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

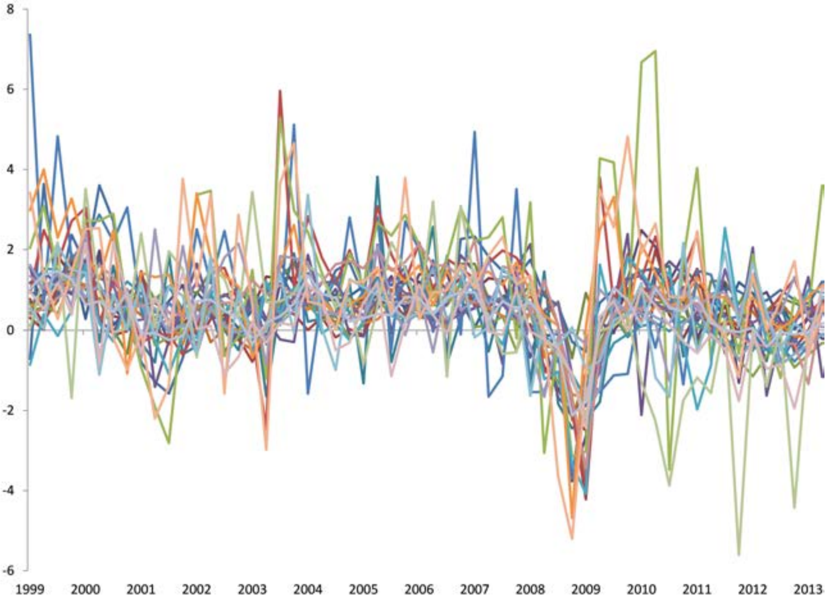
Hashem Pesaran
University of Southern California, CAFE, USA, and Trinity College, Cambridge, UK

8th ECB Workshop on Forecasting Techniques, European Central Bank, June 13-14, 2014

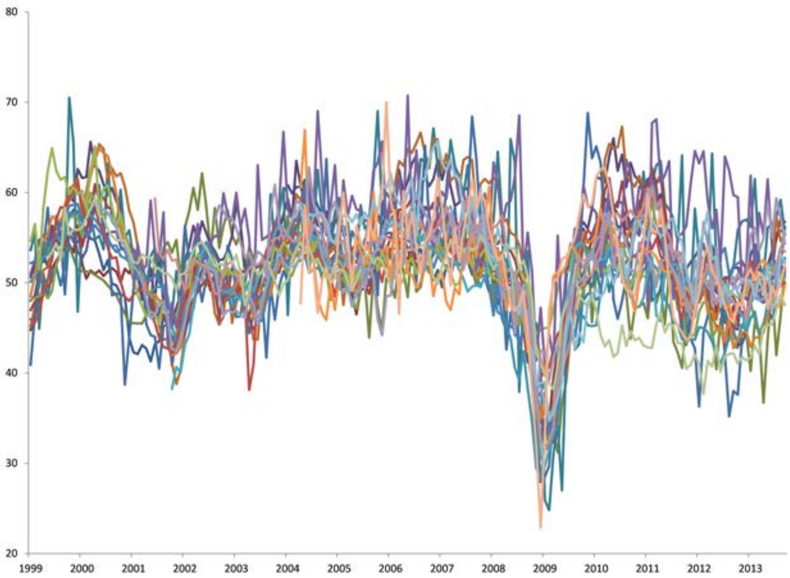
Introduction and contributions of this paper

- Global VAR (GVAR) models are used increasingly in empirical macro and finance literature and forecasting.
- In this paper, we establish conditions under which forecasts from a GVAR model uniformly converge (as $N, T \xrightarrow{j} \infty$) to infeasible optimal forecasts in the case of data generated from a factor-augmented infinite-dimensional VAR model.
- We show that the presence of a strong unobserved common factor can lead to an undetermined system.
- To solve this problem, we propose augmenting the GVAR with additional equation(s) that proxy for the factors.
- In the empirical part, we investigate the information content of the Purchasing Manager Indices (PMIs) for nowcasting and forecasting of 48 countries' growth as well as separate aggregate categories.

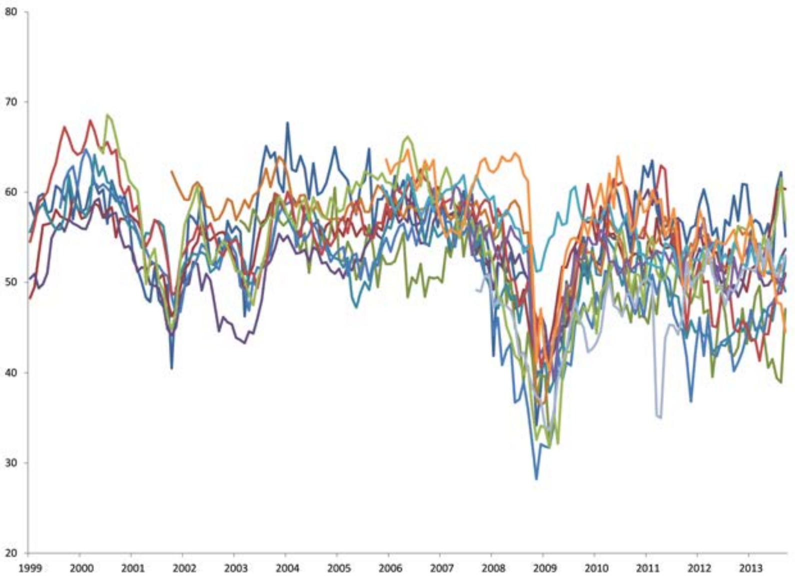
Output (1st differences of logs, $\times 100$, advanced economies)



Manufacturing PMIs (advanced economies)

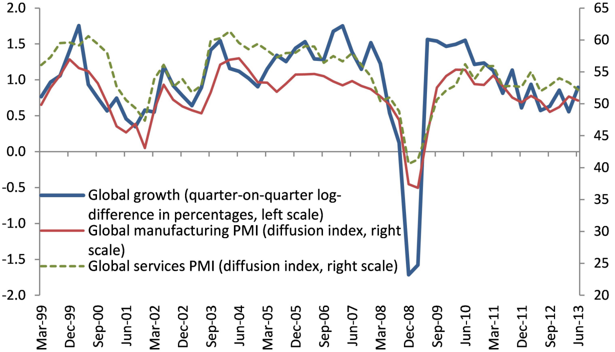


Services PMIs (all economies)



Motivation for forecasting growth with PMIs

- PMIs are timely, available for a broad range of countries, and closely watched by financial market participants as an important indicator of economic activity.
- We examine the extent to which PMIs are useful in nowcasting and forecasting individual country growths at different horizons.



An Overview of the Literature

Techniques for large datasets & the GVAR literature

- Methods for the analysis of large datasets can be classified into machine learning techniques (primarily IID observations) and time series econometric techniques (also applicable to dependent observations)
- Machine learning techniques
  - classification and regression trees
  - penalized regressions (including Lasso and Ridge regressions)
- Econometric techniques
  - Bayesian shrinkage techniques
  - factor models
  - spatiotemporal models, large-dimensional VARs and GVARs
- The GVAR approach was proposed by Pesaran, Schuermann and Weiner (2004) and used in forecasting in a number of papers, including Pesaran, Schuermann and Smith (2009), Ericsson and Reisman (2012) and Garratt, Lee and Shields (2014). Chudik and Pesaran (2014) provide a recent survey.

Nowcasting literature

- A recent survey of "nowcasting" (i.e., prediction of the present, the near future and the recent past) in economics is provided by Banbura, Giannone and Reichlin (2011) and Banbura *et al.* (2012).
- Three challenges in nowcasting and forecasting growth with PMIs:
  - "jagged" or "ragged" edge problem
  - mixed frequencies (GDP is quarterly, PMIs are monthly)
  - large dimensionality: 95 predictors of output for each economy (48 output growths and 47 PMI indices)

- The literature has dealt with these challenges in a number of ways, depending on whether the target variable and predictors are modelled as in one system (for the purpose of a real time monitoring exercise), or not.
- The most popular nowcasting model is a factor model estimated with the use of a Kalman filter and smoother (Evans, 2005, Giannone, Reichlin and Small, 2008).
- Bridge equations (Trehan, 1989, Parigi and Schlitzer, 1995, Diron, 2008, among others) and MIDAS models (Ghysels *et al.*, 2004 and 2007) have also been commonly used for nowcasting.
- We overcome the jagged edge problem by re-aligning each of the series to achieve a balanced end of the sample, and then aggregate monthly PMIs at a quarterly frequency.
- This allows us to apply any of the commonly used methods for forecasting with a large number of predictors (we implement the Lasso, Ridge, factor models, partial least squares and GVARs).

Outline

- Large-dimensional VARs and the justification of the GVAR approach
- Forecasting with large-dimensional VARs
  - Large N representation of optimal forecasts
- Using GVAR for forecasting
- Monte Carlo experiments
  - Performance of GVAR and augmented GVAR models
- Nowcasting and forecasting output growth using PMIs:
  - Data
  - Methods
  - Findings

Factor-augmented, large-dimensional VARs

- Consider the following covariance stationary factor-augmented VAR model for N cross section units collected in the vector $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{Nt})'$:

$$\mathbf{y}_t - \gamma f_t = \mathbf{\Phi} \left(\mathbf{y}_{t-1} - \gamma f_{t-1} \right) + \varepsilon_t, \quad (1)$$

$$f_t = \rho f_{t-1} + v_t,$$

  where γ is an $N \times 1$ vector of factor loadings, f_t is an unobserved common factor, $\mathbf{\Phi}$ is an $N \times N$ matrix of unknown coefficients and ε_t is an $N \times 1$ vector of idiosyncratic shocks. It is assumed that ε_t is independently distributed of v_t, both are serially uncorrelated and with zero means.
- Higher order lags/number of common factors could also be considered. We abstract, without the loss of generality, also from deterministic terms.

- A common alternative to (1) is a VAR model with factor error structure, namely
$$\mathbf{y}_t = \mathbf{\Phi}\mathbf{y}_{t-1} + \gamma f_t + \varepsilon_t. \tag{2}$$
- Because the common factor f_t is unobserved, it is unknown whether (1) or (2) should be preferred. It is therefore important to develop methods that are robust to the way the common factor is introduced in the model.
- We proceed with both models below. We show that the common factors can be in both cases approximated by cross-section averages, but the main difference between the two models is in the number of lags of cross-section averages that are required for consistent estimation and forecasting.

GVAR approach as a representation of large VARs

- Unknown parameters of models (1) and/or (2) cannot be estimated due to the well-known curse of dimensionality.
- This problem has been typically addressed in the literature by either resorting to a shrinkage of the parameter space (e.g., Large Bayesian VARs), or to a shrinkage of the data (e.g., factor models).
- We are interested in imposing a sufficient structure on (1) and/or (2) that would solve the dimensionality problem and allow consistent estimation of unit-specific equations. To this end we follow Chudik and Pesaran (2011, JoE) and impose the following assumptions.

Assumptions

Assumption 1 (Cross-sectionally weakly dependent idiosyncratic errors) Idiosyncratic errors in $\boldsymbol{\varepsilon}_t$ follow the 'spatial' or 'network' model, $\boldsymbol{\varepsilon}_t = \mathbf{R}\boldsymbol{\eta}_t$, where the $N \times N$ matrix $\mathbf{R}$ has assumed to have bounded row and column matrix norms (in N), and $\boldsymbol{\eta}_t \sim IID\left(\mathbf{0}, \mathbf{I}_N\right)$.

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

a. (Model without factor) $\gamma_i = 0$ for all $i = 1, 2, ..., N$.
b. (Model with factor) The unobserved common factor is characterized by the AR(1) model above with $|\rho| < 1$. The macro shock v_t is independently distributed of idiosyncratic errors, $E\left(v_t\right) = 0$, $E\left(v_t^2\right) = \sigma_v^2 = 1 - \rho^2$, and $E\left(v_t v_{t'}\right) = 0$ for any $t \neq t'$. The factor loadings are independently and identically distributed with nonzero mean $\gamma \neq 0$ and a finite variance. In addition, the loadings are independently distributed of v_t and $\boldsymbol{\varepsilon}_t$.

Assumption 3 (Covariance stationarity and bounded variances) There exists a small positive constant ϵ such that $\|\mathbf{\Phi}\| < 1 - \epsilon$, 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

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

and

$$\left|\phi_{ij}\right| < \frac{K}{N}, \text{ for any } j \neq i,\ i, j = 1, 2, ..., N.$$

- 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 A3 also ensures that the variance of y_{it} exists as $N \to \infty$. See Chudik and Pesaran (2011, JoE) for a related discussion.
- Assumption 4 rules out any neighbors (with the exception of own lags). This assumption can be relaxed, at the expense of expositional clarity and more complex notations, without any fundamental implications for the main results derived below.
- Following similar arguments as in Chudik and Pesaran (2011, JoE), for model (1) we obtain

$$\bar{y}_t = \mathbf{w}'\mathbf{y}_t = \mathbf{w}'\gamma f_t + \sum_{\ell=0}^{\infty} \mathbf{\Phi}^{\ell} \varepsilon_{t-\ell} = \mathbf{w}'\gamma f_t + O_p\left(N^{-1/2}\right),$$

for any weights $\mathbf{w}$ satisfying $\|\mathbf{w}\|_{\infty} = \max_i |w_i| < K/N$.

- If in addition $\sum_{i=1}^{N} w_i = 1$, then $\mathbf{w}'\gamma = \gamma + O_p\left(N^{-1/2}\right)$, and we obtain in the case of model (1), under Assumptions 1-4 (see Chudik and Pesaran, 2011, JoE) the cross-sectionally augmented AR specifications (CAAR)

$$y_{it} = \phi_{ii} y_{i,t-1} + b_{i0}\bar{y}_t + b_{i1}\bar{y}_{t-1} + \zeta_{it}, \text{ for } i \in \{1, 2, ..., N\}, \quad (3)$$

where $\zeta_{it} = \varepsilon_{it} + O_p\left(N^{-1/2}\right)$, $b_{i0} = b_{i1} = 0$ under Assumption 2.a, and

$$b_{i0} = \frac{\gamma_i}{\bar{\gamma}},\ b_{i1} = -\frac{\phi_{ii}\gamma_i}{\bar{\gamma}}, \text{ under Assumption 2.b.}$$

- Denote the corresponding least squares estimates of the unknown coefficients as $\widehat{\phi}_{ii}$, $\widehat{b}_{i0}$ and $\widehat{b}_{i1}$. Using these estimates for $i = 1, 2, ..., N$ and provided that matrix $\widehat{\mathbf{G}}_0 = \mathbf{I}_N - \widehat{\mathbf{b}}_0\mathbf{w}'$, where $\widehat{\mathbf{b}}_0 = \left(\hat{b}_{10}, \hat{b}_{20}, ..., \hat{b}_{N0}\right)'$, is invertible, one could construct the following GVAR model of $\mathbf{y}_t$:

$$\mathbf{y}_t = \widehat{\mathbf{G}}\mathbf{y}_{t-1} + \widehat{\mathbf{u}}_t, \quad (4)$$

where $\widehat{\mathbf{u}}_t = \widehat{\mathbf{G}}_0^{-1}\widehat{\zeta}_t$, $\widehat{\mathbf{G}} = \widehat{\mathbf{G}}_0^{-1}\widehat{\mathbf{G}}_1$, $\widehat{\mathbf{G}}_1 = \widehat{\mathbf{\Theta}} + \widehat{\mathbf{b}}_1\mathbf{w}'$, $\mathbf{\Theta}$ is an $N \times N$ diagonal matrix with elements $\widehat{\phi}_{ii}$ on the diagonal.

- In the case of model (2), we have

$$\bar{y}_t = \sum_{\ell=0}^{\infty} \mathbf{w}'\boldsymbol{\Phi}^{\ell} \left(\gamma f_{t-\ell} +^{\ell} \varepsilon_{t-\ell} \right) = \alpha\left(L\right) f_t + O_p\left(N^{-1/2}\right),$$

  where the polynomial $\alpha\left(L\right) = \sum_{\ell=0}^{\infty} \mathbf{w}'\boldsymbol{\Phi}^{\ell}\gamma L^{\ell}$ depends on $\boldsymbol{\Phi}$, γ and $\mathbf{w}$ (but we do not show this dependence explicitly to economize on notations)
- Note that the coefficients in the polynomial $\alpha\left(L\right)$ satisfy $\left|\alpha_{\ell}\right| = \left|\mathbf{w}'\boldsymbol{\Phi}^{\ell}\gamma\right| \leq \left\|\mathbf{w}\right\| \left\|\boldsymbol{\Phi}^{\ell}\right\| \left\|\gamma\right\| = O\left[\left(1-\epsilon\right)^{\ell}\right]$ and are thus declining at an exponential rate.

- Assuming that $a(L) = \alpha^{-1}(L)$ exists and its coefficients also decline exponentially, we obtain

$$y_{it} = \phi_{ii} y_{i,t-1} + b_i(L) \bar{y}_t + \zeta_{it}, \text{ for } i \in \{1, 2, ..., N\},$$

where $b_i(L) = \left(\gamma_i + \sum_{\ell=0}^{\infty} \boldsymbol{\phi}'_{-i} \boldsymbol{\Phi}^{\ell} \gamma\right) \alpha^{-1}(L)$ is of infinite order.

- Coefficients in the above equations can be consistently estimated by least squares with a suitable truncation lag $p = p(T)$ for cross-section averages, see Chudik and Pesaran (2013).
- Estimated unit-specific equations can then be solved in a GVAR model in the usual way. The difference between (1) and (2) is in the number of lags in the large N representations for the individual units.

Optimal forecasts for large VARs with unobserved factors

- Two approaches:
  - (plug-in approach) derive infeasible forecasts conditional on factors and then plug in estimates of the factors.
  - Use Kalman filter to derive forecasts conditional on observables variables, but this requires additional assumptions and the *full* knowledge of the model for the unobserved factors and the covariance matrix of the errors, ε_t.
- Consider the information set $\mathcal{I}_t = \{\mathbf{y}_t, \mathbf{y}_{t-1}, ..., f_t, f_{t-1}, ...\}$
- In the case of model (1), we have

$$\mathbf{y}_{th} = E\left(\mathbf{y}_{t+h} | \mathcal{I}_t\right) = \mathbf{\Phi}^h \mathbf{y}_t + \mathbf{g}_h f_t \text{ for } h = 1, 2, ..., \tag{5}$$

  where $\mathbf{g}_h = \left(\rho^h \mathbf{I}_N - \mathbf{\Phi}^h\right) \gamma$.
- In the case of model (2) that features a factor error structure, we obtain

$$\mathbf{y}_{th} = E\left(\mathbf{y}_{t+h} | \mathcal{I}_t\right) = \mathbf{\Phi}^h \mathbf{y}_t + \mathbf{g}_h^* f_t, \tag{6}$$

  where $\mathbf{g}_h^* = \sum_{\ell=0}^{h-1} \rho^{h-\ell} \mathbf{\Phi}^\ell \gamma$. Note that in both cases, the conditional forecast $\mathbf{y}_{th}$ in (5) and (6) does not depend on the covariance of ε_t.

Large N representation of optimal forecasts

- In practice, the requirement of having full knowledge of an 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, are welcome.
- The optimal forecast $\mathbf{y}_{th}$ in (5) and (6) depends on the unobserved common factor and also on a large number of unknown parameters, which cannot be estimated as $N, T \xrightarrow{j} \infty$ due to the well-known curse of dimensionality.
- We are interested in deriving a large N representation of the optimal forecast that would depend only on observables and a finite number of unknown parameters, which can be consistently estimated. In order to do so, we continue to maintain Assumptions 1-4.

- Using the expression for $\bar{y}_t$ above, we obtain, after some algebra, the following large N representation of optimal forecasts in the case of model (1)

$$y_{ith} = \begin{cases} \phi_{ii}^h y_{it} + O_p\left(N^{-1/2}\right) & \text{under A2.a} \\ \phi_{ii}^h y_{it} + \left(\rho^h - \phi_{ii}^h\right) \frac{\gamma_i}{\bar{\gamma}} \bar{y}_t + O_p\left(N^{-1/2}\right) & \text{under A2.b} \end{cases}.$$

- In the case of model (2) that features a factor error structure, we obtain the following large N representation of optimal forecasts,

$$y_{ith} = \phi_{ii}^h y_{it} + c_{hi}\left(L\right) \bar{y}_t + O_p\left(N^{-1/2}\right),$$

where the polynomial $c_{hi}\left(L\right) = 0$ under Assumption 2.a and $c_{hi}\left(L\right) = \left(g_{hi}^* + \sum_{s=0}^{\infty} \omega_{hi}' \mathbf{\Phi}^s \gamma L^s\right) \alpha^{-1}\left(L\right)$ under Assumption 2.b, in which $\omega_{hi} = \mathbf{\Phi}^{h\prime} \mathbf{e}_{Ni} - \phi_{ii}^h \mathbf{e}_{Ni}$.

GVAR approach to forecasting

- Comparing the large N representation for y_{ith} in the case of models (1) and (2), we see that the latter features infinite lag order, whereas only contemporaneous values of cross-section averages are included in the former.
- It is therefore important that proper consideration is given for the selection of lags when forecasting.
- From now on, we focus on model (1), since the only difference between the two models manifests in the number of lags.
- It is not necessarily desirable to use GVAR model (4) for forecasting under Assumption 2.b because $\mathbf{G}_0 = \mathbf{I}_N - b_0 \mathbf{w}'$, in GVAR model (4), is by construction rank deficient.

The singular case

- To illustrate this rank deficiency, consider $\mathbf{w}'\mathbf{G}_0$:

$$\begin{aligned} \mathbf{w}'\mathbf{G}_0 &= \mathbf{w}'\left(\mathbf{I}_N - \mathbf{b}_0\mathbf{w}'\right) \\ &= \mathbf{w}' - \left(\sum_{i=1}^{N} \frac{w_i \gamma_i}{\bar{\gamma}}\right)\mathbf{w}' \\ &= \mathbf{w}' - \mathbf{w}' = \mathbf{0}'. \end{aligned} \tag{7}$$

- The consequence of rank deficiency of $\mathbf{G}_0$ is that the system of CAAR equations (3) is undetermined.
- In general, it can be shown that when $\text{rank}(\mathbf{G}_0) = N - m$ for some $m > 0$, then the solution for y_{it} depends on m arbitrary stochastic processes. Therefore, the full rank condition, $rank(\mathbf{G}_0) = N$, is necessary and sufficient for the GVAR solution to be uniquely determined.

Dealing with the rank deficient case

- If $\text{rank}(\mathbf{G}_0) = N - m$ and $m > 0$, then the GVAR model would need to be augmented by m additional equations in order for $\mathbf{y}_t$ to be uniquely determined.
- In the case of system (3), $m = 1$, and augmentation by one additional equation is needed in order to obtain a unique solution for $\mathbf{y}_t$. Different options could be considered for augmentation of (3). We consider augmenting the set of conditional equations in (3) with the following marginal equation for the cross-section averages:

$$\bar{y}_t = \rho \bar{y}_{t-1} + \gamma v_t + O_p\left(N^{-1/2}\right), \tag{8}$$

and we treat $\bar{y}_t$ as a proxy for the (scaled) unobserved common factor.

- Stacking (3) and (8), we obtain the following VAR model in $\mathbf{z}_t = (\mathbf{y}_t', \bar{y}_t)'$:

$$\mathbf{A}_0 \mathbf{z}_t = \mathbf{A}_1 \mathbf{z}_{t-1} + \mathbf{u}_{zt} + O_p\left(N^{-1/2}\right), \tag{9}$$

  where $\mathbf{u}_{zt} = (\varepsilon_t', \gamma v_t)'$,

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

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

- Consider now the following forecast of $y_{i,t+h}$: $y^b_{ith} = \mathbf{e}'_{N+1,i}\mathbf{A}^{\ell}\mathbf{z}_t$, where $\mathbf{A} = \mathbf{A}_0^{-1}\mathbf{A}_1$, and $\mathbf{e}_{N+1,i}$ is an $N+1$ dimensional selection vector that selects the i-th element.
- Under Assumptions 1, 2.b, and 3-4, we show

$$y^b_{ith} = E\left(y_{i,t+h} | \mathcal{I}_t\right) + O_p\left(N^{-1/2}\right)$$

  which establishes the consistency of the forecast y^b_{ith} defined above.
- Let $\hat{y}^b_{ith} = \mathbf{e}'_{N+1,i}\hat{\mathbf{A}}^{\ell}\mathbf{z}_t$,for $i = 1, 2, ..., N$ and $h = 1, 2, ...,$ where $\hat{\mathbf{A}}$ is the least squares estimate of $\mathbf{A}$ using the CAAR regressions, (3). We also establish that

$$\left\| E\left(\mathbf{y}_{t+h} \mid \mathcal{I}_t\right) - \hat{\mathbf{y}}^b_{th} \right\|_r \xrightarrow{L_1} 0, \tag{10}$$

  $N, T \xrightarrow{j} \infty$ such that $N/T \rightarrow \varkappa$ for some $0 < \varkappa < \infty$, under Assumptions 1, 2.a or 2.b, and 3-4.

Monte Carlo experiments

DGP1: Weakly cross-sectionally dependent model.
$\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{Nt})$ for $t = -M + 1, ..., 0, 1, 2, ..., T$ is generated as

$$\mathbf{y}_t - \boldsymbol{\mu}_y = \boldsymbol{\Phi}\left(\mathbf{y}_{t-1} - \boldsymbol{\mu}_y\right) + \boldsymbol{\varepsilon}_t, \tag{11}$$

with starting values $\mathbf{y}_{-M} = \mathbf{0}$. The first $M = 100$ observations are discarded as burn-in replications.

- The matrix of coefficients $\boldsymbol{\Phi}$ is generated randomly and so that off-diagonal elements are of order $O_p\left(N^{-1/2}\right)$. In particular, $\phi_{ij} = \lambda_i \omega_{ij}$ for $i \neq j$, where $\lambda_i \sim IIDU\left(-0.2, 0.2\right)$ and $\omega_{ij} = \varsigma_{ij} / \left(\sum_{j=1}^{N} \varsigma_{ij}\right)$,with $\varsigma_{ij} \sim IIDU\left(0, 1\right)$. The diagonal elements are generated as $\phi_{ii} \sim IIDU\left(0, 0.6\right)$. This establishes that $\left\|\Phi\right\|_\infty < 0.8$ and the DGP is stationary.
- The idiosyncratic errors, $\boldsymbol{\varepsilon}_t$, are generated according to the following SAR process:

$$\boldsymbol{\varepsilon}_t = a_\varepsilon \mathbf{S} \boldsymbol{\varepsilon}_t + \boldsymbol{\eta}_t,\ 0 < a_\varepsilon < 1,\ \boldsymbol{\eta}_t \sim IIDN\left(\mathbf{0}, \mathbf{I}_N\right)$$

- The $N \times N$ dimensional spatial weights matrix $\mathbf{S}$ is given by

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

- To ensure that the idiosyncratic errors are weakly correlated, the spatial autoregressive parameter, a_ε, must lie in the range $[0, 1)$. We set $a_\varepsilon = 0.4$.
- Individual elements of $\boldsymbol{\mu}_y$ are set equal to $\mu_{yi} = \kappa\sigma_y$, where σ_y is the average of standard deviations of y_{it}. We set $\kappa = 0.8$ (in line with our sample of output growth data).

DGP2: Model featuring unobserved common factor. $\mathbf{y}_t$ and f_t, for $t = -M+1, ..., 0, 1, 2, ..., T$, are generated according to

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

$$f_t = \rho f_{t-1} + v_t$$

with the starting values $\mathbf{y}_{-M} = \mathbf{0}$, $f_{-M} = 0$, and the first $M = 100$ observations are discarded as burn-in replications.

- The coefficient matrix $\boldsymbol{\Phi}$, means $\boldsymbol{\mu}_y$, and the idiosyncratic errors in $\boldsymbol{\varepsilon}_t$ are generated in the same way as in DGP1. We set $\rho = 0.8$ and $\sigma_v^2 = 1 - \rho^2$. Factor loadings are generated as $\gamma_i \sim IIDN\left(\gamma, \sigma_\gamma^2\right)$ with $\gamma = 1$ and $\sigma_\gamma = 0.2$.
- Coefficients and innovations are generated at the beginning of each replication, and $R = 10^4$ replications were carried out for $N, T \in \{30, 50, 100, 200, 500\}$.

Average RMSFE across units

- We investigate the forecasting performance of a non-augmented GVAR model

$$\hat{y}^{a}_{ith} = \mathbf{e}'_{N,i}\widehat{\mathbf{G}}^{h}\mathbf{y}_{t},$$

and an augmented GVAR model $\hat{y}^{b}_{Th} = \mathbf{e}'_{N+1,i}\hat{\mathbf{A}}^{\ell}\mathbf{z}_{t}$ for $h = 1$ (one-step-ahead forecasts).

- We compare these forecasts with their infeasible counterparts. In particular, we compute root mean square forecast error differences

$$RMSFE = \left(\frac{1}{RN} \sum_{i=1}^{N} \sum_{r=1}^{R} \left\{ \hat{y}^{a(r)}_{iT,1} - E\left[y^{(r)}_{i,T+1} \middle| \mathcal{I}_{T}(r) \right] \right\}^{2} \right)^{\frac{1}{2}},$$

where the superscript (r) denotes individual replications. Similarly, we compute RMSFE for forecasts based on a GVAR that is augmented by an additional equation for CS averages.

Table 1: Root mean square forecast error difference between feasible and infeasible one-step-ahead forecasts in MC experiments

	RMSFE: non-augmented GVAR				*RMSFE*: augmented GVAR			
(N,T)	**30**	**50**	**100**	**200**	**30**	**50**	**100**	**200**
	DGP1: Without a common factor							
30	0.42	0.31	0.21	0.15	0.40	0.30	0.20	0.14
50	0.42	0.31	0.21	0.15	0.40	0.30	0.20	0.14
200	0.41	0.30	0.21	0.15	0.40	0.30	0.20	0.14
	DGP2: With an unobserved common factor							
30	0.89	0.68	0.56	0.51	0.49	0.38	0.29	0.23
50	1.02	0.78	0.62	0.57	0.48	0.37	0.27	0.21
200	21.89	2.48	1.13	1.00	0.47	0.34	0.24	0.17

RMSFE of aggregate

- We also investigate the forecasting performance for the aggregate variable $\bar{y}_t$:

$$RMSFE_{\bar{y}} =$$

$$\left(\frac{1}{R} \sum_{r=1}^{R} \left(\frac{1}{N} \sum_{i=1}^{N} \left\{ \hat{y}_{iT,1}^{a,(r)} - E\left[y_{i,T+1}^{(r)} \middle| \mathcal{I}_T(r) \right] \right\} \right)^2 \right)^{\frac{1}{2}},$$

where the superscript (r) denotes individual replications. Similarly, we compute RMSFE of aggregate forecasts based on a GVAR that is augmented by an additional equation for CS averages.

Table 2: Root mean square forecast error difference between feasible and infeasible one-step-ahead forecasts of aggregate variable in MC experiments

	$RMSFE_{\bar{y}}$: non-augmented GVAR				$RMSFE_{\bar{y}}$: augmented GVAR			
(N,T)	**30**	**50**	**100**	**200**	**30**	**50**	**100**	**200**
	DGP1: Without a common factor							
30	0.15	0.11	0.08	0.06	0.10	0.07	0.05	0.03
50	0.11	0.08	0.06	0.05	0.07	0.06	0.04	0.03
200	0.05	0.04	0.03	0.02	0.04	0.03	0.02	0.01
	DGP2: With an unobserved common factor							
30	0.75	0.58	0.50	0.48	0.26	0.22	0.19	0.17
50	0.87	0.68	0.56	0.54	0.23	0.20	0.16	0.14
200	20.02	2.34	1.08	0.96	0.20	0.15	0.12	0.09

Prediction of signs

- We investigate the fraction of correct positive predictions of the sign of the next period forecasts:

$$P_+ = \frac{\sum_{r=1}^{R}\sum_{i=1}^{N} I\left(\hat{y}_{iT,1}^{a,(r)} > 0 \wedge y_{i,T+1} > 0\right)}{\sum_{r=1}^{R}\sum_{i=1}^{N} I\left(\hat{y}_{iT,1}^{a,(r)} > 0 \wedge y_{i,T+1} > 0\right)},$$

- Similarly, we also compute the fraction of correct positive predictions based on a GVAR that is augmented by an additional equation for CS averages.
- We investigate the fraction of correct negative predictions of the sign of the next period forecasts, denoted by P_-.

Table 3: Fraction of correctly predicted positive signs of the next period forecasts.

	P_+: non-augmented GVAR				P_+: augmented GVAR			
(N,T)	**30**	**50**	**100**	**200**	**30**	**50**	**100**	**200**
	DGP1: Without a common factor P_+ for infeasible optimal forecast is 0.99.							
30	0.96	0.98	0.98	0.98	0.96	0.98	0.98	0.99
50	0.96	0.98	0.98	0.98	0.97	0.98	0.98	0.98
200	0.97	0.98	0.98	0.98	0.97	0.98	0.98	0.98
	DGP2: With an unobserved common factor P_+ for infeasible optimal forecast is 0.96.							
30	0.91	0.93	0.94	0.94	0.95	0.95	0.96	0.96
50	0.89	0.92	0.93	0.94	0.94	0.95	0.96	0.96
200	0.81	0.84	0.86	0.87	0.94	0.95	0.96	0.96

Table 4: Fraction of correctly predicted negative signs of the next period forecasts.

	P_-: non-augmented GVAR				P_-: augmented GVAR			
(N,T)	**30**	**50**	**100**	**200**	**30**	**50**	**100**	**200**
	DGP1: Without a common factor P_- for infeasible optimal forecast is 0.08.							
30	0.12	0.09	0.09	0.08	0.11	0.09	0.09	0.08
50	0.11	0.10	0.08	0.08	0.10	0.09	0.08	0.08
200	0.11	0.09	0.08	0.08	0.11	0.09	0.08	0.08
	DGP2: With an unobserved common factor P_- for infeasible optimal forecast is 0.25.							
30	0.30	0.29	0.27	0.27	0.26	0.27	0.23	0.24
50	0.33	0.30	0.27	0.26	0.28	0.25	0.25	0.23
200	0.38	0.33	0.35	0.33	0.28	0.25	0.26	0.24

Nowcasting and forecasting global growth

We investigate the information content of PMIs for forecasting real GDP growth of 48 individual economies across the globe.

PMI data

- PMIs are reported at a monthly frequency, are seasonally adjusted, and are reported as diffusion indices, in which a number greater than 50 indicates an expansion, and a number below 50 indicates a contraction.
- We have two types of PMIs: manufacturing PMIs and services PMIs. PMIs are not available for all countries in our dataset, and the start and end dates of the available data sets differ across countries. We have manufacturing PMI data on 34 countries, which represent 85% of world output. Country coverage of services PMI is much less comprehensive with data being available only for 13 of the 48 countries in our sample - also services PMI data goes to 1999M1 only in the case of 5 countries.
- PMIs are released in a timely manner, shortly after the reporting period.

Aggregation of PMI data

- Let us denote monthly manufacturing PMI index in country i and monthly period ℓ as $n_{i\ell}$. (We use subscript t for quarterly periods.)
- We consider two ways of aggregating monthly PMI series into quarterly observations.
  - *Sequential sampling*: For a given month $m = 1, 2, 3$, we define

$$\bar{n}_{it}^{s,m} = n_{i,3(t-1)+m}. \tag{12}$$

    This gives us three sequentially sampled quarterly series. We use the latest available monthly observation in the regressions.
  - *Temporal aggregation: We use the rolling moving average*

$$\bar{n}_{it}^{a,m} = \left(n_{i,3(t-1)+m} + n_{i,3(t-1)+m-1} + n_{i,3(t-1)+m-2}\right)/3, \tag{13}$$

    where as before $m = 1, 2, 3$ is the chosen month of a quarter, giving us three different temporally aggregated series. As in the case of sequential sampling, we select the latest available moving average at the time of forecasting.

- **Output 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 output data for all 48 countries is available and at the same time we also have a good country coverage for the PMIs. The latest available observation on output is 2013Q2. Output data is seasonally adjusted, most series by the source.
- Our data is pseudo-real time, since no truly real-time datasets seem to be available for the majority of economies in our sample.
- Output data are released with a considerable lag (from about 100 days from the beginning of the reporting quarter to about 200 days).

Forecasting procedures

- **Univariate benchmarks:**
  - Random walk (RW) benchmark (no change in growth)
  - Autoregressive (AR) benchmark model of order 1 (forecasts are computed in direct way)
  - Domestic PMI-augmented AR benchmark model (one lag)
- **Data-rich methods**: Output in each country is predicted with the available output data (both home and foreign) and PMIs.
  - Lasso regression (with 7 different options for the selection of the penalty parameter λ by cross-validation)
  - Ridge regression (with 7 different options for the selection of the shrinkage parameter λ by cross-validation)
  - Factor model (with $1, 2, 3, 4, 5$ factors from preditor variables)
  - Factor-augmented AR model (with $1, 2, 3, 4, 5$ factors)
  - Partial Least Squares (PLS, due to Herman Wold) regression (with $1, 2, 3, 4, 5$ factors from both the preditor and target variables)
  - GVAR models (standard GVAR and augmented GVAR)

- We consider 7 different cross-validation options for the selection of the penalty/shrinkage parameter in the case of Lasso and Ridge regressions.
- We also apply a Ridge shrinkage technique to estimate country-specific models in the GVAR with the penalty parameter estimated using cross validation.
- Each method is computed with either temporally aggregated PMI data, or sequentially sampled data.
- Different choices for the aggregation of monthly PMI data, the penalty/shrinkage parameter λ and the number of factors give us 64 individual data-rich forecasting models in total.

Nowcasting in the presence of structural breaks and combination of forecasts across models

- It is widely recognized that structural breaks have important consequences for forecasting of macroeconomic variables (see for instance Clements and Hendry, 1999 and 2006, and Rossi, 2006)
- There are several ways of dealing with breaks in forecasting in macroeconomics and finance. The conventional approach would be to estimate the discrete break points using one of the numerous statistical procedures developed in the literature and then to construct forecasts based on post-break periods. However, estimates of breaks are inherently uncertain, and, as emphasized by Pesaran and Timmermann (2007), the use of pre-break data, despite biasing the forecasts, can still contribute to lowering RMSFE significantly.

- Another way is to weight observations to deal with breaks. A prominent example is exponential smoothing by Holt (1957) and Brown (1959), and the optimal weights by Pesaran, Pick and Pranovich (2013). An alternative approach to weighting observations in an optimal way is to combine forecasts across different observation windows, as proposed by Pesaran and Timmermann (2007) with further results provided by Pesaran and Pick (2011). This idea has been fruitfully utilized in a number of applications (see for instance Assenmacher-Wesche and Pesaran, 2008, Pesaran, Schuermann and Smith, 2009, and Schrimpf and Wang, 2010) and do not require any knowledge about breaks.

- We utilize a combination of forecasts across different estimation windows. We choose the minimum window size $W_{\text{min}} = 20$ time periods, and consider $m_w = 5$ estimation windows. For each of the nowcasting methods, we compute the corresponding AveW forecast constructed as an arithmetic average of forecasts based on the m_w estimation windows.
- We also consider a forecast combination across different models. We compute the AveM forecast based on the simple arithmetic average of all 64 data-rich methods.
- Finally, we also compute the double average AveMAveW model by averaging 64 individual AveW data-rich forecasts.

Table 5: RMSFE of individual methods: World output, $h = 0, m = 1$

Evaluating period:	Subsample A	Subsample B	Full sample
	06Q1-08Q4	*09Q1-13Q3*	*06Q1-13Q3*
1. RW	0.966	0.517	0.753
2. AR	1.009	0.409	0.743
3. PMI-AR	0.836	0.358	0.621
4. Lasso	0.909	0.267	0.642
5. Ridge	0.747	*0.200*	*0.523*
6. FM	0.766	0.284	0.556
7. FAR	*0.743*	0.293	0.544
8. PLS	**0.722**	**0.197**	**0.507**
9. GVAR	0.797	0.427	0.622
10. augmented GVAR	0.758	0.282	0.551
AveM (4-10)	0.755	0.204	0.530
AveMAveW	0.890	0.233	0.623

Table 6: RMSFE of individual methods: World output, $h = 0, m = 2$

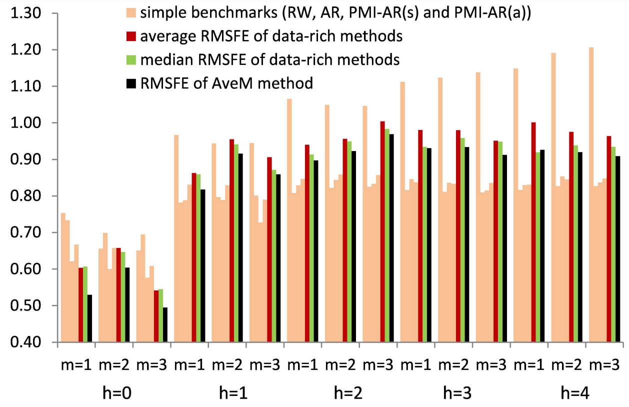
Evaluating period:	Subsample A	Subsample B	Full sample
	06Q1-08Q4	*09Q1-13Q3*	*06Q1-13Q3*
1. RW	0.933	0.260	0.656
2. AR	1.005	0.282	0.707
3. PMI-AR	0.831	0.296	0.600
4. Lasso	0.933	0.230	0.650
5. Ridge	0.867	*0.207*	0.602
6. FM	0.968	0.307	0.689
7. FAR	0.978	0.285	0.690
8. PLS	0.852	0.212	0.594
9. GVAR	*0.783*	0.312	*0.574*
10. augmented GVAR	**0.707**	0.267	**0.515**
AveM (4-10)	0.870	**0.206**	0.604
AveMAveW	0.946	0.212	0.655

Table 7: RMSFE of individual methods: World output, $h = 0$, $m = 3$

Evaluating period:	Subsample A	Subsample B	Full sample
	06Q1-08Q4	*09Q1-13Q3*	*06Q1-13Q3*
1. RW	0.926	0.257	0.651
2. AR	1.003	0.274	0.704
3. PMI-AR	0.804	0.271	0.576
4. Lasso	0.827	0.233	0.582
5. Ridge	*0.671*	0.266	0.492
6. FM	0.688	*0.232*	0.493
7. FAR	0.682	**0.221**	*0.487*
8. PLS	**0.618**	0.295	**0.469**
9. GVAR	0.765	0.288	0.556
10. augmented GVAR	0.788	0.256	0.563
AveM (4-10)	0.690	0.234	0.495
AveMAveW	0.725	0.235	0.517

RMSFE deteriorates as horizon increases

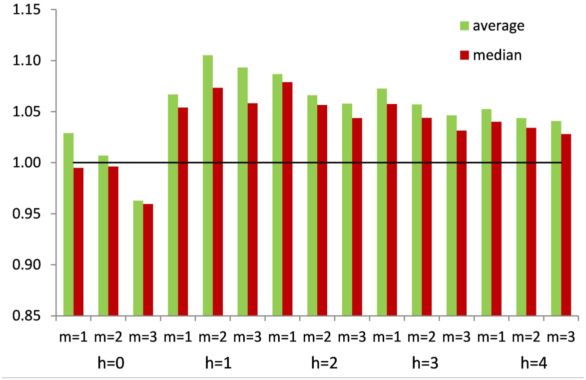
(RMSFE of global growth forecasts is computed using 06Q1-13Q3 evaluating period)



Augmented GVAR tends to outperform standard GVAR

Average and median RMSFE of the GVAR model relative to the augmented GVAR model for different horizons

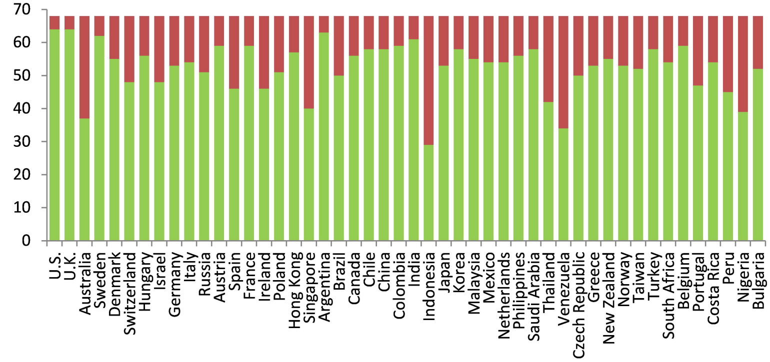
Averages and medians are computed from RMSFEs across 48 countries in the sample, based on the full 2006Q1-2013Q3 evaluating period.



Relative performance of AveM (h=0,m=1)

Number of methods (y-axis) outperformed by AveM (light/green bars) and underperformed by AveM (dark/red bars) across countries (x-axis); $h = 0$, $m = 1$

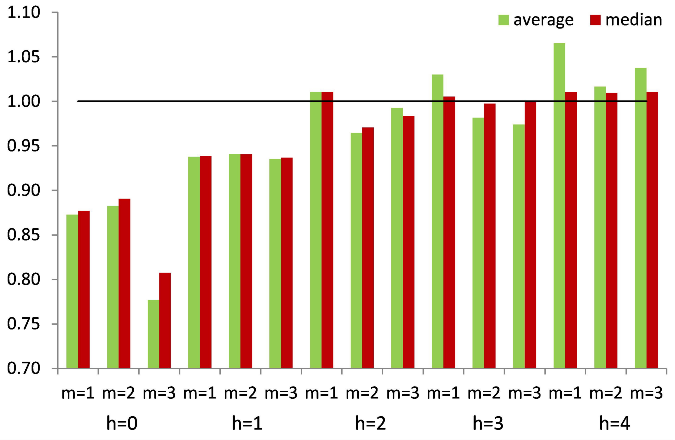
There are 68 individual methods in total on the y-axis (4 benchmarks and 64 data-rich methods). The forecasting performance is measured by RMSFE based on the full 2006Q1-2013Q3 evaluating period.



PMIs improve growth forecasts substantially, but only for horizon h=0,1

.

Average and median (across methods) of RMSFEs of global growth forecasts relative to the model without PMIs



Conclusions

- This paper shows that the presence of unobserved common factors can lead to an undetermined GVAR system.
- To solve this problem, we propose augmenting the GVAR with additional proxy equations for the unobserved factors and establish that such augmentation can produce forecasts that converge to infeasible optimal forecasts as the panel dimensions expand at similar rates.
- We empirically investigates the information content of PMIs for nowcasting and forecasting output growth across 48 countries using a variety of data-rich methods.
- We find that the augmented GVAR tends to outperform the standard GVAR model. We also find that PMIs substantially improve nowcasts of growth for horizon $h = 0$ (but not for higher horizons) - gaining about 23% reduction in the RMSE compared to models without PMI data.
- This is still a work in progress and we plan to consider techniques that down-weight observations as a way of dealing with the structural breaks - we shall also consider other forecast evaluation criteria to check the robustness of our conclusions.