International Journal of Forecasting 19 (2003) 87–94

international journal of forecasting

www.elsevier.com/locate/ijforecast

Forecasting combination and encompassing tests

Yue Fang*

University of Oregon, Lundquist College of Business, Eugene, OR 97403, USA

Abstract

In this paper we demonstrate that forecast encompassing tests are valuable tools in getting an insight into why competing forecasts may be combined to produce a composite forecast which is superior to the individual forecasts. We also argue that results from forecast encompassing tests are potentially useful in model specification. We illustrate this using forecasts of quarterly UK consumption expenditure data from three classes of models: ARIMA, DHSY and VAR models.

Keywords: ARIMA; Econometric model; Forecast encompassing; Forecast combination; VAR

1. Introduction

Many methods are used to forecast economic activity. These methods often differ in structure and in data used. They provide different insights into time series dynamics and their forecasts are not perfectly correlated with each other. A considerable literature has developed on combining competing forecasts, that is, taking some (usually linear) combination of multiple individual forecasts from different models to form a new forecast. The combined forecast often leads to increased forecast accuracy, usually being defined as a smaller root mean squared error (RMSE) than any of the individual model forecasts. See Clemen (1989) for a comprehensive bibliography review on forecast combination.

Despite its appealing aspect of easy use, this simple mechanical combination approach, however, must be exercised with caution. Firstly, a formal examination of competing forecasts should be taken to assess whether the smaller RMSE of the combined forecast can be attributed to sampling variability or whether the combined forecast is indeed superior due to the fact that no individual forecast incorporates all the relevant information so that pooling forecasts from different models leads to a better forecast. Uncertainty about the answer to this question often leads to some doubt about implementation of the forecast combination approach and complicates interpretation of combined forecast results. Secondly, the forecast combination approach implicitly acknowledges the possibility of model misspecification. In cases where models are misspecified, as many studies suggested (see, e.g., Chong and Hendry (1986), Ericsson (1989), Diebold (1989), and Clements and Hendry (1998) among others), the primary goal for forecasters should be to improve the model specification by combining information rather than to pool forecasts from misspecified models, because if an acceptable specification can be found, then optimal forecasts follow automatically. See de Menezes

*Tel.: +1-541-346-3265.

E-mail address: yfang@darkwing.uoregon.edu (Y. Fang).

0169-2070/02/$ – see front matter
PII: S0169-2070(01)00121-2

and Bunn (1998) and Taylor and Bunn (1999) for discussions on combined forecast error distributions.

In this paper we demonstrate that forecast encompassing tests are valuable tools in getting an insight into why competing forecasts may be combined to produce a composite forecast which is superior to the individual forecasts. We also argue that results from forecast encompassing tests are potentially useful in model specification. We do this for forecasts of quarterly UK consumption expenditure data from 1957(1) to 1975(4). The models include various versions of autoregressive integrated moving-average (ARIMA) models (Box, Jenkins, & Reinsel, 1994), models based on the work of Davidson, Hendry, Srba and Yeo (1978) (DHSY hereafter) and vector autoregression (VAR) models (Sims, 1980). Because all forecasting models are likely to be misspecified — they are approximations to a much more complex reality — the results in this paper should have practical value.

2. Forecasting models for consumption expenditure

The modelling of consumption is at the centre of the Keynesian economic theory and the consumer's expenditure is one of the most important aggregates to predict in macroeconomics. A great deal of research has been devoted to understanding the aggregate consumer's expenditure and progress continues at a rapid pace after more than a half century since the advances in Keynesian theory in the 1930s and 1940s. For example, we have witnessed the evolution of the nonstructural and structural approaches to macroeconomic forecasting (Diebold, 1998); recent methodological developments in econometrics such as VAR models, cointegration and Granger causality; and the ongoing debate about whether one should or should not be concerned about the 'atheoretical' nature of the VAR approach (see, e.g., Canova, 1999).

Here we consider a total of nine models for UK quarterly consumption expenditure data from 1957(1) to 1975(4): three ARIMA models, four DHSY models and two VAR models. The specifications of the three ARIMA models are based on the results of Prothero and Wallis (1976) using the quarterly UK consumption data with a shorter time period (from 1957(3) to 1967(4)). The three ARIMA models (TS1–TS3) are

$$\text{TS1: } \Delta\Delta_4 X_t = (1-\theta_4 L^4)a_t, \tag{1}$$

$$\text{TS2: } (1-\phi_1 L)\Delta_4 X_t = (1-\theta_4 L^4)a_t, \tag{2}$$

and

$$\text{TS3: } (1-\phi_1 L)(1-\Phi_4 L^4 - \Phi_8 L^8)X_t = a_t, \tag{3}$$

where X_t is the natural logarithm of UK quarterly seasonally unadjusted non-durable consumption expenditure in millions of pounds at 1970 prices (see, e.g., Charemza and Deadman, 1997) and a_t is a Gaussian white noise process. The symbols Δ and Δ_4 denote difference operators ($\Delta X_t = X_t - X_{t-1}$ and $\Delta_4 X_t = X_t - X_{t-4}$) and L is the back-shift operator ($LX_t = X_{t-1}$).

The second class of models are based on the restricted DHSY model (Davidson et al., 1978):

$$\Delta_4 X_t = \beta_1 \Delta_4 Z_t + \beta_2 \Delta\Delta_4 Z_t + \gamma\Delta_4 DO_t + \delta(X_{t-4} - Z_{t-4}) + a_t, \tag{4}$$

where Z_t is the UK personal disposable income in 1970 prices, measured in the natural logarithms. The variable DO is a 'special effects' dummy variable reflecting changes in indirect taxation (1968(1) (2) and 1973(1) (2)).[1]

We consider four versions of DHSY models based on (4). The difference lies in the way the exogenous variable Z_t is handled. Ex-ante forecasts of X_t are obtained using predicted Z_t from four ARIMA processes for the UK personal disposable income (Prothero & Wallis, 1976):

$$\Delta\Delta_4 Z_t = (1-\theta_4 L^4)a_t, \tag{5}$$

$$(1-\phi_1 L)\Delta_4 Z_t = (1-\theta_4 L^4)a_t, \tag{6}$$

$$(1-\phi_1 L)(1-\Phi_4 L^4)\Delta_4 Z_t = a_t, \tag{7}$$

and

[1] The DHSY model had an important influence on recent developments in applied econometric analysis and delineated the framework for many subsequent empirical research. See Hendry, Muellbauer and Murphy (1990) and Charemza and Deadman (1997) for discussions on the econometrics of DHSY.

$$(1 - \Phi_4 L^4 - \Phi_8 L^8)\Delta Z_t = a_t. \quad (8)$$

If (4) is used, but Z_t for the forecasting period is estimated from (5)–(8), then we refer to the DHSY models as DHSY1, DHSY2, DHSY3 and DHSY4.

We also consider the VAR method which allows for cross-variable dynamics. Since the time series examined in this paper is relatively short, we consider two restricted VAR models of order 5 focusing only on lags 1, 4 and 5, defined as follows:

$$\text{VAR1: } \Delta_1 \mathcal{Y}_t = A_0 + A_1\Delta_1 \mathcal{Y}_{t-1} + A_4\Delta_1 \mathcal{Y}_{t-4} + A_5\Delta_1 \mathcal{Y}_{t-5} + \mathcal{E}_t, \quad (9)$$

$$\text{VAR2: } \Delta_1 \mathcal{Y}'_t = A'_0 + A'_1\Delta_1 \mathcal{Y}'_{t-1} + A'_4\Delta_1 \mathcal{Y}'_{t-4} + A'_5\Delta_1 \mathcal{Y}'_{t-5} + \mathcal{E}'_t, \quad (10)$$

where $\mathcal{Y}_t = (X_t, Z_t)'$ and $\mathcal{Y}'_t = (X_t, Z_t, U_t)'$. U_t is the consumers' expenditure deflator index in 1970 prices, measured in the natural logarithms. $\mathcal{E}_t$ and $\mathcal{E}'_t$ are error terms. There are a total of 14 parameters in (9) and 30 parameters in (10).

3. Forecasts and their combinations

We use the first 10 years (1957(1)–1966(4)) as the 'fitting' period and the last 9 years (1967(1)–1975(4)) as the 'forecasting' period. For illustrative purposes, we consider only one-quarter-ahead forecasts based on models which are estimated using all data available at the forecasting time; this rolling approach allows each model to be estimated 36 times and yields 36 one-quarter-ahead forecasts.

The diagonal entries in panel A of Table 1 are RMSEs of one-quarter-ahead forecasts over the 9 years forecasting period for each of the nine models. To obtain a perspective on the robustness of the results to different evaluation measures, we also report the mean absolute errors (MAEs) and root mean squared percent errors (RMSPEs) for individual forecasts in panels B and C (diagonal entries), respectively. As can be seen, the three ARIMA, four DHSY and two VAR models have similar forecasting performances (measured by either RMSE, MAE or RMSPE). The slight differences in forecasts arise either because the models use different data or because they impose different restrictions on the reduced form.

To provide a common perspective on forecast combination for these nine models, we combine each pair of forecasts (using equal weights) and compute RMSEs (MAEs and RMSPEs) for the combined forecasts (off-diagonal entries in Table 1), regardless of whether one forecast is significantly better than the others. To make an easy evaluation, we mark the combined forecasts with asterisks if their RMSE in panel A (MAE in panel B or RMSPE in panel C) are at least 5% (say) less than those of two corresponding individual forecasts.

Results from Table 1 indicate that the combined forecasts from the ARIMA and DHSY models appear to have smaller forecast errors; as all cases in panels A through C demonstrated, the RMSE (MAE and RMSPE) of the combined forecasts from the ARIMA and DHSY models are at least 5% less than those of the corresponding individual forecasts. Some combined forecasts from VARs and ARIMA (or DHSY) models appear to be better than corresponding individual forecasts. The view that pooling forecasts from three classes of models (i.e. ARIMA models, DHSY models and VAR models) may lead to better forecasts is confirmed by the forecast encompassing analysis in Section 4. We demonstrate there that all nine models are misspecified. More specifically, several models may not effectively predict the dynamics of ΔX_t, and some ARIMA models, DHSY models and in particular, two VAR models fail in capturing all information relevant to forecasting $\Delta\Delta_4 X_t$.

4. Forecast encompassing

Forecast encompassing tests seek to evaluate whether competing forecasts may be fruitfully combined to produce a forecast superior to individual forecasts. Such tests can be implemented by regressing the actual level of X_t (or the change in X_t) on the predicted X_ts (or the predicted changes) from two models. For example, one may consider the regression model (Chong & Hendry, 1986; Ericsson, 1993)

$$X_t = \alpha_1 \hat{X}^{(I)}_{t-s,t} + \alpha_2 \hat{X}^{(II)}_{t-s,t} + u_t \quad (11)$$

Table 1
Forecast accuracy of individual and combined forecasts

	TS1	TS2	TS3	DHSY1	DHSY2	DHSY3	DHSY4	VAR1	VAR2
Panel A: RMSE									
TS1	0.0131	0.0124	0.0125	0.0115*	0.0119*	0.0120*	0.0120*	0.0121*	0.0124*
TS2		0.0130	0.0127	0.0115*	0.0123*	0.0123*	0.0120*	0.0120*	0.0121*
TS3			0.0135	0.0116*	0.0124*	0.0124*	0.0121*	0.0120*	0.0122*
DHSY1				0.0123	0.0127	0.0128	0.0128	0.0113*	0.0117
DHSY2					0.0134	0.0135	0.0133	0.0118*	0.0121*
DHSY3						0.0137	0.0135	0.0118*	0.0121*
DHSY4							0.0135	0.0117*	0.0121*
VAR1								0.0126	0.0127
VAR2									0.0135
Panel B: MAE									
TS1	0.0110	0.0099*	0.0105	0.0087*	0.0089*	0.0090*	0.0090*	0.0094	0.0098
TS2		0.0108	0.0104	0.0087*	0.0093*	0.0093*	0.0090*	0.0094	0.0093*
TS3			0.0111	0.0090*	0.0095*	0.0096*	0.0092*	0.0095	0.0095*
DHSY1				0.0096	0.0093	0.0100	0.0099	0.0088*	0.0091*
DHSY2					0.0103	0.0103	0.0102	0.0089*	0.0091*
DHSY3						0.0104	0.0103	0.0089*	0.0092*
DHSY4							0.0102	0.0089*	0.0093*
VAR1								0.0095	0.0098
VAR2									0.0103
Panel C: RMSPE									
TS1	0.00147	0.00139	0.00140	0.00129*	0.00134*	0.00134*	0.00134*	0.00136	0.00134*
TS2		0.00146	0.00142	0.00129*	0.00137*	0.00137*	0.00134*	0.00134*	0.00135*
TS3			0.00151	0.00130*	0.00139*	0.00139*	0.00136*	0.00135	0.00137*
DHSY1				0.00137	0.00142	0.00143	0.00144	0.00127*	0.00131
DHSY2					0.00150	0.00151	0.00149	0.00132*	0.00136*
DHSY3						0.00153	0.00151	0.00132*	0.00135*
DHSY4							0.00151	0.00132*	0.00136*
VAR1								0.00142	0.00143
VAR2									0.00152

In panel A * indicates that the RMSE of the combined forecast is at least 5% less than those of two individual forecasts. In panel B * indicates that the MAE of the combined forecast is at least 5% less than those of two individual forecasts. In panel C * indicates that the RMSPE of the combined forecast is at least 5% less than those of two individual forecasts.

and test for $\alpha_1 = 1$ (or $\alpha_2 = 0$) conditional on $\alpha_1 + \alpha_2 = 1$. $\hat{X}^{(I)}_{t-s,t}$ is the forecast of X_t made from model I using information available at time $t - s$ and $\hat{X}^{(II)}_{t-s,t}$ is the same thing for model II.

This is a variant of the test due to Fair and Shiller (1989, 1990), which had the form of:

$$\Delta_s X_t = \alpha_0 + \alpha_1(\hat{X}^{(I)}_{t-s,t} - X_{t-s}) + \alpha_2(\hat{X}^{(II)}_{t-s,t} - X_{t-s}) + u_t. \tag{12}$$

When $\alpha_1 = 0$ and $\alpha_2 \neq 0$, the second model forecast encompasses the first. On the contrary, the first model forecast encompasses the second if $\alpha_1 \neq 0$ and $\alpha_2 = 0$. In the case that both forecasts contain independent information for one-quarter-ahead forecasting of X_t, α_1 and α_2 should both be nonzero.

The set-up of (12) is different to that in (11): α_1 and α_2 are not subject to the constraint that α_1 and α_2 sum to one. It is usually preferable not to force α_1 and α_2 to sum to unity for as Fair and Shiller argued, (12) may be more sensitive because there are cases in which the constraint ($\alpha_1 + \alpha_2 = 1$) does not make sense. For example, if forecasts from both models are just noise, the estimates of both α_1 and α_2 should be zero. Fair and Shiller considered $\Delta_s X_t$ instead of X_t because the time series of interest in their empirical study is nonstationary. The inclusion of an intercept is also desired since it facilitates bias correction and allows biased forecasts to be evaluated.

Similar to Fair and Shiller's (1989, 1990) approach, we consider forecast encompassing tests based on (12) with $s = 1$, and in addition the following regression model:

$$\Delta\Delta_4 X_t = \alpha_0 + \alpha_1\left[\left(\hat{X}^{(I)}_{t-1,t} - X_{t-1}\right) - \left(\hat{X}^{(I)}_{t-5,t-4} - X_{t-5}\right)\right] + \alpha_2\left[\left(\hat{X}^{(II)}_{t-1,t} - X_{t-1}\right) - \left(\hat{X}^{(II)}_{t-5,t-4} - X_{t-5}\right)\right] + u_t. \tag{13}$$

We note that (12) and (13) use different regressands. The two procedures provide different insights into which information, relevant to forecasting X_t in forecasts from one model, is not in forecasts from another model.

In (12), the information contained in one model's forecast compared to that in another is assessed from a regression of the actual changes on predicted changes from two models. An advantage of using ΔX_t in comparison to, say, $\Delta\Delta_4 X_t$ as a regressand, is its simplicity. Note that if the second model forecast encompasses the first (i.e. $\alpha_1 = 0$), as shown below, $\alpha_0 \approx 0$ and $\alpha_2 \approx 1$. Hence, $\hat{X}^{(II)}_{t-1,t}$s account for almost all information contained in out-of-sample X_ts. If both forecasts contain independent information for one-quarter-ahead forecasting of X_t, then values of α_1 and α_2 build valuable intuition — the relative importance of forecasts from two models in explaining information contained in X_t. We note that although the constraint ($\alpha_1 + \alpha_2 = 1$) is not imposed, the estimated values of α_1 and α_2 often sum nicely into a number close to one for our data (see Table 2). One reason for this is that all regressors have strong correlations with the regressand (the correlation coefficients are around 0.9).

The main disadvantage of tests based on (12) is that if two models contain the same information, then the forecasts are highly corrected, and so α_1 and α_2 are not separately identified due to the severe collinearity problem. As an ad hoc solution to the collinearity problem, we use ridge regression to obtain estimation of (12).

In contrast, the regression (13) uses $\Delta\Delta_4 X_t$ as the regressand, which is obtained using both the first and the (quarterly) seasonal differences. Focusing on $\Delta\Delta_4 X_t$ allows one to go a step further to study forecast encompassing by removing more common components from two regressors. In a set-up like (13), forecast encompassing hypotheses can be tested using the standard regression methods.

Tables 2 and 3 report the estimation results of (12) and (13) for each pair of the nine models, respectively. The ridge regression estimates of α_0, α_1 and α_2 are given in Table 2.[2] Table 3 represents ordinary least squares estimates with asterisks indicating that the corresponding tests are statistically different from zero at the 5% level of significance. We correct for both heteroskedasticity and the moving average process in the estimation of the standard errors of the coefficient estimates in Table 3 using the procedure given by Hansen (1982) and White (1980). See Fair and Shiller (1990) for detailed formulae.

In examining the results from Tables 2 and 3, four general conclusions emerge. Firstly, three ARIMA models, four DHSY models and two VAR models each contain independent information relevant to forecasting X_t. Neither class encompasses the other. For example, both estimated values of α_1 and α_2 in (12) appear not to be zero for forecasts from any ARIMA models (as model I) and DHSY models (as model II); as the case of TS2 vis-à-vis four DHSY models demonstrates, the values of α-coefficient for TS2 ranges from 0.360 to 0.568, whereas those for four DHSY models are bounded by 0.338 and 0.543. This implies that for one-quarter-ahead forecasting horizon, forecast combination of the ARIMA, the DHSY and the VAR models is desirable. This result is consistent with the finding from Table 1.

Secondly, the forecast encompassing approach is complementary to the RMSE (or other criteria such as MAE and RMSPE), since, as demonstrated in results in Tables 1–3, the forecast encompassing approach can often discriminate two models even when the RMSEs are close for two forecasts and determine whether the forecast with the higher RMSE contains information not in the other forecast.

[2]The biasing constant in the ridge regression is chosen to be from 0.005 to 0.01, which produce stable regression coefficients and sufficiently small VIF values. Because the ordinary inference procedures are not applicable in ridge regression analysis due to the lack of knowledge of the exact distributional properties of test statistics, we do not give significance levels of estimated coefficients. See Neter, Kutner, and Nachtsheim (1996) for a textbook level introduction to ridge regression.

Table 2
Comparison of forecasts: estimates of (12)

	CONST	TS1	TS2	TS3	DHSY1	DHSY2	DHSY3	DHSY4	VAR1	VAR2
1	0.004	0.094	0.807							
2	0.004	0.025		0.872						
3	0.003	0.255			0.652					
4	0.001	0.148				0.753				
5	0.002	0.179					0.720			
6	0.003	0.364						0.547		
7	0.005	−0.065							1.086	
8	0.003	−0.082								1.062
9	0.003		0.355	0.549						
10	0.003		0.478		0.428					
11	0.002		0.360			0.543				
12	0.002		0.381				0.521			
13	0.003		0.568					0.338		
14	0.005		0.110						0.788	
15	0.004		−0.038							0.990
16	0.003			0.527	0.379					
17	0.002			0.416		0.487				
18	0.002			0.432			0.470			
19	0.003			0.568				0.338		
20	0.004			0.172					0.727	
21	0.004			0.009						0.942
22	0.001				0.271	0.628				
23	0.001				0.294		0.604			
24	0.002				0.632			0.269		
25	0.004				0.153				0.748	
26	0.004				−0.016					0.968
27	0.001					0.473	0.424			
28	0.001					0.798		0.098		
29	0.004					0.256			0.646	
30	0.003					0.087				0.863
31	0.001						0.788	0.105		
32	0.004						0.261		0.641	
33	0.003						0.102			0.847
34	0.004							0.095	0.804	
35	0.004							−0.058		1.010
36	0.004								0.083	0.866

For example, DHSY3 and DHSY4 encompass TS1 if one focuses on $\Delta\Delta_4 X_t$ (Table 3), although DHSY3 and DHSY4 have slightly higher RMSEs than that of TS1 (Table 1). (See also Armstrong and Collopy (1992), and Clements and Hendry (1993) on limitations of MSE error measures.)

Furthermore, our results extend and support findings reported in previous studies on the forecast accuracy comparison between different models primarily based on criteria such as RMSE, MAE or RMSPE. There were many studies on forecast accuracy comparison within the class of ARIMA models or of econometric models using similar data sets. Only a few studies on comparisons between two classes can, however, be found in the literature and the results are mixed. For example, Nelson (1972) considered one-quarter-ahead forecasts of 14 economic variables and found that the time series model performs better in ex-ante forecast comparison, but within the sample the econometric model is ahead. On the other hand, Christ (1975) reported comparisons of forecasts of real and nominal GNP in which ARIMA forecasts are ‘uniformly the poorest’. In both the Nelson and Christ studies, a simple loss function, RMSE, is used to provide an overall forecasting accuracy measure. Our results show that

Table 3
Comparison of forecasts: estimates of (13)

	CONST	TS1	TS2	TS3	DHSY1	DHSY2	DHSY3	DHSY4	VAR1	VAR2
1	−0.001	−0.901	0.147							
2	0.001	−0.657		−0.145						
3	−0.001	−0.743*			0.345*					
4	−0.001	−0.717*				0.355*				
5	−0.001	−0.699					0.336*			
6	0.001	−0.691						0.339*		
7	0.001	−1.151*							0.220	
8	0.000	−0.914								0.086
9	−0.001		1.054	−1.409						
10	−0.001		−0.414		0.380*					
11	−0.001		−0.591*			0.478*				
12	−0.001		−0.562				0.440*			
13	0.001		−0.404					0.386*		
14	0.001		−0.315						−0.022	
15	−0.001		−0.280							−0.059
16	−0.001			−0.591*	0.406*					
17	−0.002			−0.739*		0.512*				
18	−0.002			−0.696*			0.461*			
19	0.001			−0.404				0.386*		
20	−0.001			−0.512					0.033	
21	−0.001			−0.460						−0.024
22	0.001				0.256	0.122				
23	0.001				0.153		0.227			
24	−0.001				−0.059			0.425		
25	−0.001				0.408*				−0.206	
26	−0.001				0.395*					−0.185
27	0.001					−0.079	0.435			
28	−0.001					−0.013		0.380		
29	−0.001					0.442*			−0.214	
30	−0.002					0.432*				−0.198
31	−0.001						0.065	0.308		
32	0.001						0.390*		−0.166	
33	−0.002						0.389*			−0.166
34	−0.001							0.389*	−0.159	
35	0.001							0.385*		−0.156
36	−0.001								0.018	−0.135

*Statistically significant at the 5% level.

ARIMA, DHSY and VAR classes of models contain independent information on forecasting quarterly UK consumption expenditure. The forecast encompassing tests can discriminate well between them.

Finally, the finding from forecast encompassing tests can be viewed as prima facie evidence of model misspecification. In particular, one may test misspecification via different forms of regressions such as (12) and (13). For example, the results from (13) seem highly suggestive of the possibility that neither the ARIMA models, DHSY models, nor VARs are acceptable in modeling the component $\Delta\Delta_4 X_t$. We note that it may be important to consider all most relevant forecast encompassing tests because one test may be complementary to another. For example, the estimation results of (12) suggest that VAR2 encompasses all other models. The results from Table 3 based on the estimation of (13) indicate, however, that TS1 and four DHSY models encompass VAR2. The contrast is clear, depending on whether one should focus on ΔX_t or $\Delta\Delta_4 X_t$. Since both ΔX_t and $\Delta\Delta_4 X_t$ play a key role in modeling the quarterly UK consumption expenditure data, any information on possible misspecification on those two terms should prove to be valuable in improving model specification.

Acknowledgements

The paper has benefited from comments from Wilpen L. Gorr (an editor), the associate editor, and the anonymous referee, as well as from seminar participants at the 2000 ASA annual meeting.

References

Armstrong, J. S., & Collopy, F. (1992). Error measures for generalizing about forecasting methods: empirical comparisons. *International Journal of Forecasting*, *8*, 69–80.

Box, G., Jenkins, G., & Reinsel, G. (1994). In 3rd ed, *Time Series Analysis: Forecasting and Control*. New Jersey: Prentice Hall.

Canova, F. (1999). Vector autoregressive models: specification, estimation, inference, and forecasting. In Pesaran, M. H., & Wickens, M. R. (Eds.), *Macroeconomics, Handbook of Applied Econometrics*, vol. 1. Oxford: Blackwell, pp. 73–138.

Charemza, W. W., & Deadman, D. F. (1997). In 2nd ed, *New Directions in Econometric Practice*. Cheltenham: Edward Elgar.

Chong, Y. Y., & Hendry, D. F. (1986). Econometric evaluation of linear macroeconomic models. *Review of Economic Studies*, *53*, 671–690.

Christ, C. F. (1975). Judging the performance of econometric models of the US economy. *International Economic Review*, *16*, 54–74.

Clemen, R. (1989). Combining forecasts: a review and annotated bibliography. *International Journal of Forecasting*, *5*, 559–583.

Clements, M. P., & Hendry, D. F. (1993). On the limitations of comparing mean squared forecast errors. *Journal of Forecasting*, *12*, 617–637.

Clements, M. P., & Hendry, D. F. (1998). *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.

Davidson, J. E. H., Hendry, D. F., Srba, F., & Yeo, S. (1978). Econometric modelling of the aggregate time series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal*, *88*, 661–692.

de Menezes, L., & Bunn, D. (1998). The persistence of specification problems in the distribution of combined forecast errors. *International Journal of Forecasting*, *14*, 415–426.

Diebold, F. (1989). Forecast combination and encompassing: reconciling two divergent literatures. *International Journal of Forecasting*, *5*, 589–592.

Diebold, F. (1998). The past present, and future of macroeconomic forecasting. *Journal of Economic Perspectives*, *12*(Spring), 175–192.

Ericsson, N. R. (1989). Parameter constancy, mean squared forecasts errors, and measuring forecast performance: an exposition and extensions. In Ericsson, N., & Irons, J. (Eds.), *Testing Exogeneity*. Oxford: Oxford University Press.

Ericsson, N. R. (1993). Comment on 'On the limitations of comparing mean squared forecast errors' by M.P. Clements and D.F. Hendry. *Journal of Forecast*, *12*, 644–651.

Fair, R. C., & Shiller, R. (1989). The informational context of ex ante forecasts. *The Review of Economics and Statistics*, *71*, 325–331.

Fair, R. C., & Shiller, R. (1990). Comparing information in forecasts from econometric models. *The American Economic Review*, *80*, 375–389.

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50*, 1029–1054.

Hendry, D. F., Muellbauer, J., & Murphy, A. (1990). The econometrics of DHSY. In Hey, J., & Winch, D. (Eds.), *A Century of Economics: 100 years of the Royal Economic Society and the Economic Journal*. Oxford: Basil Blackwell.

Nelson, C. R. (1972). The prediction performance of the FRB-MIT-PENN model of the US economy. *The American Economic Review*, *62*, 902–917.

Neter, J., Kutner, M., Nachtsheim, C., & Wasserman, W. (1996)). In 3rd ed, *Applied Linear Regression Models*. Chicago: IRWIN.

Prothero, D. L., & Wallis, K. F. (1976). Modelling macroeconomic time series. *Journal of Royal Statistical Society A*, *139*, 468–499.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, *48*, 1–48.

Taylor, J., & Bunn, D. (1999). Investigating improvements in the accuracy of prediction intervals for combinations of forecasts: a simulation study. *International Journal of Forecasting*, *15*, 325–339.

White, H. (1980). A heteroskedasticity consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*, 817–837.

Biography: Yue FANG is assistant professor of Decision Sciences at the Lundquist College of Business, University of Oregon. He received his Ph.D. from MIT and is author of a number of academic publications on time series modeling, economic forecasting and financial econometrics.