The IMF and OECD *versus* Consensus Forecasts

by

Roy Batchelor

City University Business School, London

August 2000

Abstract: This paper compares the accuracy and information content of macroeconomic economic forecasts for G7 countries made in the 1990s by the OECD and IMF. The benchmarks for comparison are the average forecasts of private sector economists published by *Consensus Economics*. With few exceptions, the private sector forecasts are less biased and more accurate in terms of mean absolute error and root mean square error. Formal tests show that there is little information in the OECD and IMF forecasts that could be used to reduce significantly the error in the private sector forecasts.

Keywords: Forecasting, Combining Forecasts
JEL Class No. : E27, E37

Acknowledgments: I am very grateful to Patrick R. Dennis for help in assembling the data used in this study.

✉ City University Business School
Frobisher Crescent, Barbican, London EC2Y 8HB, UK.

☎ Phone:(-44-) 0-207 477 8733 💻 Fax: (-44-) 0-207 477 8881 📻 r.a.batchelor@city.ac.uk

1. Introduction

This study assesses the accuracy and value-added of the economic forecasts for G7 countries produced by the International Monetary Fund (IMF) and the Organization for Economic Cooperation and Development (OECD). The forecasts of the two agencies for the years 1990-1999 are compared with a tough but relevant benchmark, the averages of private sector forecasts published by *Consensus Economics*. The study updates the comparisons made for an earlier period by Batchelor (2000).

The quality of economic forecasts has been the subject of a broad debate in the press, the business community and government throughout the past decade. In the United States and elsewhere, the debate was stimulated by the failure of forecasters to make timely predictions of recessions of the early 1990s. In Europe, this debate was further fueled by concern with economic convergence and the synchronization of business cycles in the European Community, and by the inclusion in the Maastricht treaty of statistical criteria for key variables as preconditions for membership of the European Monetary Union. The quality of forecasts of multinational agencies, including the IMF, OECD, the United Nations and the European Community, has also come under scrutiny as part of a more general audit of the value for money that such organizations provide.

A priori it is not clear whether forecasts produced by governments and multi-national agencies should be more or less accurate than forecasts produced in the private sector by banks, business corporations, and independent consultants. Governments have two information advantages which should improve their relative accuracy. They have more timely and complete knowledge of official statistics. And they have (or should have) some insight into their own intentions and likely reactions to future events.

In the case of intergovernmental agencies, this information advantage may be attenuated by two factors. One is that forecasters are not based in the countries they are forecasting. They may not have access to various informal pieces of information and rumour that are available to home-country analysts.

The other is that governments and government-funded bodies are subject to political pressures. This may diminish the accuracy of their forecasts in several ways. At the very least it may cause some bureaucratic delays in publication. More seriously, governments may massage official statistics and projections to cast a favorable light on current policy, or to rationalize some future course of action. This problem was well illustrated by the controversies in France and Germany over the creative accounting measures taken to bring their reported budget deficits in 1997-8 closer to the Maastricht targets.

There is some evidence that in the US political bias outweighs any information advantages. The Congressional Budget Office (1996) has shown that its own independent forecasts for the US economy were consistently more accurate than those of the politically influenced Office of Management and Budget, and comparable to the consensus forecasts published by the *Blue Chip Economic Indicators* service. However, there is little corroborating work on other countries.

The performance of official forecasts in general is therefore an open and essentially empirical question. Section 2 below gives some background on the IMF and OECD data we use, and briefly reviews empirical studies of their accuracy. Section 3 introduces the *Consensus Economics* data and reviews the theoretical and practical benefits from pooling forecasts. Section 4 uses these data to compute and compare three conventional accuracy measures - bias, mean absolute error, and root mean square error in forecasts of real GDP, consumer spending, business investment, industrial production, inflation and unemployment.

Section 5 contains more formal statistical tests for relative forecast value. We test the statistical significance of differences between forecasters using the regression-based test proposed by Morgan (1940). We also conduct the test for relative information content proposed by Fair and Shiller (1990). This asks whether, even if one forecast is less accurate than another, there is nonetheless some information in the less accurate forecast which could have been used to reduce the error in the more accurate forecast. A technical innovation is that for both tests we develop GMM estimators that are robust against the heteroscedasticity and serial correlation

inevitably present in data on forecast errors. Section 6 draws some conclusions from the results.

2. The IMF and OECD Forecasts

The IMF and OECD are professional organizations with large staffs of economists. They are owned, funded and controlled by their member governments. The IMF is based in Washington and is controlled nominally by about 200 sovereign governments. Its principal purpose is to act as a lender of last resort to member governments with short and medium term balance of payments difficulties. Many senior IMF executives are political appointees, with the largest concentration from the G7 countries. The OECD is a Paris-based think-tank and inter-country liaison organization, which has as members 29 leading industrialized nations. It also is staffed by appointees from member countries.

The IMF and OECD each publish economic forecasts twice yearly. The IMF forecasts are in its *World Economic Outlook* (WEO) which is published in May and October, and the OECD forecasts are contained in *Economic Outlook* which is published in June and December. Both contain estimates/ forecasts of major economic variables in the year of publication, based on partial information about developments in that year, and forecasts for the following year.

At both organizations country forecasts are first developed internally by professional staff. The forecasts are then reviewed and discussed with the politically appointed representatives of the member governments, and often directly with the governments themselves. Anecdotal evidence, which occasionally surfaces in press reports (*International Herald Tribune*, October 24, 1988; *Financial Times*, March 30, 1996 and June 3, 1996), suggests there are occasions when the staff are at odds with national governments over the forecast, with the result that changes are made to the original numbers. Apart from any bias which it might induce, this consultation process slows the production of the forecasts. The IMF forecasts published in May and October are in fact finalized in April and September. Similarly, an elapsed time of five

weeks from staff completion to public distribution is cited as not unusual by the Chief Economist of OECD (*Wall Street Journal*, December 20, 1995).

The IMF has monitored the accuracy of its own forecasts in a series of commissioned studies, by Kenen and Schwartz (1986), Artis (1988, 1996), Barrionuevo (1993), and Loungani (2000). The performance of OECD forecasts has been evaluated independently by Ash and Smyth (1991). These are all substantial studies, covering a wide range of forecast variables, forecast horizons for many countries, and their results cannot be summarised in full here. But two points are worth noting.

First, even allowing for differences in their timing, the OECD forecasts seem a little less accurate and more biased than the IMF forecasts. However, there appears to be no significant relation between the OECD forecast error and the contemporaneous IMF forecast, and vice versa (Artis, 1988). This suggests that there is little difference in the information content of the two sets of forecasts.

Second, most of these studies compare the official forecasts to some "naive" alternative - say, a forecast of no change in interest rates, or a forecast that each year growth and inflation will proceed at the same rate as the previous year (see, for example, Theil 1958). This test is not very meaningful. While the naive no-change "random walk" model is a reasonable baseline for interest rate and stock market forecasts, these naive models do not provide credible predictions of key macroeconomic variables, and almost all IMF and OECD forecasts outperform naive alternatives very comfortably.

3. The *Consensus Economics* Forecasts

In the past decade there has been a huge growth in published economic analysis emanating from banks, corporations and independent consultants around the world, and a parallel growth in "consensus forecasting" services which gather together information from these disparate private sources. Each month since 1989, the *Consensus Economics* service has published

forecasts for major economic variables prepared by panels of 10-30 private sector forecasters, initially for the G7 countries but subsequently for over 70 other economies. Like the IMF and OECD forecasts, these predictions relate to the current year and the following year. The forecasts are published in the second week of each month, based on a survey of panelist forecasts in the previous two weeks. In contrast to the IMF and OECD forecasters, for the G7 countries virtually all *Consensus Economics* panelists are based in the country they forecast.

Below the individual forecasts for each variable, the service publishes their arithmetic average, the "consensus forecast" for that variable. Consensus forecasts are known to be hard to beat. While individual private sector forecasts may be subject to various behavioural biases (Batchelor and Dua, 1992), many of these are likely to be eliminated by pooling forecasts from several forecasters. Indeed, just as spreading investments over many assets reduces risk, so averaging forecasts across different forecasters necessarily reduces the size of the expected error, a point first formalized by Bates and Granger (1969) over 30 years ago. Since then, a large academic literature on the benefits of pooling forecasts has developed, with over 200 articles cited in the survey by Clemen (1989).

Regarding economic forecasts, Zarnowitz (1984) studied the predictions of a panel of economists tracked by the American Statistical Association and the US National Bureau of Economic Research. He concluded,

> "The group mean forecasts from a series of surveys are on average over time more accurate than most of the individual projections. This is a strong conclusion, which applies to all variables and predictive horizons covered and is consistent with evidence for different periods from other studies."

Similarly, McNees (1987) states that for a group of US macroeconomic forecasters

> "..consensus forecasts are more accurate than most, sometimes virtually all, of the individual forecasters that constitute the consensus."

Batchelor and Dua (1995) showed that in the 1980s the *Blue Chip Economic Indicators* consensus forecasts for the US outperformed about 70-80 per cent of the panel. It would of

course be helpful to identify in advance which forecasters were likely to outperform the consensus, but Batchelor (1990) showed that there is typically no consistency in individual accuracy rankings from year to year that could be used to pick the best individual forecasters.

This means that in practice the most promising alternative to official forecasts for most users of economic forecasts is not some naive model, but a consensus of private sector forecasts. This is recognized by Artis (1996), who makes a visual comparison of IMF and *Consensus Economics* forecasts for real GDP and CPI inflation, and concludes that there is "little difference between WEO and Consensus errors." In a similar vein, Loungani (2000) plots real IMF and Consensus Economics GDP forecasts for over 60 developed and developing countries in the 1990s, and notes that "the evidence points to near-perfect collinearity between private and official (multilateral) forecasts …"

In this study we look at six variables - the growth rates in real GDP, consumers expenditure, business investment and industrial production, the rate of consumer price inflation, and the unemployment rate. The first five variables are measured as growth rates from year to year, the last as a percentage of the labor force, averaged through the year. The OECD does not produce consumer price index forecasts comparable to those in Consensus Economics, and the IMF does not publish comparable figures for business investment and industrial production, so these comparisons cannot be made.

Our forecasts are for the seven G7 countries - the US, Japan, Germany, France, UK, Italy and Canada - and for ten target years, 1990-99. The figures for Germany refer to West Germany only in the years 1990-5. In the case of the IMF forecasts, we look at the predictions for the target year made at three progressively shorter horizons - in May and October of the previous year (t-1) and May of the target year t. Since these might have been available before the May and October issues of *Consensus Economics*, the IMF forecasts are compared with the *Consensus Economics* forecasts published in the second weeks of April and September. With the OECD forecasts, we look at the three sets of predictions made in the previous June and December (t-1) , and in June of the target year. But since the OECD Economic Outlook typically appears late in the month, or early in the following month, these forecasts are

compared with the June and December *Consensus Economics* figures. *Consensus Economics* was not published in May and June 1989, so there are no comparisons at the longest horizon for 1990.

A final problem to be resolved is what data to use to measure the "actual" or realized value of the forecast variables. This is less trivial than it might seem (see McNees 1989). Initial estimates of real GDP growth, for example, are often revised several times within the following year, and further revisions may occur years later as the GDP index is rebased. Here we have taken the conventional view that forecasters are judged by their ability to predict the values of variables as they appear in the middle of the following year. We use as actuals for year t the values published in the year t+1 Spring and Summer issues of Consensus Economics. The rationale is that very early "flash" estimates are too unreliable, and the years-later rebasings and redefinitions too arcane to concern either forecasters or forecast users.

4. Bias and Accuracy

In total, we make 1827 comparisons (variables x countries x target years x horizons) between IMF or OECD forecasts, and the Consensus Economics forecasts. An obvious question is - in how many cases was the private sector consensus forecast closer to the actual outturn than the forecast from the intergovernmental agency? The answer is that the consensus did better in 920 cases (50.4%) and exactly the same in 252 cases (13.8%), so the official forecasts were more accurate only in 35.8% of cases. This figure is much the same for IMF and OECD forecasts. So as expected the consensus forecasts are better, and with such a large sample the difference is statistically significant. But the IMF and OECD forecasts are clearly not completely dominated, and their relative performance deserves more detailed analysis.

We use three conventional error measures to assess relative performance. The first is forecast average *bias*, defined as the average difference between the actual value of each variable and its forecast value. A positive value for bias indicates that on average over the whole run of

forecasts for a particular variable, the actual value was under-estimated, so that the forecasts were too low. A negative bias indicates that on average the forecasts were too high.

The second statistic used is the *mean absolute error* (MAE). This is the average of all the differences between actual and forecast values, disregarding the sign of the error. So a forecast that was 1% too low (a bias of +1%) and another that was 1% too high (a bias of -1%) would both represent absolute errors of 1%.

The third statistic used is the *root mean squared error* (RMSE). This is computed by taking all errors, disregarding their signs as in the case of the MAE, and squaring them. These squared errors are then averaged to give the mean squared error (MSE). As its name suggests, the RMSE is the square root of the MSE. The MAE implicitly assumes that the seriousness of any forecasts error depends directly on the size of the error, so that an error of ±2% is treated as twice as serious as an error of ±1%. The RMSE implicitly assumes that the seriousness of any error increases sharply with square of the size of the error, so that an error of ±2% is treated as four times (2^2) as important as an error of ±1%. The RMSE therefore penalizes heavily forecasters who make a few large errors, relative to forecasters who make a larger number of small errors.

4.3 Results for G7 countries

Table 1 shows statistics on average bias for the G7 countries taken as a whole. Each figure in the table is based on errors for 7 countries x 10 target years = 70 forecasts (or 7 x 9 = 63 for the April t-1, May t-1, and June t-1 forecasts, which omit 1990). The biases are all negative for the economic activity variables (GDP, consumption, investment, production) and positive for unemployment. They are also negative for inflation. As might be expected, all biases reduce as the forecasting horizons shorten.

The generally negative bias reflects two major events of the period under study which were only partially anticipated by all forecasters. First, most countries experienced a sharp fall in

inflation in the early 1990s, and again forecasters were slow to anticipate this and were consistently a little pessimistic about inflation. Second, most economies experienced a recession or at least a slowdown in growth in the early 1990s, and in all G7 countries there were one or two years when forecasts of GDP and other output measures were much too optimistic. Fintzen and Stekler (1999) note that this has been true of all previous US recessions, and Loungani (2000) shows that the phenomenon generalizes to his larger sample of 63 countries. Of the 60 recessions in his sample, in only 3 cases did the *Consensus Economics* panel produce a negative growth forecast in the year before the recession occurred. Indeed for more than 40 cases, the consensus forecasts were still positive in April of the recession year.

In terms of the degree of bias, the *Consensus Economics* forecasts for all measures of economic activity were consistently less biased to optimism than the IMF and OECD forecasts. However, the IMF inflation forecasts at t-1 were less biased to pessimism than the *Consensus Economics* forecasts.

Tables 2 and 3 show corresponding statistics on MAE and RMSE for the G7 countries taken as a whole. As a quick test of whether the IMF and OECD forecasts outperform the *Consensus Economics* forecasts, the bottom panel of the Tables shows the ratios of the IMF and OECD error measures to the *Consensus Economics* error measure. Since a low MAE or RMSE is better, a value above 1.00 in these ratios indicates that the IMF or OECD forecast is less accurate.

The tables show similar patterns. Some variables - notably business investment and industrial production - are clearly more difficult to forecast than others, such as unemployment and inflation. However, the scale of all the errors falls as the forecast horizon shortens.

In terms of MAE and root mean square error, the Consensus Economics forecasts are more accurate than either the IMF and the OECD forecasts for almost all variables and horizons. The exceptions are in one-year ahead forecasts of inflation, where they are a little worse than the IMF, and in forecasts of unemployment, where they are a little worse than OECD. The fact that

the RMSE figures in these cases are more favorable to *Consensus Economics* than the MAE figures means that the IMF and OECD made some large errors in forecasting inflation and unemployment, which the Consensus panel collectively managed to avoid.

The superiority of the Consensus forecasts is underlined by the radar diagrams in Figure 1. The competing root mean square errors for the "year-ahead" forecasts are plotted for each variable on a separate axis. The differences are not large, but very consistent, with the line joining the Consensus RMSEs lying uniformly inside the corresponding lines for both the IMF and OECD forecasts.

4.2 Results for individual countries

Tables 1-3 report statistics which are averages across countries. As it happens, the pattern of errors across variables, time horizons and forecasters is very similar for most G7 countries, so the summary tables give a good overall picture of the relative performance of the IMF and OECD forecasts. However, it is obviously interesting to look in more detail at the country by country results. Rather than list every error measure at every forecast horizon, we have set out on Tables 4.1-4.6 comparisons of root mean square errors for the "one-year-ahead" forecasts made by the IMF and OECD towards the end of the year preceding the target year. Each table relates to forecasts of one variable, and shows RMSEs in forecasting each country, and the ratio of the RMSEs of the IMF and OECD relative to the RMSE of the *Consensus Economics* forecasts.

Tables 4.1 - 4.6 show that IMF and OECD real GDP forecasts are less accurate than the Consensus Economics forecasts for all countries except France (OECD a little better). With a few exceptions they are also less accurate in forecasting consumer spending, business investment, industrial production and unemployment. The only variable where the *Consensus Economics* forecasts do not have a clear-cut advantage is inflation. The *Consensus Economics* inflation forecasts are superior when averaged over all G7 countries. However, the IMF forecasts are more accurate for four of the individual countries, quite markedly for France and Germany.

Looking across Tables 4.1- 4.6 a final point worth noting is that the private sector forecasts outperform the IMF and OECD forecasts most markedly and consistently for the two largest economies – the US and Japan.

5. Tests of Forecast Dominance

Some of the differences in forecast performance identified above appear large - for example, the difference between OECD and *Consensus Economics* forecasts for US GDP growth, and the difference between IMF and *Consensus Economics* forecasts of inflation in Germany. Others appear small. An interesting question is - which of these comparisons shows a really significant difference in performance, and which show differences which are not really significant given the small sample size and the volatility of the target variable?

To some extent this is a subjective issue. For example, it could be argued by a user in government that the differences between OECD and *Consensus Economics* forecasters in predicting business investment are not operationally important, since all the errors are large, and the variable is not in any case a key policy target. On the other hand small differences in forecasts of policy-sensitive variables like real GDP and inflation may be of great practical importance. However, a user of macroeconomic forecasts in a multinational company might have a quite different perspective, and give a high weight to forecasts of investment and industrial production. Because of such problems in assigning weights to variables to reflect their relative importance, we do not attempt to pool forecast errors across different variables in this study. However, we do conduct two sets of tests for the statistical significance of differences in errors for each variable, using data pooled across target countries.

5.1 Significance of differences in Mean Square Errors

The first test is a variant of the procedure suggested by Morgan (1940) for testing directly the significance of differences in mean square errors, discussed by Ashley, Granger and Schmalensee (1980) and Stekler (1991).

The test involves running a regression between the differences in errors from two competing methods, and the sums of these errors. Suppose the IMF forecast of some variable in year t is IMF_t and the corresponding private sector forecast is $CONSENSUS_t$. If the actual outturn is $ACTUAL_t$, the errors from the two sets of forecasts are $ERRI_t = ACTUAL_t - IMF_t$ and $ERRC_t = ACTUAL_t - CONSENSUS_t$ respectively. We write the sum of these errors as $SUM_t = ERRI_t + ERRC_t$ and the difference between the errors as $DIFF_t = ERRI_t - ERRC_t$, and estimate the coefficients *a* and *b* in the regression:

$$DIFF_t = a + b \,.\, [SUM_t - AVSUM] + u_t \qquad (1)$$

where AVSUM is the average value of SUM_t.

It can be shown that *a* measures the difference in bias between the two forecasts, and *b* measures the difference in error variance once bias is removed. The null hypothesis of no difference in forecast performance therefore requires a = b = 0. If a is zero and b > 0, then the IMF forecast has a significantly higher error variance. If b = 0 and a > 0, then the IMF forecast has a significantly higher error variance due to greater bias. In other cases - when *b* is significantly negative, a is positive, or when the hypothesis that $a = b = 0$ cannot be rejected - we can conclude that the IMF forecasts are not significantly less accurate than the Consensus Economics forecasts.

Conventional tests for whether the parameters *a* and *b* differ from zero require the forecast errors u_t to be mutually independent and have the same variance (see Diebold and Mariano 1995, and Harvey, Leybourne and Newbold, 1997). In our data, these assumptions are not satisfied, since some years early in the period 1990-6 were more difficult to forecast than others in all countries, and in all years some countries were more difficult to forecast than others. For

this study we have developed a generalized version test which allows for these problems of correlated error variances. Full details of the test are set out below.

5.2 Differences in information content of forecasts

The second test we conduct is that proposed by Fair and Shiller (1990) to determine whether one forecast dominates another in terms of its information content. Suppose we want to compare two professional forecasts made in various years t, say IMF_t and $CONSENSUS_t$. Every year we could also make a naive forecast that the target variable took some constant value, denoted by CONSTANT. The Fair-Shiller test in effect asks us to consider a new forecast $COMBINED_t$ made by combining the naive forecast with the two professional forecasts, using a simple linear weighting scheme. The weights are chosen so as to make the combined forecast the most accurate possible in terms of minimum squared error. The combined forecast has the form:

$$COMBINED_t = [\, w_0.CONSTANT + w_1.IMF_t + w_2.CONSENSUS_t \,] \qquad (2)$$

where w_0, w_1, and w_2 are the weights which minimize squared differences between the outturns for the target variable $ACTUAL_t$ and the $COMBINED_t$ forecast.

These optimal weights can be found by running a linear regression of the form

$$ACTUAL_t = c + w_1.IMF_t + w_2.CONSENSUS_t + v_t \qquad (3)$$

where $c = w_1.CONSTANT$, and v_t is the error made at time t. The weight w_0 cannot be recovered from this regression, but the significance of the coefficient c can be tested, and this lets us assess whether there is bias in the IMF and CONSENSUS forecasts. If they are unbiased, c will not be significantly different from zero. We have already seen in Table 1 that in our data forecasts of real activity and inflation tended to be too high on average, so it is likely that the coefficient c will be negative in many cases.

The coefficients on IMF_t and $CONSENSUS_t$ effectively measure the relative information content of the two sets of forecasts. If each contains some information which is not contained in the other, the weights w_1 and w_2 will both be significantly positive. This can happen even if one forecast is uniformly less accurate than the other. For example, one might be too volatile, but give better directional signals, and this information could be used to improve the accuracy of the combined forecast.

If one of the weights is not significantly different from zero we can say that it contains no information which is not already present in the other forecast, and so adds no value to the combined forecast. If one of the weights is significantly negative, it does contain information but of a perverse kind. A negative weight means that when that forecast is raised, the optimal combined forecast should be reduced, and vice versa. If one series of forecasts, say IMF_t, was ideal in the sense that it was unbiased and dominated the baseline $CONSENSUS_t$ forecasts, we would find $c = 0$, $w_1 = 1$, and $w_2 = 0$ in Equation (3). However, as in equation (1), the error terms v_t in (3) are mutually correlated, and have non-constant variances, and this again complicate tests of whether the weights are significantly non-zero.

5.3 Testing coefficient restrictions using GMM

Since the residuals in text equations (1) and (3) are not mutually uncorrelated, nor do they have constant variance throughout the sample, ordinary least squares cannot be used to make inferences about the coefficient values. Although OLS estimates of the coefficients are unbiased, the corresponding estimates of the uncertainty surrounding these coefficients - their variance-covariance matrix - are biased and inconsistent.

The test equations (1) and (3) are linear regressions of the form

$$y_t = \beta X_t + u_t \qquad (4)$$

where y_t is the dependent variable, and X a matrix of observations a number of independent variables. For each forecast indicator, our data are annual, for seven countries i = 1, 2, .., 7 and ten target years j = 1, 2, .., 10. In estimating equations (1) and (3) we have stacked the data in blocks of seven observations, one block for each country. So the dependent variable y can be written as the vector $[y_{ij}] = [y_{11} y_{12} ... y_{1,10} \mid y_{21} y_{22} ... y_{2,10} \mid . . . \mid y_{71} y_{72} ... y_{7,10}]'$.

Ordinary least squares estimation requires that in equation (4), $var(u_t) = \sigma^2$, $cov (u_t , u_{t-k}) = 0$ $\forall$ k, and $cov(u_t , X_t) = 0$. However, the residuals from regressions estimated on our stacked data are likely to suffer from both heteroscedasticity and serial correlation.

The heteroscedasticity arises from three sources. First, the variances of forecast errors and actuals for each country are liable to be different - some countries are harder to forecast than others regardless of the year. Second, the variances of forecast errors and actuals for each year are also likely to be different - some years are harder to forecast than others in all countries. So $var(u_t) = \sigma^2_{ij.}$ Third, Harvey, Leybourne and Newbold (1996) show that in (1), under the alternative hypothesis of significant differences in forecast mean square errors, the variance of the residuals depends on the level of the independent variable, so that $cov(u_t^2 , X_t) \neq 0$.

The "serial correlation" in (4) occurs because every tenth observation represents the same target year, so shocks common to all countries will cause $cov (u_t u_{t-10}) > 0$. This is of course not genuine serial correlation, but a convenient feature of the way our panel data are organised.

Hansen (1982) has proposed a method of moments estimator which is in large samples robust to very general types of heteroscedasticity and serial correlation. The idea is to estimate (4) by OLS, and use the estimated residuals ε_t to compute a weighted variance-covariance matrix, which effectively gives less weight to the errors made in the high-variance or highly-serially-correlated observations. The variances are estimated by ε_t^2, and the k-lag covariance by $\varepsilon_t \varepsilon_{t-k}$, so that the variance-covariance matrix of coefficients becomes

$$V = (X'X)^{-1} XWX' (X'X)^{-1} \qquad (5)$$

In our case, W is a symmetric band-diagonal weighting matrix with entries $w_{t, t-k} = \phi_k \, . \, \varepsilon_t \, \varepsilon_{t-k}$, $k = 0,1, ..10,$ and the $\phi_k = 1 - k/12$ are damping factors suggested by Newey and West (1987) in order to ensure that W is positive definite. With k = 0, V reduces to the heteroscedasticity-consistent covariance matrix proposed by White (1980).

The modified variance-covariance estimator will produce different and usually higher coefficient standard errors, and lower "t-statistics", than OLS. In general, hypotheses of the form $\beta = \beta^*$ can be tested by using the result that under the null, $(\beta - \beta^*)' V (\beta - \beta^*) \sim \chi^2 (n)$ where n is the number of linear restrictions.

5.4 Results

Table 5 shows results of estimating test equation (1) for differences in mean square errors, with test statistics based on this GMM variance-covariance estimator. The data are for the "one-year-ahead" forecasts published by the IMF in October and the OECD in December of the year preceding the target year, and figures for all G7 countries are pooled as described above to obtain sufficient observations to make the comparisons meaningful.

The results show that the hypothesis a = b = 0 in (1) can be rejected for all variables except the OECD unemployment forecast. In the case of the IMF inflation forecasts this is because it is significantly less biased than the Consensus forecasts ($a < 0$). In all other cases the lower mean square errors of the Consensus forecasts are statistically significant. In some cases, where $a \neq 0$, the superiority is due to smaller bias in the Consensus forecasts - for example, IMF consumer spending forecast. In the other cases, where $b > 0$, the superiority is due to a mixture of lower bias and lower error variance.

Table 6 shows the coefficient estimates and associated GMM test statistics obtained for weighted combinations of "one-year-ahead" IMF and Consensus Economics forecasts, and OECD and Consensus Economics forecasts, as in Equation (3) above.

The results are very striking. None of the weights on IMF forecasts is significantly nonzero, and only one of the weights on OECD forecasts - for industrial production - is significantly nonzero. In contrast, all of the weights on the Consensus Economics forecasts are significantly positive, with only one exception. The Consensus Economics forecast of industrial production is dominated by the OECD forecast. In all other cases - real GDP, consumer expenditure, business investment, unemployment and consumer price inflation - the private sector consensus dominates the IMF and OECD forecasts in terms of information content.

6. Conclusions

The macroeconomic forecasts of two leading multinational agencies, the IMF and OECD, have in the 1990s generally been less accurate and less informative than the contemporaneous *Consensus Economics* forecasts, which are produced by averaging private sector predictions. On balance the information advantages of these multinational agencies do not appear sufficient to outweigh the reduction in error variance which can be achieved by pooling many individual forecasts originating from the countries concerned.

Moreover, private sector forecasters publish predictions more frequently than multinational agencies, often in response to significant pieces of news, while the agencies are constrained to publish infrequently on a fixed timetable. Hence even in cases where their accuracy is similar to that of the IMF or OECD, the Consensus Economics forecasts tend to be more timely.

These general conclusions need to be qualified in two respects. First, it should be recognised that most forecasts are joint products, and alongside their forecasts the IMF and OECD provide commentary and analysis which arguably add value to the work of private sector economists.

Second, our results are based on the ten years for which the Consensus Economics service has been running. The decade was somewhat turbulent, and presented some tough challenges to forecasters. But for each G7 country it contained only one major turning point in real growth and in inflation, and it may be dangerous to generalise from such a small sample. However, it

is noteworthy that the addition of the three "normal" years 1997-9 to the sample used in Batchelor (2000) has not changed the presumption in favour of Consensus. Indeed, in each of these years the Consensus Forecasts of GDP growth continued to show a small but consistent superiority over official forecasts.

References

Artis, Michael J., 1996, *How accurate are the IMF's short term forecasts? Another examination of the World Economic Outlook*, IMF Research Department Working Paper.

Artis, M. J., 1988, How accurate is the World Economic Outlook? A post mortem on short-term forecasting at the International Monetary Fund, *IMF Staff Studies for World Economic Outlook*, 1-49.

Ash, J. C. K, D. J. Smyth and S. M Heravi, 1990, The accuracy of OECD forecasts of the international economy: demand, output and prices, *International Journal of Forecasting*, 6, 3, 379-92.

Ashley, R., C. W. J. Granger and R. Schmalensee, 1980, Advertising and aggregate consumption: an analysis of causality, *Econometrica*, 48, 1149-1167.

Barrionuevo, J. M., 1993, How accurate are the World Economic Outlook projections?, *IMF Staff Studies for World Economic Outlook*.

Batchelor, R. A., 1990, All forecasters are equal, *Journal of Business Economics and Statistics*, 8, 143-144.

Batchelor, R. A., 2000, How useful are the forecasts of intergovernmental agencies? The OECD and IMF versus the Consensus, *Applied Economics*, forthcoming.

Batchelor, R. and P. Dua, 1995, Forecaster diversity and the benefits of combining forecasts, *Management Science*, 41, 1, 68-75.

Bates, J. M., and C. W. J. Granger, 1969, The combination of forecasts, *Operational Research Quarterly*, 20, 451-468.

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

Congressional Budget Office, 1996, *Annual report: Appendix A - Evaluating the CBO's record of economic forecasts*, (Washington, Congressional Budget Office).

Diebold, F. X. and R. S. Mariano, 1995, Comparing predictive accuracy, *Journal of Business and Economic Statistics*, 13, 253-263.

Fair, R. C. and R. J. Shiller, 1990, Comparing information in forecasts from econometric models, *American Economic Review*, 80, 375-389.

Fintzen, D. and H. O. Stekler, 1999, Why did forecasters fail to predict the 1990 recession?, *International Journal of Forecasting*, 15, 309-23.

Hansen, L-P., 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50,1029-54.

Harvey, D., S. Leybourne and P. Newbold, 1997, Testing the equality of prediction mean squared errors, *International Journal of Forecasting*, 13, 281-91.

Kenen, P. B. and A. B. Schwartz, 1986, *An assessment of macroeconomic forecasts in the International Monetary Fund's World Economic Outlook*, Princeton University Working Papers in International Economics, International Finance Section.

Loungani, P., *How accurate are private sector forecasts? Cross-country evidence from consensus forecasts of output growth*, IMF Working Paper, WP/00/77.

McNees, S. K., 1987, Consensus forecasts: tyranny of the majority?, *New England Economic Review*, Nov/ Dec, 15-21.

McNees, S. K., 1989, Forecasts and actuals: the tradeoff between timeliness and accuracy, *International Journal of Forecasting*, 5, 3, 409-416.

Morgan, W. A., 1940, A test for the significance of the difference between two variances in a sample from a normal bivariate population, *Biometrika*, 31, 13-19.

Newey, W. K. and K. D. West, 1987, A simple positive semi-definite heteroskedasticity and autocorrelation consistent variance-covariance estimator, *Econometrica*, 55, 703-8

Stekler, H. O., 1991, Macroeconomic forecast evaluation techniques, *International Journal of Forecasting*, 7, 375-384.

Theil, H., 1958, *Economic Forecasts and Policy*, (Amsterdam: North Holland), Ch. 2.

White, H., 1980, A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity, *Econometrica*, 50, 1-25.

Zarnowitz, V., 1984, The accuracy of individual and group forecasts from business outlook surveys, *Journal of Forecasting*, 3, 11-26.

Table 1. Bias in IMF, OECD and *Consensus Economics* Forecasts: all G7 countries

	IMF			**Consensus**		
	May t-1	*Oct t-1*	*May t*	*April t-1*	*Sep t-1*	*April t*
Real GDP	-1.04	-0.74	-0.14	-0.79	-0.57	-0.11
Consumer Expenditure	-0.71	-0.52	0.05	-0.43	-0.26	0.04
Business Investment						
Industrial Production						
Consumer Price Index	-0.30	-0.19	-0.07	-0.40	-0.32	-0.05
Unemployment Rate	0.26	0.04	0.00	0.19	0.07	-0.01

	OECD			**Consensus**		
	June t-1	*Dec t-1*	*June t*	*June t-1*	*Dec t-1*	*June t*
Real GDP	-0.94	-0.49	-0.13	-0.77	-0.30	-0.08
Consumer Expenditure	-0.61	-0.17	0.00	-0.39	-0.07	0.01
Business Investment	-2.40	-1.29	-0.08	-1.95	-1.02	-0.64
Industrial Production	-2.07	-1.45	-0.86	-1.80	-1.16	-0.59
Consumer Price Index						
Unemployment Rate	0.24	0.07	0.03	0.33	0.05	0.03

Notes: Figures are per cent per annum (per cent of labor force in the case of unemployment).
Results for Industrial Production and Unemployment exclude Germany.
Forecasts are for years 1990-9; but *April t-1, May t-1, June t-1* figures omit 1990.

Table 2. Mean Absolute Errors of IMF, OECD and *Consensus Economics* Forecasts: all G7 countries

	IMF			**Consensus**		
	May t-1	*Oct t-1*	*May t*	*April t-1*	*Sep t-1*	*May t*
Real GDP	1.53	1.40	0.85	1.42	1.28	0.87
Consumer Expenditure	1.35	1.16	0.86	1.13	1.04	0.76
Business Investment						
Industrial Production						
Consumer Price Index	0.87	0.60	0.35	0.71	0.62	0.30
Unemployment Rate	0.85	0.64	0.35	0.75	0.58	0.33

	OECD			**Consensus**		
	June t-1	*Dec t-1*	*June t*	*June t-1*	*Dec t-1*	*June t*
Real GDP	1.49	1.22	0.75	1.41	1.12	0.68
Consumer Expenditure	1.18	1.02	0.67	1.11	0.97	0.60
Business Investment	4.82	3.97	3.15	4.57	3.53	2.59
Industrial Production	2.98	2.47	1.89	2.79	2.33	1.56
Consumer Price Index						
Unemployment Rate	0.61	0.42	0.27	0.71	0.43	0.27

MAE Ratios	**IMF/ Consensus**			**OECD/ Consensus**		
	June t-1	*Dec t-1*	*June t*	*May t-1*	*Oct t-1*	*May t*
Real GDP	1.08	1.09	0.98	1.06	1.10	1.10
Consumer Expenditure	1.20	1.11	1.13	1.06	1.05	1.11
Business Investment				1.06	1.12	1.22
Industrial Production				1.07	1.06	1.21
Consumer Price Index	1.22	0.97	1.17			
Unemployment Rate	1.14	1.10	1.06	0.85	0.98	1.02

Notes: Figures are per cent per annum (per cent of labor force in the case of unemployment).
Results for Industrial Production and Unemployment exclude Germany.
Forecasts are for years 1990-9; but *April t-1, May t-1, June t-1* figures omit 1990.

Table 3. Root Mean Square Errors of IMF, OECD and *Consensus Economics* Forecasts: all G7 countries

	IMF			**Consensus**		
	May t-1	*Oct t-1*	*May t*	*April t-1*	*Sep t-1*	*May t*
Real GDP	2.01	1.76	1.03	1.85	1.62	0.98
Consumer Expenditure	1.73	1.49	1.06	1.54	1.35	0.91
Business Investment						
Industrial Production						
Consumer Price Index	0.92	0.85	0.45	0.91	0.83	0.40
Unemployment Rate	1.03	0.79	0.43	0.91	0.72	0.38

	OECD			**Consensus**		
	June t-1	*Dec t-1*	*June t*	*June t-1*	*Dec t-1*	*June t*
Real GDP	1.91	1.51	0.90	1.83	1.35	0.80
Consumer Expenditure	1.53	1.29	0.85	1.47	1.20	0.74
Business Investment	6.43	5.25	4.11	5.99	4.72	3.43
Industrial Production	3.87	3.10	2.36	3.77	3.00	1.92
Consumer Price Index						
Unemployment Rate	0.78	0.52	0.35	0.86	0.55	0.30

RMSE Ratios	**IMF/ Consensus**			**OECD/ Consensus**		
	June t-1	*Dec t-1*	*June t*	*May t-1*	*Oct t-1*	*May t*
Real GDP	1.08	1.09	1.05	1.04	1.12	1.12
Consumer Expenditure	1.12	1.10	1.16	1.04	1.08	1.15
Business Investment				1.07	1.11	1.20
Industrial Production				1.03	1.03	1.23
Consumer Price Index	1.01	1.02	1.11			
Unemployment Rate	1.12	1.10	1.12	0.91	0.95	1.15

Notes: Figures are per cent per annum (per cent of labor force in the case of unemployment).
Results for Industrial Production and Unemployment exclude Germany.
Forecasts are for years 1990-9; but *April t-1, May t-1, June t-1* figures omit 1990.

Table 4.1 Root Mean Square Errors: Real GDP

	IMF *Oct t-1*	CE *Sep t-1*	OECD *Dec t-1*	CE *Dec t-1*	Ratios to CE IMF	OECD
US	1.39	1.20	1.34	1.10	1.16	1.22
Japan	2.36	2.13	2.10	1.73	1.11	1.22
Germany	1.73	1.53	1.36	1.26	1.13	1.08
France	1.67	1.54	1.29	1.31	1.09	0.98
UK	1.66	1.62	1.43	1.35	1.02	1.06
Italy	1.43	1.41	1.18	1.13	1.01	1.04
Canada	1.89	1.71	1.68	1.49	1.10	1.12
G7	1.76	1.62	1.51	1.35	1.09	1.12

Table 4.2 Root Mean Square Errors: Consumer Expenditure

	IMF *Oct t-1*	CE *Sep t-1*	OECD *Dec t-1*	CE *Dec t-1*	Ratios to CE IMF	OECD
US	1.37	1.29	1.27	1.15	1.06	1.11
Japan	1.60	1.42	1.57	1.21	1.13	1.30
Germany	1.29	1.00	1.00	0.99	1.29	1.01
France	1.14	0.96	0.87	0.86	1.19	1.01
UK	1.47	1.73	1.65	1.64	0.85	1.01
Italy	1.83	1.59	1.33	1.22	1.15	1.08
Canada	1.59	1.28	1.12	1.15	1.25	0.97
G7	1.49	1.35	1.29	1.20	1.10	1.08

Table 4.3 Root Mean Square Errors: Business Investment

	IMF *Oct t-1*	CE *Sep t-1*	OECD *Dec t-1*	CE *Dec t-1*	Ratios to CE IMF	OECD
US			3.79	2.50		1.52
Japan			6.80	5.97		1.14
Germany			5.97	5.61		1.06
France			3.69	3.98		0.93
UK			4.08	3.76		1.08
Italy			4.85	4.22		1.15
Canada			6.54	5.91		1.11
G7			5.25	4.72		1.11

Table 4.4 Root Mean Square Errors: Industrial Production

	IMF	CE	OECD	CE	Ratios to CE	
	Oct t-1	*Sep t-1*	*Dec t-1*	*Dec t-1*	IMF	OECD
US			1.92	1.43		1.34
Japan			4.25	4.07		1.04
Germany						
France			3.52	3.40		1.04
UK			2.39	2.29		1.04
Italy			3.29	3.47		0.95
Canada			2.62	2.59		1.01
G7			3.10	3.00		1.03

Table 4.5 Root Mean Square Errors: Consumer Price Index

	IMF	CE	OECD	CE	Ratios to CE	
	Oct t-1	*Sep t-1*	*Dec t-1*	*Dec t-1*	IMF	OECD
US	0.59	0.66			0.89	
Japan	0.81	0.66			1.24	
Germany	0.57	0.60			0.95	
France	0.52	0.68			0.77	
UK	1.42	1.30			1.09	
Italy	0.97	0.69			1.41	
Canada	0.68	0.98			0.70	
G7	0.85	0.83			1.02	

Table 4.6 Root Mean Square Errors: Unemployment Rate

	IMF	CE	OECD	CE	Ratios to CE	
	Oct t-1	*Sep t-1*	*Dec t-1*	*Dec t-1*	IMF	OECD
US	0.70	0.50	0.58	0.40	1.38	1.47
Japan	0.35	0.32	0.27	0.20	1.08	1.30
Germany						
France	0.84	0.78	0.53	0.60	1.08	0.87
UK	0.98	0.88			1.10	
Italy	0.94	0.80	0.59	0.75	1.17	0.78
Canada	0.77	0.85	0.59	0.62	0.91	0.95
G7	0.79	0.72	0.52	0.55	1.10	0.95

Table 5. Significance of Differences in Mean Square Error: all G7 countries

IMF error - Consensus error	*a*	*b*	$\chi^2(2)$
Real GDP	0.1750* (4.94)	0.0268* (2.21)	49.72*
Consumer Expenditure	0.2600* (3.95)	0.0259 (0.91)	16.10*
Consumer Price Index	-0.1314* (5.41)	0.0405 (1.19)	47.98*
Unemployment Rate	-0.0041 (0.10)	0.0529* (2.28)	5.18

OECD error - Consensus error	*a*	*b*	$\chi^2(2)$
Real GDP	0.1931* (6.10)	0.0407* (2.26)	39.83*
Consumer Expenditure	0.0971* (2.56)	0.0336 (1.90)	13.16*
Business Investment	0.2686 (1.01)	0.0508* (3.98)	15.84*
Industrial Production	0.2757* 2.58)	-0.0025 (-0.16)	6.99*
Unemployment Rate	-0.0428 (1.04)	0.1428 (1.40)	2.48

Notes: Estimates of coefficients of text Equation (1). Figures in parentheses are t-statistics testing the coefficients against zero. Coefficients significantly different from zero at the 95% level are shown by *. $\chi^2(2)$ is a test statistic for the joint hypothesis $a = b = 0$. Cases where this is rejected are shown by *.

Table 6. Fair-Shiller Tests for Forecast Dominance: all G7 countries

	Constant c	**IMF** w_1	**Consensus** w_2	$\chi^2(2)$
Real GDP	-0.39	-0.73	1.71*	25.53*
	(-0.68)	(-1.08)	(2.97)	
Consumer Expenditure	-0.17	0.16	0.78*	1.53
	(-0.34)	(0.42)	(2.39)	
Consumer Price Index	-0.73*	-0.01	1.14*	45.80*
	(-5.13)	(-0.01)	(2.41)	
Unemployment Rate	-0.13	0.14	0.88*	1.78
	(-0.56)	(0.96)	(6.02)	

	Constant c	**OECD** w_1	**Consensus** w_2	$\chi^2(2)$
Real GDP	-0.20	-0.79*	1.81*	13.17*
	(-0.63)	(-1.86)	(4.43)	
Consumer Expenditure	0.09	-0.19	1.12*	0.74
	(0.33)	(-0.65)	(3.92)	
Business Investment	-3.09*	0.32	1.22*	23.11*
	(-4.37)	(1.63)	(5.38)	
Industrial Production	-2.00*	0.92*	0.34	16.73*
	(-3.84)	(2.24)	(0.89)	
Unemployment Rate	-0.05	-0.09	1.10*	0.90
	(-0.37)	(-0.36)	(4.69)	

Notes: Estimates of coefficients of text Equation (3). Figures in parentheses are t-statistics testing the coefficients against zero. Coefficients significantly different from zero at the 95% level are shown by *. $\chi^2(2)$ is a Wald test of the restriction $(c, w_1, w_2) = (0, 0, 1)$, which is implied by the joint hypothesis that (a) the Consensus forecast is unbiased, and (b) the IMF/ OECD forecast contains no additional information. Cases where this is rejected are shown by *

Figure 1. Radar graphs of root mean square errors
% per annum
Source: Table 3