Evaluating Individual and Mean Non-Replicable Forecasts*

Chia-Lin Chang
Department of Applied Economics
National Chung Hsing University
Taichung, Taiwan

Philip Hans Franses
Econometric Institute
Erasmus School of Economics
Erasmus University Rotterdam
The Netherlands

Michael McAleer
Econometric Institute
Erasmus School of Economics
Erasmus University Rotterdam
The Netherlands
and
Tinbergen Institute
The Netherlands
and
Department of Quantitative Economics
Complutense University of Madrid

Revised: April 2011

* For financial support, the first author wishes to thank the National Science Council, Taiwan, and the third author wishes to thank the Australian Research Council, National Science Council, Taiwan, and the Japan Society for the Promotion of Science.

Abstract

Macroeconomic forecasts are often based on the interaction between econometric models and experts. A forecast that is based only on an econometric model is replicable and may be unbiased, whereas a forecast that is not based only on an econometric model, but also incorporates expert intuition, is non-replicable and is typically biased. In this paper we propose a methodology to analyze the qualities of individual and means of non-replicable forecasts. One part of the methodology seeks to retrieve a replicable component from the non-replicable forecasts, and compares this component against the actual data. A second part modifies the estimation routine due to the assumption that the difference between a replicable and a non-replicable forecast involves measurement error. An empirical example to forecast economic fundamentals for Taiwan shows the relevance of the methodological approach using both individuals and mean forecasts.

Keywords: Individual forecasts, mean forecasts, efficient estimation, generated regressors, replicable forecasts, non-replicable forecasts, expert intuition.

JEL Classifications: C53, C22, E27, E37.

1. Introduction

Econometric models are frequently used to provide base-level forecasts in macroeconomics. Usually, these model-based forecasts are adjusted by experts who possess intuition. For example, Franses, Kranendonk and Lanser (2011) document that this holds for all forecasts such as GDP and inflation generated from the large macroeconomic model created at the CPB Netherlands Bureau for Economic Policy Analysis. The difference between the pure model-based forecast and the final forecast is often called intuition or judgment. Intuition is a trade secret owned by a forecaster, as it is rarely written down, but it can have significant value in forecasting key economic fundamentals.

A forecast that is based on an econometric model is replicable and may be unbiased, whereas a forecast that is not based on an econometric model is non-replicable and is typically biased. In practice, most macroeconomic forecasts (from CPB, but also from the Federal Reserve, the World Bank, OECD and IMF) are non-replicable. In some cases, model-based forecasts are available and one can then derive their link with the final expert forecasts, but in many cases only the final forecast is available.

Indeed, CPB's forecasts are only available in their final form, and only by re-running the model could Franses, Kranendonk and Lanser (2011) quantify the expert intuition. In many cases, however, it may be unknown to the analyst if the forecaster has relied on the outcome of an econometric model, or even if an econometric model has been used. The analyst usually has only a forecast of an economic variable, and the analyst must then evaluate its quality. Various recent studies like Fildes, et al. (2009), Franses and Legerstee (2010) and Eroglu and Croxton (2010) have indicated that it is important to examine the behaviour of experts prior to evaluating forecast accuracy. In this paper, we pursue this line of research.

In this paper we examine the evaluation of the quality of a range of available non-replicable forecasts, with a specific focus on the individuals and mean values of potentially biased forecasts. For this, we propose a methodology that approaches this issue from two different angles. The first aims to de-bias the non-replicable forecast by retrieving and comparing their replicable components. The second approach modifies the estimation method.

In order to illustrate, we use data from Taiwan for three reasons. First, a consistent data set is available for the government and two professional quarterly forecasts of economic fundamentals over an extended period. Second, no previous comparison seems to have been made of the individual and mean competing forecasts. Third, there does not seem to have been any comparison of individual and mean forecasts based on an optimal subset of the alternative forecasts.

The plan of the remainder of the paper is a follows. Section 2 presents the econometric model specification, analyses replicable and non-replicable forecasts, considers optimal forecasts and efficient estimation methods, compares individual replicable forecasts with alternative means of replicable forecasts, and presents a direct test of an experts' added value. The data analysis and a relevant empirical example of alternative individual and mean forecasts of economic fundamentals for Taiwan are discussed in Section 3. Some concluding comments are given in Section 4.

2. Model Specification

In this section we present a method to evaluate non-replicable forecasts. First we deal with individual forecasts, and then we consider alternative mean forecasts.

2.1. Individual Forecasts

Consider a variable y as a T x 1 vector of observations to be explained (typically, an economic fundamental, such as the inflation rate or the real GDP growth rate), and assume that there are m forecasts X_i for this variable y, where $i = 1,2,\ldots,m$. In order to evaluate the quality of each individual forecast, one can consider the auxiliary regression

$$y = \alpha_i + \beta_i X_i + u_i \qquad (1)$$

where the error term has mean zero and common variance $\sigma_{u_i}^2$. The interest lies in the estimated values of α_i and β_i, where the true parameters are 0 and 1, respectively.

When the forecasts, X_i, would be fully based on an econometric model, then one can apply ordinary least squares (OLS) to (1) to estimate the parameters, α_i and β_i, and test their values against 0 and 1, respectively. However, when X_i is the end-product of the interaction between model output and an expert's intuition, OLS is not valid (see Franses et al. (2009)).

There are now two possible strategies to approach this issue. The first is to replace the X_i by a model-based forecast created by the analyst. Assume that this analyst has access to publicly available information contained in the T x k_i matrix W_i. The analyst can now run the regression

$$X_i = W_i \delta_i + \eta_i \tag{2}$$

where it is assumed that the first column of W_i concerns the intercept, and where the error term has mean zero and common variance $\sigma^2_{\eta_i}$. Applying OLS to (2) yields $\hat{X}_i$. In a next step, the analyst can replace (1) by

$$y = \alpha_i + \beta_i \hat{X}_i + u_i \tag{3}$$

As $\hat{X}_i$ in (3) is a generated regressor, the error term in (3) also contains a term with the measurement error η_i in (2), and hence when OLS is used, it is essential that the appropriate covariance matrix is computed, see Franses et al. (2009) for further details. An alternative is to apply OLS to (3) and to incorporate the Newey-West HAC covariance matrix estimator (see, for example, Smith and McAleer (1994)).

A second approach is to replace (1) by

$$y = \alpha_i + \beta_i (W_i \delta_i + \eta_i) + u_i \tag{4}$$

which can be written as

$$y = \alpha_i + \beta_i X_i + u_i - \beta_i \eta_i \tag{5}$$

for which it is clear that OLS is inconsistent for (5) as X_i is correlated with η_i. A simple solution is to use the Generalized Methods of Moments (GMM estimator).

2.2. Alternative Mean Forecasts

An alternative to evaluating the *m* forecasts individually is to use alternative mean forecasts, such as

$$\sum_{i=1}^{m} \lambda_i X_i \tag{6}$$

where λ_i are known constants. Typical constants would be $\lambda_i = \frac{1}{m}$, but also other variants are possible. The equivalent of (1) now becomes

$$y = \alpha + \beta \sum_{i=1}^{m} \lambda_i X_i + u \tag{7}$$

where the error term has mean zero and common variance σ_u^2.

The equivalent of (3) now becomes

$$y = \alpha + \beta \sum_{i=1}^{m} \lambda_i \hat{X}_i + \varepsilon \tag{8}$$

with

$$\varepsilon = u + \beta \sum_{i=1}^{m} \lambda_i (X_i - \hat{X}_i) \tag{9}$$

Given (2), we have

$$\hat{X}_i = W_i (W_i' W_i)^{-1} W_i' X_i = P_i X_i \tag{10}$$

Substituting (10) into (9) gives

$$\varepsilon = u + \beta \sum_{i=1}^{m} \lambda_i (W_i \delta_i - P_i X_i)$$

or equivalently

$$\varepsilon = u - \beta \sum_{i=1}^{m} \lambda_i P_i \eta_i \tag{11}$$

The covariance matrix of ε is given by

$$V = E(\varepsilon\varepsilon') = \sigma_u^2 I + \beta^2 \sum_{i=1}^{m} \lambda_i^2 \sigma_i^2 P_i \tag{12}$$

if u and η_i are uncorrelated for all i = 1,2,..., m. If OLS is used to estimate (8), the covariance matrix should be based on (12).

Defining

$$H = [1; \sum_{i==1}^{m} \lambda_i \hat{X}_i] \tag{13}$$

and

$$\theta' = (\alpha, \beta)$$

then (8) can be written as

$$y = H\theta + \varepsilon \tag{14}$$

so that the covariance matrix of $\hat{\theta}$ is given by

$$Var(\hat{\theta}) = (H'H)^{-1} H'VH(H'H)^{-1} \tag{15}$$

When V in (12) is substituted in (15), one has

$$Var(\hat{\theta}) = \sigma_u^2 (H'H)^{-1} + \beta^2 (H'H)^{-1} H' \left(\sum_{i=1}^{m} \lambda_i \sigma_i^2 P_i \right) H(H'H)^{-1} \tag{16}$$

which shows that the standard OLS covariance matrix of $\hat{\theta}$, namely the first term on the right-hand side of (16), leads to a downward bias in the covariance matrix and a corresponding upward bias in the corresponding t-ratios. The covariance matrix in (16) can be consistently estimated by the Newey-West HAC covariance matrix. Smith and McAleer (1994) evaluate the finite sample properties of the HAC estimator for purposes of testing hypotheses and constructing confidence intervals in the case of generated regressors. Their analysis also applies to the case of forecasts considered in the present paper.

Again, an alternative approach builds on (5) and is given as

$$y = \alpha + \beta \sum_{i=1}^{m} \lambda_i (X_i - \eta_i) + u$$

or

$$y = \alpha + \beta \sum_{i=1}^{m} \lambda_i X_i + \left(u - \beta \sum_{i=1}^{m} \lambda_i \eta_i \right) \tag{17}$$

As $\sum_{i=1}^{m} \lambda_i X_i$ is correlated with $\sum_{i=1}^{m} \lambda_i \eta_i$, one again needs to apply GMM.

3. Data and Empirical Analysis

Since 1978, actual data and three sets of updated forecasts of the inflation rate and real GDP growth rate have been released by the Government of Taiwan, Republic of China (for further details, see Chang et al. (2009)). The unemployment rate is not regarded as a key economic fundamental in Taiwan. In this paper, we use the most recent revised government forecasts. The government forecasts (F1) and actual values of the inflation rate and real GDP growth rate are obtained from the Quarterly National Economic Trends, Directorate-General of Budget, Accounting and Statistics, Executive Yuan, Taiwan, 1980-2009. The forecasts from the two private forecasting institutions are obtained from the Chung-Hua Institution for Economic Research (F2) and Taiwan Institute of Economic Research (F3).

In addition to comparing actual data on both the inflation rate and real growth rate with three sets of individual forecasts, four alternative mean forecasts are also considered, namely the mean of all three forecasts and of three pairs of mean forecasts. In the Tables, M refers to the mean of all three forecasts, M12 refers to the mean of F1 and F2, M13 refers to the mean of F1 and F3, and M23 refers to the mean of F2 and F3.

As the actual values of the inflation rate and real GDP growth rate are available, the accuracy of the government and two private (that is, individual) forecasts, as well as the effects of econometric model versus intuition, can be compared and tested. The sample period used for the actual values and the three sets of individual forecasts of seasonally unadjusted quarterly inflation rate and real growth rate of GDP is 1995Q3-2009Q2, for a total of 56 observations.

We have analyzed the data on unit roots and structural breaks. The diagnostics for unit roots (which are unreported) indicate that we can work with the growth rates data, as in Figures 1 and 2. Visual inspection from the same graphs does not suggest potential structural breaks, and there is also no evidence of structural breaks caused by any changes in measurement methods at the government agency and two private forecasting institutions in Taiwan.

The inflation rate and the three individual forecasts, F1, F2 and F3, are given in Figure 1, and the corresponding plots of the real GDP growth rate and the three individual forecasts are given in Figure 2. Figure 3 gives the inflation rate, the mean of the three forecasts, and the means of pairs of forecasts, while the corresponding plots of the real GDP growth rate, the mean of the three forecasts, and the means of pairs of forecasts are given in Figure 4.

Table 1 gives the correlations of the inflation rate, three individual forecasts, the mean of three forecasts, the means of pairs of forecasts (and their replicable counterparts, which are obtained from Tables 4 and 5 (to be discussed below) , with the corresponding plots of the real GDP growth rate given in Table 2. In these two tables, hats (circumflex) denote their replicable counterparts. In Tables 1 and 2, the highest correlations for both the actual inflation rate and the real GDP growth rate are with F1, followed by M13; for both variables, F1 is highly correlated with M12, M13 and M23, F2 is highly correlated with M12 and M23, F3 is highly correlated with M23, M is highly correlated with M12 and M13, M12 is highly correlated with M13, and M13 is highly correlated with M23. The correlations are generally higher between the original variables than between their fitted counterparts.

The goodness-of-fit measures, namely root mean square error (RMSE) and mean absolute deviation (MAD), of the replicable and non-replicable forecasts are given in Table 3 for both variables. For the non-replicable forecasts, in the upper panel of Table 3, the single forecast, F1, is best for both variables using RMSE and MAD, while the mean of two forecasts, M13, is second best for the inflation rate, and M12 is second best for the real GDP growth rate. A similar outcome holds for the replicable forecasts, with $\hat{F}1$ best for both variables using RMSE and MAD, while $\hat{M}13$ is second best for both variables using RMSE and MAD.

These results suggest that, in general, the first individual forecast is best in terms of both RMSE and MAD, followed by a mean combination of the first and third individual forecasts, for both the inflation rate and real GDP growth rate, regardless of whether a non-replicable or replicable forecast is used. Table 3 also shows that the biased non-replicable forecasts are apparently much more accurate than their replicable counterparts. Hence, the added intuition of experts seems to lead to substantial improvement. This improvement is most evident for F1, where RMSE for the replicable forecast is about twice as large as for the non-replicable forecast.

In Tables 4a-4b and Tables 5a-5b, we report on the retrieval of a replicable part from the non-replicable forecasts based on public information for the inflation rate and real GDP growth rate, respectively. This public information is set at one-period lagged real growth, one-period lagged inflation, one period lagged forecast for forecaster 1, one period lagged forecast for forecaster 2 and one period lagged forecast for forecaster 3.

It is evident that the lagged values of the forecasts of all three forecasters are insignificant in all four tables, so the forecasters do not seem to include each other's predictions. The one-period lagged real GDP growth rate is significant for all seven forecasts for both the inflation rate and real GDP growth rate. Apart from the significant case of F1 in Table 4a, the one-period lagged inflation rate is not significant in capturing expertise for any of the seven forecasts for either variable. The F tests for the significance of the replicable part in Tables 4a-4b and Tables 5a-5b indicate clearly that the expertise in equation (3) is captured by the one-period lagged variables, specifically the one-period lagged real GDP growth rate.

In order to examine if the replicable forecasts are unbiased, we consider equations (3) and (8) for three forecasts and four mean forecasts, which are given in Tables 6a-6b for the inflation rate and real GDP growth rate. As the replicable forecasts lead to generated regressors, the appropriate Newey-West HAC standard errors are calculated for valid inferences. The F test is a test of the null hypothesis $H_0: \alpha = 0, \beta_i = 1$ for *i* = 1,2,3. If the null hypothesis is not rejected, then the model via the replicable forecast can predict the actual value, whereas rejection of the null means that expert intuition could triumph over the model in case the non-replicable forecasts are not biased. Except for F1 and F2 for the real GDP growth rate in Table 6a, the null hypothesis is rejected in all cases, which makes it clear that intuition is significant in explaining actual values, and hence dominates the econometric model. This supports the RMSE and MAD scores in Table 3.

Tables 7a-7b and Tables 8a-8b focus on the accuracy of the non-replicable forecasts for three individual forecasts and four mean forecasts in equations (5) and (17) for the inflation rate and real GDP growth rate. As the non-replicable forecasts are correlated with the measurement errors, GMM is necessary for valid inference, where the instrument list for GMM for forecaster *i* includes one-period lagged real growth, one-period lagged inflation, one-period lagged forecast for forecaster 1, and one-period lagged forecast for forecaster 2 and one period lagged forecast for forecaster 3. The F test is a test of the null hypothesis $H_0: \alpha = 0, \beta_i = 1$ for *i* = 1,2,3. Conditional on the information set, if the null hypothesis is not rejected, then the non-replicable forecast can accurately predict the actual value, whereas rejection of the null hypothesis means that the non-replicable forecast is biased.

Except in one case, namely GMM estimation of M for the inflation rate in Table 7b, the null hypothesis is rejected for all individual forecasts and mean forecasts. Thus, conditional on the information set, the non-replicable forecast cannot predict the actual inflation rate. Ignoring the OLS results in Tables 8a-8b, mirroring the results in Tables 7a-7b, except for one case, namely GMM estimation of F1 for the real GDP growth rate in Table 8a, the null hypothesis is rejected for all individual forecasts and mean forecasts. Thus, conditional on the information set, the non-replicable forecast cannot predict the actual real GDP growth rate. If we compare the F test values in Tables 7 and 8 with those in Table 6, we see that the non-replicable forecasts have greater bias than the replicable forecasts. Again, the non-replicable forecasts are much more accurate than the replicable forecasts, which means that the intuition possessed by the forecasters greatly improves any model-based forecasts.

It is instructive to note that using alternative mean forecasts can be beneficial. For inflation, we see that the GMM-based results in Table 7b indicate the M delivers unbiased forecasts. For GDP growth, the situation is somewhat different. There we see that the non-replicable F1 is unbiased (Table 8a), and Table 3 also suggests it has the smallest forecast error. However, Table 8b clearly shows that using alternative mean forecasts is not sensible as all the alternatives examined in Table 8b lead to biased forecasts.

4. Concluding Remarks

A forecast that is based on an econometric model is replicable and may be unbiased, whereas a forecast that is not based on an econometric model is non-replicable and is typically biased. Government and professional forecasters alike can, and do, provide both replicable and non-replicable forecasts. Both types of forecasts can be considered in alternative mean forecasts, including trimmed mean forecasts.

Many forecasts are only available in their final form, so that it can be difficult to quantify expert intuition. In many cases, it may be unknown to the analyst if the forecaster has relied on the outcome of an econometric model, or even if an econometric model has been used. The analyst usually has only a forecast of an economic variable, and the analyst must then evaluate its quality. It has been shown in the literature that it is important to examine the behaviour of experts prior to evaluating forecast accuracy.

This paper pursued such a line or research by developing a methodology to evaluate individual and alternative mean forecasts using efficient estimation methods, and compared individual replicable forecasts with alternative mean forecasts. An empirical example to forecast economic fundamentals for Taiwan showed the relevance of the methodological approach proposed in the paper. The empirical analysis showed that replicable and non-replicable forecasts could be distinctly different from each other, that efficient and inefficient estimation methods, as well as consistent and inconsistent covariance matrix estimates, could lead to significantly different outcomes, alternative mean forecasts could yield different forecasts from their individual components, and the relative importance of econometric model versus intuition could be evaluated in terms of forecasting performance.

It was shown that individual forecasts could perform quite differently from the mean forecasts of two or three individual forecasts, that intuition was significant in explaining actual values, and hence dominated the model, and that expert intuition that has been used to obtain the non-replicable forecasts of the inflation rate and real GDP growth rate was not sufficient to forecast accurately the actual values.

One of the major findings of the paper is that a proper analysis of alternative mean forecasts could suggest a weaker dominance of other forecasts, as is typically documented in the literature. The GMM-based analysis shows that the alternative forecasts could well be found to be biased, while the OLS-based analysis did not give any such warning signals.

References

Chang, C.-L., P.H. Franses and M. McAleer (2009), How Accurate are Government Forecasts of Economic Fundamentals? The Case of Taiwan, International Journal of Forecasting, to appear. Available at SSRN: http://ssrn.com/abstract=1431007.

Eroglu, C. and K.L. Croxton (2010), Biases in Judgmental Adjustments of Statistical Forecasts: The Role of Individual Differences, International Journal of Forecasting, 26, 116-133.

Fildes, R, P. Goodwin, M. Lawrence, and K. Nikolopoulos (2009), Effective Forecasting and Judgemental Adjustments: An Empirical Evaluation and Strategies for Improvement in Supply-Chain Planning, International Journal of Forecasting, 25, 3-23.

Franses, P.H., H. Kranendonk, and D. Lanser (2011), One Model and Various Experts: Evaluating Dutch Macroeconomic Forecasts, International Journal of Forecasting, 27, 482-495.

Franses, P.H. and R. Legerstee (2010), Do Experts' Adjustments on Model-based SKU-level Forecasts Improve Forecast Quality?, Journal of Forecasting, 29, 331-340

Franses, P.H., M. McAleer and R. Legerstee (2009), Expert Opinion Versus Expertise in Forecasting, Statistica Neerlandica, 63, 334-346.

Smith, J. and M. McAleer (1994), Newey-West Covariance Matrix Estimates for Models with Generated Regressors, Applied Economics, 26, 635-640.

Figure 1. Inflation Rate and Three Individual Forecasts,

1995Q3-2009Q2

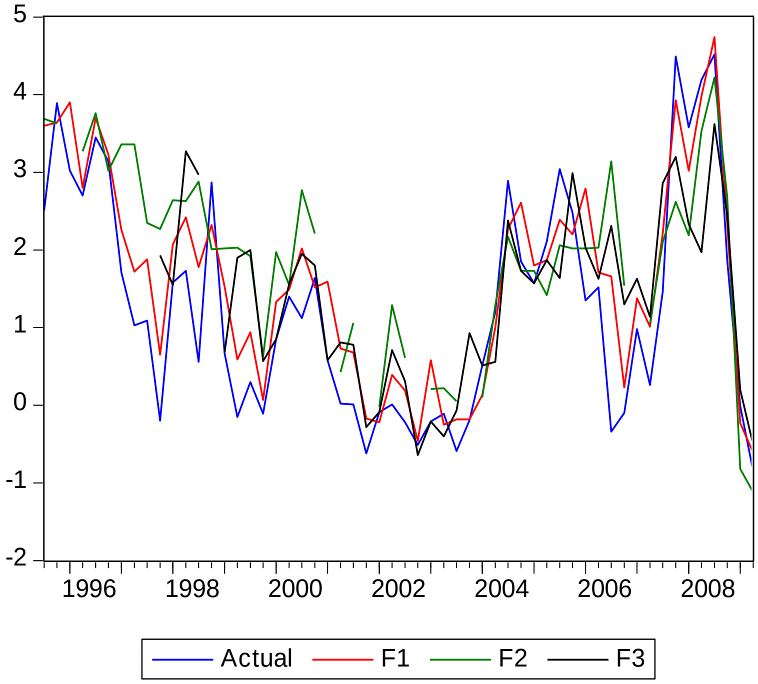

Figure 2. Real GDP Growth Rate and Three Individual Forecasts,

1995Q3-2009Q2

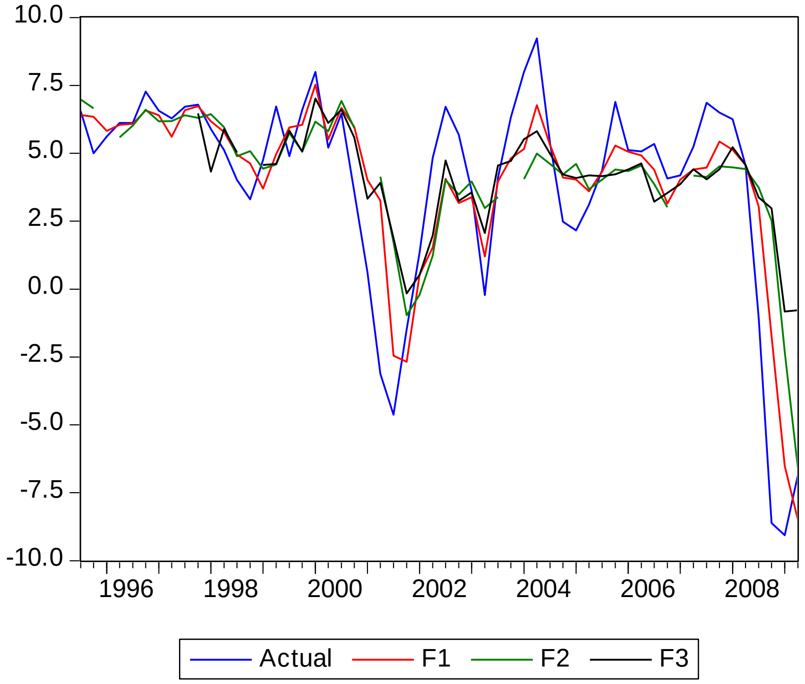

Figure 3. Inflation Rate, Mean of Three Individual Forecasts, Means of Pairs of Individual Forecasts, 1995Q3-2009Q2

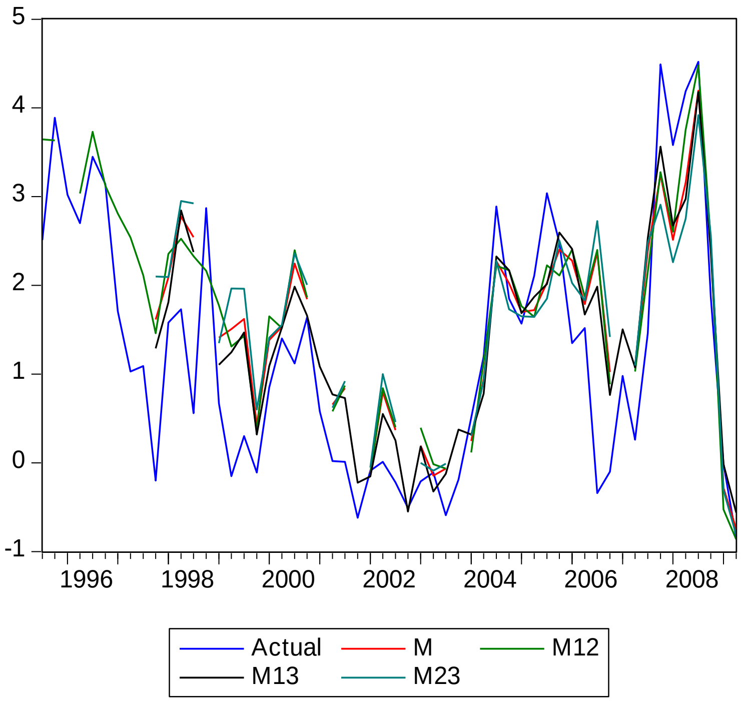

Figure 4. Real GDP Growth Rate, Mean of Three Individual Forecasts, Means of Pairs of Individual Forecasts, 1995Q3-2009Q2

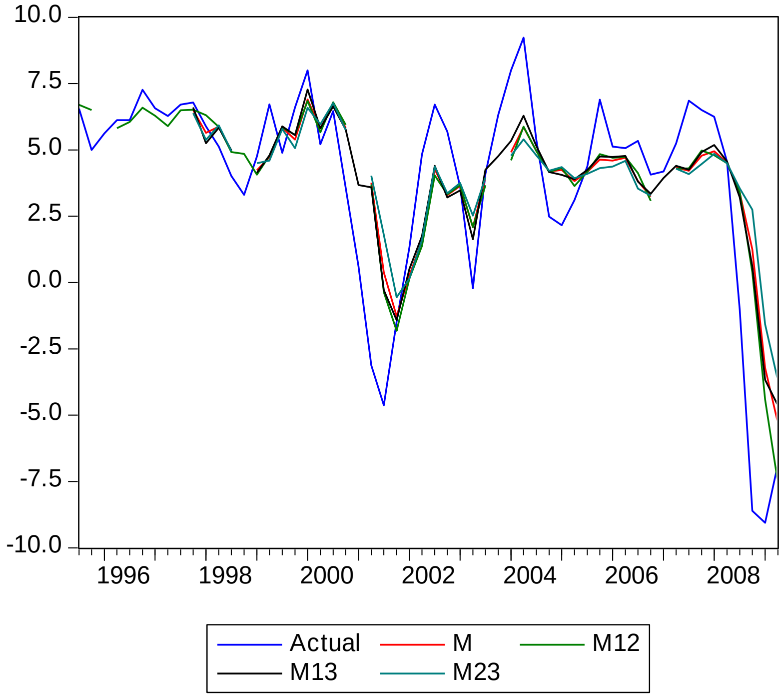

Table 1. Correlations of Inflation Rate, Three Individual Forecasts, Mean of Three Individual Forecasts, Means of Pairs of Individual Forecasts, and their Replicable Counterparts

	Actual	F1	F2	F3	M	M12	M13	M23	$\hat{F}1$	$\hat{F}2$	$\hat{F}3$	$\hat{M}$	$\hat{M}12$	$\hat{M}13$	$\hat{M}23$
Actual	1.000														
F1	0.915	1.000													
F2	0.656	0.839	1.000												
F3	0.678	0.826	0.850	1.000											
M	0.803	0.947	0.947	0.939	1.000										
M12	0.828	0.964	0.953	0.873	0.987	1.000									
M13	0.845	0.964	0.883	0.946	0.987	0.966	1.000								
M23	0.693	0.865	0.964	0.960	0.981	0.950	0.950	1.000							
$\hat{F}1$	0.783	0.853	0.741	0.741	0.829	0.835	0.840	0.771	1.000						
$\hat{F}2$	0.699	0.778	0.822	0.769	0.836	0.833	0.810	0.828	0.901	1.000					
$\hat{F}3$	0.709	0.793	0.793	0.789	0.838	0.827	0.828	0.822	0.942	0.966	1.000				
$\hat{M}$	0.760	0.834	0.805	0.777	0.854	0.855	0.845	0.823	0.970	0.978	0.981	1.000			
$\hat{M}12$	0.766	0.840	0.802	0.770	0.853	0.857	0.845	0.817	0.974	0.974	0.971	0.999	1.000		
$\hat{M}13$	0.769	0.843	0.775	0.771	0.846	0.846	0.848	0.804	0.991	0.942	0.978	0.990	0.989	1.000	
$\hat{M}23$	0.710	0.791	0.817	0.784	0.844	0.838	0.824	0.833	0.925	0.994	0.987	0.988	0.981	0.965	1.000

Notes: F1: DGBAS: Directorate General of Budget, Accounting and Statistics (Government), F2: Chung-Hua: Chung-Hua Institution for Economic Research, F3: Taiwan: Taiwan Institute of Economic Research, M: Mean of three forecasts, M12: Mean of F1 and F2, M13: Mean of F1 and F3, M23: Mean of F2 and F3. Hats (circumflex) denote the replicable counterparts.

Table 2. Correlations of Real GDP Growth Rate, Three Individual Forecasts, Mean of Three Individual Forecasts, Means of Pairs of Individual Forecasts, and their Replicable Counterparts

	Actual	F1	F2	F3	M	M12	M13	M23	$\hat{F}1$	$\hat{F}2$	$\hat{F}3$	$\hat{M}$	$\hat{M}12$	$\hat{M}13$	$\hat{M}23$
Actual	1.000														
F1	0.898	1.000													
F2	0.736	0.942	1.000												
F3	0.758	0.916	0.921	1.000											
M	0.832	0.984	0.978	0.960	1.000										
M12	0.842	0.990	0.980	0.931	0.996	1.000									
M13	0.866	0.990	0.953	0.964	0.995	0.988	1.000								
M23	0.760	0.950	0.986	0.973	0.990	0.979	0.976	1.000							
$\hat{F}1$	0.814	0.931	0.916	0.862	0.932	0.938	0.925	0.911	1.000						
$\hat{F}2$	0.702	0.898	0.950	0.874	0.931	0.933	0.907	0.936	0.963	1.000					
$\hat{F}3$	0.753	0.918	0.941	0.874	0.938	0.941	0.922	0.933	0.986	0.990	1.000				
$\hat{M}$	0.765	0.924	0.941	0.881	0.940	0.944	0.925	0.932	0.991	0.990	0.997	1.000			
$\hat{M}12$	0.771	0.925	0.939	0.875	0.940	0.944	0.925	0.930	0.993	0.988	0.997	0.999	1.000		
$\hat{M}13$	0.797	0.930	0.927	0.870	0.937	0.942	0.927	0.921	0.999	0.975	0.994	0.996	0.997	1.000	
$\hat{M}23$	0.718	0.906	0.949	0.878	0.935	0.937	0.913	0.937	0.972	0.999	0.995	0.995	0.993	0.983	1.000

Notes: F1: DGBAS: Directorate General of Budget, Accounting and Statistics (Government), F2: Chung-Hua: Chung-Hua Institution for Economic Research, F3: Taiwan: Taiwan Institute of Economic Research, M: Mean of three forecasts, M12: Mean of F1 and F2, M13: Mean of F1 and F3, M23: Mean of F2 and F3. Hats (circumflex) denote the replicable counterparts.

Table 3

Goodness-of-fit of Replicable and Non-Replicable Forecasts for Three Individual Forecasts, Means of Three Individual Forecasts, Means of Pairs of Individual Forecasts, 1995Q3-2009Q2

Non-replicable Forecasts	Inflation Rate		Real GDP Growth Rate	
	RMSE	MAD	RMSE	MAD
F1	0.413	0.524	3.795	1.323
F2	1.409	0.943	8.079	1.888
F3	1.082	0.758	9.919	2.123
M	0.856	0.726	7.433	1.865
M12	0.790	0.715	5.568	1.584
M13	0.627	0.619	6.383	1.744
M23	1.201	0.836	9.690	2.130
Replicable Forecasts	**Inflation Rate**		**Real GDP Growth Rate**	
	RMSE	MAD	RMSE	MAD
$\hat{F}1$	0.895	0.754	6.209	1.946
$\hat{F}2$	1.325	0.964	9.678	2.262
$\hat{F}3$	1.108	0.851	10.51	2.217
$\hat{M}$	1.064	0.841	8.364	2.112
$\hat{M}12$	1.061	0.838	7.691	2.082
$\hat{M}13$	0.946	0.777	7.666	2.020
$\hat{M}23$	1.222	0.917	10.01	2.245

Note: RMSE and MAD denote root mean square error and mean absolute deviation, respectively.

Table 4a

Retrieving Replicable Components from the three Individual Non-Replicable Forecasts

Included Variables	Inflation Rate		
	F1	F2	F3
Intercept	0.092	0.401	0.176
	(0.235)	(0.243)	(0.246)
Real GDP Growth(*t*-1*)*	0.127	0.156	0.103
	(0.030)***	(0.030)***	(0.031)***
Inflation(*t*-1)	0.544	0.133	0.119
	(0.228)**	(0.225)	(0.240)
F1(*t*-1)	0.040	0.266	0.255
	(0.368)	(0.373)	(0.383)
F 2(*t*-1)	-0.155	0.167	0.175
	(0.263)	(0.261)	(0.274)
F 3(*t*-1)	0.312	-0.079	0.072
	(0.224)	(0.213)	(0.240)
Adj. R^2	0.684	0.620	0.538
F test	17.89***	12.08***	9.840***

Notes: The regression model is (2) where *i* = 1 for F1 forecast (government), *i* = 2 for F2 forecast (Chung-Hwa institution), and *i* = 3 for F3 forecast (Taiwan institution). W_i in (2) for the forecast for forecaster 1 is approximated by one-period lagged real growth, one-period lagged inflation, one period lagged forecast for forecaster 1, one period lagged forecast for forecaster 2 and one period lagged forecast for forecaster 3. The F test is a test of expertise. Standard errors in parentheses.

** and *** denote significance at the 5% and 1% levels, respectively.

Table 4b

Retrieving Replicable Components from the Four Non-Replicable Mean Forecasts

Included Variables	Inflation Rate			
	M	M12	M13	M23
Intercept	0.304	0.291	0.153	0.347
	(0.221)	(0.229)	(0.218)	(0.226)
Real GDP Growth(*t*-1*)*	0.135	0.149	0.116	0.130
	(0.027)***	(0.029)***	(0.028)***	(0.028)***
Inflation(*t*-1)	0.274	0.312	0.353	0.146
	(0.204)	(0.211)	(0.212)	(0.209)
F1(*t*-1)	0.222	0.214	0.152	0.237
	(0.337)	(0.351)	(0.339)	(0.345)
F 2(*t*-1)	0.034	-0.040	0.002	0.190
	(0.236)	(0.246)	(0.242)	(0.242)
F 3(*t*-1)	0.035	0.090	0.157	-0.032
	(0.198)	(0.200)	(0.212)	(0.203)
Adj. R^2	0.682	0.682	0.665	0.639
F test	15.15***	15.55***	16.12***	12.68***

Notes: The regression model is (2) where *i* = 1 for mean of 3 forecasters, *i* = 2 for mean of F1 and F2, *i* = 3 for mean of F1 and F3, and *i* = 4 for mean of F2 and F3. W_i in (2) for the forecast for forecaster 1 is approximated by one-period lagged real growth, one-period lagged inflation, one period lagged forecast for forecaster 1, one period lagged forecast for forecaster 2 and one period lagged forecast for forecaster 3. The F test is a test of expertise. Standard errors in parentheses.

*** denotes significance at the 1% level.

Table 5a

Retrieving Replicable Components from the Three Individual Non-Replicable Forecasts

Included Variables	Real GDP Growth Rate		
	F1	F2	F3
Intercept	0.495	0.765	2.077
	(0.761)	(0.502)	(0.546)***
Real GDP Growth(*t*-1)	0.664	0.246	0.222
	(0.141)***	(0.095)**	(0.102)**
Inflation(*t*-1)	-0.172	-0.093	-0.035
	(0.160)	(0.108)	(0.116)
F1(*t*-1)	0.131	0.383	0.220
	(0.382)	(0.256)	(0.275)
F2(*t*-1)	0.407	0.577	0.126
	(0.446)	(0.307)*	(0.321)
F3(*t*-1)	-0.344	-0.400	-0.069
	(0.386)	(0.259)	(0.277)
Adj. R^2	0.844	0.885	0.725
F test	45.52***	59.74***	22.05***

Notes: The regression model is (2) where i = 1 for F1 forecast (government), i = 2 for F2 forecast (Chung-Hwa institution), and i = 3 for F3 forecast (Taiwan institution). W_i in (2) for the forecast for forecaster i is approximated by one-period lagged real growth, one-period lagged inflation, one period lagged forecast for forecaster 1, one period lagged forecast for forecaster 2 and one period lagged forecast for forecaster 3. The F test is a test of expertise. Standard errors in parentheses.

* , ** and *** denote significance at the 10%, 5% and 1% levels, respectively.

Table 5b

Retrieving Replicable Components from the Four Non-Replicable Mean Forecasts

Included Variables	Real GDP Growth Rate			
	M3	M12	M13	M23
Intercept	1.053	0.577	1.283	1.391
	(0.554)*	(0.613)	(0.597)**	(0.477)***
Real GDP Growth(*t*-1)	0.392	0.471	0.447	0.235
	(0.106)***	(0.116)***	(0.111)***	(0.091)**
Inflation(*t*-1)	-0.072	-0.110	-0.099	-0.050
	(0.120)	(0.132)	(0.127)	(0.103)
F1(*t*-1)	0.200	0.212	0.168	0.291
	(0.284)	(0.313)	(0.301)	(0.244)
F2(*t*-1)	0.461	0.569	0.272	0.402
	(0.339)	(0.374)	(0.351)	(0.292)
F3(*t*-1)	-0.331	-0.418	-0.210	-0.271
	(0.286)	(0.315)	(0.303)	(0.246)
Adj. R^2	0.865	0.875	0.834	0.859
F test	48.55***	53.98***	41.21***	46.10***

Notes: The regression model is (2) where i = 1 for mean of 3 forecasters, i = 2 for mean of F1 and F2, i = 3 for mean of F1 and F3, and i = 4 for mean of F2 and F3. W_i in (2) for the forecast for forecaster i is approximated by one-period lagged real growth, one-period lagged inflation, one period lagged forecast for forecaster 1, one period lagged forecast for forecaster 2 and one period lagged forecast for forecaster 3. The F test is a test of expertise. Standard errors in parentheses.

*, ** and *** denote significance at the 10%, 5% and 1% levels, respectively.

Table 6a

Are Replicable Forecasts for Three Individual Forecasts Accurate?

Estimation Method	Inflation Rate					
	Intercept	F1	F2	F3	Adj. R^2	F Test
OLS	-0.340 (0.248)	1.035 (0.135)***			0.598	3.58**
HAC	[0.156]***	[0.115]***				
OLS	-0.729 (0.358)**		1.126 (0.185)**		0.493	6.17***
HAC	[0.305]***		[0.180]***			
OLS	-0.673 (0.328)**			1.249 (0.191)***	0.517	5.03**
HAC	[0.237]***			[0.176]***		

Estimation Method	Real GDP Growth Rate					
	Intercept	F1	F2	F3	Adj. R^2	F Test
OLS	-0.374 (0.591)	1.081 (0.127)			0.637	0.20
HAC	[0.710]	[0.128]***				
OLS	-1.107 (0.909)		1.220 (0.209)***		0.447	0.56
HAC	[1.094]		[0.209]***			
OLS	-4.396 (1.216)***			1.982 (0.288)***	0.531	5.63***
HAC	[1.434]***			[0.296]***		

Notes: The regression model is (3) where i = 1 for F1 forecast (government), i = 2 for F2 forecast (Chung-Hwa institution), and i = 3 for F3 forecast (Taiwan institution). Standard errors in parentheses. Newey-West HAC standard errors are in brackets.

** and *** denote significance at the 5% and 1% levels, respectively.
The F test is a test of the null hypothesis $H_0: \alpha = 0, \beta_i = 1$, i = 1,2,3.

Table 6b

Are Replicable Forecasts for Four Mean Forecasts Accurate?

Estimation Method	Inflation Rate						
	Intercept	M	M12	M13	M23	Adj. R^2	F Test
OLS	-0.693 (0.306)**	1.195 (0.167)***				0.562	4.55**
HAC	[0.264]**	[0.179]***					
OLS	-0.632 (0.295)**		1.134 (0.157)***			0.568	4.38**
HAC	[0.257]**		[0.167]***				
OLS	-0.534 (0.276)*			1.171 (0.157)***		0.583	4.39**
HAC	[0.190]***			[0.145]***			
OLS	-0.788 (0.351)**				1.216 (0.190)***	0.505	4.50**
HAC	[0.325]**				[0.225]***		
Estimation Method	**Real GDP Growth Rate**						
	Intercept	M	M12	M13	M23	Adj. R^2	F Test
OLS	-1.576 (0.823)*	1.353 (0.190)***				0.548	1.93
HAC	[1.215]	[0.208]***					
OLS	-0.784 (0.719)		1.172 (0.161)***			0.559	0.65
HAC	[1.074]		[0.176]***				
OLS	-1.830 (0.771)**			1.412 (0.177)***		0.605	2.30*
HAC	[1.100]			[0.186]***			
OLS	-2.314 (1.043)**				1.500 (0.244)***	0.472	2.47*
HAC	[1.572]				[0.286]***		

Notes: The regression model is (8) where *i* = 1 for mean of 3 forecasters, *i* = 2 for mean of F1 and F2, *i* = 3 for mean of F1 and F3, and *i* = 4 for mean of F2 and F3. Standard errors in parentheses. Newey-West HAC standard errors are in brackets.

*, **, and *** denote significance at the10%, 5% and 1% levels, respectively.
The F test is a test of the null hypothesis $H_0: \alpha = 0, \beta_i = 1$, *i* = 1,2,3,4.

Table 7a

Examining Bias in Non-Replicable Forecasts for Three Individual Forecasts

Estimation Method	Inflation Rate					
	Intercept	F1	F2	F3	Adj. R^2	F Test
OLS	-0.357 (0.118)***	1.009 (0.056)***			0.853	9.29***
GMM	-0.306 (0.092)***	0.993 (0.060)***			0.838	11.33***
OLS	-0.206 (0.280)		0.822 (0.124)***		0.467	7.77***
GMM	-0.394 (0.273)		0.747 (0.174)***		0.314	10.05***
OLS	-0.231 (0.235)			0.902 (0.135)***	0.492	3.41**
GMM	-0.323 (0.201)			0.738 (0.186)***	0.400	10.44***

Notes: The regression model is (5) where *i* = 1 for F1 forecast (government), *i* = 2 for F2 forecast (Chung-Hwa institution), and *i* = 3 for F3 forecast (Taiwan institution). The instrument list for GMM for forecaster *i* includes one-period lagged real growth, one-period lagged inflation, one-period lagged forecast for forecaster 1, and one-period lagged forecast for forecaster 2 and one period lagged forecast for forecaster 3. Standard errors in parentheses.

*** denotes significance at the 1% level.

The F test is a test of the null hypothesis $H_0: \alpha = 0, \beta_i = 1$, *i* = 1,2,3.

Table 7b

Examining Bias in Non-Replicable Forecasts for Four Mean Forecasts

Estimation Method	Inflation Rate						
	Intercept	M	M12	M13	M23	Adj R2	F Test
OLS	-0.471 (0.231)**	1.044 (0.124)***				0.636	4.67**
GMM	-0.410 (0.249)	1.210 (0.128)***				0.577	1.44
OLS	-0.455 (0.203)**		1.010 (0.094)***			0.700	7.64***
GMM	-0.382 (0.191)*		0.893 (0.133)***			0.631	8.69***
OLS	-0.440 (0.168)**			1.065 (0.096)***		0.730	5.68***
GMM	-0.326 (0.152)**			0.828 (0.145)***		0.659	11.73***
OLS	-0.324 (0.286)				0.925 (0.152)***	0.472	3.90**
GMM	-0.262 (0.242)				0.666 (0.184)***	0.321	8.98***

Notes: The regression model is (17) where *i* = 1 for mean of 3 forecasters, *i* = 2 for mean of F1 and F2, *i* = 3 for mean of F1 and F3, and *i* = 4 for mean of F2 and F3. The instrument list for GMM for forecaster *i* includes one-period lagged real growth, one-period lagged inflation, one-period lagged forecast for forecaster 1, and one-period lagged forecast for forecaster 2 and one period lagged forecast for forecaster 3. Standard errors in parentheses.

** and *** denote significance at the 5% and 1% levels, respectively.
The F test is a test of the null hypothesis $H_0: \alpha = 0, \beta_i = 1$, *i* = 1,2,3,4.

Table 8a

Examining Bias in Non-Replicable Forecasts for Three Individual Forecasts

Estimation Method	Real GDP Growth Rate					
	Intercept	F1	F2	F3	Adj R^2	F Test
OLS	-0.565 (0.429)	1.118 (0.085)***			0.760	1.03
GMM	0.177 (0.324)	0.960 (0.050)***			0.768	0.35
OLS	-1.160 (0.788)		1.217 (0.164)***		0.516	1.09
GMM	-8.903 (2.396)***		2.845 (0.559)***		-0.586	7.47***
OLS	-3.720 (1.789)***			1.789 (0.239)***	0.550	6.26***
GMM	-11.72 (2.098)***			3.515 (0.497)***	-0.098	15.8***

Notes: The regression model is (5) where *i* = 1 for F1 forecast (government), *i* = 2 for F2 forecast (Chung-Hwa institution) and *i*= 3 for F3 forecast (Taiwan institution). The instrument list for GMM for forecaster *i* includes one-period lagged real growth, one-period lagged inflation, one-period lagged forecast for forecaster 1, one-period lagged forecast for forecaster 2 and one period lagged forecast for forecaster 3.Standard errors in parentheses.

*** denotes significance at the 1% level.

The F test is a test of the null hypothesis $H_0: \alpha = 0, \beta_i = 1$, *i* = 1,2,3.

Table 8b

Examining Bias in Non-Replicable Forecasts for Four Mean Forecasts

Estimation Method	Real GDP Growth Rate						
	Intercept	M	M12	M13	M23	Adj R2	F Test
OLS	-1.845 (0.720)**	1.411 (0.160)***				0.647	3.59**
GMM	-6.926 (1.469)***	2.439 (0.345)***				0.187	11.5***
OLS	-1.012 (0.577)*		1.209 (0.117)***			0.674	1.72
GMM	-5.328 (1.240)***		2.068 (0.293)***			0.241	10.1***
OLS	-2.019 (0.632)***			1.447 (0.140)***		0.703	5.56***
GMM	-5.978 (1.215)***			2.232 (0.287)***		0.426	12.5***
OLS	-2.473 (2.521)**				1.529 (0.586)***	0.534	3.38**
GMM	-11.26 (2.521)***				3.410 (0.586)***	-0.514	10.2***

Notes: The regression model is (17) where *i* = 1 for mean of 3 forecasters, *i* = 2 for mean of F1 and F2, *i* = 3 for mean of F1 and F3, and *i* = 4 for mean of F2 and F3. The instrument list for GMM for forecaster *i* includes one-period lagged real growth, one-period lagged inflation, one-period lagged forecast for forecaster 1, and one-period lagged forecast for forecaster 2 and one period lagged forecast for forecaster 3. Standard errors in parentheses.

*, ** and *** denote significance at the 10%, 5% and 1% levels, respectively.
The F test is a test of the null hypothesis $H_0: \alpha = 0, \beta_i = 1$, *i* = 1,2,3,4.