Comment:
"Forecasting Economic and Financial Variables with Global VARs" by M. Hashem Pesaran, Till Schuermann and L. Venessa Smith.

by

Kajal Lahiri,
University at Albany, SUNY, Albany, NY 12222.
klahiri@albany.edu

This paper reports the near term forecasting power of a large Global Vector Autoregressive (GVAR) model originally developed by Pesaran, Schuermann and Weiner, PSW, (2004) and subsequently fine-tuned and re-estimated over 1979Q1-2003Q4 by Dees, di Mauro, Pesaran and Smith, DdPS, (2007).[1] The GVAR model explicitly specifies interdependencies between different countries and sub-regions in terms of three transparent channels: i) domestic variables are related to corresponding trade-weighted foreign variables to match the international trade pattern of the country under consideration; ii) non-zero pair wise correlations in residuals between countries and equations are allowed to capture a certain amount of dependence in idiosyncratic shocks, and iii) common observed shocks (*viz*, oil prices) that can affect all countries simultaneously are permitted. The key feature of the GVAR model as compared to factor structural VAR models is the direct introduction of observed country-specific foreign variables in the individual country models to deal with pervasive dependencies in the world economy in a more flexible manner. The authors first estimate the individual county-specific vector error correcting models, then these individual models are combined in an internally consistent manner to generate forecasts for all variables of all countries simultaneously. The 1-quarter and 4-quarter ahead forecasts for the eight quarters 2004Q1-2005Q4 are analyzed for 134 variables from 26 regions made up of 33

[1] See also Baltagi (2004), Dennis and Lopez (2004), Johansen (2004) and Wallis (2007), and the rejoinder by Pesaran et al. (2004).

countries covering about 90% of world output. The seven macro variables considered are real output, inflation, short term and long term interest rates, stock market indices, exchange rates, and crude oil prices.

The forecasting work reported in this paper represents a massive hallmark effort by the authors to study how modeling international interdependencies in a more comprehensive and flexible manner can improve near term macroeconomic forecasts. Since the specification and diagnostic tests of the model used in this paper have been meticulously developed and reported in Dees et al. (2007), I will restrict my comments only on the reported quality of the forecasts generated by the model.

The current paper reports two major innovations. First, in order to handle possible structural breaks in different variables at different times in any of the countries, the authors adopt a very pragmatic approach of forecast averaging, and analyze predictive errors from 19 alternative models over 10 different estimation windows, generating 190 alternative predictions. Secondly, since the number of out-of-sample forecast periods is only 8, they generalize the predictive superiority test of Diebold and Mariano (1995) to a panel data of many countries. Their central finding is that the "AveAve" forecasts from the GVAR, computed as the double averages of forecasts from different models estimated over different observation windows are in general better than forecasts from a single GVAR model estimated over a single estimation window or from some simple benchmark models like the random walk and AR(1) models with and without drifts.

More specifically, Pesaran et al. (2009) compare the root mean squared forecast error (RMSFE) of the average forecast with forecasts from an individual model using a panel version of the Diebold-Mariano (1995) test. However, in what follows, I will argue that the true underlying uncertainty of the average forecast that is needed to construct its confidence interval is not solely determined by the squared error of the average forecast; rather it is determined by the average variance of the individual forecast errors. The wedge between the two is the variance of point forecasts of individual model forecasts – the price one inadvertently pays while pooling diverse forecasts. The underlying variability of the individual point forecasts do not get reflected in the squared forecast error associated with the average forecast. Simply put, the average of the variances of forecast errors from different models is not the same as the variance of the average error.

Note that RMSFE of *h*-period ahead individual forecasts can be decomposed as

$$\frac{1}{N}\sum_{i=1}^{N}(A_t - F_{ith})^2 = (A_t - F_{.th})^2 + \frac{1}{N-1}\sum_{i=1}^{N}(F_{ith} - F_{.th})^2 \qquad (1)$$

where $F_{.th} = \frac{1}{N}\sum_{j=1}^{N} F_{jth}$ is the average forecast from N models. The first term on the right hand side of equation (1) is the squared error associated with the average forecast, and the last term is the variance of the individual forecasts (d_{th}). Taking expectations conditional on a given a set of forecasts at time t, we get the following relationship between the aggregate forecast uncertainty $U_{th} = \mathrm{E}\left(\frac{1}{N}\sum_{i=1}^{N}\left(A_t - F_{ith}\right)^2\right)$, the variance of consensus forecast errors and disagreement among individual forecasts:

$$U_{th} = \mathrm{E}(A_t - F_{.th})^2 + d_{th}. \qquad (2)$$

Beginning with Zarnowitz and Lambros (1987), recent research on characterizing the appropriate measure of overall forecast uncertainty (U_{th}) of an aggregate forecast suggests that the average of the individual forecast error variances, i.e., $U_{th} \equiv \frac{1}{N}\sum_{i=1}^{N} U_{ith}$ where $U_{ith} = \mathrm{E}(A_t - F_{ith})^2$, is the correct measure and can be interpreted as the confidence an outside observer or a policy maker will have in a randomly drawn typical individual forecast from a panel of forecasts.[2] Note that Pesaran et al. (2009) consider only the first term $\mathrm{E}(A_t - F_{.th})^2$ in their comparisons, which will necessarily underestimate the true underlying variability of the pooled forecast compared to a forecast based on a single benchmark model. Note also that the first term on the right hand side of (2) can be written as

$$\mathrm{E}(A_t - F_{.th})^2 = \frac{1}{N^2}\mathrm{E}\left[\sum_{i=1}^{N}(A_t - F_{ith})^2\right] + \frac{1}{N^2}\mathrm{E}\left[\sum_{i=1}^{N}\sum_{j\neq i}^{N}(A_t - F_{ith})(A_t - F_{jth})\right], \quad (3)$$

where the second term on the right hand side of equation (3) is the average covariance of the individual idiosyncratic forecast errors. Using single factor decomposition of forecast errors, Lahiri and Sheng (2009) have shown that, with a reasonably large *N*, equation (3) reduces to $\sigma^2_{\lambda|th}$, the anticipated variance of the forthcoming aggregate shock over the forecast horizon. Thus, we can derive a very simple relationship between aggregate forecast uncertainty, variance of aggregate shock and disagreement as:

$$U_{th} = \sigma^2_{\lambda|th} + d_{th}. \quad (4)$$

[2] *cf.* Davies and Lahiri (1995), Lahiri and Sheng (2009), Boero et al. (2008), and Giordani and Söderlind (2003). See also Reifschneider and Tulip (2007).

The above equation explains that the RMSFE of the average forecast will be significantly less that the average RMSFE of the individual forecasts when the variability of the individual forecasts is big. Thus, when Pesaran et al. (2009) find that for real GDP growth or inflation, the RMSFE of AveAve forecast is the very low compared with all individual RMSFEs, it simply means that for these variables the individual models generated more diverse forecasts. It does not imply that the precision of the average forecast from the stand point of a policy maker is correspondingly less than that of forecasts from a single model. However, what still remains intriguing is why RMSFE for 4-quarter forecasts reported in the paper are uniformly and substantially lower than those for 1-quarter ahead forecasts across almost all country groups and target variables.

Another way of looking at the unfairness of the comparison of RMSFE of AveAve forecasts with individual benchmark forecasts is that the uncertainty of the estimated parameters are not being factored in these comparisons. Since the AveAve forecasts are based on a huge number of additional parameter estimates compared to a single model, it is expected that the true forecast variability of the average forecast will be considerably higher than the RMSFE of average forecasts. A true forecast superiority comparison should incorporate these considerations. Given the size and diversity of models and the paucity of available out-of-sample periods, it is possibly difficult to derive asymptotic RMSFE expressions, but an appropriate forecast comparison can be accomplished by model simulation and by bootstrapping forecast variability, see West (2006). A simpler alternative will be to take the square root of the sum of the variance of individual model forecasts and the MSFE of the combined forecast before comparing them with RMSFE of forecasts from individual models. Without these adjustments, the claimed forecasting superiority of the GVAR model can not be established.

Having said this, it will be very interesting to compare forecasts from the best GVAR model specification from either PSW (2004) or DdPS (2007) with the chosen benchmark forecasts using the panel Diebold-Mariano test. This way, one can study the relative forecasting efficiency of a preferred GVAR model rather than merely establishing the power of forecast combination. As the authors correctly noted, compared to a well specified GVAR model, naïve benchmark models have very little value for counter-factual policy analysis. By the same token, policy simulation based on an indiscriminate use of 190 models rather than a well specified single GVAR model will be a hard sell to a policy maker. We also suggest that the power characteristics of this new panel Diebold-Mariano test should be examined carefully in view of the assumed cross-sectional independence of the candidate forecasts – a task that, I believe, can be easily accomplished by the authors.

References:

Baltagi, B.H. 2004, Comment, Journal of Business and Economic Statistics, 22, 163-164.

Boero G, J. Smith and K.F. Wallis, 2008, Uncertainty and disagreement in economic prediction: the Bank of England Survey of External Forecasters, *Economic Journal* 118: 1107-1127.

Davies A. and K. Lahiri, 1995, A new framework for analyzing survey forecasts using three-dimensional panel data, *Journal of Econometrics,* 68: 205-227.

Diebold, F.X. and R. Mariano, 1995, Comparing predictive efficiency, *Journal of Business and Economic Statistics*, 13, 253-265.

Giordani, P. and P. Söderlind, 2003, Inflation forecast uncertainty. *European Economic Review,* 47: 1037-1059.

Johansen, S, (2004), Comment, *Journal of Business and Economic Statistics*, 22, 169-172.

Lahiri, K. and X. Sheng, (2009), Measuring uncertainty by disagreement: the missing link. *Journal of Applied Econometrics,* Forthcoming.

Pesaran, M.H., T. Schuermann and S.C. Weiner, (2004), Modeling regional interdependencies using a global error-correcting macroeconomic model, *Journal of Business and Economic Statistics*, 22, 129-162.

Pesaran, M.H., T. Schuermann and S.C. Weiner, (2004), Rejoinder, *Journal of Business and Economic Statistics*, 22, 175-181.

Reifschneider, D. and P. Tulip, 2007, Gauging the uncertainty of the economic outlook from historical forecasting errors. Federal Reserve Board, Finance and Economics Discussion Series: 60.

Richard, D. and J.A. Lopez, 2004, Comment, *Journal of Business and Economic Statistics*, 22, 165-1169.

Wallis, K.F, 2004, Comment. *Journal of Business and Economic Statistics*, 22, 172-174.

West, K. D., (2006), Forecast Evaluation, In Elliott G. et al. (Eds.), *Handbook of Economic Forecasting*, Chapter 3, North-Holland, 99-134.

Zarnowitz, V. and L.A. Lambros, (1987), Consensus and uncertainty in economic prediction, *Journal of Political Economy*, 95, 591-621.