WP/00/77

IMF Working Paper

How Accurate Are Private Sector Forecasts? Cross-Country Evidence from Consensus Forecasts of Output Growth

Prakash Loungani

INTERNATIONAL MONETARY FUND

 WP/00/77

IMF Working Paper

Research Department

How Accurate Are Private Sector Forecasts? Cross-Country Evidence from Consensus Forecasts of Output Growth

Prepared by Prakash Loungani[1]

Authorized for distribution by Eduardo Borensztein

April 2000

Abstract

This paper evaluates the performance of *Consensus Forecasts* of GDP growth for industrialized and developing countries from 1989 to 1998. The questions addressed are (1) How do forecast errors differ across industrialized and developing countries? (2) How well do forecasters predict recessions? (3) Are forecasts efficient and unbiased? (4) How does private sector performance compare with that of international organizations? (5) Is forecaster discord a reliable predictor of forecast accuracy? Two key results emerge. First, the record of failure to predict recessions is virtually unblemished. Second, there is high degree of similarity between private forecasts and those of international organizations.

JEL Classification Numbers: C5, E3, D8, G2, L2

Keywords: Forecasting, Recessions, GDP, Consensus, IMF.

Author's E-Mail Address: Ploungani@imf.org

---

[1] I thank Tam Bayoumi, Eduardo Borensztein, Peter Clark, Kajal Lahiri, and Ratna Sahay for valuable suggestions and support. I am also grateful to seminar participants at the IMF and the Federal Reserve Board and to numerous other individuals for comments and data. Grace Juhn provided outstanding research assistance.

Contents Page

I. INTRODUCTION

How well does the private sector forecast output growth? This paper provides evidence on this question using data from the publication *Consensus Forecasts*. The evidence is useful for a number of reasons. First, many industrialized countries are in the midst of long expansions, and the question of when these expansions will come to an end is being debated. Many private forecasters have issued rosy predictions for growth in the major industrialized countries and for several developing countries as well. Can these predictions be trusted? How good are forecasters at predicting recessions? Second, private sector capital flows have supplanted official funds as the dominant form of external financing for many countries. Hence, private sector assessments of the relative macroeconomic outlook for various countries play a role in guiding the allocation of capital across the globe.

Third, many in the "official" sector are increasingly relying on these forecasts as a summary of the private sector's assessment of the macroeconomic outlook. In addition to being extensively used in the multilateral institutions for this purpose[2], *Consensus Forecasts* are used by national government agencies, as revealed for example in the following quote from a speech by New Zealand's central bank governor Donald Brash (1998):

> "We do not ourselves make forecasts of the international economy, but instead use the monthly *Consensus Forecasts* ... We certainly have no reason to believe that we could produce better forecasts for our overseas markets than can the forecasters 'on the ground' in the countries concerned."

This paper evaluates the performance of *Consensus Forecasts* of real GDP growth for a large number of industrialized and developing countries for the time period October 1989 to December 1998.[3] The questions addressed are the following:

- How do forecast errors differ across industrialized and developing countries?
- How well do forecasters predict recessions?
- Are the forecasts efficient and unbiased?

---

[2] Publications such as the IMF's World Economic Outlook (WEO), the World Bank's Global Economic Prospects (GEP) and the OECD's Economic Outlook (EO) contain references to the *Consensus* forecasts. See, for instance, WEO: Interim Assessment (December 1997, pp. 34-6), Staff Studies for the WEO (December 1997, pp. 23-25) and GEP (1999, p. 9).

[3] Despite the increasing visibility of *Consensus Forecasts*, there has been very little independent analysis of their accuracy. To my knowledge, the only studies are by Artis (1997), Batchelor (1997), Harvey, Leybourne and Newbold (1999), and Gallo, Granger and Jeon (1999). The first two restrict attention to the G7 countries, the third to the United Kingdom, and the last to the United States, the United Kingdom and Japan.

- How does private sector performance compare with that of multilateral agencies?

- Is forecaster discord a reliable predictor of forecast accuracy?

Readers interested in the bottom-line will find answers to these questions in the concluding section. The remainder of the paper consists of a description of the data in Section II, evidence on accuracy and ability to forecast recessions in Section III, evidence on efficiency and bias in Section IV, a comparison with official sector forecasts in Section V and an analysis of the relationship between discord and accuracy in Section VI.

II. DATA

Consensus Forecasts has provided macroeconomic forecasts for industrialized countries on a monthly basis since October 1989. Over time, the coverage has expanded to encompass many developing countries; forecasts for these countries are reported in the publication's off-shoots, viz., *Latin American Consensus Forecasts* (published bi-monthly since 1993), *Asia Pacific Consensus Forecasts* (monthly since 1995), and *Eastern Europe Consensus Forecasts* (bi-monthly since 1998). Each of these publications surveys a number of prominent financial and economic analysts, and reports their individual forecasts as well as simple statistics summarizing the distribution of forecasts. The focus of this paper is on the mean forecast (the "consensus") and the standard deviation across forecasters.

Table 1 provides a list of the 63 countries used in the analysis, the sample period over which forecasts are available for each, and whether they are classified as an "industrialized" country or a "developing" country. As noted, for the industrialized countries the forecasts start in October 1989 (with a single exception). For the developing countries, the starting dates are more varied: in about 25 percent of the cases, the starting date is between October 1989 and October 1993; in the remaining cases, the starting date is January 1995. Most of results in this paper are based on an "unbalanced" panel (i.e., countries enter the sample at different dates) in order to make use of all available information.

The "event" being forecast is annual average real GDP growth. In principle, one could study the entire sequence of monthly (or bi-monthly) forecasts of this event. However, this preliminary study focuses on forecasts made in April and October of each year. The rationale for choosing the April forecasts is that by this time a "settled" estimate of the real GDP outcome for the previous year is available for almost all countries and this information can be incorporated in the forecast for the current year; by October, about half-a-year's worth

Table 1. List of Countries

Country Group	Industrialized Countries	Developing Countries		
		Asia-Pacific Economies	Latin American Economies	Transition & Other Economies
List of Countries	Austria (1) Australia (1) Belgium (1) Canada (1) Denmark (1) Finland (1) France (1) Germany (1) Greece (3) Ireland (1) Israel (1) Italy (1) Japan (1) Netherlands (1) New Zealand (1) Norway (1) Portugal (1) Spain (1) Sweden (1) Switzerland (1) United Kingdom (1) United States (1)	Bangladesh (4) China (4) Hong Kong (2) India (4) Indonesia (2) Korea (1) Malaysia (1) Pakistan (4) Philippines (4) Singapore(1) Sri Lanka (4) Taiwan (1) Thailand (2) Vietnam (4)	Argentina (4) Bolivia (4) Brazil (1) Chile (4) Colombia (4) Costa Rica (4) Dominican Rep. (4) Ecuador (4) Mexico (1) Panama (4) Paraguay (4) Peru (4) Uruguay (4) Venezuela (4)	Bulgaria (4) Czech Republic (4) Hungary (2) Poland (2) Romania (4) Russia (4) Slovakia (4) Slovenia (4) Ukraine (4) Egypt (4) Saudi Arabia (4) South Africa (3) Turkey (4)
Sample Period	(1) From Oct. 1989 (3) From Oct. 1993	(1) From Oct. 1989 (2) From April 1991 (4) From Jan. 1995	(1) From Oct. 1989 (4) From Jan. 1995	(2) From April 1991 (3) From Oct. 1993 (4) From Jan. 1995

Source: Author's calculations.

of data is available, but there is still a significant amount of uncertainty about the eventual outcome; the choice of these months also facilitates comparison with the IMF's forecasts.[4] For each of these months, both *year-ahead forecasts* and *current-year forecasts* are studied. In summary, there is a sequence of four forecasts for each event; the labels "Apr(t-1)", "Oct(t-1)", "Apr(t)", "Oct(t)" are used to refer to the year-ahead and current-year forecasts, respectively.

III. ACCURACY AND ABILITY TO PREDICT RECESSIONS

A. Accuracy of Forecasts

The forecast errors are given by:

$$(1) \qquad e(t) = F(t) - A(t)$$

where A(t) is a vector of growth outcomes (the "actuals"), and F(t) is the corresponding vector of forecasts. A perennial issue in the forecasting literature is whether the "actual" value should correspond to the early releases of the data or later revisions. I chose to use the former, persuaded by the argument that subsequent releases of the data often incorporate information (such as revisions of weights, changes in methods of construction, etc.) that forecasters simply could not have been aware of at the time of the forecast.[5]

Three measures of forecast accuracy are used. The first is the Mean Absolute Error (MAE), which is the average across all countries and over all years of the differences between actual and forecast values, disregarding the sign of the error. The second is the root mean square error (RMSE). To compute RMSE, the forecast errors are squared and averaged over the sample to get the mean squared error (MSE); RMSE is the square root of MSE.

The third measure, Theil's inequality coefficient (TIC), is defined as follows:

$$(2) \qquad TIC(t) = RMSE(t)/[((A(t) - A(t-1))^2/n)^{**}(½)]$$

---

[4] July forecasts will be used in the comparison with OECD forecasts, whereas the comparison with World Bank forecasts will require some other months, as discussed later.

[5] This choice was implemented by using the real GDP data as reported in the May WEO of the following year. For example, the 1990 forecast was compared to the realization as reported in the May 1991 WEO. In cases where this was not possible because the data were not reported, I attempted to use the first available realization reported in the WEO.

TIC accomplishes two things. It scales RMSE by the variability of the underlying data and it offers a way of evaluating forecasting performance relative to a "naive" forecast of no change in the growth rate between t-1 and t. TICs of less than 1 are said to beat the naive forecast.

Two main findings emerge when these measures are computed (see Figures 1 to 3). First, as one would expect, the magnitude of the forecast error declines as the forecast horizon gets shorter. In Figure 1, MAE for all countries is 2 percent for the year-ahead forecast Apr(t-1) and declines to just under 1 percent for the current-year forecast Oct(t). Note that real GDP growth averaged about 3 percent a year over this period (2.3 percent for industrialized countries; 3.6 percent for developing). In Figure 2, RMSE declines from 3 percent in Apr(t-1) to just over 1 in Oct(t). In both figures, the pattern is similar for industrialized and developing countries.

Second, while in absolute terms errors are larger for developing countries than for the industrialized countries (as shown in Figures 1 and 2), taking account of the variability of the underlying data reverses this conclusion (Figure 3). Values of TIC are always a bit higher for the industrialized country sample than for developing countries. Another noteworthy result is that year-ahead forecasts either do not beat the naive forecast of an unchanged growth rate, as in the case of the Apr(t-1) forecast, or just barely beat it, as for the Oct(t-1) forecast.

B. Forecasting Recessions

There were a total of 60 episodes of "recession" in the sample (Table 2).

Since I define "recession" as any year in which real GDP declined, the term encompasses cyclical downturns (as in the case of the United States in 1991), declines in output associated with transition from planned economies to market economies (as in the case of Hungary and Poland), and declines associated with crises of various kinds (e.g., the ERM crisis in 1992-93, the Mexican crisis in 1995 and the Asian crisis in 1997-98).

The properties of forecasts during recession years are summarized in Table 3. The first two rows show qualitative properties of the forecasts. As shown in the first row, as of April of the preceding year, the consensus forecast was for positive growth in 58 of the 60 episodes of recession (the only two episodes for which negative growth was forecast were Russia and Ukraine in 1996). By October of the preceding year, the forecast was still for positive growth in all but three cases; however, in quite a few cases the forecast was lowered between April and October by large magnitudes, suggesting that forecasters were picking up some indications that there may be trouble ahead. By April of the year of the recession, a downturn was forecast in about a third of the cases, and by October a recession was correctly called in the vast majority of the cases. However, while forecasters do recognize recessions in the year in which they occur, the results in the second row show that the magnitude of the downturn was almost always underpredicted. For instance, in October the forecasted growth exceeded actual growth in 50 out of the 60 episodes.

FIGURE 1. MEAN ABSOLUTE ERRORS (MAE)

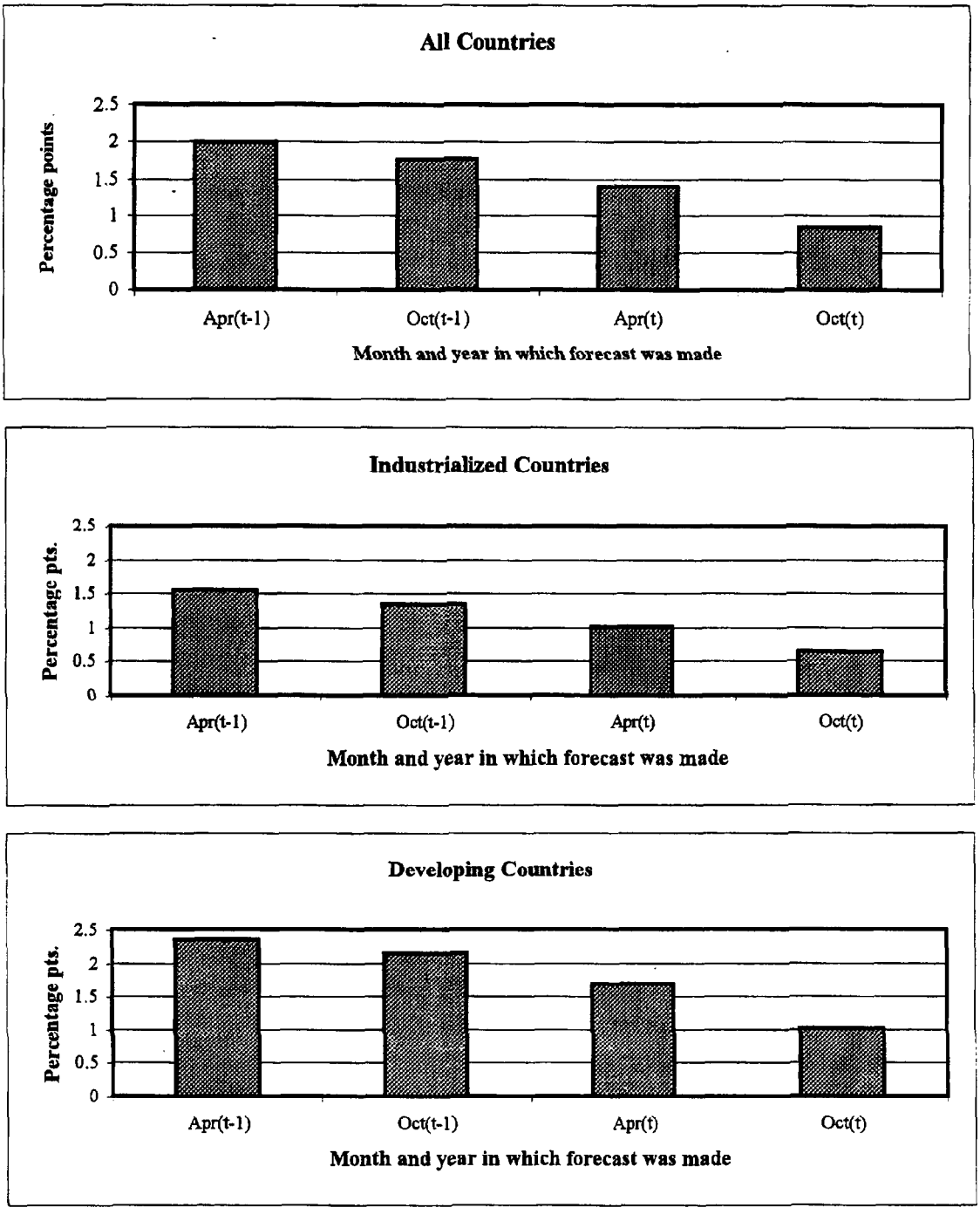
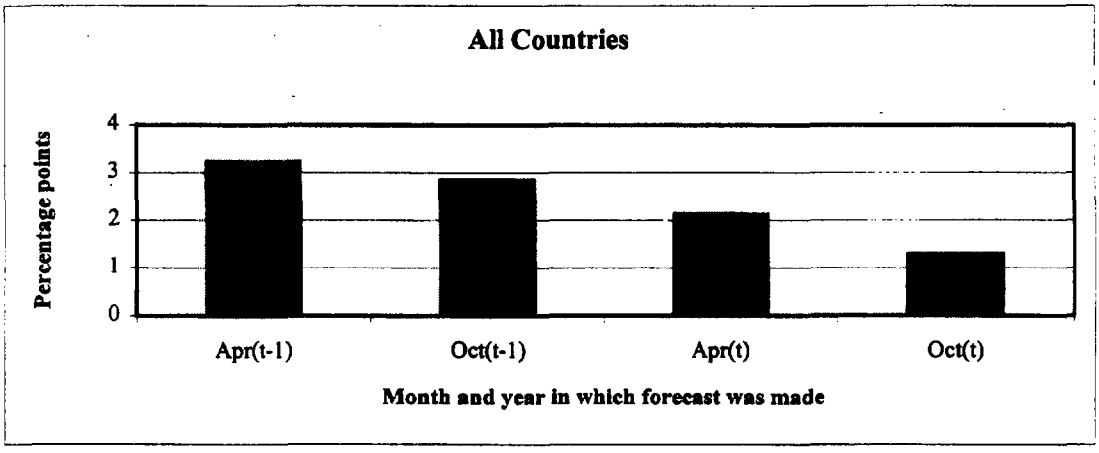
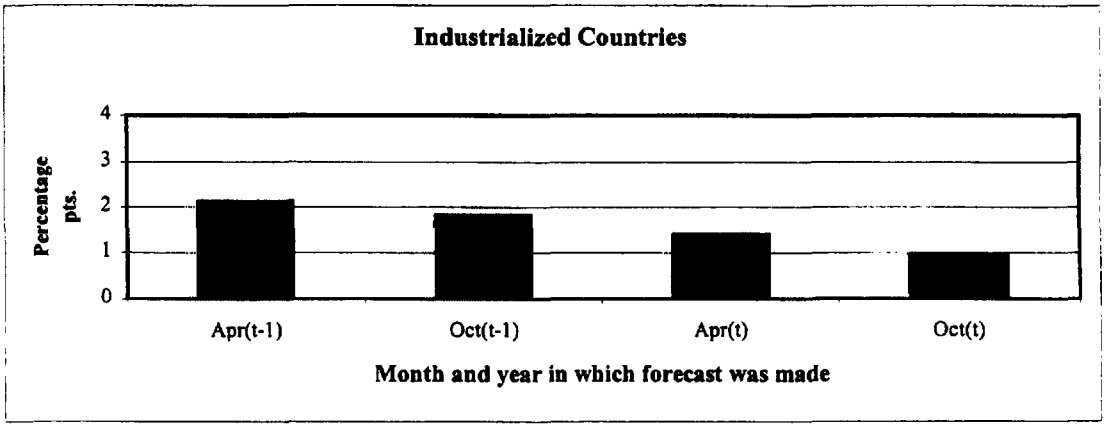


FIGURE 2. ROOT MEAN SQUARE ERRORS (RMSE)





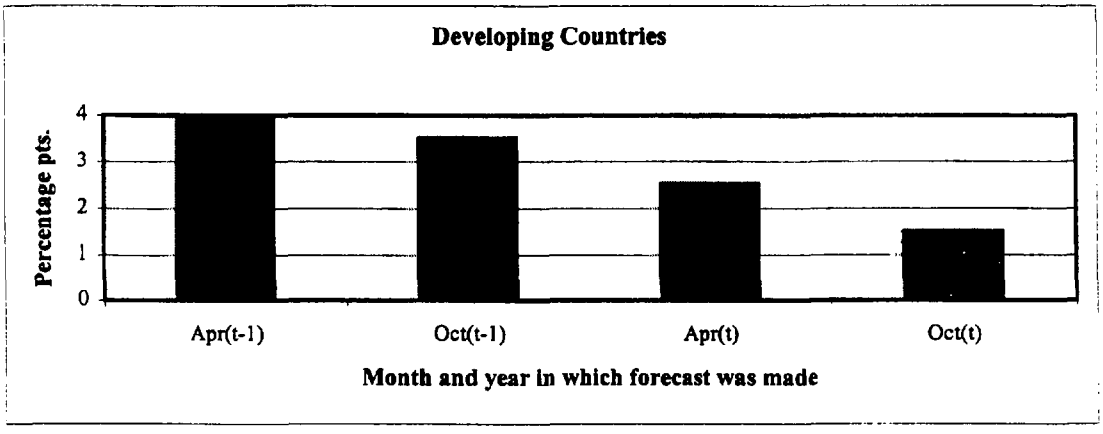


FIGURE 3. THEIL'S INEQUALITY COEFFICIENT (TIC)

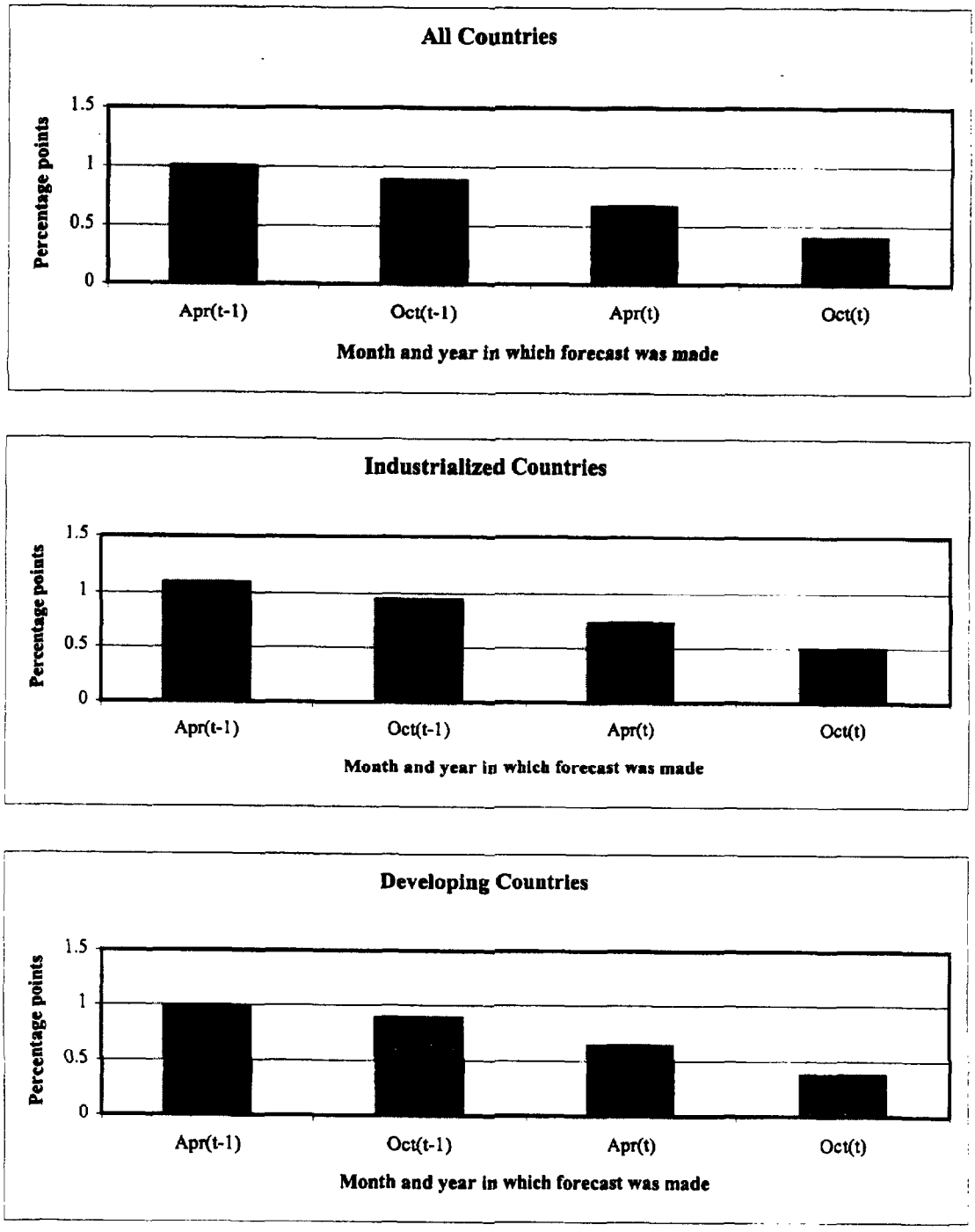


Table 2. List of Episodes of Recessions

Country Group	Industrialized Countries	Developing Countries		
		Asia-Pacific Economies	Latin American Economies	Transition & Other Economies
List of Countries, Recession year(s)	**Austria** *1993* **Australia** *1991* **Belgium** *1993* **Canada** *1991* Denmark **Finland** *1991, '92, '93* **France** *1993* **Germany** *1993* **Greece** *1993* Ireland Israel **Italy** *1993* **Japan** *1998* Netherlands **New Zealand** *1989, '90, '91, '98* Norway **Portugal** *1993* **Spain** *1993* **Sweden** *1991, '92, '93* **Switzerland** *1991, '92, '93*, '96 **UK** *1991, '92* **USA** *1991*	Bangladesh China **Hong Kong** *1998* India **Indonesia** *1998* **Korea** *1998* **Malaysia** *1998* Pakistan **Philippines** *1998* Singapore Sri Lanka Taiwan **Thailand** *1997, '98* Vietnam	**Argentina** *1995* Bolivia **Brazil** *1990, '92* Chile Colombia **Costa Rica** *1996* Dominican Rep. Ecuador **Mexico** *1995* Panama Paraguay Peru **Uruguay** *1995* **Venezuela** *1996, '98*	**Bulgaria** *1996, '97* **Czech Rep.** *1998* **Hungary** *1991, '92, '93* **Poland** *1991* **Romania** *1997, '98* **Russia** *1995, '96, '98* Slovakia Slovenia **Ukraine** *1995, '96, '97, '98* Egypt **Saudi Arabia** *1995* South Africa Turkey

Source: Author's calculations.

The final row shows the average forecast *error* at the four forecast horizons over all 60 episodes and also for the industrialized countries and developing countries separately. There is a large upward bias in the year-ahead April forecasts that only slowly dissipates over time. The bias is larger for the developing country sample, but the qualitative pattern is similar across countries.

Table 3: Forecast Performance During Recession Episodes
[Total Number of Episodes = 60]

	Apr. (t-1)	Oct. (t-1)	Apr. (t)	Oct. (t)
Number of episodes where a recession was forecast (Forecast < 0)	2	3	20	47
Number of episodes where forecast was too optimistic (Forecast > Actual)	60	60	59	50
Average forecast error, all 60 episodes	5.76	4.87	2.91	0.89
Industrialized countries	3.77	3.15	1.84	0.81
Developing countries	10.31	8.62	4.89	1.05

Source: Author's calculations.

The failure to predict recessions has been a notable feature of forecasts of the U.S. economy. Analyzing forecasts for the U.S. economy for the period 1953 to 1986, Zarnowitz (1986) concluded that major "failures of forecasting are related to the incidence of slowdowns and contractions in general economic activity. Forecasts ... go seriously wrong when such setbacks occur." Likewise, Fintzen and Stekler (1999)--building on work by McNees (1991) and Zarnowitz (1991)--note that the last three U.S. recessions were "not recognized even as they occurred, ... [though] the forecasts did indicate that the economy would be slowing down." The results here suggest that inability to predict recessions is a ubiquitous feature of growth forecasts.[6]

[6] There are two classes of theories for why recessions are not forecast. The first is that the *information* needed is lacking: forecasters either do not have access to reliable real-time information or lack reliable models for translating available information into predictions of a recession. The second is that the *incentives* for producing an "outlier" forecast (a recession or a strong boom) are lacking. For instance, some researchers and private forecasters argue that the incentives are tilted towards not predicting a recession. For instance, Zarnowitz argues that "predicting a general downturn is always unpopular, and predicting it prematurely ahead of others may prove quite costly to the forecaster and his customers" (1986, p. 9). Gary Shilling, a private forecaster, states: "Most economists are paid to be cheerleaders. Whistle blowers are unemployable" [Smalhout (2000)].

C. Accuracy: Recession vs. Non-Recession Years

The findings about forecasting performance during recessions raise the question of whether the pattern displayed in Figures 1 to 3 would disappear if recession years were deleted from the sample. Figures 4 to 6 suggest that this is not the case. As shown in the left-hand-side panels of these figures, the gain in accuracy from deleting the recession years is noticeable but not overwhelming. Note in particular that the Theil statistic for the current-year forecasts is quite similar for the full sample and the non-recession years.

IV. Efficiency and Bias

Two tests of efficiency are presented. The basic idea is to test whether or not the forecast errors contain systematic information that could be used to improve the forecast. The first test exploits the fact that the data set contains more than one forecast of the same event. A property of an efficient forecast is that forecast revisions should be uncorrelated; put differently, if the forecast revisions were correlated, forecast errors would contain systematic information that could be used to improve the forecast. The second test--the one more commonly used--involves a regression of the actual values on a constant and the forecasted values. Efficiency requires that the estimate of the constant is zero and that of the slope coefficient is unity.

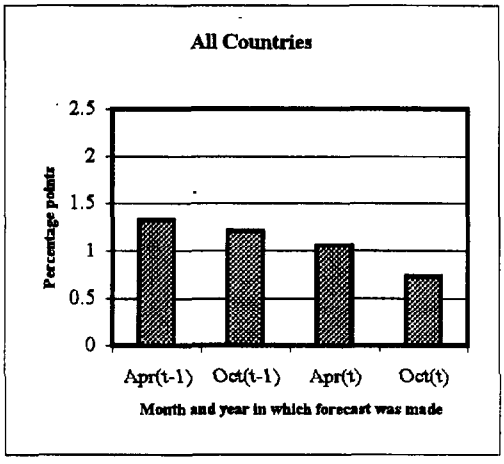
The first test of efficiency is based on work by Nordhaus (1987). The basic idea behind the test is that a sequence of forecasts of the same event must follow a martingale. Or, as Nordhaus puts it:

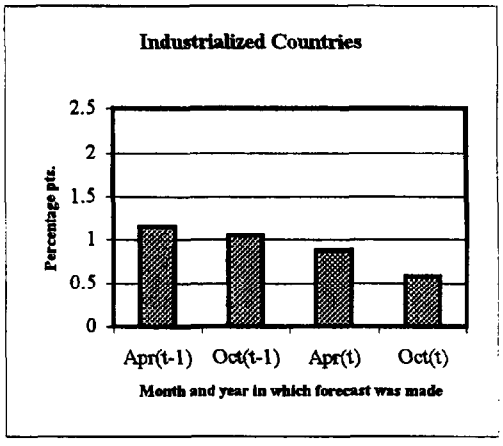
> "If I could look at your most recent forecasts and accurately say, "Your next forecast will be 2 percent lower than today's," then you can surely improve your forecasts."

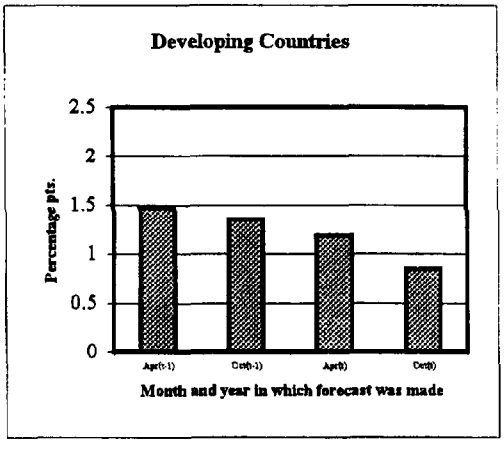
To implement this test on the data used here, define the *initial revision* of the forecast as the change in the forecast between Oct.(t-1) and Apr.(t-1), the *middle revision* as the change between Apr.(t) and Oct.(t-1), and the *final revision* as the change between Oct.(t) and Apr.(t). The results from regressions of later revisions on earlier ones are shown in Table 4. The first set of regressions is for all countries. In all four regressions, there is evidence of a strong positive correlation among forecast revisions. As shown in the next two sets of regressions, efficiency can be rejected for both industrialized and developing countries. However, the evidence against efficiency is somewhat weak is when the revisions are non-adjacent (i.e. between the final and initial revisions).

FIGURE 4. MEAN ABSOLUTE ERRORS (MAE)

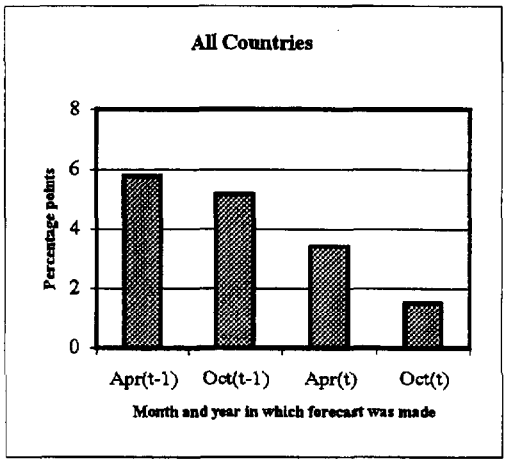
Non-Recession Years

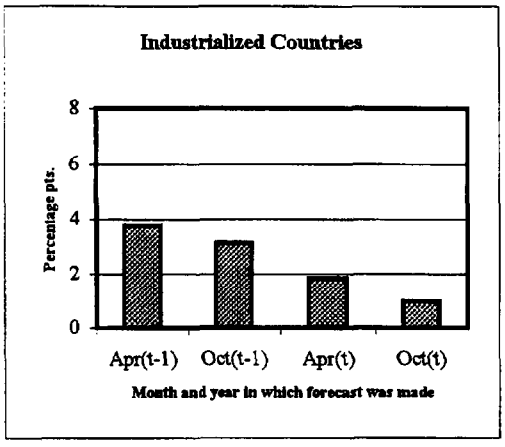






Recession Years





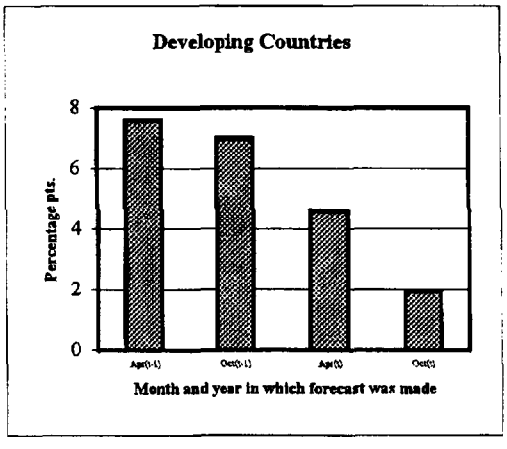
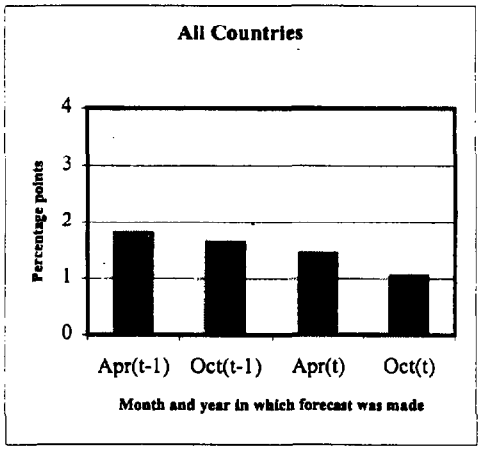
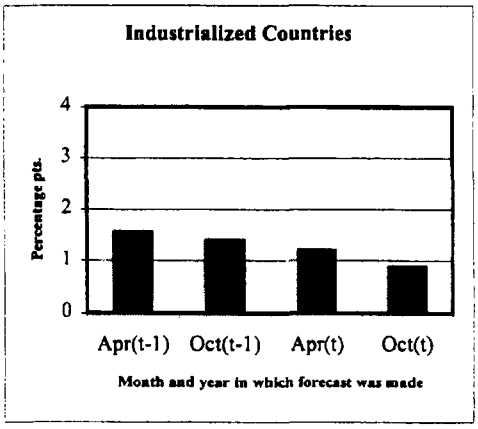
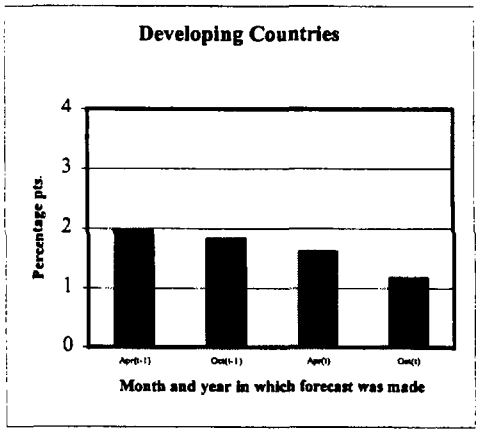


FIGURE 5. ROOT MEAN SQUARE ERRORS (RMSE)

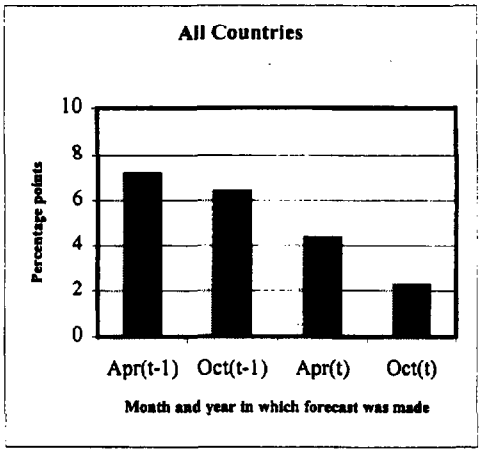
Non-Recession Years

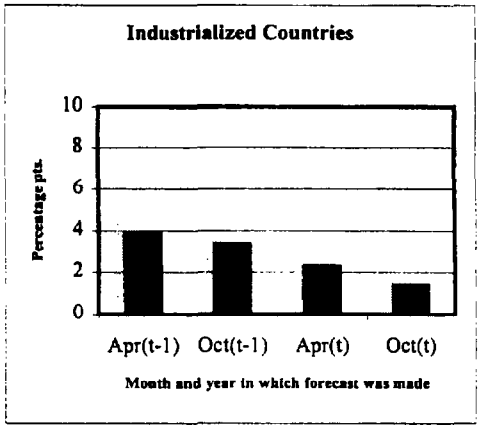






Recession Years





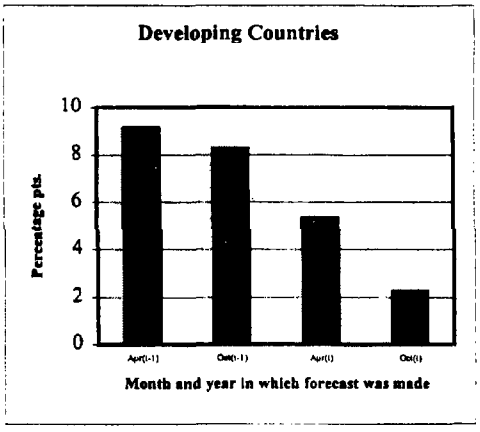
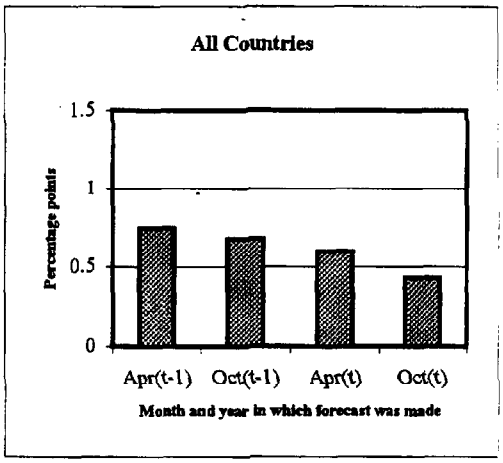
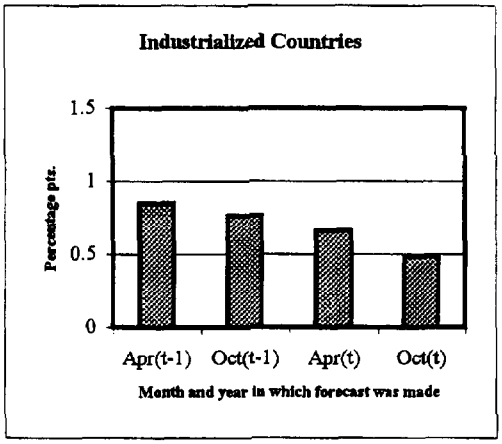
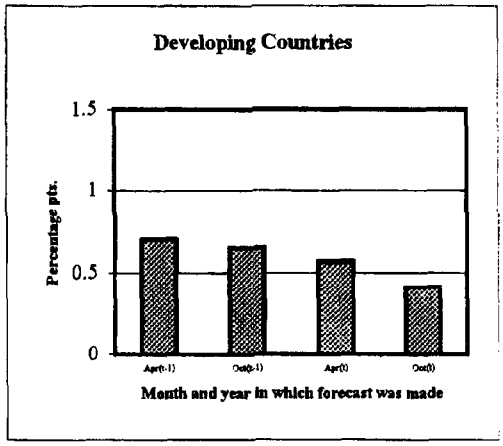


FIGURE 6. THEIL'S INEQUALITY COEFFICIENT (TIC)

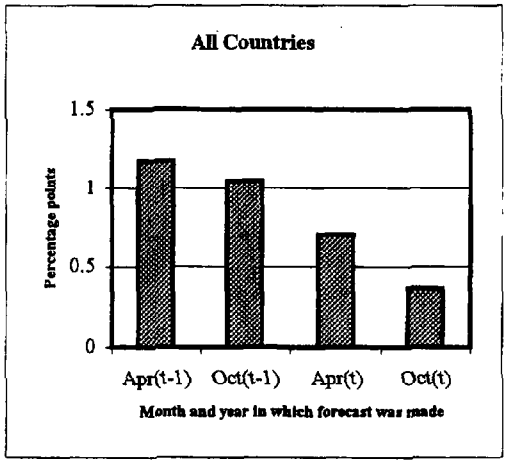
Non-Recession Years

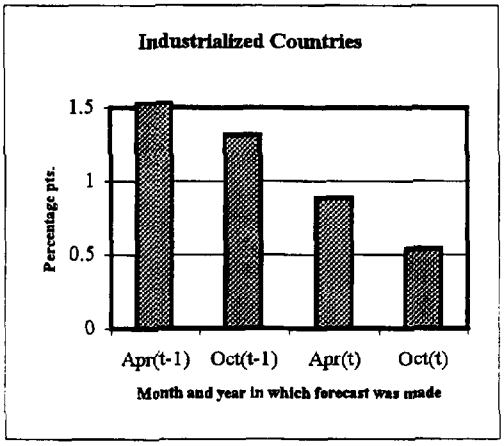






Recession Years





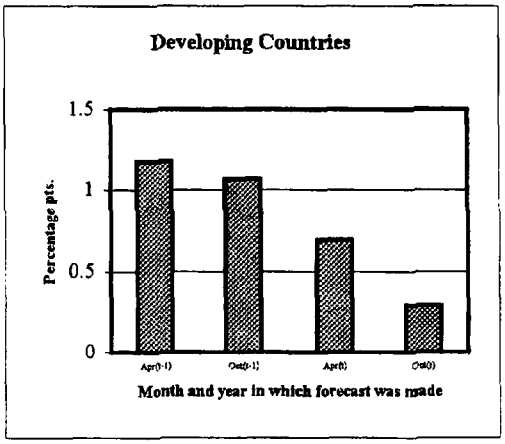


Table 4. Test of Efficiency based on Forecast Revisions

Country Group	Dependent Variable	Independent Variable(s)			Adj. R^2
		Middle Revision	Initial Revision	Intercept	
All	Final Revision	0.64	.	0.07	0.34
		(0.09)		(0.05)	
All	Final Revision	.	0.58	0.25	0.09
			(0.17)	(0.06)	
All	Final Revision	0.61	0.09	0.07	0.35
		(0.09)	(0.11)	(0.06)	
All	Middle Revision	.	0.80	0.30	0.19
			(0.16)	(0.05)	
Industrialized	Final Revision	0.52	.	-0.05	0.22
		(0.08)		(0.05)	
Industrialized	Final Revision	.	0.19	0.09	0.02
			(0.10)	(0.06)	
Industrialized	Final Revision	0.60	-0.18	-0.06	0.25
		(0.09)	(0.10)	(0.06)	
Industrialized	Middle Revision	.	0.61	0.25	0.24
			(0.10)	(0.05)	
Developing	Final Revision	0.65	.	0.21	0.36
		(0.09)		(0.09)	
Developing	Final Revision	.	0.75	0.41	0.12
			(0.22)	(0.11)	
Developing	Final Revision	0.59	0.23		0.38
		(0.10)	(0.15)		
Developing	Middle Revision	.	0.88	0.21	0.19
			(0.22)	(0.09)	

Source: Author's calculations.

Overall, there appears to be a tendency for "forecast smoothing", that is, a tendency for a revisions in one direction to be followed by further revisions in the same direction.[7]

The second test of efficiency is based on the following regression:

(3) $$A(t) = a_0 + a_1 F(t) + u(t)$$

Efficiency of forecasts holds when the intercept is zero, the slope is 1 and the u(t) are random. Regressions of A(t) on a constant and F(t) are reported in Table 5.

Table 5. Test of Efficiency

		Dependent Variable: "Actual" Real GDP Growth			
		(1)	(2)	(3)	(4)
Country Group	Independent Variables	Apr. (t-1)	Oct. (t-1)	Apr. (t)	Oct. (t)
All	Constant	-0.39	-0.93	-0.90	-0.25
		(0.41)	(0.29)	(0.21)	(0.13)
	Forecast	0.86	1.06	1.19	1.09
		(0.11)	(0.07)	(0.05)	(0.03)
	F-Statistic	15.37	12.96	9.35	6.22
	(p-value)	(0.0000)	(0.0000)	(0.0001)	(0.0022)
Industrial	Constant	-1.37	-1.83	-0.84	-0.39
		(0.81)	(0.48)	(0.24)	(0.13)
	Forecast	1.29	1.58	1.35	1.20
		(0.31)	(0.18)	(0.10)	(0.05)
	F-Statistic	7.89	8.53	6.49	7.55
	(p-value)	(0.0006)	(0.0003)	(0.0019)	(0.0007)
Developing	Constant	-0.65	-1.71	-1.40	-0.30
		(0.75)	(0.59)	(0.37)	(0.21)
	Forecast	0.88	1.15	1.23	1.09
		(0.15)	(0.10)	(0.06)	(0.04)
	F-Statistic	9.33	8.66	7.37	3.39
	(p-value)	(0.0001)	(0.0003)	(0.0008)	(0.0355)

Source: Author's calculations.
Notes: Robust standard errors are reported.

[7] Forecast smoothing has been found in many other studies of forecasting performance. Nordhaus (1987) advances a couple of conjectures to explain this finding. The first is that forecasters are fearful that "jumpy" or "jagged" forecasts will be treated as inconsistency by their bosses or customers. Second, "studies from behavioral psychology suggest that people tend to hold on to prior views too long."

In all cases, the joint hypothesis of a zero constant and a slope coefficient of unity is rejected, as indicated by the F-statistic and associated p-values reported in the last row of each block of regressions. The rejection is marginal in the case of the Oct(t) forecasts for developing countries.

Since the data are pooled across countries and over time, there is reason to suspect that the u(t) would not be random. I attempt to control for some of the possible correlations by augmenting the regression to include year fixed effects and a fixed effect for whether the country is industrialized or developing. The idea is that some years may be harder to forecast than others for all countries, and industrialized countries may be easier to forecast in all years than developing countries; these differences are picked up by the fixed effects. The results are reported in Table 6 (however, the estimated fixed effects are not reported). Once again, while there are a few cases to the contrary, the bulk of the evidence indicates rejection of efficiency.

Table 6. Test of Efficiency (Fixed Effects Included)

		Dependent Variable: "Actual" Real GDP Growth			
		(1)	(2)	(3)	(4)
Country Group	Independent Variables	Apr. (t-1)	Oct. (t-1)	Apr. (t)	Oct. (t)
All	Constant	-1.98	-1.56	-1.27	-0.33
		(0.82)	(0.68)	(0.49)	(0.31)
	Forecast	0.94	1.24	1.23	1.10
		(0.13)	(0.09)	(0.05)	(0.04)
	F-Statistic	9.36	3.56	9.19	3.70
	(p-value)	(0.0001)	(0.0293)	(0.0001)	(0.0255)
Industrial	Constant	0.27	0.20	-1.45	-0.58
		(0.79)	(0.48)	(0.41)	(0.32)
	Forecast	1.30	1.69	1.35	1.22
		(0.32)	(0.18)	(0.10)	(0.06)
	F-Statistic	6.56	31.01	7.13	7.91
	(p-value)	(0.0019)	(0.0000)	(0.0011)	(0.0005)
Developing	Constant	0.28	-0.51	-1.16	1.04
		(0.92)	(0.57)	(1.19)	(0.21)
	Forecast	0.88	1.13	1.21	1.08
		(0.14)	(0.10)	(0.06)	(0.04)
	F-Statistic	0.44	0.91	6.21	4.34
	(p-value)	(0.6465)	(0.4060)	(0.0024)	(0.0141)

Source: Author's calculations.
Notes: Robust standard errors are reported.

Forecasts are said to be unbiased if forecast errors average to zero.[8] Evidence on whether this is the case with the consensus forecasts is provided in Figure 7. As shown in the top panel, the year-ahead forecasts are marked by optimism--of about a percentage point, on average--which dissipates by October. But there is a marked difference across the two groups. The bias is much greater for developing countries than for industrial countries. In addition, the current-year forecasts for industrial countries show essentially no bias. Table 7 shows the standard errors associated with these estimates of bias. As shown, the null hypothesis of no bias cannot be rejected for the industrialized countries; for developing countries (and for the sample of all countries), the null is rejected only by the time of the Oct.(t) forecasts.

Table 7. Test of Bias

	Dependent Variable: Consensus Forecast Errors			
	(1)	(2)	(3)	(4)
Country Group	Apr. (t-1)	Oct. (t-1)	Apr. (t)	Oct. (t)
All	0.92	0.71	0.27	-0.04
	(0.17)	(0.14)	(0.11)	(0.06)
Industrial	0.60	0.37	0.05	-0.06
	(0.16)	(0.13)	(0.11)	(0.07)
Developing	1.20	1.01	0.44	-0.01
	(0.28)	(0.25)	(0.17)	(0.10)

Source: Author's calculations.
Notes: Each cell reports the results of a regression of forecast errors on a constant. Robust standard errors are reported.

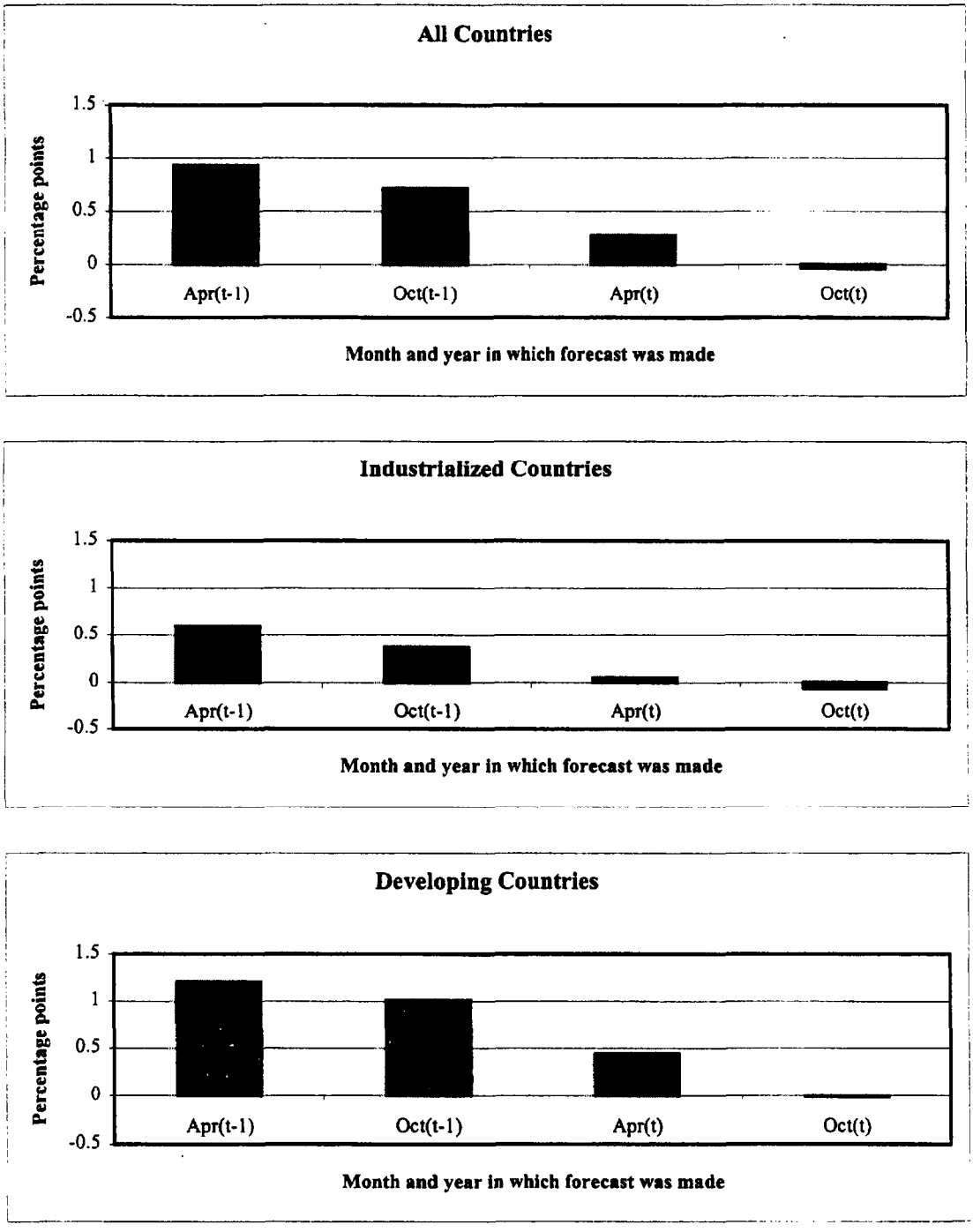
V. COMPARISON OF PRIVATE AND OFFICIAL FORECASTS

The IMF's World Economic Outlook (WEO), published in May and October of each year, reports current-year forecasts for member countries.[9] For about 50 of the largest countries--accounting for 90 percent of world output--the forecasts are updated for each

[8] The second test of efficiency used here is sometimes also used as a test of unbiasedness. However, as Holden and Peel (1990) demonstrate, this test provides a sufficient, but not a necessary condition for unbiasedness. They suggest that "correct inferences concerning unbiasedness can be obtained by testing whether the forecast error has a mean of zero" (p. 124).

[9] Year-ahead forecasts are reported for industrialized countries; for developing countries the reporting of year-ahead forecasts is less systematic.

FIGURE 7. BIAS (AVERAGE FORECAST ERROR)



WEO exercise; these countries are referred to as "Group A" countries. For the other countries, the WEO forecasts are from the most recent Article IV consultation or IMF program document, but they are "incrementally adjusted to reflect changes in assumptions and global economic conditions."[10] The May WEO forecasts are compared with the April consensus and the October WEO forecasts with the October consensus. Given the high degree of correlation in monthly consensus forecasts, it is unlikely that using the April consensus would make much difference to the results.

The upper panels of Figure 8 show the current-year forecasts and forecast errors, respectively, made by the IMF and Consensus. As should be evident from the scatter plots, the correlation between the forecasts is in excess of 0.9; the correlation between the forecast errors is in the range 0.70 to 0.96. Restricting attention to the Group A countries, which receive greater attention during the WEO exercise, does not alter the conclusion that these forecasts are virtually identical.

The comparison between OECD forecasts and Consensus is based on the June *Economic Outlook* published by the former and the July issue of the latter. While the OECD's focus is on assessing the outlook for industrialized countries, it does present forecasts for developing countries as well. Figure 9 shows the very high correlation between forecasts--and forecast errors--from the two sources for both industrialized countries (top panels) and developing countries (bottom panels).

In the case of the World Bank, only three sets of forecasts were available and these were matched to Consensus forecasts from the October 1995, January 1996 and January 1998 issues. Once again, the correlation between forecasts from the two sources exceeds 0.9 (see Figure 10).

Overall, the evidence points to near-perfect collinearity between private and official (multilateral) forecasts in the case of growth forecasts for the period under study. This conclusion needs to be tested more extensively for other variables and other time periods (and also with higher frequency data).[11] Another point to keep in mind is that the forecasts

---

[10] Preface to October 1998 WEO.

[11] However, if the near-perfect collinearity results holds up, it can be useful in evaluating the validity of theories on the sources of bias in economic forecasts. For instance, Beach, Schavey and Isidro (1999) have alleged that IMF growth forecasts are too optimistic in cases where the countries have IMF programs because the odds of program failure are not adequately reflected in the forecasts. Since the private sector is not subject to the same pressures, it is difficult to understand why its forecasts end up so close to IMF forecasts. Conversely, there is a large literature that attempts to explain bias in private sector forecasts in terms of the incentives that these forecasters face and strategic behavior among forecasters. Since IMF (and OECD and World Bank) forecasters do not face the same incentives and are not engaged in strategic behavior vis-a-vis the set of private forecasters, it

(continued...)

are only part of the product offered by private and multilateral publications; often the "story" being told to explain the outlook and the attendant risks can be more important than the point forecasts.

VI. FORECASTER DISCORD AND FORECAST ACCURACY

In addition to the mean forecast, users of *Consensus Forecasts* often look at the degree of discord across forecasters, as measured by the standard deviation of the individual forecasts. For instance, the high standard deviation of forecasts in the case of Japan over the last few years is taken to be a signal that developments in Japan have been particularly difficult to forecast in recent times. The high standard deviation could serve as a warning that the forecast accuracy for Japan's growth may be low. Conversely, a low level of forecaster discord might suggest that the country's growth prospects are relatively easy to forecast, and hence that one should expect that the forecast error will be low.

How reliable is forecaster discord as a predictor of accuracy? To investigate this issue, I estimate a regression of the (absolute value of the) forecast error on the standard deviation of the forecast and other variables. The sample consists of 26 countries, listed in Table 8, for which individual forecasts are available. The data are pooled for these countries for four years (1995 to 1998) and for two forecast horizons (the current-year April forecast and the current-year October forecast)[12]; this yields a total of 208 observations.

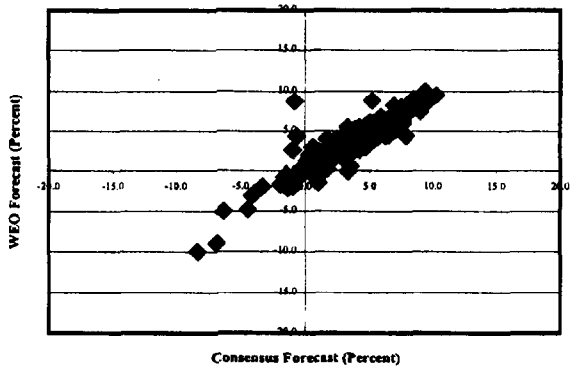
---

is once again difficult to explain the collinearity in forecasts. My conjecture, therefore, is that the near-perfect collinearity comes about because both private and multilateral forecasters are essentially reliant on official (government) forecasts, and lack either the information or the incentive to deviate too much from the government forecast. This conjecture can be tested by extending the analysis to examine the correlations among private sector, multilateral and government forecasts.

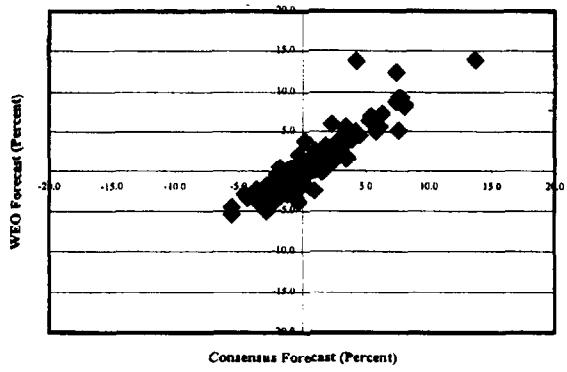
[12] In principle, one could carry out a similar analysis for the year-ahead standard deviation as well.

FIGURE 8: COMPARISON OF CONSENSUS WITH IMF'S WORLD ECONOMIC OUTLOOK (WEO) FORECASTS

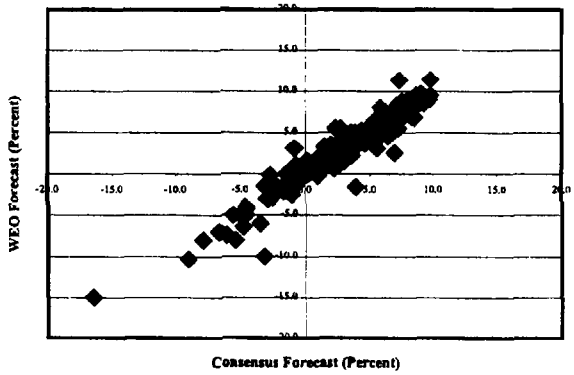
April Consensus Forecasts
and May WEO Forecasts



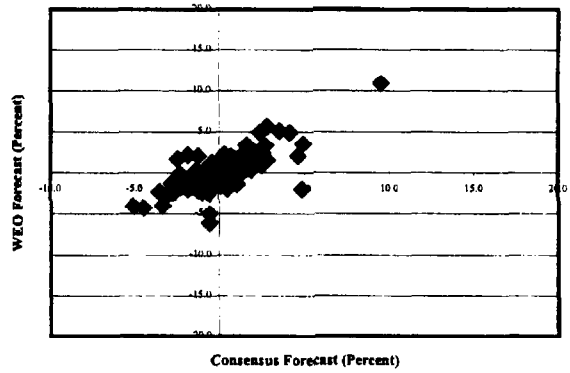
April Consensus Forecast Errors
and May WEO Forecast Errors



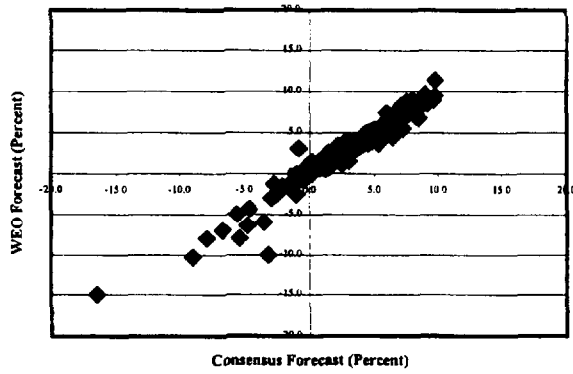
October Consensus Forecasts
and October WEO Forecasts



Oct. Consensus Forecast Errors
and Oct. WEO Forecast Errors



October Consensus Forecasts
and October WEO Forecasts
"Group A" countries



Oct. Consensus Forecast Errors
and Oct. WEO Forecast Errors
"Group A" countries

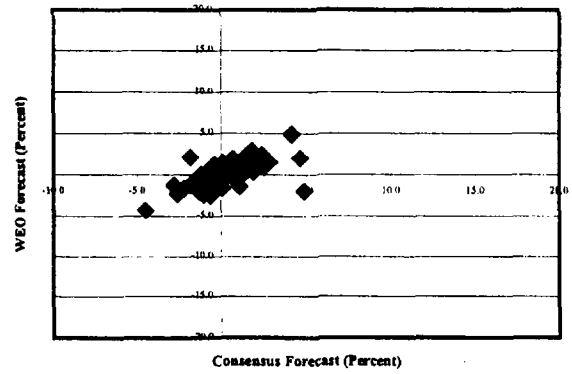
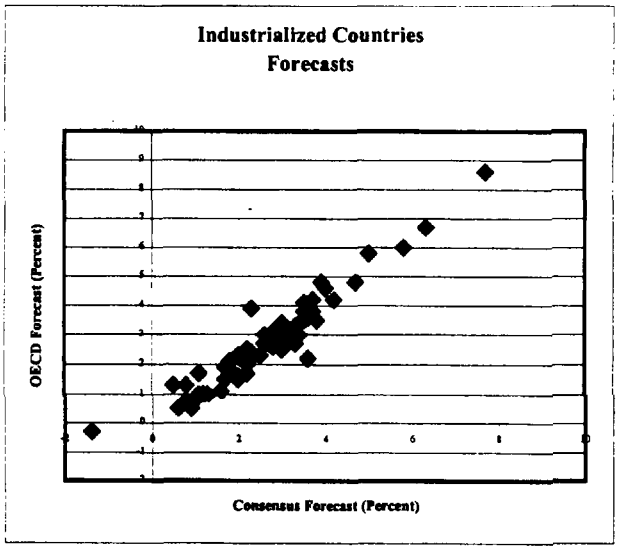
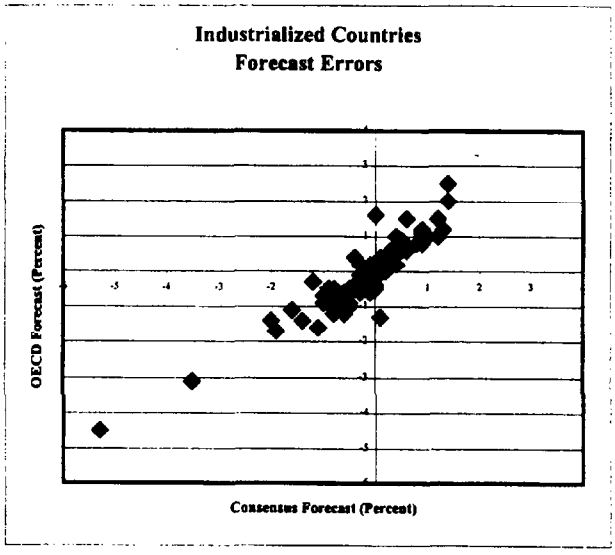
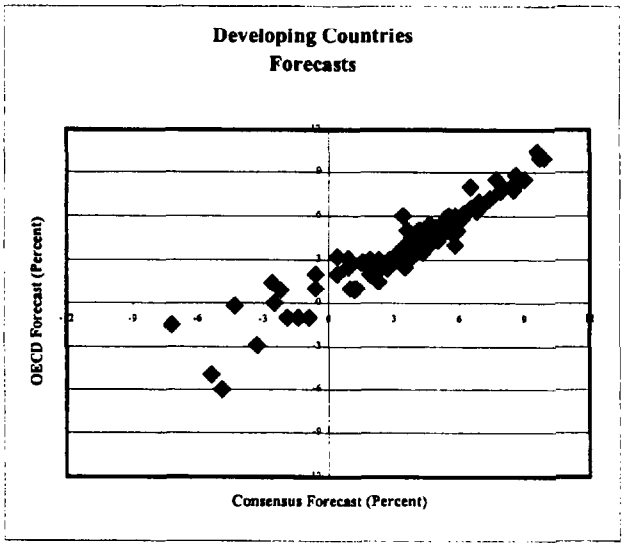


FIGURE 9: COMPARISON OF CONSENSUS WITH OECD'S ECONOMIC OUTLOOK FORECASTS







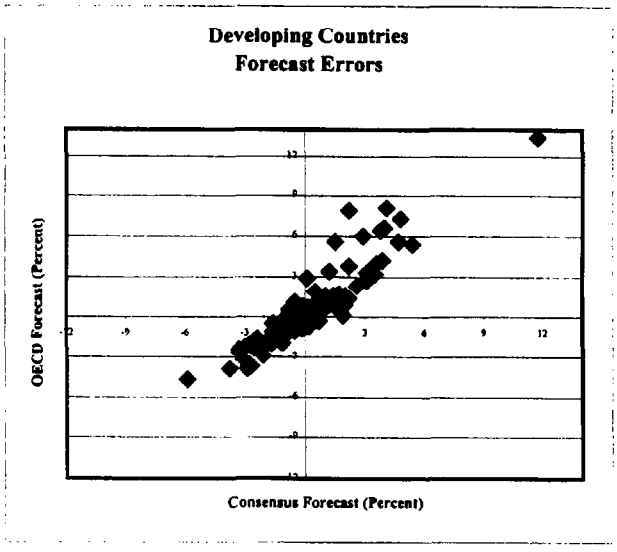
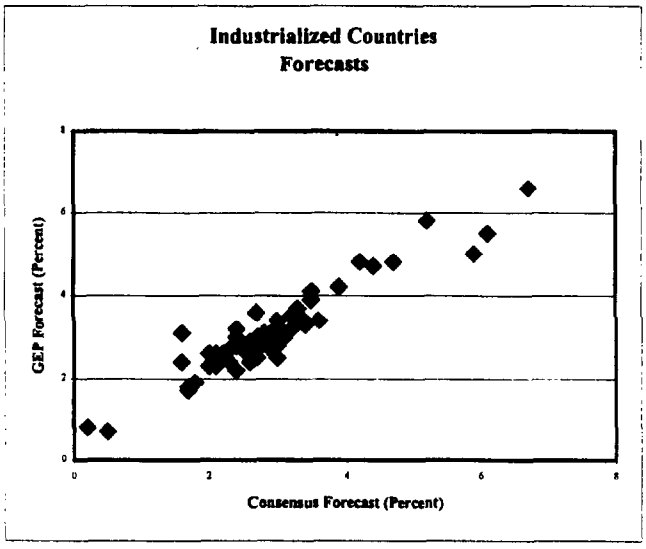
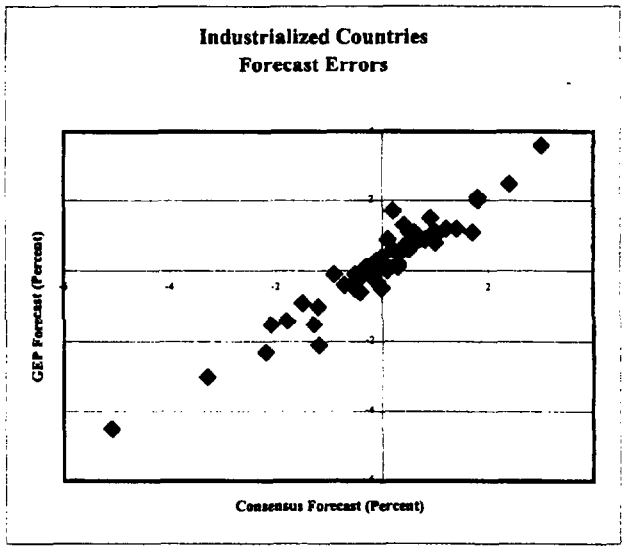
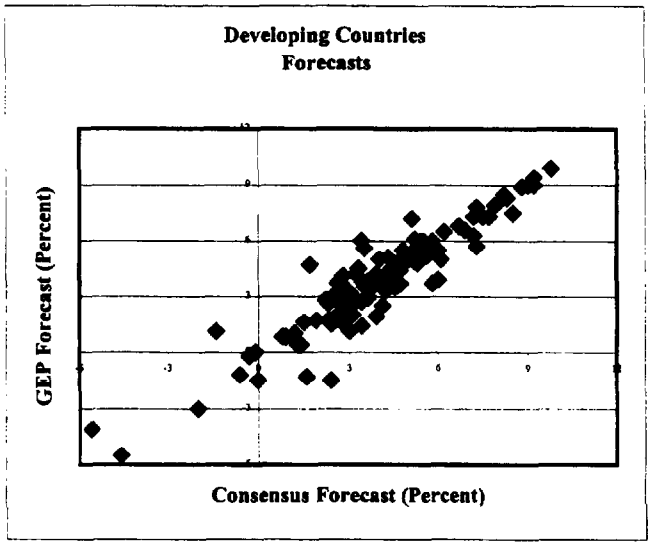


FIGURE 10. COMPARISON OF CONSENSUS WITH WORLD BANK'S GLOBAL ECONOMIC PROSPECTS (GEP) FORECASTS







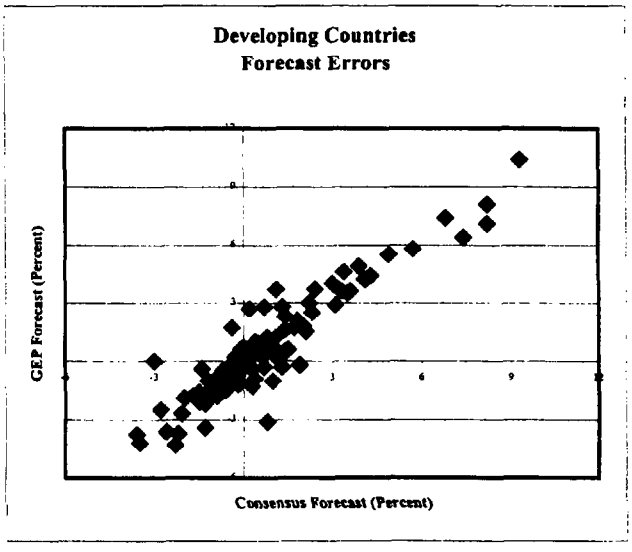


Table 8. List of Countries used in Forecaster Discord Regressions

Industrialized Countries	Asia-Pacific Economies	Latin American Economies
Canada	Australia	Argentina
France	China	Brazil
Germany	Hong Kong	Chile
Italy	India	Mexico
Japan	Indonesia	Venezuela
Netherlands	Korea	
Spain	Malaysia	
Sweden	New Zealand	
United Kingdom	Singapore	
United States	Taiwan	
	Thailand	

Source: Author's calculations.

The results of the estimation are given in Table 9. In addition to the standard error, the following explanatory variables are included: (1) dummy variables for each region and each year; (2) dummy variables to test whether the results are due to outliers; and (3) a dummy variable which takes the value 1 if the forecast was made in April and 0 if the forecast was made in October. Since, as was shown above, the forecast error is higher earlier on in the year than later, the expected sign of the coefficient on this dummy variable is positive.

The initial regression in column (1) shows that higher forecaster discord is indeed associated with higher forecast errors, controlling for the month of the forecast (April or October). The coefficient estimate is positive and significantly different from zero.

Table 9. Forecast Errors and Forecaster Discord
Dependent variable: Current-Year Forecast Error, Absolute Value (AFE)

Independent Variables	(1)	(2)	(3)	(4)	(5)
Standard deviation of forecasts	2.0	1.9	1.8	0.7	0.6
	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)
(0,1) Dummy for April or Oct. forecast	0.5	0.5	0.5	0.3	0.2
(April=1)	(0.14)	(0.14)	(0.14)	(0.1)	(0.1)
(0,1) Dummy for Industrial countries	.	-0.4	-0.5	-0.4	-0.2
		(0.2)	(0.2)	(0.1)	(0.15)
(0,1) Dummy for Asia-Pacific countries	.	-0.2	-0.2	-0.2	0.07
		(0.2)	(0.2)	(0.13)	(0.14)
(0,1) Dummy for 1995	.	.	-0.1	-0.1	-0.1
			(0.2)	(0.14)	(0.1)
(0,1) Dummy for 1996	.	.	-0.5	-0.25	-0.2
			(0.2)	(0.14)	(0.15)
(0,1) Dummy for 1997	.	.	-0.1	-0.02	-0.1
			(0.2)	(0.14)	(0.15)
(0,1) Dummy for very high AFE	.	.	.	4.0	.
(mean + 2 times s.d)				(0.3)	
(0,1) Dummy for high AFE	.	.	.	.	2.9
(mean + s.d)					(0.2)
Intercept	-0.2	0.1	0.4	0.6	0.4
	(0.13)	(0.2)	(0.3)	(0.2)	(0.2)
Adjusted R^2	0.37	0.37	0.38	0.72	0.73
Number of observations	208 (26 countries x 4 years x 2 forecast horizons)				

Source: Author's calculations.
Note: Numbers in parentheses are standard errors

Adding on region-specific fixed effects [column (2)] and year-specific fixed effects does not materially affect the strength of the positive correlation between (absolute) forecast error and forecaster discord. The inclusion of dummy variables to pick up the effects of outliers [columns (4) and (5)] attenuates the correlation, but it remains positive and significantly different from zero.

Overall, these results provide some support for the common practice of using the standard deviation of the forecast as an rough indication of the difficulty of the forecasting "terrain", and consequently as one determinant of the magnitude of the forecast error.[13]

VII. CONCLUSIONS

This paper has assembled some evidence on private sector growth forecasts for a large sample of countries. The main questions addressed, and the stylized facts that emerge, are as follows:

- ***How do forecast errors differ across industrialized and developing countries?***

  In absolute terms, the magnitude of the errors tends to be larger for developing than for industrialized countries. However, growth is more variable for developing countries; if one adjusts for this fact, say by scaling the forecast error for a country by the variability of its growth, the errors are a bit smaller for developing countries.

- ***How well do forecasters predict recessions?***

  The simple answer is: "Not very well." Only two of the 60 recessions that occurred over the sample were predicted a year in advance, two-thirds remained undetected by the April of the year in which the recession occurred, and in about a quarter of the cases the forecast in October was still for positive growth (albeit small). In eighty percent of the cases, the forecast made in October of the year of the recession underestimated its extent. This predictive failure is well-known in the case of U.S. recessions; the contribution of the work here is to show that it is a ubiquitous feature of growth forecasts. This predictive failure could arise either because forecasters lack the requisite *information* (in terms of reliable real-time data or reliable models) or because they lack the *incentives* to predict recessions; further work would be needed to discriminate between these two classes of theories.

- ***Are the forecasts efficient?***

  Efficiency of the forecasts can be rejected on the basis of a standard test, viz., a regression of outcomes on a constant and the forecasts. Exploiting the particular nature of this data set, I also document another rejection of efficiency: forecasts revisions appear to be correlated (i.e., forecasts are "too smooth"). The "smoothness" suggests that there are behavioral reasons for changing forecasts slowly or that incentives are such that smooth

[13] Zarnowitz and Lambros (1987) and Gallo, Granger and Jeon (1999) caution against using the standard deviation (across analysts) as a measure of the standard deviation of the consensus.

forecasts are preferred to "jerky" ones (for instance, the former may be more palatable to customers and bosses).

- ***Are the forecast errors unbiased?***

The forecast errors show upward bias, with the extent of the bias being much higher for the developing country sample than for industrialized countries. The bias is strongest for the year-ahead forecast made in April; it declines steadily and has vanished by the time of the current-year April forecast in the case of industrialized countries, and by the current-year October forecast for developing countries.

- ***How does the performance of the private sector compare with that of multilateral agencies?***

Private sector and official sector forecast errors have a correlation of 0.9 or better. "Horse races" between the two, as far as statistics of accuracy, bias and efficiency are concerned, are likely to result in a statistical photo-finish. This "near-perfect collinearity" result needs to be tested for other variables and other time periods. But if it holds up, it can be used to evaluate theories about sources of bias in macroeconomic forecasts. For instance, it is sometimes alleged that IMF forecasts for program countries are subject to upward bias because they do not adequately reflect the odds of program failure. However, since private forecasters are not subject to the same pressures, these allegations are difficult to reconcile with the near-perfect collinearity result.

- ***Is forecaster discord a reliable predictor of forecast accuracy?***

There is a positive relationship between the two: when there is greater discord across forecasters, the forecast error tends to be larger, on average. At the same time, the relationship is not overwhelmingly strong. This means that forecast discord can be used as *one* element in trying to gauge the likely magnitude of the forecast error, but it cannot be used as the *only* element.

REFERENCES

Artis, Michael, "How Accurate are the IMF's Short-term Forecasts? An Examination of the World Economic Outlook", Staff Studies for the WEO, 1997.

Batchelor, Roy, "How Useful are the Forecasts of Intergovernmental Agencies?" City University Business School, 1997. (Provided by Consensus Forecasts, Inc.)

Beach, William W., Aaron B. Schavey, and Isabel M. Isidro, "How Reliable are IMF Economic Forecasts?," Heritage Foundation Report 99-05, August 27, 1999.

Brash, Donald T., 1998, "Reserve Bank Forecasting: Should We Feel Guilty?," An address to The New Zealand Society of Actuaries (http://www.rbnz.govt.nz/speeches/sp981021.htm).

Fintzen, David and H.O. Stekler, "Why Did Forecasters Fail to Predict the 1990 Recession," International Journal of Forecasting 15 (1999), 309-323.

Gallo, Giampiero M., Clive Granger and Yongil Jeon, "The Impact of the Use of Forecasts in Information Sets", Deutsche Bank Research, Research Notes in Economics and Statistics, September 1999.

Harvey, David I., Stephen J. Leybourne and Paul Newbold, "Analysis of a Panel of UK Macroeconomic Forecasts," Loughborough University, Department of Economics Research Paper No. 99/8, July 1999.

Holden, Ken and D.A. Peel, "On Testing for Unbiasedness and Efficiency of Forecasts," The Manchester School Vol. LVIII No. 2, June 1990, 120-27.

Kenen, Peter and Stephen B. Schwartz, "An Assessment of Macroeconomic Forecasts in the International Monetary Fund's World Economic Outlook," Princeton University, Working Papers in International Economics G-86-04, December 1986.

McNees, S.K. (1991), Forecasting Cyclical Turning Points: The Record in the Past Three Recessions. In K. Lahiri and G.H. Moore, *Leading Economic Indicators: New Approaches and Forecasting.* New York: Cambridge University Press.

Nordhaus, William D. "Forecasting Efficiency: Concepts and Applications," The Review of Economics and Statistics, 69, 1997, 667-74.

Smalhout, Jim, "A Load of Crystal Balls," Euromoney, February 2000, Issue 370, 14-16.

Zarnowitz, Victor, "The Record and Improvability of Economic Forecasting," NBER Working Paper No. 2099, December 1986.

Zarnowitz, Victor, "Has Macro-Forecasting Failed?," NBER Working Paper No. 3867, 1991.

Zarnowitz, Victor and Louis A. Lambros, "Consensus and Uncertainty in Economic Prediction", Journal of Political Economy, 1985, vol. 95, No. 3, 591-621.