}^{aug}_{i,t+h|t} = \mathbf{y}_{i,t+h|t} + O_p\left(N^{-1/2}\right), \tag{A.8}$$

uniformly in i under Assumptions 2.a or 2.b, which establishes the consistency of the forecast $\mathbf{y}^{aug}_{i,t+h|t}$. ■

Proof of Theorem 1. We prove the theorem in two parts under model M_a, given by (1). First, we provide a proof for the case when $n^* = m$. For $h = 1$ we can write the infeasible optimal forecast, given by

(7) under $\mathbf{\Gamma}_1 = -\mathbf{\Phi}\mathbf{\Gamma}_0$, as

$$E\left(\mathbf{y}_{T+1} \mid \mathcal{I}_T, \mathcal{F}_T\right) = \mathbf{\Phi}\left(\mathbf{y}_T - \mathbf{\Gamma}_0\mathbf{f}_T\right) + \mathbf{\Gamma}_0\mathbf{\Pi}\mathbf{f}_T = \mathbf{\Theta}\mathbf{y}_T - \mathbf{\Theta}\mathbf{\Gamma}_0\mathbf{f}_T + \mathbf{\Gamma}_0\mathbf{\Pi}\mathbf{f}_T + \left(\mathbf{\Phi} - \mathbf{\Theta}\right)\left(\mathbf{y}_T - \mathbf{\Gamma}_0\mathbf{f}_T\right), \quad \text{(A.9)}$$

and the feasible AugGVAR forecast as (see (43))

$$\hat{\mathbf{y}}^{aug}_{T+1|T} = \widehat{\mathbf{\Theta}}\mathbf{y}_T - \widehat{\mathbf{\Theta}}\hat{\mathring{\mathbf{B}}}_0\bar{\mathbf{y}}_{wT} + \hat{\mathring{\mathbf{B}}}_0\hat{\mathbf{\Pi}}_{\bar{y}}\bar{\mathbf{y}}_{wT}, \quad \text{(A.10)}$$

where $\bar{\mathbf{y}}_{wT} = \mathbf{W}'\mathbf{y}_T$, and $\hat{\mathring{\mathbf{B}}}_0 = \left(\widehat{\mathbf{B}}'_{0,1}, \widehat{\mathbf{B}}'_{0,2}, ..., \widehat{\mathbf{B}}'_{0,N}\right)'$. Subtracting (A.10) from (A.9) yields

$$\begin{aligned} E\left(\mathbf{y}_{T+1} \mid \mathcal{I}_T, \mathcal{F}_T\right) - \hat{\mathbf{y}}^{aug}_{T+1|T} &= \left(\mathbf{\Theta} - \widehat{\mathbf{\Theta}}\right)\mathbf{y}_T + \left(\mathbf{\Gamma}_0\mathbf{\Pi}\mathbf{f}_T - \hat{\mathring{\mathbf{B}}}_0\hat{\mathbf{\Pi}}_{\bar{y}}\bar{\mathbf{y}}_{wT}\right) \\ &\quad - \left(\mathbf{\Theta}\mathbf{\Gamma}_0\mathbf{f}_T - \widehat{\mathbf{\Theta}}\hat{\mathring{\mathbf{B}}}_0\bar{\mathbf{y}}_{wT}\right) + \left(\mathbf{\Phi} - \mathbf{\Theta}\right)\left(\mathbf{y}_t - \mathbf{\Gamma}_0\mathbf{f}_T\right). \end{aligned} \quad \text{(A.11)}$$

We now investigate the properties of the individual elements on the right side of (A.11). First, consider $\left(\mathbf{\Theta} - \widehat{\mathbf{\Theta}}\right)\mathbf{y}_T$ and note that the second moments of $\mathbf{y}_{iT}$ exist. To establish the existence of second moments, consider

$$E\left(\mathbf{y}_{iT}\mathbf{y}'_{iT}\right) = E\left(\tilde{\mathbf{y}}_{iT}\tilde{\mathbf{y}}'_{iT}\right) + E\left(\mathbf{\Gamma}_{0,i}\mathbf{f}_T\mathbf{f}'_T\mathbf{\Gamma}'_{0,i}\right),$$

where $\tilde{\mathbf{y}}_{iT} = \mathbf{y}_{iT} - \mathbf{\Gamma}_{0,i}\mathbf{f}_T$ and the equality holds because $\tilde{\mathbf{y}}_{iT} = \mathbf{E}'_{i,n}\sum_{\ell=0}^{\infty}\mathbf{\Phi}^{\ell}\boldsymbol{\varepsilon}_{T-\ell}$ is independently distributed of $\mathbf{f}_T = \sum_{\ell=0}^{\infty}\mathbf{\Pi}^{\ell}\mathbf{v}_{T-\ell}$. Taking the spectral norm, we obtain

$$\begin{aligned} \left\|E\left(\mathbf{y}_{iT}\mathbf{y}'_{iT}\right)\right\| &\leq \left\|E\left(\tilde{\mathbf{y}}_{iT}\tilde{\mathbf{y}}'_{iT}\right)\right\| + \left\|E\left(\mathbf{\Gamma}_{0,i}\mathbf{f}_T\mathbf{f}'_T\mathbf{\Gamma}'_{0,i}\right)\right\|, \\ &\leq \left\|\mathbf{R}\right\|^2\sum_{\ell=0}^{\infty}\left\|\mathbf{\Phi}\right\|^{2\ell} + \left\|E\left(\mathbf{v}_t\mathbf{v}'_t\right)\right\|^2\left\|\mathbf{\Gamma}_{0,i}\right\|^2\sum_{\ell=0}^{\infty}\left\|\mathbf{\Pi}\right\|^{2\ell}, \\ &\leq K, \end{aligned}$$

where the upper bound K does not depend on i nor on N, $\left\|\mathbf{R}\right\|^2 \leq \left\|\mathbf{R}\right\|_1\left\|\mathbf{R}\right\|_{\infty} < K$ by Assumption 1, $\sum_{\ell=0}^{\infty}\left\|\mathbf{\Phi}\right\|^{2\ell} < K$ by Assumption 3, and by Assumption 2.b we have $\left\|E\left(\mathbf{v}_t\mathbf{v}'_t\right)\right\|^2 < K$, $\left\|\mathbf{\Gamma}_{0,i}\right\|^2 < K$, and $\sum_{\ell=0}^{\infty}\left\|\mathbf{\Pi}\right\|^{2\ell} < K$. This establishes that $\left\|E\left(\mathbf{y}_{iT}\mathbf{y}'_{iT}\right)\right\|$ is uniformly bounded and therefore

$$E\left\|\mathbf{y}_T\right\|_{\infty} < K. \quad \text{(A.12)}$$

Chudik and Pesaran (2011, Theorem 1) establish that in the full rank case with $n^* = m$, the least squares estimates of (30) satisfy

$$\sqrt{T}\left(\hat{\mathbf{\Phi}}_{ii} - \mathbf{\Phi}_{ii}\right) = O_p\left(1\right), \text{ and } \sqrt{T}\left(\hat{\mathbf{B}}_{is} - \mathbf{B}_{is}\right) = O_p\left(1\right), \quad \text{(A.13)}$$

for $s = 0, 1$, uniformly in i as $N, T \xrightarrow{j} \infty$, where $\hat{\mathbf{\Phi}}_{ii}$ and $\hat{\mathbf{B}}_{is}$, for $s = 0, 1$, are the least squares estimates of $\mathbf{\Phi}_{ii}$ and $\mathbf{B}_{is}$ in (30). The first part of (A.13) implies $E\left\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\right\|_{\infty} \to 0$ and together with (A.12) we obtain

$$E\left\|\left(\mathbf{\Theta} - \widehat{\mathbf{\Theta}}\right)\mathbf{y}_T\right\|_{\infty} \to 0. \quad \text{(A.14)}$$

Consider next

$$
\begin{aligned}
\mathbf{\Gamma}_0\mathbf{\Pi}\mathbf{f}_T - \widehat{\mathring{\mathbf{B}}}\hat{\mathbf{\Pi}}_{\bar{y}0}\bar{\mathbf{y}}_{wT} &= \left(\mathbf{\Gamma}_0\mathbf{\Pi}\mathbf{f}_T - \widehat{\mathring{\mathbf{B}}}_0\mathbf{\Pi}_{\bar{y}}\bar{\mathbf{y}}_{wT}\right) - \widehat{\mathring{\mathbf{B}}}_0\left(\hat{\mathbf{\Pi}}_{\bar{y}} - \mathbf{\Pi}_{\bar{y}}\right)\bar{\mathbf{y}}_{wT}, \\
&= \left(\mathbf{\Gamma}_0\mathbf{\Pi}\mathbf{f}_T - \mathring{\mathbf{B}}_0\mathbf{\Pi}_{\bar{y}}\bar{\mathbf{y}}_{wT}\right) - \left(\widehat{\mathring{\mathbf{B}}}_0 - \mathring{\mathbf{B}}_0\right)\mathbf{\Pi}_{\bar{y}}\bar{\mathbf{y}}_{wT} - \widehat{\mathring{\mathbf{B}}}_0\left(\hat{\mathbf{\Pi}}_{\bar{y}} - \mathbf{\Pi}_{\bar{y}}\right)\bar{\mathbf{y}}_{wT}.
\end{aligned}
$$

Since

$$
\begin{aligned}
\bar{\mathbf{y}}_{wT} &= \bar{\mathbf{\Gamma}}_{W,0}\mathbf{f}_T + \mathbf{W}'\sum_{\ell=0}^{\infty}\mathbf{\Phi}^{\ell}\boldsymbol{\varepsilon}_{T-\ell}, \\
&= \bar{\mathbf{\Gamma}}_{W,0}\mathbf{f}_t + O_p\left(N^{-1/2}\right),
\end{aligned}
$$

it can be shown using the same arguments as in Chudik and Pesaran (2011) that $\hat{\mathbf{\Pi}}_{\bar{y}}$ is a consistent estimator of $\mathbf{\Pi}_{\bar{y}} = \bar{\mathbf{\Gamma}}_{W,0}\mathbf{\Pi}\bar{\mathbf{\Gamma}}_{W,0}^{-1}$ and $E\left\|\hat{\mathbf{\Pi}}_{\bar{y}} - \mathbf{\Pi}_{\bar{y}}\right\|_{\infty} \to 0$. Furthermore, (A.12) and (8) imply $E\left\|\bar{\mathbf{y}}_{wT}\right\|_{\infty} = E\left\|\mathbf{W}'\mathbf{y}_t\right\|_{\infty} < K$ and the right part of (A.13) implies $E\left\|\widehat{\mathring{\mathbf{B}}}_0 - \mathring{\mathbf{B}}_0\right\|_{\infty} \to 0$. It therefore follows that

$$
E\left\|\mathbf{\Gamma}_0\mathbf{\Pi}\mathbf{f}_T - \widehat{\mathring{\mathbf{B}}}_0\hat{\mathbf{\Pi}}_{\bar{y}}\bar{\mathbf{y}}_{wT}\right\|_{\infty} \to 0. \tag{A.15}
$$

Similarly, using both parts of (A.13), which imply $E\left\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\right\|_{\infty} \to 0$ and $E\left\|\widehat{\mathring{\mathbf{B}}}_0 - \mathring{\mathbf{B}}_0\right\|_{\infty} \to 0$, and using the same arguments as in the proof of (A.15), we obtain

$$
E\left\|\mathbf{\Theta}\mathbf{\Gamma}_0\mathbf{f}_T - \widehat{\mathbf{\Theta}}\widehat{\mathring{\mathbf{B}}}_0\bar{\mathbf{y}}_{wT}\right\|_{\infty} \to 0. \tag{A.16}
$$

Consider now the rows of $(\mathbf{\Phi} - \mathbf{\Theta})(\mathbf{y}_t - \mathbf{\Gamma}_0\mathbf{f}_T)$ corresponding to the i-th cross-section unit, denoted as

$$
\begin{aligned}
\boldsymbol{\vartheta}_{iT} &\equiv \mathbf{E}'_{i,n}(\mathbf{\Phi} - \mathbf{\Theta})(\mathbf{y}_T - \mathbf{\Gamma}_0\mathbf{f}_T) = \mathbf{\Phi}'_{-i}(\mathbf{y}_T - \mathbf{\Gamma}_0\mathbf{f}_T), \\
&= \mathbf{\Phi}'_{-i}\sum_{\ell=0}^{\infty}\mathbf{\Phi}^{\ell}\mathbf{R}\boldsymbol{\eta}_{T-\ell},
\end{aligned}
$$

where $\mathbf{\Phi}'_{-i} = (\mathbf{\Phi} - \mathbf{\Theta})'\mathbf{E}_{i,n}$ satisfies condition (8) under Assumption 4, and $\mathbf{y}_T - \mathbf{\Gamma}_0\mathbf{f}_T = \sum_{\ell=0}^{\infty}\mathbf{\Phi}^{\ell}\mathbf{R}\boldsymbol{\eta}_{T-\ell}$. The elements of $E\left(\boldsymbol{\vartheta}_{iT}\boldsymbol{\vartheta}'_{iT}\right)$ are uniformly of order $O\left(N^{-1}\right)$, in particular

$$
\begin{aligned}
\left\|E\left(\boldsymbol{\vartheta}_{iT}\boldsymbol{\vartheta}'_{iT}\right)\right\| &= \sum_{\ell=0}^{\infty}\mathbf{\Phi}'_{-i}\mathbf{\Phi}^{\ell}\mathbf{R}\mathbf{R}'\mathbf{\Phi}^{\ell\prime}\mathbf{\Phi}_{-i}, \\
&\leq \left\|\mathbf{\Phi}_{-i}\right\|^2\left\|\mathbf{R}\right\|^2\sum_{\ell=0}^{\infty}\left\|\mathbf{\Phi}\right\|^{2\ell}, \\
&\leq \frac{K}{N}, \qquad \text{(A.17)}
\end{aligned}
$$

where $\left\|\mathbf{\Phi}'_{-i}\right\|^2 < KN^{-1}$ by Assumption 4, and as before $\left\|\mathbf{R}\right\|^2 \leq \left\|\mathbf{R}\right\|_1\left\|\mathbf{R}\right\|_{\infty} < K$ by Assumption 1, and $\sum_{\ell=0}^{\infty}\left\|\mathbf{\Phi}\right\|^{2\ell} < K$ by Assumption 3. (A.17) implies

$$
E\left\|(\mathbf{\Phi} - \mathbf{\Theta})(\mathbf{y}_T - \mathbf{\Gamma}_0\mathbf{f}_T)\right\|_{\infty} \to 0. \tag{A.18}
$$

Using (A.14)-(A.16), and (A.18) in (A.11), we obtain

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

as desired. Result (47) in the case of $n = m^*$ and for $h > 1$ can be established in a similar way.

Consider now the case when $n^* > m$. As compared to the case $n^* = m$, the main difference is that $\bar{\mathbf{y}}_{wT} = \bar{\mathbf{\Gamma}}_{W,0}\mathbf{f}_t + O_p\left(N^{-1/2}\right)$ is asymptotically singular since n^* ($> m$) cross-section averages, $\bar{\mathbf{y}}_{wT}$, span the same space as the m unobserved common factors as $N \rightarrow \infty$. Let $\mathbf{S}_a$ be any $m \times n^*$ selection matrix that selects m linearly independent rows of the full column rank matrix $\bar{\mathbf{\Gamma}}_{W,0}$, and $\mathbf{S}_b$ be an $m \times (n^* - m)$ selection matrix that selects the remaining rows. Moreover, let $\bar{\mathbf{y}}_{waT} \equiv \mathbf{S}_a\bar{\mathbf{y}}_{wT}$ and $\bar{\mathbf{y}}_{wbT} \equiv \mathbf{S}_b\bar{\mathbf{y}}_{wT}$. In this notation, $\bar{\mathbf{y}}_{waT}$ contains a selection of m cross-section averages from $\bar{\mathbf{y}}_{wT}$, $\bar{\mathbf{y}}_{wbT}$ is an $(n^* - m) \times 1$ vector of the remaining $n^* - m$ cross-section averages, and $\bar{\mathbf{y}}_{waT} = \mathbf{S}_a\bar{\mathbf{\Gamma}}_{W,0}\mathbf{f}_t + O_p\left(N^{-1/2}\right) = \bar{\mathbf{\Gamma}}_{Wa,0}\mathbf{f}_t + O_p\left(N^{-1/2}\right)$, where $\bar{\mathbf{\Gamma}}_{Wa,0} = \mathbf{S}_a\bar{\mathbf{\Gamma}}_{W,0}$ is a full rank (invertible) $m \times m$ matrix. Clearly, $\mathbf{S}_a$ exists (by the assumption of full column rank of $\bar{\mathbf{\Gamma}}_{W,0}$), but it is not unique. Similar to the first part of this proof, we focus on the individual elements on the right side of (A.11), which shows the difference between the infeasible optimal forecast $E\left(\mathbf{y}_{T+1} \mid \mathcal{I}_T, \mathcal{F}_T\right)$ and the feasible forecast AugGVAR $\hat{\mathbf{y}}_{T+1|T}^{aug}$. Consider first $\left(\mathbf{\Theta} - \widehat{\mathbf{\Theta}}\right)\mathbf{y}_T$. Uniform consistency of $\hat{\mathbf{\Phi}}_{ii}$ is established in Chudik and Pesaran (2011) also for the case $n^* > m$, and therefore (A.14) continues to hold when $n^* > m$. The last term on the right side of (A.11), namely $(\mathbf{\Phi} - \mathbf{\Theta})(\mathbf{y}_t - \mathbf{\Gamma}_0\mathbf{f}_T)$, does not depend on n^* and therefore (A.18) continues to hold also for $n^* > m$. This leaves us with the two middle terms on the right side of (A.11). Consider first the CALS regression (30), which we write here for all the available time observations as

$$\mathbf{Y}_i = \mathbf{Y}_{i,-1}\mathbf{\Phi}_{ii}' + \bar{\mathbf{Y}}_w\mathbf{B}_{i0}' + \bar{\mathbf{Y}}_{w,-1}\mathbf{B}_{i1}' + \mathbf{U}_{\xi i},$$

where $\mathbf{Y}_i = (\mathbf{y}_{i,2}, \mathbf{y}_{i,3}, ..., \mathbf{y}_{i,T})'$ is a $(T-1) \times n_i$ matrix of observations on the dependent variables for the i-th cross-section unit, $\mathbf{Y}_{i,-1} = (\mathbf{y}_{i,1}, \mathbf{y}_{i,2}, ..., \mathbf{y}_{i,T-1})'$, $\bar{\mathbf{Y}}_w = (\bar{\mathbf{y}}_{w2}, \bar{\mathbf{y}}_{w3}, ..., \bar{\mathbf{y}}_{wT})'$, $\bar{\mathbf{Y}}_{w,-1} = (\bar{\mathbf{y}}_{w1}, \bar{\mathbf{y}}_{w2}, ..., \bar{\mathbf{y}}_{wT-1})'$, and $\mathbf{U}_{\xi i} = (\boldsymbol{\xi}_{i2}, \boldsymbol{\xi}_{i3}, ..., \boldsymbol{\xi}_{iT})'$. The fit of this regression is

$$\mathbf{P}_{a+b}\mathbf{Y}_i = \mathbf{Y}_{i,-1}\hat{\mathbf{\Phi}}_{ii}' + \bar{\mathbf{Y}}_w\hat{\mathbf{B}}_{i0}' + \bar{\mathbf{Y}}_{w,-1}\hat{\mathbf{B}}_{i1}',$$

where

$$\mathbf{P}_{a+b} = \mathbf{X}_{a+b}\left(\mathbf{X}_{a+b}'\mathbf{X}_{a+b}\right)^{+}\mathbf{X}_{a+b}$$

is the orthogonal projector matrix that projects onto the column space of $\mathbf{X}_{a+b}$, and $\mathbf{X}_{a+b} = \left(\mathbf{Y}_{i,-1}, \bar{\mathbf{Y}}_w, \bar{\mathbf{Y}}_{w,-1}\right)$ is the observation matrix on the full set of regressors. Note that, due to the asymptotic perfect multicollinearity problem of regressors in $\bar{\mathbf{Y}}_w$ as well as regressors in $\bar{\mathbf{Y}}_{w,-1}$, the moments of $\hat{\mathbf{B}}_{i0}$ and $\hat{\mathbf{B}}_{i1}$ need not exist, but the fit of the regression is asymptotically unaffected (as $N \rightarrow \infty$) by the inclusion of $n^* - m$ asymptotically redundant variables $\bar{\mathbf{y}}_{wbt}$ and $\bar{\mathbf{y}}_{wb,t-1}$. Using the properties of orthogonal projections, we have the following identity (for any N),

$$\mathbf{P}_{a+b}\mathbf{Y}_i = \mathbf{P}_a\mathbf{Y}_i + \mathbf{P}_b\mathbf{M}_a\mathbf{Y}_i, \tag{A.19}$$

where

$$\begin{aligned}
\mathbf{P}_a &= \mathbf{X}_a \left(\mathbf{X}_a'\mathbf{X}_a\right)^+ \mathbf{X}_a,\ \mathbf{M}_a = \mathbf{I}_{T-1} - \mathbf{P}_a, \\
\mathbf{X}_a &= \left(\mathbf{Y}_{i,-1}, \bar{\mathbf{Y}}_{wa}, \bar{\mathbf{Y}}_{wa,-1}\right),\ \bar{\mathbf{Y}}_{wa} = \bar{\mathbf{Y}}_w \mathbf{S}_a',\ \bar{\mathbf{Y}}_{wa,-1} = \bar{\mathbf{Y}}_{wa,-1}\mathbf{S}_a', \\
\mathbf{P}_b &= \mathbf{X}_b \left(\mathbf{X}_b'\mathbf{X}_b\right)^+ \mathbf{X}_b, \\
\mathbf{X}_b &= \left(\bar{\mathbf{Y}}_{wb}, \bar{\mathbf{Y}}_{wb,-1}\right),\ \bar{\mathbf{Y}}_{wb} = \bar{\mathbf{Y}}_w \mathbf{S}_b',\ \bar{\mathbf{Y}}_{wb,-1} = \bar{\mathbf{Y}}_{wb,-1}\mathbf{S}_b'.
\end{aligned}$$

Identity (A.19) decomposes the fit $\mathbf{P}_{a+b}\mathbf{Y}_i$ into the partial fit $\mathbf{P}_a\mathbf{Y}_i$ from the regressors in $\mathbf{X}_a$ and the contribution $\mathbf{P}_b\mathbf{M}_a\mathbf{Y}_i$ to the overall fit from the subset of additional (asymptotically redundant) regressors in $\mathbf{X}_b$. It is easy to verify that identity (A.19) holds by noting that $\mathbf{P}_b\mathbf{M}_a = \mathbf{P}_{a+b}\mathbf{M}_a$ (since $\mathbf{X}_a$ is included in $\mathbf{X}_{a+b}$), and

$$\mathbf{P}_a + \mathbf{P}_{a+b}\mathbf{M}_a = \mathbf{P}_a + \mathbf{P}_{a+b}\left(\mathbf{I}_{T-1} - \mathbf{P}_a\right) = \mathbf{P}_{a+b}.$$

The partial fit $\mathbf{P}_a\mathbf{Y}_i$ corresponds to the case with m cross-section averages in the first part of the proof, and all of the convergence results in the first part of the proof therefore apply to the partial model with $\mathbf{X}_a$ regressors. In particular, $\mathbf{P}_a\mathbf{Y}_i$ converges to the fit of the true model, uniformly in i, as $N, T \xrightarrow{j} \infty$. Moreover,

$$E\left\|\mathbf{\Gamma}_0\mathbf{f}_T - \hat{\mathring{\mathbf{B}}}_{(a),0}\bar{\mathbf{y}}_{waT}\right\|_\infty \to 0, \tag{A.20}$$

where $\hat{\mathring{\mathbf{B}}}_{(a),0}$ is the matrix of coefficient estimates corresponding to the regressors $\bar{\mathbf{Y}}_{wa}$ in the partial model with regressors $\mathbf{X}_a$ only. Now consider $E\left\|\hat{\mathring{\mathbf{B}}}_{(a),0}\bar{\mathbf{y}}_{waT} - \hat{\mathring{\mathbf{B}}}_0\bar{\mathbf{y}}_{wT}\right\|_\infty$. Relying again on Theorem 1 in Chudik and Pesaran, $\mathbf{M}_a\mathbf{Y}_i$ is a consistent estimate of the idiosyncratic innovations $\mathbf{U}_\varepsilon = (\boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3, ..., \boldsymbol{\varepsilon}_T)'$ (as $N, T \xrightarrow{j} \infty$). Moreover, $\bar{\mathbf{y}}_{wbt} = \bar{\mathbf{\Gamma}}_{Wb,0}\mathbf{f}_t + O_p\left(N^{-1/2}\right)$, and $\mathbf{f}_t$ is independently distributed of $\boldsymbol{\varepsilon}_{t'}$ for all $t, t' = 1, 2, ..., T$. It therefore follows that

$$E\left\|\hat{\mathring{\mathbf{B}}}_{(a),0}\bar{\mathbf{y}}_{waT} - \hat{\mathring{\mathbf{B}}}_0\bar{\mathbf{y}}_{wT}\right\|_\infty \to 0,$$

as $N, T \xrightarrow{j} \infty$ and, using (A.20), we obtain

$$E\left\|\mathbf{\Gamma}_0\mathbf{f}_T - \hat{\mathring{\mathbf{B}}}_0\bar{\mathbf{y}}_{wT}\right\|_\infty \to 0, \tag{A.21}$$

as $N, T \xrightarrow{j} \infty$. Using (A.21) and the uniform consistency of $\hat{\mathbf{\Phi}}_{ii}$, we have

$$E\left\|\mathbf{\Theta}\mathbf{\Gamma}_0\mathbf{f}_T - \widehat{\mathbf{\Theta}}\hat{\mathring{\mathbf{B}}}_0\bar{\mathbf{y}}_{wT}\right\|_\infty \leq E\left\|\mathbf{\Theta}\mathbf{\Gamma}_0\mathbf{f}_T - \mathbf{\Theta}\hat{\mathring{\mathbf{B}}}_0\bar{\mathbf{y}}_{wT}\right\|_\infty + E\left\|\left(\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\right)\hat{\mathring{\mathbf{B}}}_0\bar{\mathbf{y}}_{wT}\right\|_\infty \to 0,$$

as $N, T \xrightarrow{j} \infty$. Using similar arguments, it can be shown that $E\left\|\mathbf{\Gamma}_0\mathbf{\Pi}\mathbf{f}_T - \hat{\mathring{\mathbf{B}}}_0\hat{\mathbf{\Pi}}_{\bar{y}}\bar{\mathbf{y}}_{wT}\right\|_\infty \to 0$, as $N, T \xrightarrow{j} \infty$, which concludes the proof of (47) for the case $n^* > m$ and $h = 1$. Result (47) for $h > 1$ can be obtained in a similar way. The singularity of $\mathbf{G}_0 = \mathbf{I}_N - \mathring{\mathbf{B}}_0\mathbf{W}'$ is implied by (32). ∎

References

Andreou, E., E. Ghysels, and A. Kourtellos (2013). Should macroeconomic forecasters use daily financial data and how? *Journal of Business and Economic Statistics 31*(2), 240–251.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics 120*(1), 387–422.

Chudik, A., V. Grossman, and M. H. Pesaran (2014). A multi-country approach to forecasting output growth using PMIs. Federal Reserve Bank of Dallas, Globalization and Monetary Policy Institute Working Paper No. 213.

Chudik, A. and M. H. Pesaran (2011). Infinite-dimensional VARs and factor models. *Journal of Econometrics 163*(1), 4–22.

Chudik, A. and M. H. Pesaran (2013). Econometric analysis of high dimensional VARs featuring a dominant unit. *Econometric Reviews 32*(5-6), 592–649.

Chudik, A. and M. H. Pesaran (2014). Aggregation in large dynamic panels. *Journal of Econometrics 178*(2), 273–285.

Chudik, A. and M. H. Pesaran (2015a). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics 188*(2), 393–420.

Chudik, A. and M. H. Pesaran (2015b). Theory and practice of GVAR modelling. *forthcoming in Journal of Economic Surveys*. doi:10.1111/joes.12095.

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.

di Mauro, F. and M. H. Pesaran (Eds.) (2013). *The GVAR Handbook: Structure and Applications of a Macro Model of the Global Economy for Policy Analysis*. Oxford University Press. ISBN-10: 0199670080, ISBN-13: 978-0199670086.

Diebold, F. X. and R. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics 13*(3), 253–263.

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), 1–38.

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

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*(5), 603–620.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2004). 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.

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.

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.

Ledoit, O. and M. Wolf (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis 88*(2), 365–411.

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.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Pesaran, M. H. (2015). *Time Series and Panel Data Econometrics*. Oxford University Press. ISBN-978-0-19-873691-2.

Pesaran, M. H., T. Schuermann, and L. V. Smith (2009). Forecasting economic and financial variables with global VARs. *International Journal of Forecasting 25*(4), 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 Economic Statistics 22*(2), 129–162.

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