



WP/21/216

IMF Working Paper

An Evaluation of World Economic Outlook Growth Forecasts—2004-17

by Oya Celasun, Jungjin Lee, Mico Mrkaic, and Allan Timmermann

IMF Working Papers describe research in progress by the author(s) and are published to elicit comments and to encourage debate. The views expressed in IMF Working Papers are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

I N T E R N A T I O N A L M O N E T A R Y F U N D

IMF Working Paper**Research Department****An Evaluation of World Economic Outlook Growth Forecasts—2004-17****Prepared by Oya Celasun, Jungjin Lee, Mico Mrkaic, Allan Timmermann¹**

Authorized for distribution by Oya Celasun

August 2021

***IMF Working Papers* describe research in progress by the author(s) and are published to elicit comments and to encourage debate.** The views expressed in IMF Working Papers are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

Abstract

This paper examines the performance of World Economic Outlook (WEO) growth forecasts for 2004-17. Short-term real GDP growth forecasts over that period exhibit little bias, and their accuracy is broadly similar to those of Consensus Economics forecasts. By contrast, two- to five-year ahead WEO growth forecasts in 2004-17 tend to be upward biased, and in up to half of countries less accurate than a naïve forecast given by the average growth rate in the recent past. The analysis suggests that a more efficient use of available information on internal and external factors—such as the estimated output gap, projected terms of trade, and the growth forecasts of major trading partners—can improve the accuracy of some economies’ growth forecasts.

JEL Classification Numbers: C12, C40, C53

Keywords: Forecasting, forecasting bias and efficiency

Author’s E-Mail Address: ocelasun@imf.org, mmrkaic@imf.org, jlee2@imf.org, atimmermann@ucsd.edu.

¹ The authors thank Gian Maria Milesi-Ferretti for very useful discussions, Bryan Zou and Ritong Qu for excellent research assistance, and Luisa Calixto and Daniela Rojas for excellent editorial support. The analysis in this paper is based on the statistical tests presented in Timmermann (2018), provided as an online annex of this working paper.

I. INTRODUCTION

In its semi-annual World Economic Outlook (WEO) report, the Fund presents macroeconomic projections covering the current year and the next five years for almost all of its 190 member countries.² Given their pivotal role for the IMF's surveillance activities, and as a key resource on the global economic outlook more broadly, WEO forecasts are periodically evaluated and ways to improve their performance are explored.

This study updates the previous evaluations of WEO real GDP growth forecasts and compares forecasting performance in 2004-17 to that in 1990-2003, the period covered by the most recent evaluation of WEO forecasts.³ The analysis covers longer-horizon forecasts (two-to-five years ahead) in addition to the current- year and one-year-ahead forecasts that previous evaluations had focused on.

Performance is evaluated along three dimensions—accuracy, bias, and efficiency. Accuracy relates to the overall magnitude of forecast errors, whereas bias relates to whether outcomes are systematically over or under predicted. Past evaluations have found that growth tends to be overpredicted and inflation underpredicted in WEO forecasts. Efficiency relates to whether forecasts can be improved by a better use of available information; forecasts would be inefficient if the errors can be predicted using information available at the time of forecasting. We compare the accuracy and bias of the 2004-17 WEO forecasts against three benchmarks: WEO forecasts for 1990-2003, forecasts derived from simple time-series models, and forecasts published by Consensus Economics (CE henceforth). We also present some tests on the sources and efficiency of WEO growth forecast errors.

The evaluation seeks to answer the following questions:

1. How does the accuracy of WEO growth forecasts differ across country groups, forecast horizons, and how has it changed over time? To what extent can the relative difficulty of the forecasting environment explain differences in accuracy across countries?
2. How large are forecasting biases and how do they vary across groups, regions, and forecast horizons? How have biases changed since the last evaluation?
3. To what extent can errors in forecasting external factors such as the terms of trade, commodity prices, and growth in large trading partners explain WEO GDP forecast errors?
4. Can errors be predicted based on information available at the time of forecast preparation? Are forecast errors serially correlated, and are forecasts of external factors or the output gap correlated with the errors?

² Interim updates of real GDP growth forecasts for the current year and next year are published in WEO Updates in January and July for a subset of countries.

³ This study is the fifth in a series of evaluations of WEO forecasts, following Artis (1988 and 1997); Barrionuevo (1993); and Timmermann (2007).

5. How does the accuracy of WEO forecasts compare to those of CE forecasts?

The rest of the paper is structured as follows. Section 2 describes the data. Section 3 documents the predictive accuracy of WEO GDP growth forecasts, comparing performance across different horizons, countries, and over time; it also contrasts the accuracy of WEO growth forecasts to those of a naïve forecast based on the historical average of past growth. Section 4 examines the bias in growth forecasts and how it has changed over time. Section 5 presents a set of regressions aimed at understanding the sources of errors and determining whether information on the external environment is used efficiently, including information on the terms of trade, the output gap, and growth forecasts for China, the euro area, and the United States. Section 6 compares the performance of the WEO forecasts of GDP growth to a similar set of forecasts produced by Consensus Economics. Section 7 summarizes the main findings and discusses steps through which WEO forecast performance could be enhanced.

II. DATA

A. Forecasts and Outturns of GDP Growth

Timing conventions. In April (Spring) and October (Fall) of each year the IMF's World Economic Outlook (WEO) reports forecasts of IMF member countries' economic performance. The forecasts result from a process stretching several weeks back from the publication date. In each round, WEO forecasts are reported for the current year as well as for each of the next five years. Hence, the forecasts are available for six different years running from $h=0$ to $h=5$, where $h=0$ corresponds to the current-year, $h=1$ to the next year, etc. With two forecast rounds within a year, forecasts are made for 12 different horizons in total. Current-year fall ($h=0, F$) and spring ($h=0, S$) forecasts differ from the forecasts with longer horizons since economic outcomes for part of the year targeted by these forecasts are observed at the time the forecast is produced. For example, the current-year fall forecasts have the advantage that preliminary data for at least half of the current year will typically have been observed. Hence, the current-year forecasts, and particularly those reported in the fall, are really a hybrid of a nowcast and a more traditional forecast. As such, we would expect forecasts for the current-year to be substantially more accurate than the longer-term forecasts.

Our full data set on forecasts and matched actual values goes back to 1990 for the current year forecasts and 1995 for the five year ahead ones, and ends in 2017, giving us a sample of 28 outturn observations for the current-year forecast horizon and 23 outturn observations for the five-year horizon. Most of the calculations in this paper will be using the sample period from 2004 till 2017, with comparisons calculated for 1994-2003.

Real GDP growth data are subject to revisions, requiring a choice on which vintage of the outturn to use to measure the "actual" value. We follow the convention of using the actual value for a given year (t) as reported in the following year's ($t + 1$) Fall issue of the WEO.⁴

⁴ Using revised data from the latest data vintage available would lead us to mix errors with data revisions, reducing our ability to assess real-time forecasting performance.

A natural starting point is to analyze how close the forecast was to the outcome. We define the h -step-ahead forecast error for country i as the difference between the outcome and the h -step-ahead forecast:

$$e_{it|t-h} = y_{it} - \hat{y}_{it|t-h}, \quad (1)$$

where y_{it} denotes GDP growth in country i in year t , $\hat{y}_{it|t-h}$ is the forecast of y_{it} produced in year $t-h$, where $h \in \{0, 1, 2, 3, 4, 5\}$ is approximately the forecast horizon measured in years.⁵ A negative error means that growth was overpredicted.

B. Defining Groups of Economies

Forecasting performance is presented for 12 groups: the entire sample (World); a total of 10 subgroups based on either income level, fuel-exporter status, or geographic region following the groupings in the October 2017 WEO; and IMF-program status (i.e. whether the country was implementing an IMF program in the year targeted by the forecast). The three groups by income level are: Advanced Economies (AE), Emerging Market Economies (EME) and Low-Income countries (LIC). The union of the last two groups corresponds to the Emerging Market and Developing Economy (EMDE) group in the WEO. The EMDE group is then further split into Fuel Exporters (EMDE_FE) and Fuel Importers (EMDE_FI), or by region: Emerging and Developing Europe (EEUR); Emerging and Developing Asia (DASIA); Latin America and the Caribbean (LAC); Middle East, North Africa, Afghanistan, and Pakistan and Commonwealth of Independent States (MENAP_CIS); and Sub-Saharan Africa (SSA). The last group is formed by pooling the country-years under IMF-supported programs (Program).

C. Outliers

The distribution of a given country's real GDP growth rates tends to be asymmetric, skewed toward the left. This asymmetry manifests itself in the mean growth outturn for a country typically falling short of the median growth outturn over the same period (Table 1), and reflects that many economies are occasionally set back by unpredictable and costly events such as natural disasters, cross-border or internal conflict, or severe economic crises, whereas only a few experience substantial upside surprises such as a major natural-resource discovery. The presence of a tail of very weak or negative growth outturns could partly explain why the average forecast error tends to be negative (implying a tendency for overpredicting growth) in the raw data.⁶ Before proceeding with the analysis of forecast performance, we drop outliers associated with outcomes that are essentially unpredictable, so that the findings are more informative about forecasting practices and how they can be improved. In particular, we drop observations for years with major natural disasters and cross-border conflict, and apparent data entry errors. We also drop the forecast errors for 2009, a year marked by a rare global economic crisis and severe growth shortfalls many economies. However, we also report the main results when the 2009 errors are kept in the sample. Table 1 shows that dropping 2009 has, by far, the

⁵ The WEO forecasts are generated in the spring and fall which means that the forecast horizon for the current year effectively involves a fraction of a year.

⁶ Box 1.5 in the October 2018 World Economic Outlook presents stylized facts on large output collapses (cumulative declines of per capita GDP of 20 percent or more), showing that they are not rare. That analysis identifies 58 collapse episodes with a trough in 1994 or later).

largest effect on both the weighted average skew as well as on the percent of countries with a negative skew. Still, dropping country-year observations with conflict, natural disasters, or entry errors helps to meaningfully reduce outliers in our data.

Annex 1 provides more details on the criteria for selecting the outliers and presents the mean and median of errors before and after eliminating them. The mean forecast error typically increases, getting closer to the median error, after dropping observations for years marked by natural disasters and conflict (Table 1).

D. Forecasting Instruments

Our analysis of the sources and efficiency of WEO growth forecast errors uses WEO forecasts of the output gap for the advanced economies in the sample, as well as forecasts of the terms of trade and commodity terms of trade for all countries. The forecasts of terms of trade are based on WEO projections of import and export prices for each country.

For the output gap and terms of trade variables, just like for real GDP growth, we have separate forecasts from the Spring and Fall WEO issues and forecasts covering horizons from $h = 0$ (current-year) through $h = 5$ (five-years ahead), with a total of 12 forecast horizons.

III. PREDICTIVE ACCURACY

A. Predictive Accuracy for Different Country Groups and Forecast Horizons

To gauge predictive accuracy, we report the most commonly used measure, the root mean squared forecast error (RMSE). A measure of “absolute” forecast accuracy, the RMSE indicates by how many units (e.g., percentage points of GDP growth) the forecast differed from the outcome on average over the sample period. It is given by the square root of the average squared error. For a sample $[t_0: t_1]$ and a forecast horizon of h , the RMSE for country i is computed as:

$$RMSE_{i,h} = \sqrt{(t_1 - t_0 + 1)^{-1} \sum_{t=t_0}^{t_1} e_{it|t-h}^2} \quad (2)$$

Figure 1 displays the inter-quartile ranges, medians, and GDP-weighted means of the RMSE values for each of the 12 country groups.⁷ Moving from left to right along the horizontal-axis of each chart, we show the RMSE in increasing order of length of the forecast horizon, namely the RMSEs of current-year Fall ($h = 0, F$), current-year Spring ($h = 0, S$), next-year Fall ($h = 1, F$), and next-year Spring ($h = 1, S$) forecasts, followed by the two-, three-, four-, and five-year Fall WEO forecasts ($h=2, F$) to ($h=5, F$).

In the period 2004–17, AEs had the most accurate forecasts, with the median RMSE of the group starting at about 0.7 percentage point for the current-year Fall forecast and rising to about 1.8 percentage points by the next-year Spring forecast, and staying close to that level for

⁷ The weighted average is calculated by weighing the forecast error of a given country in year t with the GDP of that country in year t at purchasing-power parity and dividing by total PPP GDP in year t for that group. The same weighting method is used in calculating real GDP growth rates for the groups and regions in the *World Economic Outlook*, Table 1.1.

the two- to five-year horizons. In the case of EMEs and LICs, the median RMSE starts at about 1.2–1.5 percentage points for the current-year Fall horizon, rising slightly above 2 percentage points by the next-year Fall forecast, and staying around that level over the longer horizons. While the median RMSE of EMEs and LICs are very close, the LICs have a slightly wider interquartile range, indicating more diverse WEO forecast accuracy in that group. Among EMEs and LICs, fuel exporters (EMDE_FE) tend to have larger forecast errors, with the median RMSE close to 2 percentage points for the current-year Fall forecasts and around 3.7 percentage points for the three- to five-year forecasts. Across EMDE regions, forecasts tend to be least accurate in MENAP_CIS, reflecting a high share of fuel exporters in the group. Median RMSE values are broadly comparable across the other EMDE regional groups, and for the sample of IMF program observations.

As expected, the RMSE values of current-year forecasts ($h = 0, F$ and $h = 0, S$) are notably lower than those of forecasts of longer horizons. Current-year forecasts are more accurate because indicators for the outcome in the target year, such as industrial production or payroll and employment reports, are observed by the forecasters. For the current-year Fall forecast, preliminary values of GDP growth will also have been observed for part of the year. The availability of such pertinent information almost mechanically improves forecast accuracy.

Accuracy drops as the horizon lengthens, but nonlinearly. The greatest losses in accuracy occur as the horizon extends from three months (the current-year Fall forecasts) to nine months (current-year Fall forecasts) and then to 15 months (Fall next-year forecasts). To document how RMSEs for individual countries evolve as the forecasting horizon lengthens, Figure 2 shows the distribution of the ratio of the RMSE value for a given forecast horizon relative to the RMSE value for the closest shorter forecast horizon. There is clear evidence of large improvements, sometimes over 100 percent, in predictive accuracy as we move from current-year Spring WEO forecasts ($h = 0, S$) to current-year Fall WEO forecasts ($h = 0, F$). Gains are still sizable, but smaller, when moving from the next-year Fall WEO forecasts ($h = 1, F$) to the current-year Spring WEO forecasts ($h = 0, S$). Smaller gains, and in some cases small losses, in predictive accuracy are observed as we switch from forecasts for the next-year made in the Spring to those made in the Fall (comparing ($h=1, S$) versus ($h=1, F$)), or when comparing forecasts for the outer years. Put differently, accuracy doesn't change nearly as sharply as the horizon lengthens beyond one and a half years (that is, beyond the next-year Spring forecasts).

B. How Has Forecasting Performance Changed Over Time?

This section analyzes how the predictive accuracy of forecasts for the period 2004–17 compare with those for 1990–2003, the subsample used in the last comprehensive evaluation of WEO forecasts (Timmermann, 2007). Shifts in the underlying economic environment—for instance, due to major supply shocks or financial crises, or the introduction of new monetary or fiscal policy frameworks—could drive changes in forecasting performance. Changes may also result from the adoption of new forecasting methods or by the emergence of new datasets that allow forecasters to better monitor and predict economic outcomes.

Comparing the predictive accuracy across the two subsamples, however, only allows us to determine how the accuracy of WEO forecasts have changed in an *absolute* sense. Since the

volatility and predictability of real GDP growth cannot be expected to be the same in the two subsamples (or across countries), our findings should not necessarily be interpreted as evidence that the WEO forecasts have become better or worse over time. We turn to that issue in the next section.

Figure 3 shows the difference in the RMSE values calculated for two sub-samples, those for 1990–2003 minus those for 2004–17 (a positive median difference, shown by the red dotted curve, indicates an improvement in forecast accuracy in 2004–17 relative to 1990–2003 for more than half of the countries).

For more than half of the countries in the overall sample, forecasts for most horizons were indeed more accurate (i.e. RMSE values were lower) in 2004–17 relative to 1994-2003. The median decline in RMSEs for the World sample is about 0.7 percentage point for the shortest, current-year Fall projection and drops to about 0.3 percentage point for the four- and five-year ahead horizons. Within the AE and EMDE_FE groups, only about half of the countries saw improvements in accuracy at the four-to five-year forecast horizons, and within the MENAP_CIS and SSA groups only about half have seen improvements at the five-year horizon. In all other cases, the median of the difference between the 1990-2003 RMSE and the 2004-2017 RMSE are positive, meaning improved accuracy over time. Annex Figure 1 shows that the increases in accuracy tend to be mostly statistically significant (especially within the LAC, MENAP_CIS, and SSA groups) while the worsening in accuracy (increases in RMSEs) are often statistically insignificant.⁸ Importantly, there are virtually no instances in which increases in RMSE-values for the current- and next year forecasts between the first and second subsamples are statistically significant. In the AE group, the lack of improvements in accuracy for the outer horizons reflect the relatively large errors for the years following the Global Financial Crisis, including during the euro area crisis. The large improvements in forecast accuracy for the EEUR group, in turn, reflect the large output declines and forecasting errors made in the early 1990s following the collapse of the Soviet Union. The range of RMSE changes for Program episodes are generally similar to those in the EME and LIC groups; close to 1 percentage point through the four-year horizon, and less than half a percentage point in the five-year horizon.

How does including the errors made for year 2009 alter the findings on predictive accuracy? Annex Figure 2 shows that for the World sample, accuracy for all horizons up to four years improves for most countries in 2004–17 even if errors for 2009 are kept in the sample, as indicated by the median positive difference in RMSEs. Zooming into groups, the AE group is an exception. The median difference for countries in the AE group is negative for the one-year ahead Spring forecast and for forecasts with longer horizons, reflecting the large errors in that group for the year 2009.

In sum, we find evidence of small but notable improvements in the predictive accuracy of the WEO short-term GDP growth forecasts for a clear majority of countries in 2004–17 relative to 1990–2003, and for slightly more than half of all countries for the longer forecast horizons as

⁸ Statistical significance is determined using a bootstrap test, described in section 3.3.1 of Timmermann (2018), provided as an online annex of this paper.

well. All in all, the findings suggest improvements in the IMF country teams' ability to forecast growth over the one and a half years, with a more mixed record for the longer horizon forecasts.

C. Taking Underlying Variability into Account

RMSE-values discussed so far are estimates of "absolute" forecast accuracy that do not put forecasting performance in the context of how difficult it was to predict the outcome in the first place. We would expect it to be easier to predict GDP growth for more stable, developed economies than for emerging markets or low-income countries with less diversified economies. It is also possible that some time periods coincide with a more challenging forecasting environment than others.

One way to take underlying variability into account in assessing accuracy is to scale the mean squared error of the WEO growth forecasts by the variance of the predicted variable. The variance is traditionally computed around the sample mean. However, as the mean for the entire sample period would not have been known in real time, we compare the outcome to the recursively updated historical average (prevailing mean) using the values known as of the time of forecasting. The resulting ratio, known as the Theil U -statistic, shows the proportion of the variance of the outcome that was predicted by WEO forecasts at a given horizon, with the modification that the variance estimate in the denominator is computed using a recursively updated mean estimate:

$$U_{i,h} = \frac{\sum_{t=2004}^{2017} (y_{it} - \hat{y}_{i,t|t-h}^{WEO})^2}{\sum_{t=2004}^{2017} (y_{it} - \bar{y}_{i,t|t-h})^2} \quad (3)$$

By taking underlying growth variability into account, the U -statistic allows for a fairer comparison of predictive accuracy across countries.

Notice that the U -statistic is also a measure of how the WEO forecasts' accuracy compare with those of a "naïve" benchmark forecast—the simple recursive average of the variable. Values of the U -statistic below unity suggest that WEO forecasts are relatively more accurate than the historical average, whereas values above unity suggest that the historical average is more accurate than the WEO forecast.

Figure 4 shows values of the Theil U -statistic computed over the sample 2004–17. At the shortest two forecast horizons ($h = 0, F$) and ($h = 0, S$), these values are small—below about 0.4 and 0.6 respectively, for three quarters of economies globally. WEO forecasts thus appear to incorporate valuable information during the current year that facilitates substantially more accurate forecasting than simply using the historical average of growth outturns.

As the forecast horizon expands, however, the ability of the WEO forecasts to dominate the historical average clearly deteriorates. Like for RMSE values, much of the worsening in the Theil U -statistic generally occurs as the horizon lengthens from three months ($h=0, F$) to about two years ($h=1, S$); for about a quarter of economies the U -statistic climbs above unity as of the next-year Spring forecast ($h=1, S$). The median U -Statistic for the World sample remains only modestly below unity for the three to five year ahead forecasts, meaning that for close to

half of the economies longer term WEO forecasts are less accurate than a recursively-computed mean at these horizons. The rate of underperformance seems particularly strong among AEs where the median reaches unity by the next year Spring forecast, in LAC and SSA where it is around unity from the three-year ahead forecasts onwards, and in DASIA where it is very close to unity from the next-year Spring forecast onwards.

Accounting for growth volatility significantly changes the rankings of forecast accuracy across country groups—suggesting more uniform performance across groups. The fuel-exporter dominated EMDE_FE and MENAP_CIS groups, where absolute forecast accuracy was the lowest on the basis of RMSEs, score the strongest on relative accuracy as reflected by their lower median U -statistics. That is, the forecasts of fuel-exporting economies are more accurate than what would be predicted by their high degree of GDP growth volatility.

To get a sense of how the relative accuracy of WEO forecasts may have changed over time, Figure 5 shows differences in the Theil- U statistic for 1990–2003 and 2004–16. The charts show that relative accuracy has improved for most countries for horizons of up to two years, with noticeable gains for many countries in the LIC group. The exception is the AE group; relative accuracy has declined in most countries in the AE group (except for the current year forecasts, for which about half of the countries have seen improvements and the other half declines). Relative accuracy has also declined for the next year and longer-horizon forecasts for most economies in the DASIA group, and for the four- and five-year ahead forecasts for the SSA and Program groups. All in all, the results noted earlier on the changes in absolute accuracy over time—improvements in the shorter horizons (with the notable exception of the AE countries) and a mixed record for the longer ones—generally holds for relative accuracy.

IV. BIAS AND EFFICIENCY TESTS

A. Biases in the Forecasts

The bias of a forecast is a measure of its tendency to systematically over- or under-predict the outcome. Past evaluations have found the WEO real GDP growth forecasts to be upward biased.

For each group of economies, g , the bias over some sample $[t_0; t_1]$ equals the mean of the forecast error, i.e.,

$$bias_{g,t_0:t_1} = \frac{1}{t_1-t_0+1} \sum_{t=t_0}^{t_1} e_{g,t|t-h}. \quad (4)$$

Based on equation (1), positive (negative) values of the bias correspond to under predictions (overpredictions) of growth.

Figure 6 displays the median, weighted mean, and interquartile range of the bias over 2004–17 for different country groups. Whereas the results show no significant tendency in the World sample for upward or downward bias at the same- and next-year horizons, they do point to a tendency for overprediction at the two-year and longer horizons. For the World sample, the median forecast error is small and positive (about 0.1) for the current-year and close to zero

for the next-year forecasts, meaning that growth was on average underpredicted by a small amount for about half of the countries. However, a clear tendency for overprediction emerges as the horizon lengthens. The median forecast error declines to about -0.3 percentage point for the two-year ahead forecasts and further to about -0.5 percentage point for the three-to-five year ahead forecasts. The growth forecasts of one quarter of all countries are biased upward by more than 1 percentage point at the three- to five-year forecast horizons.

Turning to the income groups, while the median bias is similar for the AE, EME, and LIC groups for the two-year and longer horizons, the LIC group exhibits some tendency for optimism also at the shorter horizons (with a median bias at about 0.35 percentage point for the next-year spring and fall forecasts (and a quarter of LICs having biases in excess of about 0.75-1 percentage point at those two horizons, respectively). Biases are also more diverse in size and direction for EME and LIC countries as suggested by wider inter-quartile ranges of forecast biases compared with AEs. Overoptimism is more typical for smaller EM and LIC countries than for the larger ones, as seen in the “GDP-weighted mean” of the bias being consistently larger the median (indicating that a greater tendency for overpredictions for countries with lower GDP levels).

Though the median bias among EMDE Fuel Exporters is generally smaller than the one for EMDE Fuel Importers, the range of biases is wider for the former – especially at the short and medium horizons - consistent with the typically larger terms-of-trade shocks and output volatility experienced by fuel-exporting economies.

Biases differ meaningfully across the EMDE geographical regions. Growth forecasts have been mostly optimistic at all but the shortest horizons for countries in the EEUR and SSA regions, while they have often been pessimistic in DASIA. Forecasts for countries in LAC and MENAP_CIS exhibit a slight tendency for underprediction at the shorter horizons, but they turn optimistic as the horizon extends to two-years and beyond.

Finally, the median bias in the Program group is similar to the median biases for the broader EME and LIC groups at the four- to five-year horizons, but slightly more optimistic at the one- and two-year ahead horizons.

The biases discussed in the previous paragraphs should be of concern depending on how systematic they are at the country level. This issue can be addressed by examining the statistical significance of the bias. The share of countries where the bias is statistically greater or smaller than zero are reported in Table 2.⁹ The results show growth tends to be underpredicted at the shorter horizons (relatively high shares in the first 3 columns in the upper part of Table 2) and to be overpredicted at the longer ones (relatively high shares in the last three columns in the bottom part of Table 2). Growth for the current year is systematically underpredicted for close to a fifth of World economies and overpredicted for about 3-6 percent of economies, whereas growth for three- to five-years ahead are under predicted for about 6 percent of economies and over predicted for more than a quarter of them. The EEUR countries have the highest share of

⁹ We evaluate the statistical significance of the bias using a one-sided test with heteroskedasticity and autocorrelation consistent standard errors to compute p-values. Here and elsewhere in the paper, we use Newey-West heteroskedasticity and autocorrelation consistent (hac) standard errors to compute t-statistics. We set the truncation of the Bartlett window equal to $\max(h, 1)$ where h is the forecast horizon.

overpredictions for the longer horizons (about two thirds of countries have statistically significant upward biases for the three- to five-year ahead forecasts. Program countries' shares of statistically significant biases are similar to those estimated for the broader EM and LIC groups. Fuel Exporters have relatively high shares (about 10-20 percent) of systematically upward or downward biases in shorter horizons only, whereas Fuel Importers have high shares of upward biased forecasts for the longer horizons (25-30 percent of Fuel Importer EMDEs have statistically significant upward biases for three- to five-year ahead growth).

In sum, country-level WEO growth forecasts in 2004–17 were modestly downward biased at the current- and next-year horizons (except in the Fuel Exporter and to some extent the LIC and SSA groups), and displayed a comparatively stronger tendency for overprediction for the two-year and longer horizons for all groups except the EMDE Fuel Exporter group.

A. Formal Tests of Shifts in the Bias

We next compare the biases in 2004–17 to those in 1990–2003. Figure 7 depicts the interquartile range of the change in bias, calculated for each country by subtracting the mean forecast error in 2004–17 from that in 1990–2003. For the clear majority of subgroups and horizons, the median of the difference is negative, indicating a lessened tendency for overpredictions in the 2004–17 period. For the EME and LIC countries, the decline in the median bias is relatively uniform across the forecast horizons, broadly in the range of 0.4-1.0 percentage point, suggesting that forecasts became less optimistic (i.e., biased upward) for more than half of countries, at all horizons. By contrast, within the AE group, many countries saw a reduced tendency for over prediction in the one- to two-years ahead forecasts, little change for same year forecasts (where biases were small to begin with), but more upward bias in forecasts at the three-year and longer horizons for a slight majority of countries.

All in all, the findings are consistent with the earlier finding that forecasts have become more accurate over time for most countries at the shorter horizons, and that, at least for some economies, the improved accuracy reflects reduced bias. That said, for a few groups, namely for AE, MENAP_CIS, and SSA, the median error in forecasting growth four-to-five years out is largely unchanged or larger than in the 1990–2003 period. This means that in these groups, biases have increased for slightly more than half of countries for the four- and five-year ahead forecasts. For the AE countries, this likely reflects the unforeseen persistent weakness of growth relative to the forecasts made before the Global Financial Crisis, in part because of the Euro Area sovereign debt crisis. Biases in growth projections under IMF programs have generally declined: optimism in Program forecasts declined for 75 percent of cases for the same year, next-year, and two-year-ahead horizons, and for more than half of cases for three to five-year forecasts.

The decline in the bias is naturally smaller if we include in the sample forecast errors for 2009—a year when growth fell dramatically short of previous forecasts as a result of the Global Financial Crisis (Annex Figure 3). The greatest impact is for Advanced Economies, which were the epicenter of the crisis. Once the 2009 errors are included, the median decline in overprediction in this group becomes very small and confined to the current- and next-year forecasts. Similarly, including errors for the year 2009 reduces the decline in the bias for EMDE Fuel Exporters. For other groups, the change in the bias remains broadly similar to that

obtained when excluding the 2009 errors from the sample, suggesting that the main conclusions of improved accuracy and reduced bias are not driven by dropping outturns for 2009 from the sample.

To test how systematic the changes in biases are, we carry out a bootstrap permutation test of the null hypothesis of equal absolute biases in the two subsamples:¹⁰

$$H_0 : E[|bias_{1990-2003}|] = E[|bias_{2004-2016}|]. \quad (5)$$

The results are summarized in aggregated (weighted) form in Annex Figure 4. The sum of the negative and positive bars corresponds to the change in the weighted mean bias in Figure 7; the darker colored segments correspond to the weighted average of the statistically significant changes, the lighter colored ones to the statistically insignificant ones. For all groups, we see (from the dominant size of the dark red bars) that for most horizons many countries have significantly reduced biases. For the World sample, the weighted average of significant declines in overprediction bias outweigh the statistically significant increases for all of the horizons except the current year ones (where biases are generally smaller to begin with). Some clear exceptions are in the Advanced Economy (at two-to-five year horizons, consistent with the unexpectedly weak growth in the aftermath of 2009 and during the euro area sovereign debt crisis), and DASIA groups for the zero-to two year horizons. All in all, the findings confirm the statistically significant declines in overprediction bias in many economies, especially the larger ones.

B. Serial Correlation in Forecast Errors

Under squared error loss, forecast errors should be serially uncorrelated whenever forecast horizons are non-overlapping. For example, the error in predicting year t growth in the Fall WEO of year t should be uncorrelated with the error made in predicting year $t-1$ growth in the Fall of year $t-1$. This is because the year $t-1$ error should be observed in the Fall of year t and thus be fully taken into account in preparing the year t forecast.

In general, we can compute an estimate of the serial correlation in the forecast error of group g from a linear regression with no intercept:

$$e_{g,\tau|\tau-h} = \rho_g e_{g,\tau-1|\tau-1-h} + \varepsilon_{g,\tau}. \quad (6)$$

Computed over the sample period $t_0 : t_1$ this yields a coefficient estimate:

$$\hat{\rho}_g = \frac{\sum_{2004}^{2017} e_{g,\tau|\tau-h} e_{g,\tau-1|\tau-1-h}}{\sum_{2004}^{2017} e_{g,\tau-1|\tau-1-h}^2}. \quad (7)$$

¹⁰ A statistical analysis is tempered by the short data sample, which could invalidate inference relying on asymptotic distributions and well-known finite-sample distortions of standard test statistics that correct for heteroskedasticity and autocorrelation in the regression residuals are well known (see Timmermann 2007 for references). To at least partially address the problem that standard critical levels may not provide a reliable guide to inference based on simple t - and F -statistics in small samples, we rely on bootstrap methods in our analysis.

We can then test the null $H_0 : \rho_g = 0$ against a two-sided alternative, $\rho_g \neq 0$. Positive values of $\hat{\rho}_g$ indicate that forecast errors of the same sign are more likely to follow each other. This suggests a tendency for over - or under-predictions to persist through time. Conversely, negative values of $\hat{\rho}_g$ imply that the forecast errors tend to reverse in consecutive years.

Figure 8 shows the share of countries for which the estimated serial correlation coefficient is statistically significant, with either a positive or negative sign. The test is carried out for the current year Fall and next-year Fall forecasts (forecast horizons $(h = 0, F)$ and $(h = 1, F)$, so that overlapping data is not an issue and the first error is almost certainly observed by the time the second forecast is made. The results reveal the presence of serial correlation in the forecast errors associated with the Fall current-year and Fall next-year WEO vintages in 15–30 percent of countries except in the AE group. These rejection rates are significantly higher than the 5 percent we would expect under the null of no serial correlation, especially for the LIC group. This exercise reveals that forecast errors—and biases—could be reduced if forecasts react more strongly to the recently observed errors.

C. Local Serial Correlation

The previous section looks at evidence of autocorrelation in the individual countries' forecast errors over the full sample. However, it is possible that forecast errors became serially correlated only during certain periods such as the Global Financial Crisis. Such events may have induced a sequence of over or underpredictions if the underlying forecasting methods did not adapt to the shock sufficiently fast. Analyzing evidence of persistence that is more "local" in time is difficult for individual countries for which we only have annual outcomes. However, we can pool estimates of serial correlation across countries in a particular group of economies so as to get more robust local estimates of "average" serial correlation within these economies. To this end, we consider the following "local" estimator for the N^g countries in group g :

$$\hat{\rho}_{g,t} = \frac{\frac{1}{4} \sum_{i=1}^{N^g} (e_{i,t-2}e_{i,t-1} + e_{i,t-1}e_{i,t} + e_{i,t}e_{i,t+1} + e_{i,t+1}e_{i,t+2})}{\frac{1}{5} \sum_{i=1}^{N^g} (e_{i,t-2}^2 + e_{i,t-1}^2 + e_{i,t}^2 + e_{i,t+1}^2 + e_{i,t+2}^2)}. \quad (8)$$

By using the covariance between the forecast error at time t , $e_{i,t}$ and the past ($e_{i,t-1}$) and future ($e_{i,t+1}$) forecast errors, along with two adjacent cross products of forecast errors, this estimator captures serial correlation that is "local in time". On its own, this estimator would be very noisy, using only five adjacent observations, but averaging over all countries within a group tends to smooth the resulting estimates. This covariance is scaled by the average of the neighboring squared forecast errors so as to get a correlation-type measure that is easier to interpret.

Figure 9 plots the resulting local serial correlation estimates at the current- and next-year forecast horizons for the world economy so that the effect of cross-sectional averaging is strong (given that the number of countries is large). Local serial correlation of errors in forecasting

world GDP growth temporarily plunges from positive levels (which was strongest around 2005) to negative values in the aftermath of the Global Financial crisis. At all four forecast horizons, the local serial correlation estimate increases to positive levels again towards the end of the sample. This pattern indicates that overpredictions of current- and one-year-ahead GDP growth during the Global Financial Crisis were followed by underpredictions during the recovery, pointing to a possible tendency to “compensate” with pessimism following large overprediction errors.

V. SOURCES OF FORECAST ERRORS

This section carries out several exercises to better understand the sources of WEO forecast errors and how they can be improved. The WEO forecast preparation process puts considerable emphasis on integrating predictions across countries, regions, and variables so as to produce coherent and globally consistent projections of economic activity. Our analysis thus stresses factors that are important elements of the global economic environment.

In the first exercise, we look at the extent to which WEO growth forecast errors can be traced to errors in predicting external factors—forecast errors for systemically-important economies and the terms of trade. To the extent that assumptions about certain key drivers such as future commodity prices are "hard wired" in the forecasting process, good or bad assumptions or projections for these variables might help explain forecasting performance.

In a second exercise, we analyze whether some of the procedures currently in place to ensure global consistency have their intended effect. In particular, we test for the informational efficiency of forecasts, using a range of indicators of global economic activity. Such tests build on the condition that indicators observed at the time of forecasting (such as the country’s own terms of trade forecast or large-country growth forecasts) should not be able to predict the errors. Finally, in the same spirit as the second exercise, we examine whether forecast errors can be predicted by the output gap estimate made in the same round of forecasting. The previous evaluation (Timmermann 2007) documented a tendency for growth to be over predicted in years when output was estimated to be below potential.

A. Contemporaneous Errors in Forecasts of External Factors

International linkages in financial, goods, and labor markets mean that GDP growth in major countries or economic areas influence GDP growth in other countries. To assess the extent to which forecast errors for major economies are aligned with the growth forecast errors for individual economies, we regress the h -step-ahead forecast error in economy g , $e_{g,t,h}$, on the same-period US GDP forecast error, $e_{US,t|t-h}$, or the China and Euro Area forecast errors :

$$e_{g,t|t-h} = \alpha_g^h + \beta_g^h e_{US,t|t-h} + \varepsilon_{g,t,h}. \quad (9)$$

$$e_{g,t|t-h} = \alpha_g^h + \beta_g^h e_{China,t|t-h} + \varepsilon_{g,t,h}. \quad (10)$$

$$e_{g,t|t-h} = \alpha_g^h + \beta_g^h e_{EU,t|t-h} + \varepsilon_{g,t,h}. \quad (11)$$

Note that this is not a predictive regression as we are using the contemporaneous forecast error for the major economies. Hence, the results from this regression can only be used to gauge whether larger forecast errors for e.g. US GDP growth are associated with larger forecast errors for GDP growth in economy g . We run these regressions using forecast error data for four forecast horizons— $(h = 0, F)$, $(h = 1, F)$, $(h = 2, F)$, and $(h = 5, F)$, respectively.

Results from these regressions are reported in Table 3. The first 12 rows of Table 3 contain calculations based on estimations of equation (9) – using US growth forecast errors. The first four columns report the median of estimated βg for US forecast errors and the four selected forecast horizons. The median coefficient on the US forecast error varies significantly across country groups, but is mostly positive, indicating that overpredictions (underpredictions) of US growth spills into over predictions (underpredictions) of the growth rates of other countries. For instance, the median pass through of the error made in predicting US growth in same year growth forecasts to other countries same year Fall forecasts is about 13 percent. Columns 5 through 12 show that the share of countries where the impact of the US forecast error on other countries' errors is positive and statistically significant is generally small (5 to 16 percent of countries—as seen line 1 for columns 5-8) but these shares are typically larger than those of countries for which the coefficient is negative and statistically significant (2-8 percent—line 1 of columns 9-12).¹¹ Among countries for which the coefficient is statistically significant, the US forecast errors can explain about a third of the variation in country forecast errors (columns 13-21).¹²

Unsurprisingly, errors in predicting Chinese GDP growth matter significantly to the forecast errors of many economies. The results of regressing country growth forecast errors on China's growth forecast errors (equation 10) are shown in the middle block of Table 3. The median error pass-through coefficient is positive for all country groups, and 27-46 percent globally across the four forecast horizons (first line of columns 1-4). For a sizable fraction of economies (about 30-40 percent globally) the coefficients are positive and statistically significant (columns 5-8). There is virtually no region for which the share of countries with positive significant coefficients is less than 10 percent. The spillovers from China's growth forecast errors seem particularly strong for the EMDE Fuel Exporter, MENAP_CIS, and LAC groups. This is not surprising since those groups include many commodity-dependent economies and China accounts for an important share of global commodity demand. By contrast to elevated shares of countries with positive spillovers, the share of countries where the coefficients are negative are generally very small (columns 9-12). The R-squared of the regressions where the estimated coefficient on China's growth forecast error is positive are sizable, generally between 40 and 50 percent. Interestingly, and in sharp contrast with the US forecast errors, Chinese forecast errors are statistically significant and positive for a particularly large fraction of world economies (close to 40 percent) at the five-year horizon, consistent with Chinese growth being an important driver of the growth of many countries.

¹¹ Statistical significance is determined by looking at the statistical significance of the βg -estimate using a one-sided test conducted at the 10 percent level.

¹² The same regressions for the 1990–2016 period, as compared with the 2004–17 period we focus on here, yield a somewhat greater share of significant correlations (Timmermann, 2018), reflecting the larger relative size of the US economy in the 1990s and its greater importance in driving unanticipated movements in the growth rates of other economies in that period.

Euro Area growth forecast errors also have an impact on other economies, especially for the current- and next-year horizons. The median pass through to other economies is sizable—50 and 37 percent globally for the same and next-year Fall horizons (last block of Table 3). But the share of countries where the impact is positive and statistically significant is less than those for China and closer to those for the United States—10-25 percent globally depending on the forecast horizon (line 1 of columns 5-8 in Table 3). Growth errors for the Euro Area are particularly important for the growth forecast errors and Eastern European countries (with pass through rates around 100 percent) and Advanced Economies. The explanatory power of Euro Area growth forecast errors for the growth forecast errors of other economies is generally sizable (columns 9-15).

All in all, the evidence in Table 3 suggests that more accurate forecasts for the Euro Area, United States, and especially China would help improve the accuracy of the global growth forecast not only directly, given the large weight of these economies in the global economy, but also indirectly, since errors for these economies spill over into the errors of others as they proxy for shocks to growth factors that are common to the global economy.

Our final check is for how WEO forecast errors of individual countries' terms of trade affect growth forecast errors. To this end, the terms of trade for country i in year t is denoted by tot_{it} while the h -year-ahead forecast of $tot_{i,t}$ is denoted by $\widehat{tot}_{it|t-h}$. We define terms of trade forecast errors as the actual value in the final vintage of our database (Fall 2017) minus the forecast made h -periods previously for that year, i.e.:

$$e_{it|t-h}^{tot} = tot_{it} - \widehat{tot}_{it|t-h}, \quad (12)$$

To explore the linkage between forecast errors for GDP growth in year t for country i , denoted $e_{i,t|t-h}$, and errors in forecasting the terms of trade for that country, we use the following regression specification:

$$e_{1,t|t-h} = \alpha_{i,h} + \beta_{i,h} e_{i,t|t-h}^{tot} + \varepsilon_{i,t,h}, \quad (13)$$

using the four forecast horizons as for the US, China, and Euro Area forecasts. Data on tot_{it} is from the same WEO database vintage as the growth forecasts (with 2000 being the base year for TOT indices). We would expect to find positive estimates of $\beta_{i,h}$ since underpredictions of the terms of trade (a positive value of $e_{it|t-h}^{tot}$) plausibly translate into underpredictions of GDP growth (a positive value of $e_{it|t-h}$) and vice versa.

The median impact of the terms of trade forecast error on growth forecast errors is typically positive (Table 4, columns 1-4). The shares of countries with positive and significant correlations between TOT and growth forecast errors (columns 5-8) are typically well larger than those with negative correlations (columns 9-12). The EMDE Fuel Exporter, MENAP_CIS, and LAC groups—all of which include a high share of resource-intensive economies where terms of trade shocks tend to matter the most—have high shares of countries where the correlations are positive and significant. The median correlations between TOT and

growth forecast errors are generally small (always less than 0.05), typically smaller than the coefficients on the major economy growth forecast errors. That said, the magnitude of TOT forecast errors is larger than those of growth errors for the US, China, and the Euro Area, so the overall impacts of growth and TOT forecast errors on the growth forecast errors are comparable. Likewise, TOT forecast errors can explain a meaningful fraction (30-50 percent) of growth forecast errors.

In general, we find the TOT errors to be a significant predictor of the growth forecasts of up to 56 percent of economies in the EMDE Fuel Exporter group, and typically 5-15 percent of countries for other EMDE. TOT forecast errors seem to be significant for a larger fraction of countries for the longer forecast horizons such as two and five years.

All in all, the analysis confirms the importance of striving for accurate forecasts of the global economic environment, since errors in predicting external variables noticeably affect the growth forecast errors of a meaningful share of individual economies.

B. Are WEO Growth Forecast Errors Predictable?

Forecast errors reflect the surprise component in the outcome. To the extent that large common supply or demand factors affect broad sets of economies, we would expect forecast errors for major economies to be significantly correlated with forecast errors in other economies. This is indeed what we examined in the previous section.

Conversely, we should not expect the forecasts themselves in a given forecast round to possess predictive power over future forecast errors. At each point in time, the forecasts should simply reflect the information that is available in that period and this should already be efficiently incorporated into the individual country forecasts of GDP growth (e.g. the projected GDP growth rates of large economies or the projected terms of trade for the country itself). To see if this condition holds, we estimate a similar set of regressions as in (9)–(11), now using the GDP growth forecasts for the US, China, and euro area (rather than their forecast errors) as predictors:

$$e_{g,t|t-h} = \alpha_g^h + \beta_g^h \hat{y}_{t|t-h}^{US} + \varepsilon_{g,t,h}. \quad (14)$$

$$e_{g,t|t-h} = \alpha_g^h + \beta_g^h \hat{y}_{t|t-h}^{CHINA} + \varepsilon_{g,t,h}. \quad (15)$$

$$e_{g,t|t-h} = \alpha_g^h + \beta_g^h \hat{y}_{t|t-h}^{EU} + \varepsilon_{g,t,h}. \quad (16)$$

A positive coefficient in equation (14) for country g means that the US forecast is more correlated with the actual growth of country g than it is with its projected growth. Put differently, the impact of near-term economic strength or weakness in the US economy is not sufficiently taken into account in preparing the contemporaneous forecasts of these economies. A negative coefficient means the opposite—the US forecast is more highly correlated the growth forecast of country g than its actual growth, suggesting that the forecast “overreacts” to the projected US growth rate.

The regression results do suggest some degree of inefficiency in the way US, China, and Euro Area growth forecasts are considered in the forecasting process. The first 12 rows and columns 5-8 in Table 5 suggest that for 10-25 percent of world economies a higher US growth forecast predicts a higher growth forecast error. By contrast, the share of countries where the coefficient is negative and statistically significant—meaning that higher projected US growth is followed by an overprediction of growth—is very low (columns 9-12).

In the case of China's growth forecasts, it's more often the case that forecasts are overly sensitive to the China growth forecast than they are insufficiently sensitive. This can be seen by the higher shares of significant and negative coefficients shown in columns 9-12 in the second set of rows of Table 5 than in columns 5-8. For almost one third of economies globally, a higher five-year ahead projected growth rate for China has typically coincided with growth overpredictions (column 12).

Forecasts of Euro Area growth can predict the growth forecast errors of other economies as well (bottom set of rows in Table 5). Like in the case of the US, the coefficients are much more often positive than negative (comparing columns 5-8 with columns 9-12), with an especially elevated share of forecasts under-reacting to Euro Area growth forecasts for two- and five-year ahead horizons. For instance, growth forecast errors of about 28 and 33 percent of all economies are correlated with the Euro Area growth forecast at the two- and five-year ahead horizons.

For examining the efficiency of the use of terms-of-trade forecasts, the following regression specifications are used:

$$e_{i,t|t-h} = \alpha_{i,h} + \beta_{i,h} \text{tot}_{it|t-h} + \varepsilon_{i,t,h}, \quad (17)$$

Terms of trade forecasts are also significant predictors of individual countries' growth forecast errors for many groups and horizons (Table 6). A positive correlation between growth forecast errors and the terms of trade forecast means that growth turns out to be systematically underpredicted when the terms of trade is projected to be strong, and overpredicted when the terms of trade is projected to be weak. While we find a nonnegligible share of countries where the growth forecasts fail to respond strongly enough to the terms of trade outlook (columns 5-8 in Table 6) there are many cases where the forecast overreacts to the terms of trade forecast as well. All in all, this evidence suggests that accounting for the terms-of-trade outlook more carefully can help improve the forecasts of 10-30 percent of world economies (summing up the shares reported in the first row of columns 8-11 and column 12-15 of Table 6, for a given forecast horizon).

C. Output Gap

As a final step, we test whether estimated output gaps are systematically correlated with growth forecast errors. We use the following specification, using the output gap in group or economy g for period t predicted at time $t-h$, $GAP_{g,t|t-h}$:

$$e_{g,t|t-h} = \alpha_g^h + \beta_g^h GAP_{g,t|t-h} + \varepsilon_{g,t}. \quad (18)$$

We run this regression only for Advanced Economies since output gap data is sparse for other groups in the 2004-17 period. We find a strong negative correlation of -0.6 (significant at the one percent level) for the AE group as a whole (i.e., regressing the weighted mean growth forecast for AEs on the weighted mean output gap) for the Spring next year forecasts ($h=1, S$). The correlations are also negative, but not significant, for the same-year forecasts ($h=0, F, S$) and the next-year Fall forecasts ($h=1, F$). Running the regression for each country individually, we find that about one third of the countries in the AE group have a significant negative correlation between the output gap and the forecast error. This means that for a third of countries growth is systematically overpredicted when the economy is estimated to have spare capacity, possibly reflecting an assumption that the output gap would close (with actual growth exceeding the country's potential growth rate) over the WEO forecasting horizon.

VI. COMPARISON WITH CONSENSUS ECONOMICS FORECASTS

No economic forecast can be expected to be perfect and it is useful to have benchmarks for how accurate we should reasonably expect the WEO forecasts to be. This section compares the predictive accuracy of WEO growth forecasts to growth forecasts produced by Consensus Economics (CE), an organization that surveys private-sector forecasters.

Consensus Economics (CE) generates monthly updates to their next-year and current-year forecasts and so produces a sequence of 24 forecasts of the same outcome. For example, GDP growth in 2016 would be predicted from January 2015 through December 2016. We label the monthly vintages of the current-year forecasts $\{h=0, m\}_{m=1}^{12}$ while the next-year forecasts are labeled $\{h=1, m\}_{m=1}^{12}$, where $m=1$ is the January forecast, while $m=12$ is the December forecast.

To provide a meaningful comparison, the timing of the two sets of forecasts in the comparison should be as close as possible. Otherwise, one forecast may appear to be better than the other simply because it uses more up-to-date information. We pair the current-year March ($h=0, m=3$) and September ($h=0, m=9$) CE forecasts with the WEO current-year Spring ($h=0, S$) and Fall ($h=0, F$) forecasts, respectively. Similarly, we pair next-year March ($h=1, m=3$) and September ($h=1, m=9$) CE forecasts with the WEO forecasts for $h=1, S$ and $h=1, F$, respectively.

Because CE reports forecasts for individual countries, our analysis compares country-level forecasts of GDP growth rather than analyzing forecasts at the more aggregate/group level.

A. Relative Forecasting Performance

The first performance measure we use to compare the accuracy of the WEO and CE forecasts is the ratio of their RMSE values for a given forecast horizon, denoted by $RMSE(WEO_h)$ and

$RMSE(CE_h)$, respectively. Specifically, for each of the four forecast horizons $h = 0, S, h = 0, F, h = 1, S$, and $h=1, F$, we compute the ratio:

$$U_h = \frac{RMSE(WEO_h)}{RMSE(CE_h)} - 1. \quad (19)$$

Values $U_h > 0$ indicate that the RMSE of the WEO forecasts exceeds the RMSE of the CE forecasts and that the CE forecasts, on average, were more accurate during the sample. Cases for which $U_h < 0$ indicate the opposite. Moreover, the amount by which U_h differs from zero quantifies the relative performance of one forecast versus the other.

To investigate statistical significance of the differences in predictive accuracy, we also conduct formal tests of the null that the MSE values of the WEO and CE forecasts are identical in expectation. Our null of equal predictive accuracy takes the form:

$$H_0 : E \left[\left(e_{i,t|t-h}^{WEO} \right)^2 \right] = E \left[\left(e_{i,t|t-h}^{CE} \right)^2 \right], \quad (20)$$

where $e_{i,t|t-h}^{WEO}$ is the h -step-ahead forecast error from the WEO forecast of GDP growth in country i at time t , while $e_{i,t|t-h}^{CE}$ is the corresponding forecast error associated with the CE forecast. Defining the forecast error loss differential $dif_{i,t|t-h} = \left(e_{i,t|t-h}^{CE} \right)^2 - \left(e_{i,t|t-h}^{WEO} \right)^2$,

we follow Diebold and Mariano (1995) and test this null hypothesis by regressing the loss differential on an intercept

$$dif_{i,t|t-h} = \alpha_{ih} + \varepsilon_{i,t|t-h}. \quad (21)$$

Positive and significant estimates of α_{ih} (using a t -test) suggest that the h -step-ahead forecast produced by CE generated significantly higher mean squared error values than the WEO forecasts. Conversely, significantly negative estimates of α_{ih} suggest that the WEO forecasts were significantly less accurate than the CE forecasts.

The first four columns of Table 7 show results for the RMSE ratios (given in equation 19) computed for the individual countries covered by both CE and the WEO, comparing the Spring and Fall WEO forecasts with the September and March Consensus Economics forecasts. The top three rows present the 25th, 50th, and 75th percentiles of the cross-country distribution of RMSE ratios, calculated across these countries. For a majority of countries, the WEO forecasts generate lower RMSE values than the CE forecasts for the Fall and current year Spring forecasts, as evidenced by the negative RMSE ratios for the median first row). The proportion of countries for which the RMSE value of the WEO forecasts is lower than that of the CE ranges between 56 percent and 66 percent for these horizons. Moreover, the 25th percentiles are generally more negative than the 75th percentiles are positive, suggesting that the

advantage in the WEO forecasts' precision over the CE forecasts are slightly larger in magnitude than the shortfall where the CE forecasts are more accurate than the WEO forecasts. The Diebold-Mariano test statistics for the significance of differences are significant only in a few cases—close to the size (5 percent) of the test. That said, for Spring next-year forecasts, WEO forecasts are more accurate than Consensus Economics forecasts only about 40 percent of the time.

When comparing the Spring and Fall WEO forecasts with the April and October Consensus Economics forecasts (columns 5-8 of Table 7), we find that the latter are more accurate in the majority of cases, showing that incorporating further data for even a few additional weeks can make a marked difference for accuracy.

VII. CONCLUSION

The main takeaways from the analysis are as follows:

Outliers

- WEO growth forecasts tend to be biased upward in part because unpredicted growth booms are much rarer than severe growth collapses (due to natural disasters, wars, and systemic financial crises such as the Global Financial Crisis of 2008-09). Removing unforecastable growth collapses from the data sample helps to get a better sense of forecasting performance under normal circumstances.

Accuracy

- In 2004-17, forecast accuracy was highest for advanced economies and lower for, and comparable between, emerging-market and low-income economies. Among the latter two groups, forecasts were more accurate for fuel importers than for fuel-exporters.
- Predictive accuracy drops as the forecasting horizon lengthens, but not linearly. Reductions in predictive accuracy are largest moving from the three-month horizon (Fall forecasts for the current year) to the two-year horizon (Spring forecasts for the year ahead). Reductions in predictive accuracy tend to be quite small as the horizon lengthens beyond two years and the accuracy of two-year and five-year-ahead forecasts are not very different.
- Short-term forecasts were more accurate in 2004-17 than in 1990-2003 for well more than half of the countries. At the longer horizons, accuracy has increased for most fuel-importing emerging market and developing economies, while only about half of countries saw improved accuracy among advanced economies and fuel-exporter emerging market and low-income economies (the other half saw modest declines in accuracy).
- Growth forecast errors tend to be broadly proportional to the volatility of growth—a measure of the difficulty of forecasting. Once the volatility of growth is accounted for, accuracy was similar across all country groups in 2004-17.

- WEO forecasts are generally more accurate than a naïve forecast based on a recursively-computed historical average. In more than half of countries globally, WEO forecasts for all forecast horizons, especially those up to 2-3 years, had lower root mean squared errors than a forecast based on a historical average. However, there are some areas of concern, especially in the case of longer-term forecasts. For about half of the countries among advanced economies and in Latin America and the Caribbean, Sub-Saharan Africa, and Developing Asia, naïve forecasts were more accurate than WEO forecasts for horizons longer than 2-3 years.

Bias

- Whereas there is no discernible upward or downward bias in growth forecasts for the current-year and next-year (except for some optimism in the low-income country group), growth forecasts for the two-year and longer horizons tend to be upward biased. The median overprediction of growth across all countries is about 0.3 percentage point for the two-year horizon and about 0.5 percentage point for the three- to five-year horizons. The median bias in the Program group is similar to the median biases for the broader emerging market and low income country groups at the four- to five-year horizons, but slightly higher (more optimistic) at the one- and two-year ahead horizons.
- The tendency for optimism (overprediction) has generally declined: bias was lower in 2004-2017 than in 1990-2003 for at least half of the countries at all horizons, with declining biases being more pronounced among emerging market and developing economies.

Sources and Predictability of forecast errors

- Forecast errors for the United States, the Euro Area, and especially China are correlated with the forecast errors of many other economies. Errors in projecting the terms of trade also contributes to growth forecast errors, especially for commodity exporters.
- For about 8 percent of advanced economies and 20-30 percent of emerging market economies, current-year growth overpredictions or underpredictions tend to persist through time. This means that for about one in ten countries globally near-term forecasts can be improved by learning from recent errors.
- Forecasts can be improved for up to a quarter of economies by more efficiently incorporating information on the growth forecasts of the United States. For up to 30 of countries—especially commodity exporters—GDP forecasts can also be improved by more efficiently taking into account the country’s own terms of trade forecast.
- Growth tends to be systematically overpredicted when output is estimated to be below potential, but less so than in the last evaluation (Timmermann 2007).

Comparison with Consensus Economics Forecasts

- The accuracy of the Spring and Fall current-year and next-year WEO forecasts is broadly comparable to the accuracy of the corresponding Consensus Economics

forecasts reported for March and September. The timing of WEO forecasts vary from year to year, sometimes preceding and sometimes following Consensus Forecasts; any differences in accuracy appear to reflect timing and informational advantage rather than systematic differences in forecasting performance.

These findings suggest a number of ways in which the performance of the WEO forecasts can be improved.

- Efforts to reduce optimism bias should focus on the long-term, multi-year forecasts and in some cases even the one-year-ahead forecast reported in the Spring WEO. Two-to five years ahead growth forecasts that are significantly higher than the average of recent growth rates of a country should be closely scrutinized.
- Changes to the way information on the output gap is incorporated into the forecasts could lead to improvements in predictive accuracy, including by not imposing an assumption that the gap will be closed (through above-potential growth) over the forecast horizon.
- Given the often-significant relation between errors in forecasting GDP growth on the one hand and errors in predicting individual countries' terms of trade, strategies for improving the accuracy of forecasts of these terms of trade measures could improve on forecasts of output growth. The same holds for the growth forecasts of China and to some extent the Euro Area and the United States. Moreover, the accuracy of some countries' forecasts can be improved by better incorporating information on the US growth forecast. An assessment of the uncertainty surrounding terms of trade and major economy growth forecasts would be helpful in evaluating the uncertainty of GDP growth forecasts.
- While the accuracy of the WEO and Consensus Economics forecasts of GDP growth are broadly similar, there are some major economies for which the Consensus Economics forecasts appear to dominate the WEO forecasts. Inspecting the reasons for the somewhat weaker performance of the WEO GDP growth forecasts for these countries might be useful.
- Forecast errors—and biases—could be reduced if forecasts react more strongly to the recently observed errors (given the finding of serially correlated errors in many countries). At the same time, after large downturns that imply large past overpredictions of growth, care should be taken not to systematically underpredict growth to compensate for past optimism.

VIII. REFERENCES

- [1] Artis, Michael J., 1988, How Accurate Is the World Economic Outlook? A Post Mortem on Short-Term Forecasting at the International Monetary Fund. Staff Studies for the World Economic Outlook (Washington, International Monetary Fund), 1-49.
- [2] Artis, Michael J., 1997, How Accurate Are the WEO's Short-Term Forecasts? An Examination of the World Economic Outlook. Staff Studies for the World Economic Outlook (Washington, International Monetary Fund).
- [3] Barrionuevo, J.M., 1993, How Accurate Are the World Economic Outlook Projections? Staff Studies for the World Economic Outlook (Washington, International Monetary Fund).
- [4] Chong, Y. Y., and D. F. Hendry. 1986. Econometric evaluation of linear macro-economic models. *Review of Economic Studies* 53:671-90
- [5] Diebold, F. X., and R. S. Mariano. 1995. Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13:253-63.
- [6] Elliott, Graham, and Allan Timmermann, 2016. *Economic Forecasting*. Princeton University Press.
- [7] Patton, Andrew J., and Allan Timmermann, 2010, Monotonicity in Asset Returns: New Tests with Applications to the Term Structure, the CAPM, and Portfolio Sorts. *Journal of Financial Economics* 98, 605-625.
- [8] Rossi, Barbara, 2013, Advances in Forecasting under Instability. Chapter 21, pages 1203-1324 in G. Elliott and A. Timmermann (eds.) *Handbook of Economic Forecasting* vol 2B. North-Holland.
- [9] Timmermann, Allan, 2007, An Evaluation of the World Economic Outlook Forecasts. *IMF Staff Papers* vol 54, 1, 1-33.
- [10] Timmermann, Allan, 2018, An Analysis of the Forecasting Performance of the World Economic Outlook, manuscript (provided as an online annex of this paper).
- [11] Timmermann, Allan, and Yinchu Zhu, 2017, Monitoring Forecasting Performance. Unpublished manuscript, UCSD and University of Oregon.

ANNEX 1. Limiting Outliers in the Forecast Evaluation Exercise

When assessing forecasting performance, data points heavily affected by factors outside the forecasting process (outliers) are removed to sharpen the analysis. The filtering of outliers is described in steps below. Table 1 in the main text shows some descriptive statistics before and after the filters are applied.

1. Starting Point

GDP forecasts from past WEO vintages are used to construct a database with 12 series corresponding to 6 horizons and two bi-annual forecasting exercises (Spring and Fall) for the period 1990-2016.

2. Filter one: Conflict

Using the UCDP Georeferenced Event Dataset, a conflict indicator is set to one in years when 100 or more people are killed due to conflict per 1 million population. We then drop years when conflict=1 and the adjacent years for certain countries where it is evident that economic performance was severely affected by the conflict. These additional country (year) pairs to be excluded are Angola (1998), Georgia (1994), Iraq (1992-1993), Libya (2012-2013), Moldova (1993-1994), Ukraine (2015), and Tajikistan (1994). In the case of Sudan, errors for 2011 are excluded given the separation between Sudan and South Sudan.

3. Filter two: Disasters

Using EM-DAT the International Disaster Database, we drop country-year observations where total disaster damage exceeds 3 percent of GDP and growth is negative. In addition, we drop country-year observations affected by the Ebola epidemic (2014 and 2015 for Guinea, Sierra Leone, and Liberia).

4. Filter three: Ad-Hoc

Errors for Ireland in 2015 are dropped due to a jump in GDP that year from the relocation of intellectual property by large multinational companies. Also, we dropped Equatorial Guinea due to large volatility surrounding oil discovery and extraction. We drop year 2009 from the subsample of 2004-2016 because of the Global Financial Crisis.

Note that the previous filters drop all forecasts made for a given country-year based on the outturn being unforecastable. The next two filters apply to different horizons separately.

5. Filter four: Zero forecast errors

We drop forecast horizons with precisely zero forecast errors as it is highly unlikely that forecasts match outturns exactly. Note that since we use growth estimates for year t recorded

in the Fall WEO of year $t+1$ as outturns, it is possible those outturns are carried-over forecasts rather than actual statistics, resulting in zero forecast errors.

6. Filter five: Minimum number of observations in subsamples

With the analysis divided into two subsamples (years 1990-2003 and 2004-2016), we drop forecast horizons with observations less than eight in a given period. As expected, this filter affects the 1990-2003 subsample more than the 2004-2016 subsample.

Figure 1. Forecast accuracy: median, weighted mean, and interquartile range of root mean squared errors of WEO real GDP growth forecasts

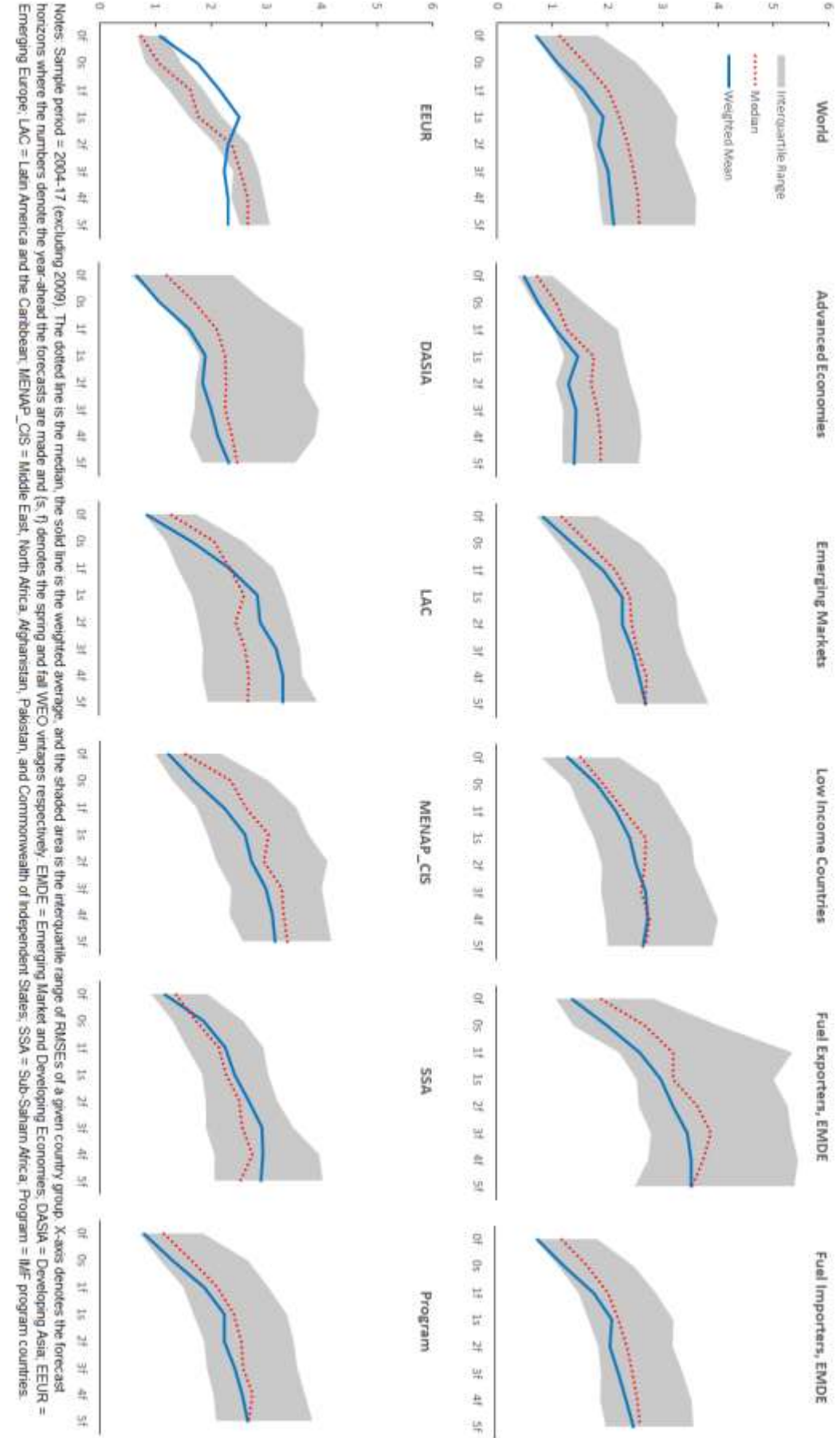
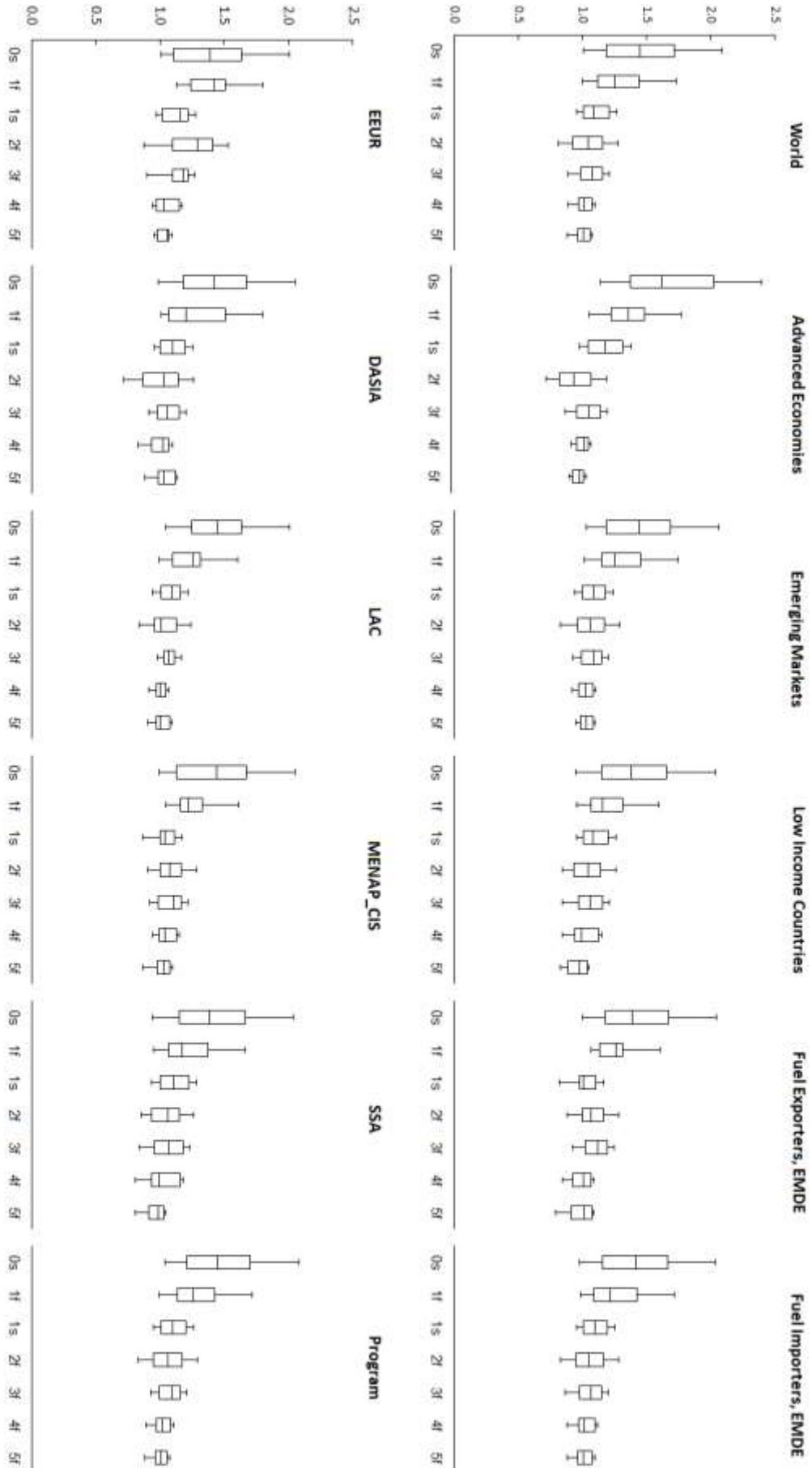
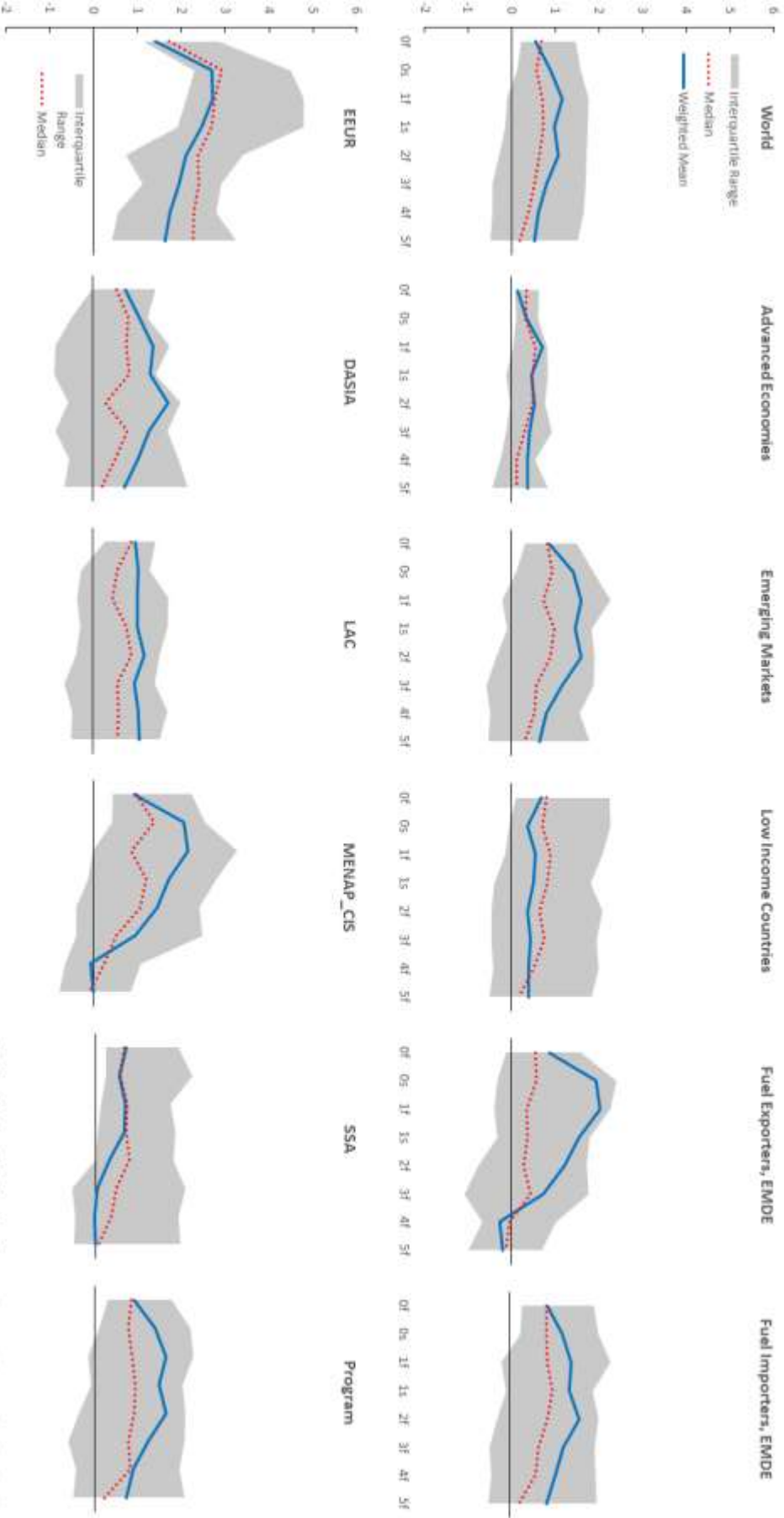


Figure 2 . Root Mean Square Error Ratios



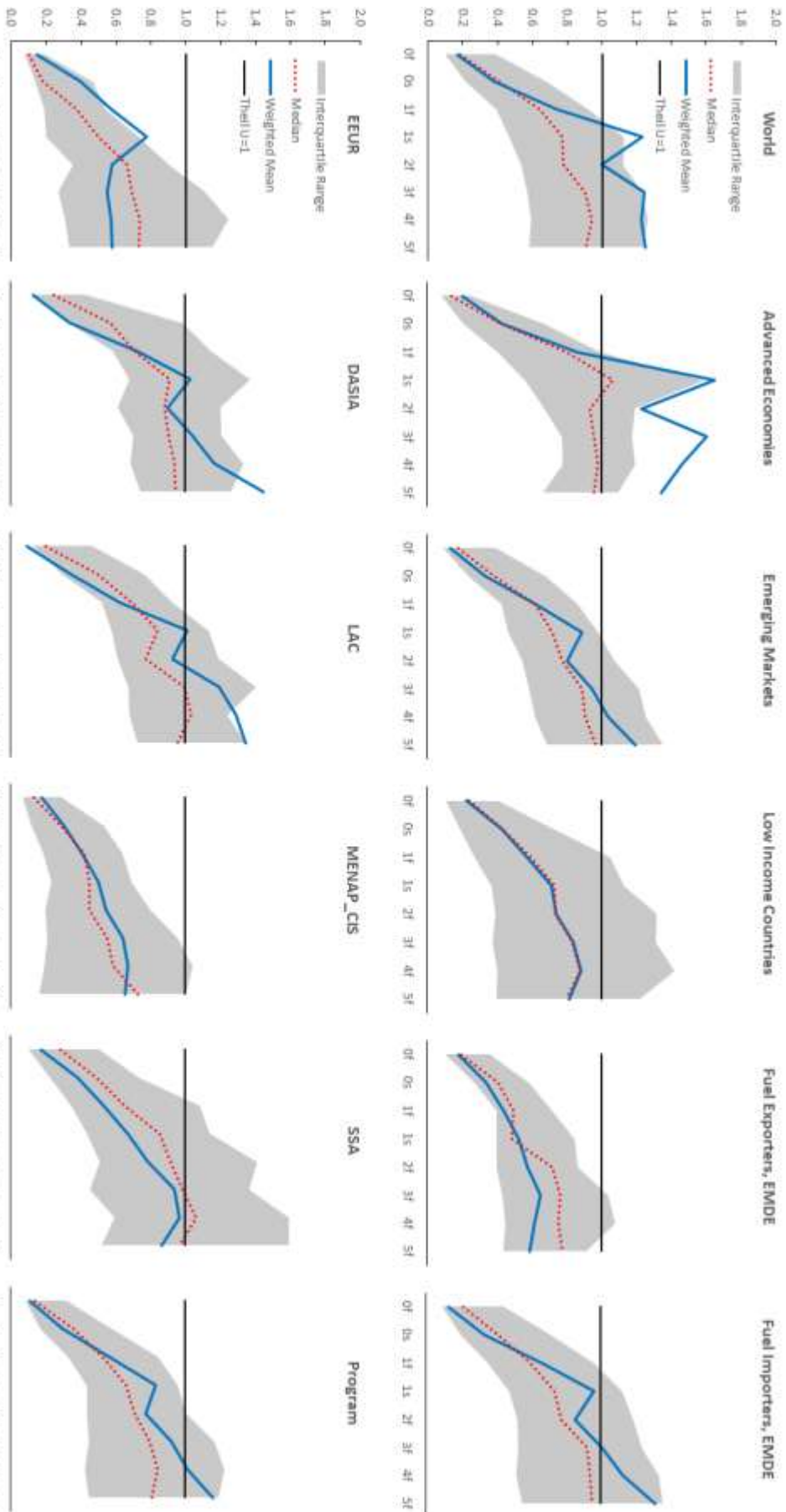
Note: Sample period = 2004-17. Distributions of ratios of RMSEs for WEO GDP growth forecasts for two consecutive forecasting horizons across regions and across forecast horizons. The boxes show the p25-p75 quantile range and the horizontal lines show the p10-p90 quantile range. The median of the country group is the centerline in the box. EMDE = Emerging Market and Developing Economies; DASIA = Developing Asia; EUR = Emerging Europe; LAC = Latin America and the Caribbean; MENAP_CIS = Middle East, North Africa, Afghanistan, Pakistan, and Commonwealth of Independent States; SSA = Sub-Saharan Africa; Program = IMF program countries.

Figure 3. Difference in the RMSE values calculated for two sub-samples, those for 1990–2003 minus those for 2004–17



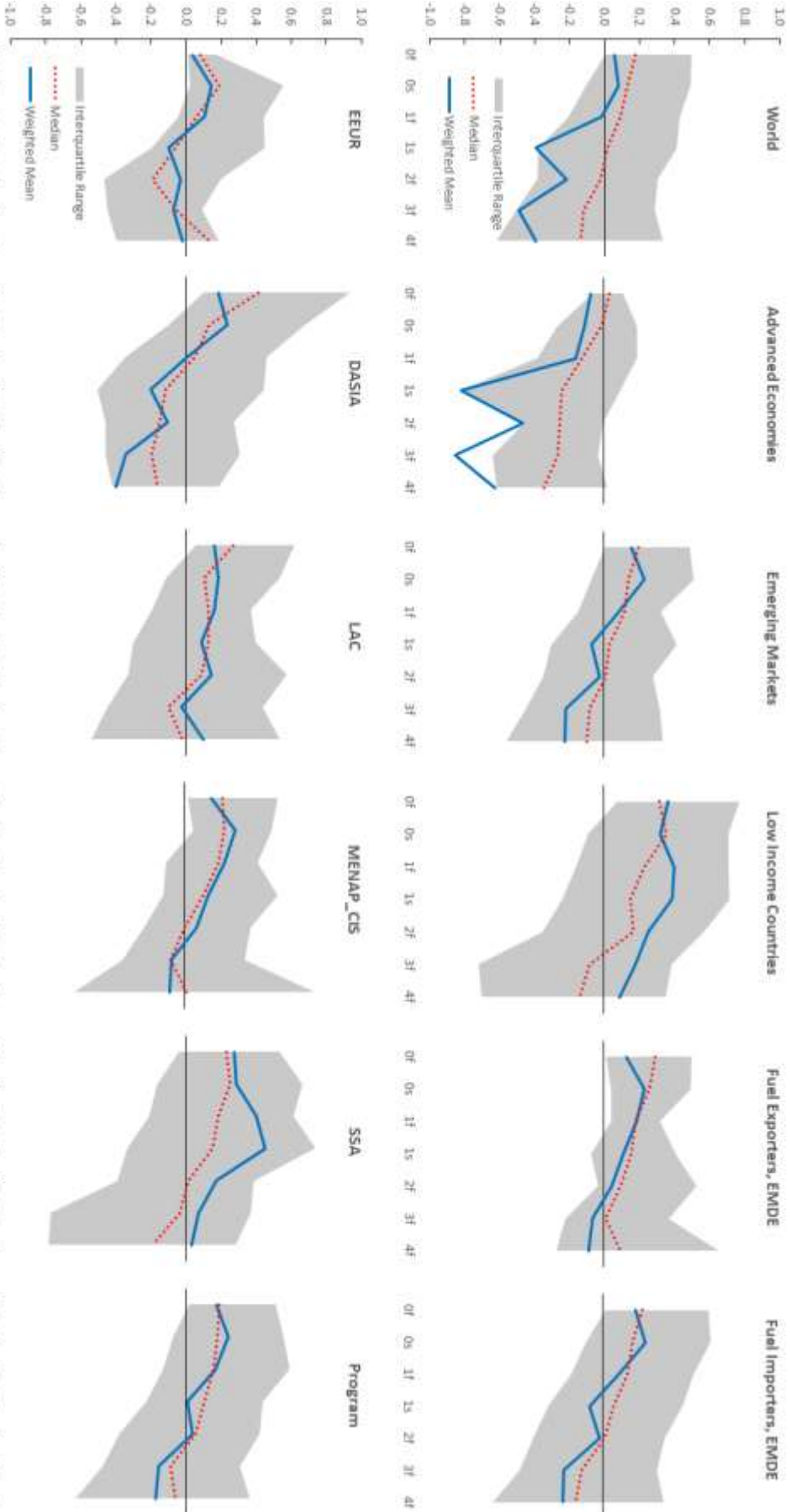
Note: EMDE = Emerging Market and Developing Economies; DASIA = Developing Asia; EUR = Emerging Europe; LAC = Latin America and the Caribbean; MENAP_CIS = Middle East, North Africa, Afghanistan, Pakistan, and Commonwealth of Independent States; SSA = Sub-Saharan Africa; Program = Iltf program countries.

Figure 4. Values of the Theil U-statistic computed over the sample 2004–17



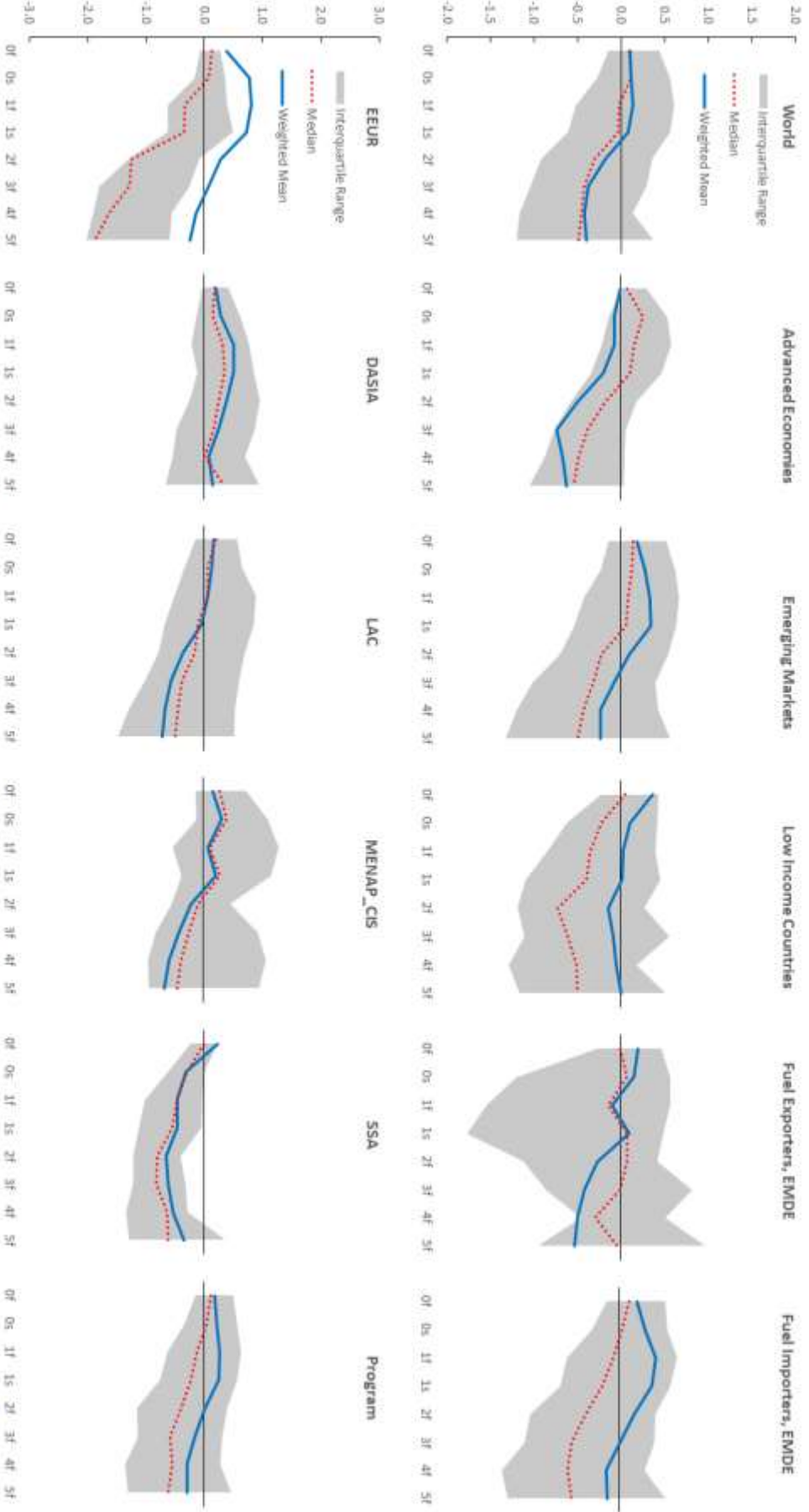
Note: The Figure shows estimates of the Theil U-statistic computed as the ratio of the mean squared error (MSE) of the recursively updated historical average GDP growth (i.e. the "naive" forecast). Estimates are based on the sample 2004–17, with the naive forecast errors calculated by averaging that started in 1993. Values below unity indicate that the WEO forecasts are relatively more accurate than the historical average forecasts in an absolute and relative sense, while values above unity suggest that the historical average is most accurate. EMDE = Emerging Market and Developing Economies, DASIA = Developing Asia, EUR = Emerging Europe, LAC = Latin America and the Caribbean, MENAP_CIS = Middle East, North Africa, Afghanistan, Pakistan, and Commonwealth of Independent States, SSA = Sub-Saharan Africa, Program = IMF program countries.

Figure 5. Difference in the Theil-U Statistic, those for 1990-2003 minus those for 2004-17



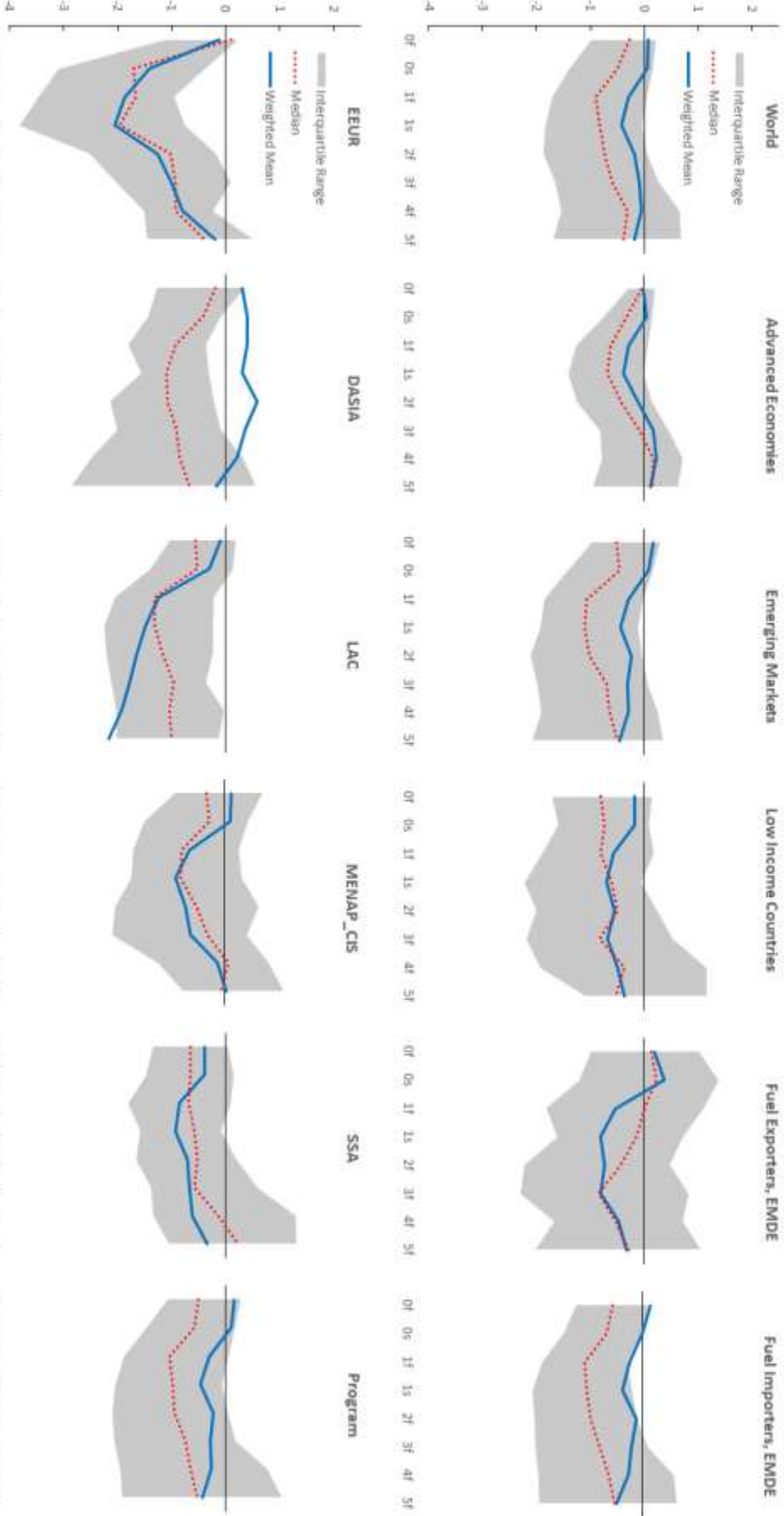
Note: EMDE = Emerging Market and Developing Economies; DASIA = Developing Asia; EUR = Emerging Europe; LAC = Latin America and the Caribbean; MENAP_CIS = Middle East, North Africa, Afghanistan, Pakistan, and Commonwealth of Independent States; SSA = Sub-Saharan Africa; Program = IMF program countries.

Figure 6. Forecast bias: median, weighted mean, and interquartile range of WEO real GDP growth forecast biases



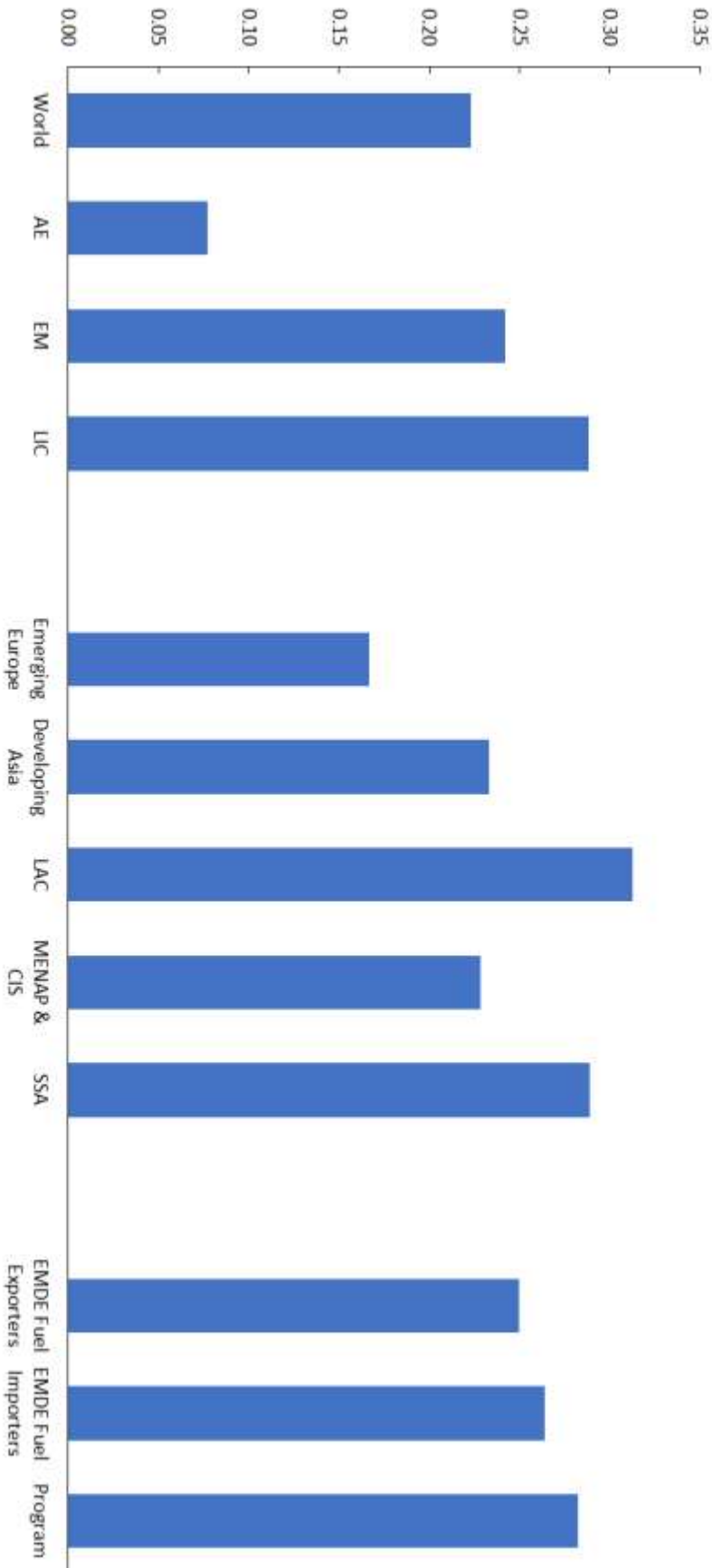
Notes: Sample (2004-17), shows biases of WEO GDP growth forecasts for different country groups and different forecast horizons. Biases are calculated as the sample mean differences between the actual GDP growth and the predicted value, with positive values indicating underpredictions while negative values indicate overpredictions (i.e. optimism) of the outcome. EMDE = Emerging Market and Developing Economies; DASIA = Developing Asia; EUR = Emerging Europe; LAC = Latin America and the Caribbean; MIENAP_CIS = Middle East, North Africa, Afghanistan, Pakistan, and Commonwealth of Independent States; SSA = Sub-Saharan Africa; Program = IJF program countries.

Figure 7. Differences in bias: those for 1990-2003 minus those for 2004-17



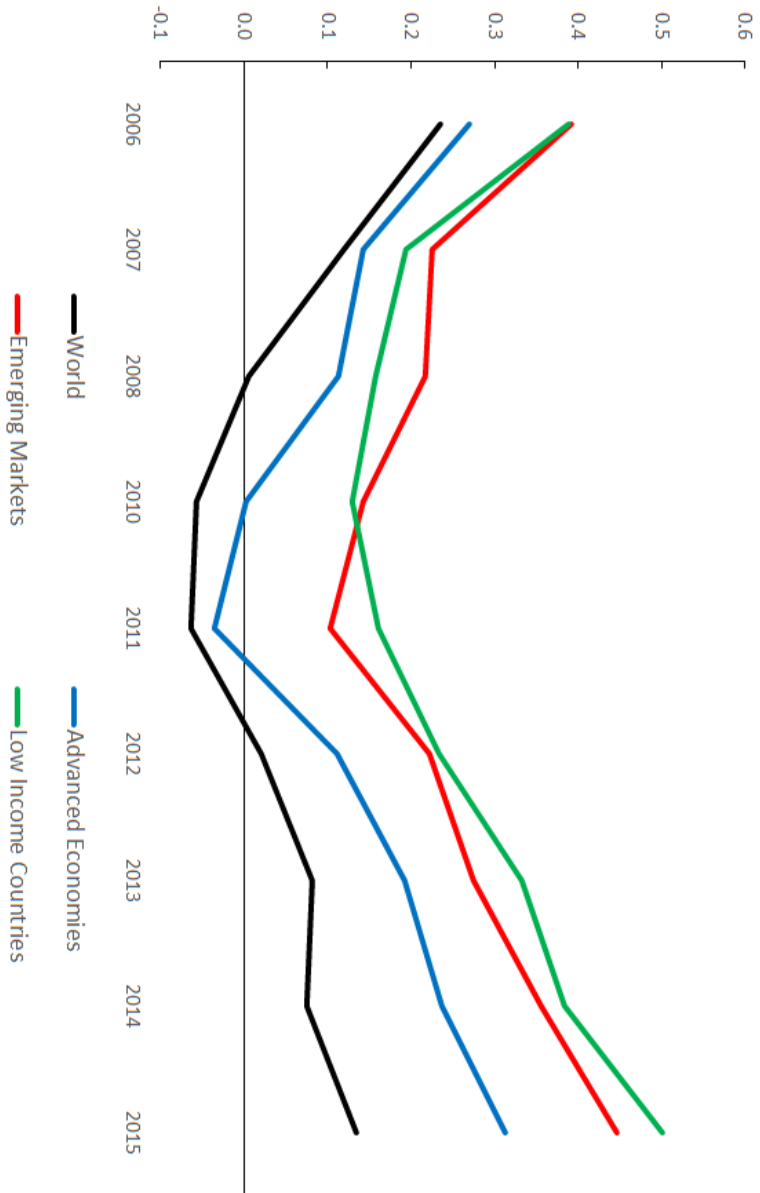
Notes: Differences in biases of WEO GDP growth forecasts between two periods, 1990-03 and 2004-17, shown across regions and across forecast horizons. Biases are calculated as the sample mean differences between the actual GDP growth and the predicted value, with positive values indicating underpredictions while negative values indicate overpredictions (i.e. optimism) of the outcome. Differences in biases are then calculated for the two periods. EMIDE = Emerging Market and Developing Economies; DASA = Developing Asia; EEUR = Emerging Europe; LAC = Latin America and the Caribbean; MENAP_CIS = Middle East, North Africa, Afghanistan, Pakistan, and Commonwealth of Independent States; SSA = Sub-Saharan Africa; Program = IMF program countries.

Figure 8. Share of countries with serially correlated forecast errors



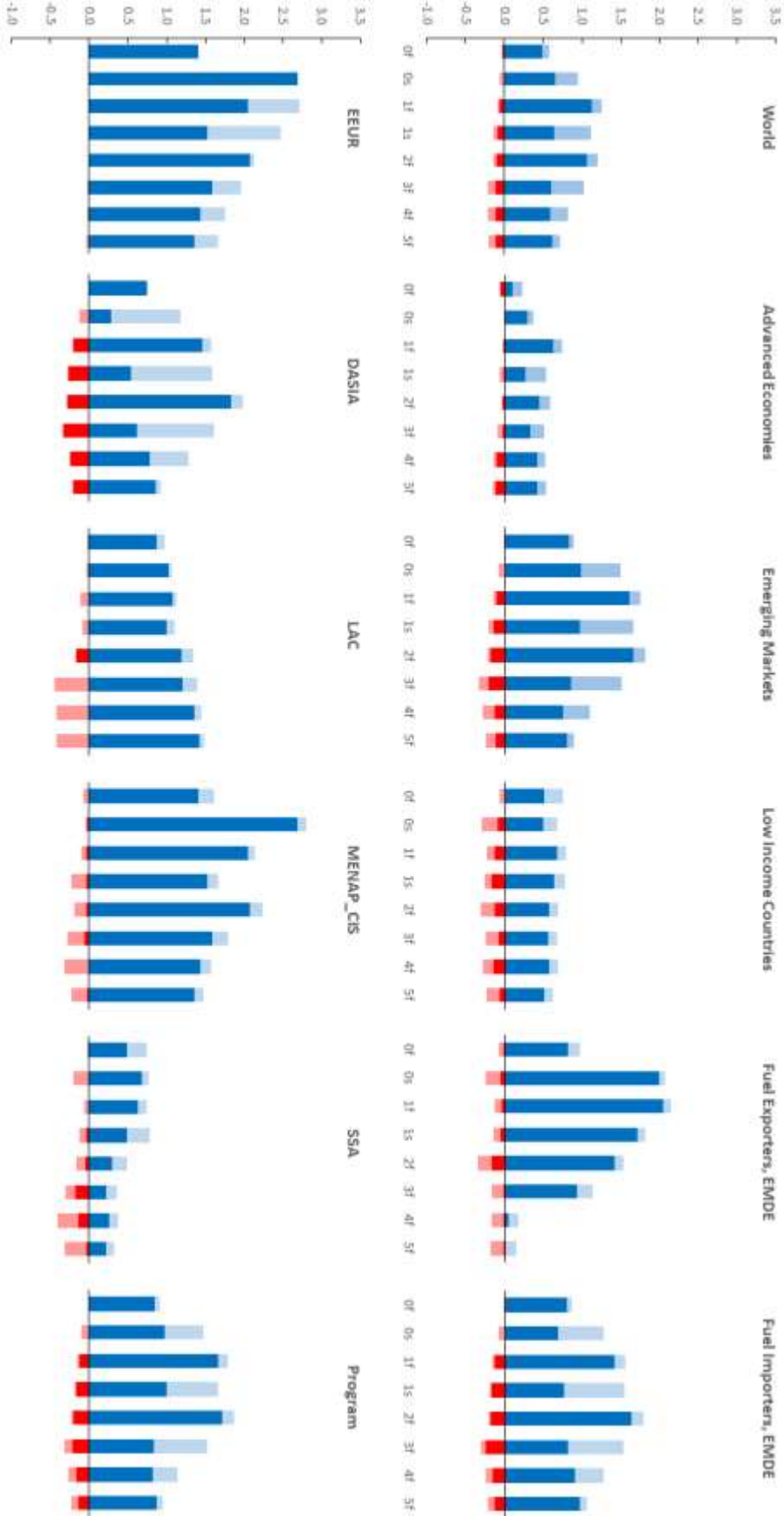
Note: Sample = 2004:17. An estimate of the serial correlation of forecast errors for country i is calculated from a linear regression without the intercept term. The regression contains the the h -year-ahead forecast error at time t on the LHS and the h -year-ahead forecast error at time $t-1$ on the RHS. The regression runs over the pre-specified time period for each country separately. The significance of the regression parameter is used to test for the presence of serial correlations of forecast errors. EMDE = Emerging Market and Developing Economies; LIC = Low-Income Countries; EM = Emerging Markets (EMDE - LIC); DASIA = Developing Asia; EEUR = Emerging Europe; LAC = Latin America and the Caribbean; MENAP & CIS = Middle East, North Africa, Afghanistan, Pakistan, and Commonwealth of Independent States; SSA = Sub-Saharan Africa; Program = IMF program countries.

Figure 9. Local serial correlation in forecast errors



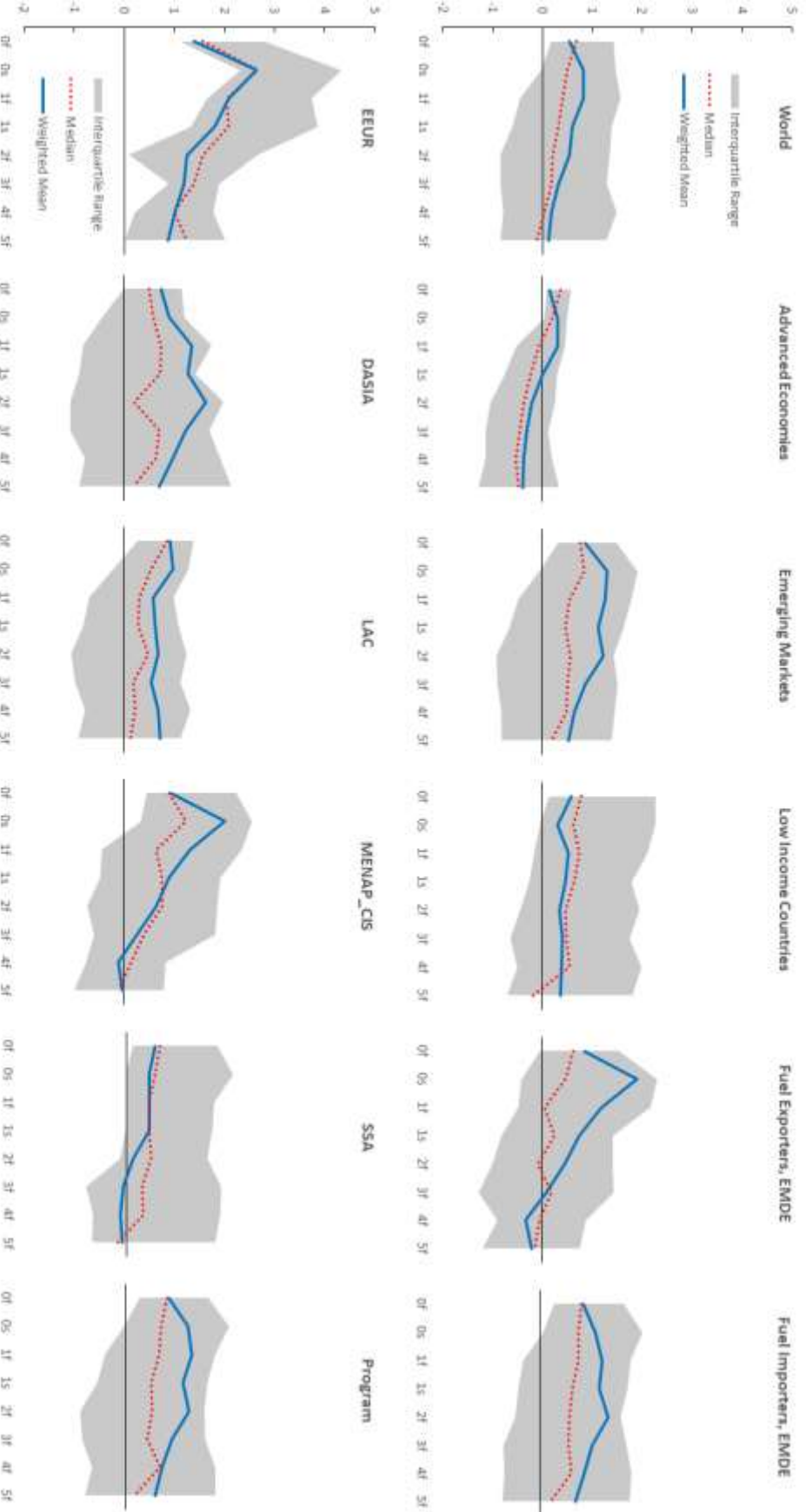
Notes: Local serial correlation over time for select country groups, 2006-15. Pooling forecast errors over a group of countries allows us to get more robust estimates of the average serial correlations of errors. Note that these are averages over countries for each point in time. In this sense, they are measures of correlations that is local in time.

Annex Figure 1. Significant share of positive and negative changes in RMSEs



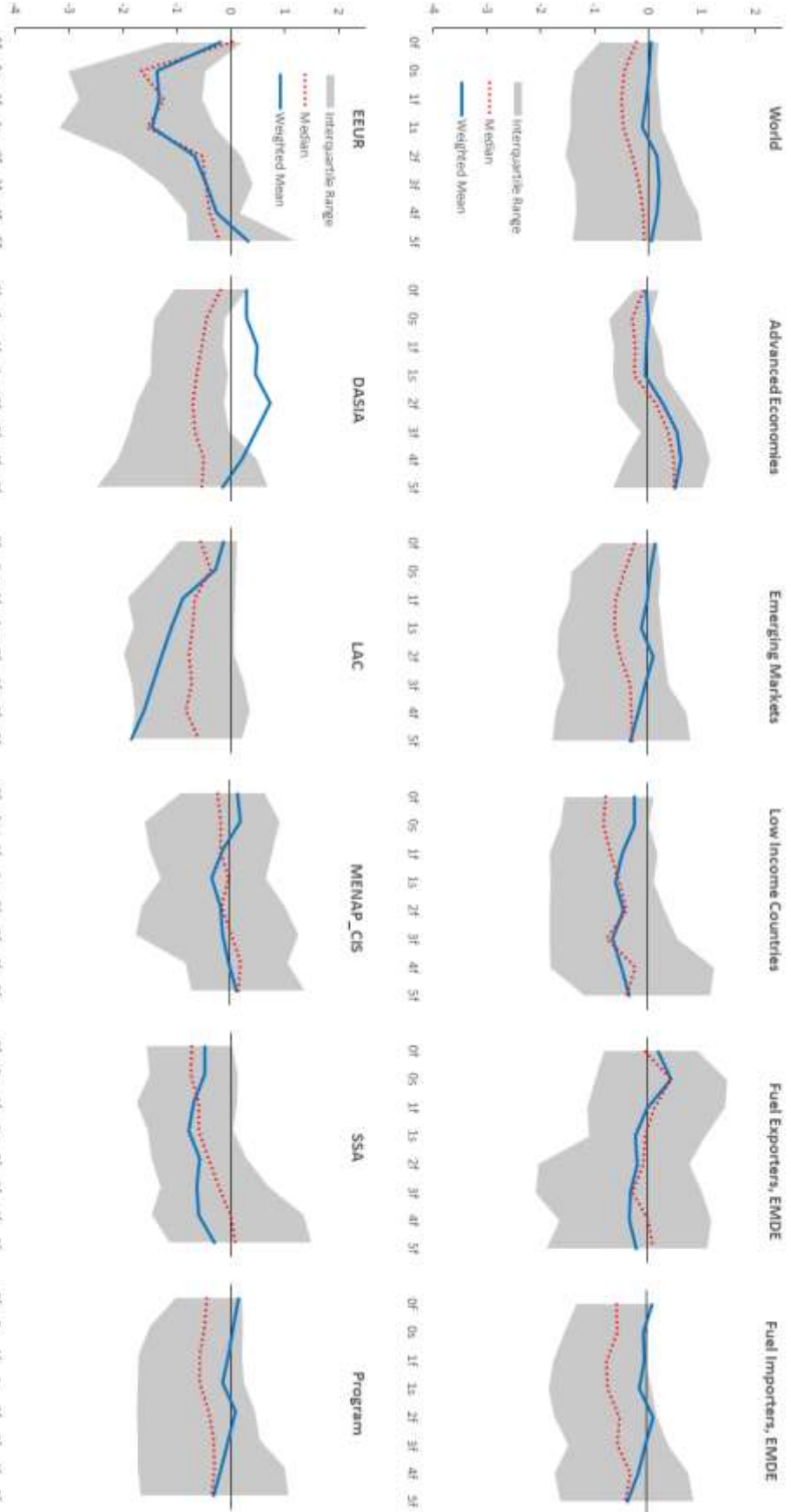
Note: Differences of RMSEs of growth forecasts between two periods, 1990-03 and 2004-17, shown across regions and across forecast horizons. Darker colors indicate significant share and lighter colors insignificant share, based on the p-value obtained from bootstrapping. The bootstrapping is conducted under the null that the expected squared forecast errors are the same across the two periods. A small p-value indicates that the 2004-17 RMSE value is significantly smaller than the 1990-2003 RMSE value. EMDE = Emerging Market and Developing Economies, DASIA = Developing Asia, EUR = Emerging Europe, LAC = Latin America and the Caribbean, MENAP_CIS = Middle East, North Africa, Afghanistan, Pakistan, and Commonwealth of Independent States, SSA = Sub-Saharan Africa, Program = IMF program countries.

Annex Figure 2. Change in accuracy when 2009 is included



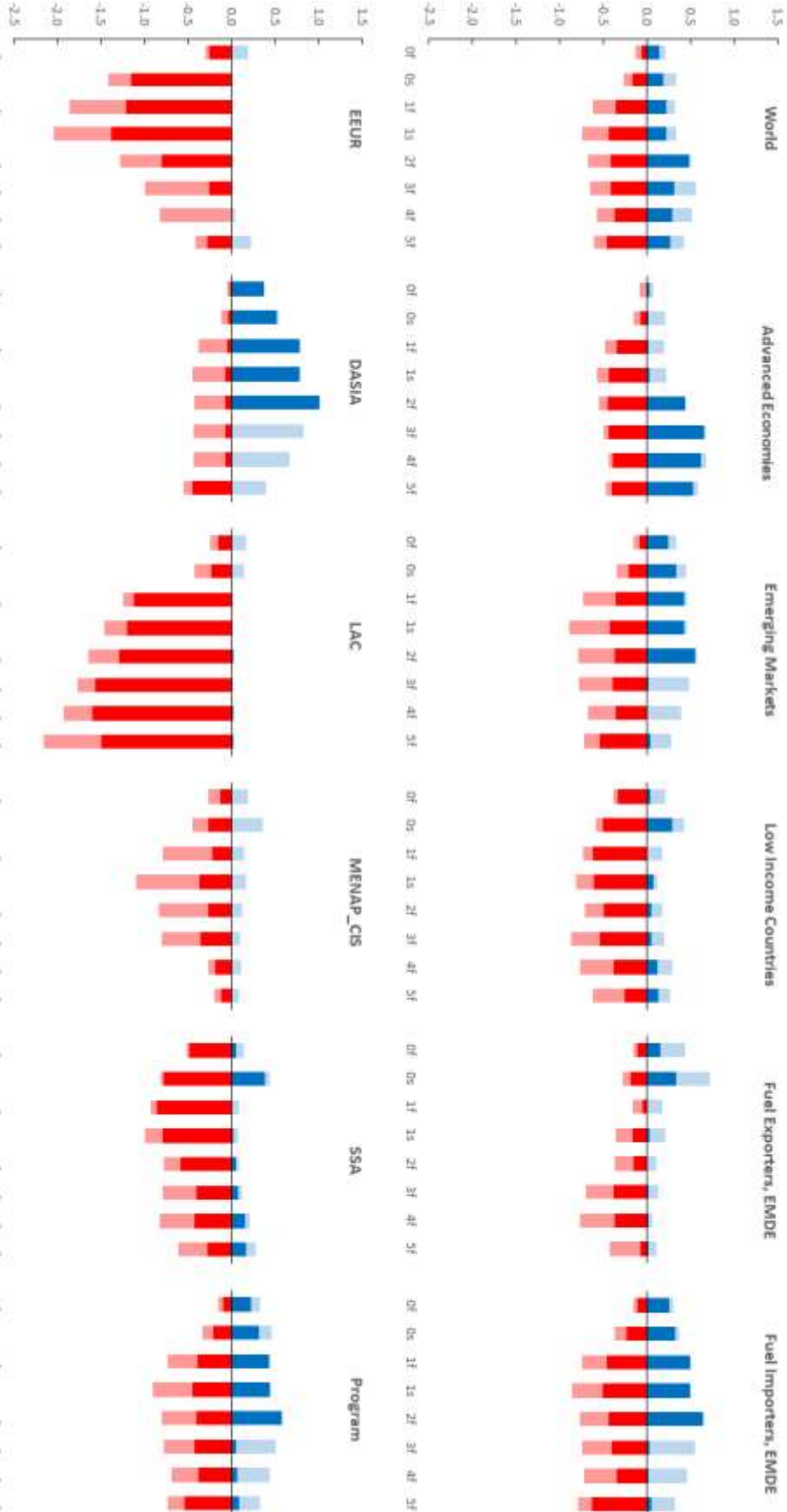
Note: EMDE = Emerging Market and Developing Economies; DASIA = Developing Asia; EUR = Emerging Europe; LAC = Latin America and the Caribbean; MENAP_CIS = Middle East, North Africa, Afghanistan, Pakistan, and Commonwealth of Independent States; SSA = Sub-Saharan Africa; Program = IMF program countries.

Annex Figure 3. Change in bias when 2009 is included



Note: Differences in biases of WEO GDP growth forecasts between two periods, 1990-03 and 2004-17, shown across regions and across forecast horizons. Biases are calculated as the sample mean differences between the actual GDP growth and the predicted value, with positive values indicating underpredictions while negative values indicate overpredictions (i.e. optimism) of the outcome. Differences in biases are then calculated for the two periods. EMDE = Emerging Market and Developing Economies, DASA = Developing Asia, EUR = Emerging Europe, LAC = Latin America and the Caribbean, MENAP_CIS = Middle East, North Africa, Afghanistan, Pakistan, and Commonwealth of Independent States, SSA = Sub-Saharan Africa, Program = IMF program countries.

Annex Figure 4. Shares of statistically significant positive and negative changes in Bias



Note: Bars denote weighted averages of 1990-2003 biases subtracted by those of 2004-2017. Darker blue bars denote significant decreases, lighter blue bars denote insignificant decreases, darker red bars denote significant increases, lighter red bars denote insignificant increases; 2004-2017 biases are calculated without 2009 value. EMDE = Emerging Market and Developing Economies; DASIA = Developing Asia; EUR = Emerging Europe; LAC = Latin America and the Caribbean; MIENAP_CIS = Middle East, North Africa, Afghanistan, Pakistan, and Commonwealth of Independent States; SSA = Sub-Saharan Africa; Program = IMF program countries.

Table 1. Difference between the mean and median of real GDP growth¹, 1994-2017

	Full sample		After dropping years with conflict		After dropping years with natural disasters		After dropping entry errors		After dropping 2009	
	Weighted average ²	Percent w/ negative values	Weighted average	Percent w/ negative values	Weighted average	Percent w/ negative values	Weighted average	Percent w/ negative values	Weighted average	Percent w/ negative values
World	-0.27	72.4	-0.27	71.9	-0.27	70.3	-0.28	71.2	-0.17	62.8
Advanced Economies	-0.19	82.1	-0.19	82.1	-0.19	82.1	-0.19	84.6	-0.06	64.1
Emerging Market Economies	-0.36	70.2	-0.36	70.2	-0.37	69.1	-0.37	69.9	-0.29	62.4
Low-Income Economies	-0.24	69.5	-0.21	67.8	-0.20	64.4	-0.20	64.4	-0.08	62.7
EMDE Fuel Exporters	-0.60	53.6	-0.61	60.7	-0.64	60.7	-0.65	63.0	-0.51	48.1
EMDE Fuel Importers	-0.28	74.2	-0.28	71.8	-0.28	69.4	-0.28	69.4	-0.21	66.1
EEUR	-0.81	75.0	-0.82	75.0	-0.90	75.0	-0.90	75.0	-0.69	58.3
DASIA	-0.20	72.4	-0.20	72.4	-0.20	69.0	-0.20	69.0	-0.13	62.1
LAC	-0.28	75.0	-0.28	75.0	-0.32	75.0	-0.32	75.0	-0.31	71.9
MENAP_CIS	-0.64	62.9	-0.61	60.0	-0.60	57.1	-0.60	57.1	-0.48	51.4
SSA	-0.02	68.9	-0.09	68.9	-0.08	66.7	-0.13	68.2	0.03	65.9
Program ³	-0.51	73.6	-0.50	72.5	-0.52	70.8	-0.53	72.3	-0.42	66.9

Note: EMDE = Emerging Market and Developing Economies; DASIA = Developing Asia; EEUR = Emerging Europe; LAC = Latin America and the Caribbean; MENAP_CIS =

Middle East, North Africa, Afghanistan, Pakistan, and Commonwealth of Independent States; SSA = Sub-Saharan Africa; Program = IMF program countries.

¹ Real GDP growth used is outturns used to construct forecast errors, fall growth in the following year the forecast is made.

² Difference between mean and median is calculated for each country, then weighted averages and shares are calculated for respective regions.

³ A country is classified as program country when cross-country averages are calculated if a country was in program at least once within periods.

Table 2. Share of countries with statistically-significant positive or negative biases

	<i>(h=0,F)</i>	<i>(h=0,S)</i>	<i>(h=1,F)</i>	<i>(h=1,S)</i>	<i>(h=2,F)</i>	<i>(h=3,F)</i>	<i>(h=4,F)</i>	<i>(h=5,F)</i>
Percent of countries with statistically-significant downward biases								
World	17.1	18.9	13.1	10.9	6.9	6.3	5.7	6.9
Advanced Economies	13.9	16.7	8.3	5.6	0.0	0.0	0.0	0.0
Emerging Market Economies	15.7	22.9	14.5	13.3	7.2	6.0	4.8	6.0
Low-Income Economies	21.4	14.3	14.3	10.7	10.7	10.7	10.7	12.5
EMDE Fuel Exporters	16.7	20.8	16.7	16.7	8.3	4.2	4.2	8.3
EMDE Fuel Importers	18.3	19.1	13.9	11.3	8.7	8.7	7.8	8.7
DASIA	20.0	16.0	16.0	12.0	16.0	16.0	16.0	20.0
EEUR	0.0	33.3	11.1	0.0	0.0	0.0	0.0	0.0
LAC	31.3	28.1	25.0	18.8	12.5	9.4	6.3	9.4
MENAP_CIS	20.0	30.0	20.0	23.3	10.0	10.0	10.0	10.0
SSA	9.3	4.7	2.3	2.3	2.3	2.3	2.3	2.3
Program	18.6	19.6	14.4	10.3	5.2	4.1	3.1	4.1
Percent of countries with statistically-significant upward biases								
World	3.4	6.3	10.3	10.3	21.1	25.7	28.0	25.7
Advanced Economies	0.0	2.8	5.6	5.6	27.8	33.3	33.3	30.6
Emerging Market Economies	4.8	7.2	8.4	8.4	13.3	22.9	26.5	25.3
Low-Income Economies	3.6	7.1	16.1	16.1	28.6	25.0	26.8	23.2
EMDE Fuel Exporters	8.3	16.7	20.8	20.8	12.5	12.5	8.3	12.5
EMDE Fuel Importers	3.5	5.2	9.6	9.6	20.9	26.1	30.4	27.0
DASIA	0.0	4.0	8.0	4.0	4.0	12.0	16.0	16.0
EEUR	0.0	0.0	0.0	0.0	44.4	66.7	66.7	66.7
LAC	9.4	12.5	12.5	12.5	15.6	18.8	25.0	25.0
MENAP_CIS	3.3	3.3	6.7	6.7	6.7	16.7	16.7	13.3
SSA	4.7	9.3	18.6	20.9	34.9	30.2	32.6	27.9
Program	4.1	7.2	12.4	13.4	26.8	29.9	32.0	27.8

Note: Newey West standard errors; EMDE = Emerging Market and Developing Economies; DASIA = Developing Asia; EEUR = Emerging Europe; LAC = Latin America and the Caribbean; MENAP_CIS = Middle East, North Africa, Afghanistan, Pakistan, and Commonwealth of Independent States; SSA = Sub-Saharan Africa; Program = IMF program countries.

Table 3. Spillovers from Forecast Errors of United States, China, and Euro Area

	Median coefficient on growth forecast error			Percent of countries with statistically significant positive coefficients			Percent of countries with statistically significant negative coefficients			Average R-square of regressions with statistically significant positive coefficients		Average R-square of regressions with statistically significant negative coefficients			
	(h=0,F)	(h=1,F)	(h=2,F)	(h=0,F)	(h=1,F)	(h=2,F)	(h=0,F)	(h=1,F)	(h=2,F)	(h=0,F)	(h=1,F)	(h=2,F)	(h=0,F)	(h=1,F)	(h=2,F)
Spillover from US Forecast Error															
World	0.13	0.32	-0.10	0.07	10.3	16.1	10.9	5.2	4.6	2.3	8.0	3.4	0.34	0.39	0.35
Advanced Economies	0.29	0.32	0.17	0.07	11.4	22.9	17.1	11.4	2.9	5.7	0.0	0.0	0.28	0.34	0.37
Emerging Market Economies	0.14	0.33	-0.17	0.12	6.0	13.3	10.8	3.6	6.0	1.2	12.0	6.0	0.30	0.37	0.34
Low-Income Economies	-0.01	0.37	-0.19	-0.10	16.1	16.1	7.1	3.6	3.6	1.8	7.1	1.8	0.39	0.46	0.36
EMDE Fuel Exporters	0.28	0.54	0.26	0.17	8.3	12.5	12.5	0.0	0.0	0.0	4.2	0.0	0.32	0.40	0.27
EMDE Fuel Importers	0.07	0.29	-0.20	0.04	10.4	14.8	8.7	4.3	6.1	1.7	11.3	5.2	0.36	0.41	0.37
DASIA	0.01	0.32	0.15	0.23	8.0	20.0	16.0	8.0	4.0	4.0	8.0	8.0	0.39	0.42	0.41
EEUR	0.43	-0.12	-0.94	-0.44	11.1	11.1	11.1	0.0	0.0	0.0	33.3	0.0	0.23	0.32	0.42
LAC	0.17	0.44	-0.10	0.10	0.0	12.5	6.3	3.1	3.1	0.0	9.4	9.4	n.a.	0.45	0.30
MENAP_CIS	0.20	0.34	-0.46	0.13	10.0	6.7	13.3	3.3	10.0	3.3	10.0	0.0	0.32	0.32	0.26
SSA	0.00	0.55	-0.20	-0.02	18.6	18.6	4.7	2.3	4.7	0.0	7.0	2.3	0.38	0.42	0.42
Program	0.06	0.25	-0.27	-0.02	12.4	13.4	7.2	3.1	5.2	3.1	11.3	3.1	0.36	0.40	0.34
Spillover from China Forecast Error															
World	0.39	0.46	0.37	0.27	29.9	33.9	33.3	39.1	2.9	0.6	2.9	5.2	0.45	0.42	0.43
Advanced Economies	0.41	0.50	0.33	0.16	47.2	58.3	38.9	27.8	0.0	0.0	0.0	2.8	0.46	0.41	0.46
Emerging Market Economies	0.53	0.64	0.52	0.46	31.7	36.6	40.2	50.0	1.2	0.0	0.0	2.4	0.43	0.40	0.42
Low-Income Economies	0.15	0.19	0.13	0.16	16.1	14.3	19.6	30.4	7.1	1.8	8.9	10.7	0.47	0.53	0.44
EMDE Fuel Exporters	0.60	0.82	0.44	0.62	29.2	25.0	25.0	58.3	0.0	0.0	0.0	0.0	0.41	0.35	0.31
EMDE Fuel Importers	0.39	0.39	0.36	0.29	24.6	28.1	33.3	38.6	4.4	0.9	4.4	7.0	0.45	0.44	0.44
DASIA	0.32	0.37	0.22	0.27	33.3	20.8	33.3	37.5	0.0	0.0	0.0	8.3	0.43	0.51	0.44
EEUR	0.39	0.41	0.48	0.27	22.2	22.2	33.3	11.1	11.1	0.0	0.0	0.0	0.32	0.42	0.40
LAC	0.58	0.65	0.46	0.41	28.1	43.8	40.6	43.8	0.0	0.0	3.1	3.1	0.52	0.40	0.41
MENAP_CIS	0.59	0.70	0.63	0.58	30.0	40.0	40.0	70.0	3.3	0.0	3.3	0.0	0.40	0.39	0.41
SSA	0.12	0.18	0.12	0.12	16.3	11.6	18.6	30.2	7.0	2.3	7.0	11.6	0.44	0.53	0.46
Program	0.37	0.43	0.38	0.28	24.7	27.8	30.9	30.9	5.2	1.0	4.1	6.2	0.45	0.42	0.41
Spillover from Euro Area Forecast Error															
World	0.50	0.37	0.04	-0.03	16.0	25.0	11.5	11.5	3.2	0.6	4.5	6.4	0.37	0.38	0.37
Advanced Economies	0.88	0.60	0.44	0.42	47.1	58.8	35.3	23.5	0.0	0.0	0.0	0.0	0.33	0.46	0.37
Emerging Market Economies	0.78	0.50	0.04	0.03	15.7	25.3	8.4	13.3	3.6	1.2	3.6	3.6	0.37	0.37	0.39
Low-Income Economies	-0.07	0.14	-0.16	-0.35	7.1	14.3	8.9	5.4	3.6	0.0	7.1	12.5	0.43	0.32	0.36
EMDE Fuel Exporters	0.43	-0.18	-0.71	-0.74	4.2	4.2	0.0	4.2	4.2	4.2	12.5	12.5	0.32	0.33	n.a.
EMDE Fuel Importers	0.49	0.41	0.05	-0.01	13.9	24.3	10.4	11.3	3.5	0.0	3.5	6.1	0.39	0.36	0.37
DASIA	0.36	0.47	-0.08	0.01	12.0	28.0	12.0	4.0	8.0	0.0	8.0	8.0	0.32	0.36	0.39
EEUR	1.20	0.99	0.86	1.16	44.4	66.7	22.2	55.6	0.0	0.0	0.0	0.0	0.40	0.43	0.37
LAC	0.77	0.37	0.09	-0.11	12.5	21.9	12.5	9.4	3.1	0.0	0.0	9.4	0.30	0.29	0.33
MENAP_CIS	0.25	0.21	-0.12	-0.13	6.7	13.3	0.0	6.7	3.3	3.3	6.7	0.0	0.45	0.34	0.53
SSA	0.24	0.12	-0.18	-0.49	9.3	11.6	7.0	7.0	2.3	0.0	11.6	11.6	0.48	0.38	0.42
Program	0.57	0.37	-0.06	-0.08	14.1	26.1	10.9	12.0	4.3	0.0	6.5	8.7	0.41	0.36	0.37

Note: The calculations are based on country-by-country regressions of growth forecast errors on the US growth forecast error for the same horizon. EMDE = Emerging Market and Developing Economies; DASIA = Developing Asia; EEUR = Emerging Europe; LAC = Latin America and the Caribbean; MENAP_CIS = Middle East, North Africa, Afghanistan, Pakistan, and Commonwealth of Independent States; SSA = Sub-Saharan Africa; Program = IMF program countries.

Table 4. Spillovers from Terms-of-Trade Forecast Errors

	Median coefficient on terms-of-trade forecast error			Percent of countries with statistically significant positive coefficients			Percent of countries with statistically significant negative coefficients			Average R-square of regressions with statistically significant positive coefficients			Average R-square of regressions with statistically significant negative coefficients							
	(h=0,F)	(h=1,F)	(h=2,F)	(h=0,F)	(h=1,F)	(h=2,F)	(h=0,F)	(h=1,F)	(h=2,F)	(h=0,F)	(h=1,F)	(h=2,F)	(h=0,F)	(h=1,F)	(h=2,F)					
World	0.001	0.006	0.007	0.006	7.6	10.6	17.1	23.5	5.9	3.5	5.9	8.2	0.34	0.46	0.44	0.41	0.44	0.42	0.44	0.42
Advanced Economies	-0.009	0.025	0.017	0.010	2.9	5.7	8.6	28.6	11.4	2.9	5.7	5.7	0.32	0.45	0.31	0.42	0.49	0.48	0.29	0.36
Emerging Market Economies	0.004	0.003	0.005	0.010	9.9	9.9	19.8	23.5	6.2	3.7	4.9	8.6	0.36	0.43	0.48	0.39	0.42	0.43	0.61	0.64
Low-Income Economies	0.001	0.005	0.005	0.002	7.4	14.8	18.5	20.4	1.9	3.7	7.4	9.3	0.31	0.50	0.41	0.43	0.32	0.38	0.33	0.42
EMDE Fuel Exporters	0.005	0.020	0.038	0.036	8.7	17.4	47.8	56.5	4.3	4.3	4.3	4.3	0.45	0.49	0.48	0.41	0.50	0.33	0.57	0.52
EMDE Fuel Importers	0.001	0.003	0.000	0.001	8.9	10.7	13.4	15.2	4.5	3.6	6.3	9.8	0.32	0.45	0.43	0.41	0.38	0.43	0.46	0.55
DASIA	0.000	0.008	0.002	0.000	13.6	18.2	27.3	18.2	0.0	0.0	4.5	18.2	0.44	0.41	0.50	0.38	n.a.	n.a.	0.88	0.72
EEUR	-0.009	-0.017	0.001	0.046	0.0	0.0	11.1	11.1	22.2	0.0	0.0	0.0	n.a.	n.a.	0.37	0.43	0.45	n.a.	n.a.	n.a.
LAC	0.005	0.005	0.001	0.010	9.4	12.5	15.6	28.1	3.1	6.3	6.3	3.1	0.33	0.43	0.48	0.39	0.50	0.48	0.50	0.33
MENAP_CIS	0.005	0.005	0.018	0.026	6.7	10.0	26.7	26.7	6.7	3.3	6.7	13.3	0.27	0.39	0.42	0.41	0.38	0.33	0.42	0.54
SSA	0.003	0.006	0.006	0.001	9.5	11.9	14.3	19.0	2.4	4.8	7.1	7.1	0.32	0.57	0.44	0.43	0.25	0.38	0.35	0.40
Program	0.001	0.003	0.005	0.003	6.4	7.4	14.9	19.1	5.3	4.3	6.4	7.4	0.31	0.46	0.42	0.40	0.40	0.40	0.40	0.42

Note: The calculations are based on country-by-country regressions of growth forecast errors on the terms-of-trade forecast error for the same horizon. EMDE = Emerging Market and Developing Economies; DASIA = Developing Asia; EEUR = Emerging Europe; LAC = Latin America and the Caribbean; MENAP_CIS = Middle East, North Africa, Afghanistan, Pakistan, and Commonwealth of Independent States; SSA = Sub-Saharan Africa; Program = IMF program countries.

Table 5. Correlations between Growth Forecast Errors and Growth Forecasts of United States, China, and Euro Area

	Median coefficient on growth forecast					Percent of countries with statistically significant positive coefficients					Percent of countries with statistically significant negative coefficients				
	(h=0,F)	(h=1,F)	(h=2,F)	(h=5,F)		(h=0,F)	(h=1,F)	(h=2,F)	(h=5,F)		(h=0,F)	(h=1,F)	(h=2,F)	(h=5,F)	
Correlation to US Growth Forecast															
World	0.24	0.30	0.94	0.69		24.7	16.1	25.3	10.9		0.0	3.4	1.1	2.9	
Advanced Economies	0.21	0.46	0.36	1.07		31.4	17.1	11.4	25.7		0.0	0.0	2.9	0.0	
Emerging Market Economies	0.35	0.47	1.58	0.86		27.7	19.3	37.3	12.0		0.0	1.2	0.0	2.4	
Low-Income Economies	0.14	-0.17	0.44	0.38		16.1	10.7	16.1	0.0		0.0	8.9	1.8	5.4	
EMDE Fuel Exporters	0.77	0.56	1.57	-0.34		25.0	16.7	33.3	0.0		0.0	4.2	0.0	0.0	
EMDE Fuel Importers	0.23	0.20	0.93	0.72		22.6	15.7	27.8	8.7		0.0	4.3	0.9	4.3	
DASIA	0.25	-0.05	0.85	0.28		24.0	12.0	12.0	4.0		0.0	4.0	0.0	4.0	
EEUR	0.24	0.34	1.39	1.36		22.2	22.2	55.6	22.2		0.0	0.0	0.0	0.0	
LAC	0.40	0.42	1.23	0.50		37.5	25.0	31.3	12.5		0.0	0.0	0.0	3.1	
MENAP_CIS	0.33	0.70	1.93	0.90		23.3	26.7	50.0	6.7		0.0	0.0	0.0	0.0	
SSA	0.12	-0.21	0.51	0.42		11.6	2.3	16.3	2.3		0.0	11.6	2.3	7.0	
Program	0.21	0.23	0.94	0.53		18.6	16.5	27.8	8.2		0.0	4.1	0.0	4.1	
Correlation to China Growth Forecast															
World	0.05	0.05	-0.08	-0.59		8.0	10.9	9.2	4.0		2.3	8.6	18.4	31.6	
Advanced Economies	-0.03	-0.21	-0.42	-0.42		0.0	0.0	0.0	0.0		0.0	13.9	33.3	30.6	
Emerging Market Economies	0.08	0.12	-0.08	-0.98		6.1	12.2	6.1	2.4		1.2	9.8	20.7	40.2	
Low-Income Economies	0.08	0.16	0.11	-0.18		16.1	16.1	19.6	8.9		5.4	3.6	5.4	19.6	
EMDE Fuel Exporters	0.10	0.34	0.56	-1.05		8.3	20.8	20.8	0.0		8.3	4.2	4.2	45.8	
EMDE Fuel Importers	0.07	0.08	-0.05	-0.65		10.5	12.3	9.6	6.1		1.8	7.9	16.7	28.9	
DASIA	0.11	0.11	0.07	-0.45		25.0	16.7	12.5	8.3		0.0	4.2	8.3	16.7	
EEUR	-0.05	-0.26	-1.03	-0.86		0.0	0.0	0.0	0.0		0.0	22.2	66.7	22.2	
LAC	0.13	0.14	0.16	-0.56		9.4	15.6	6.3	3.1		0.0	12.5	21.9	34.4	
MENAP_CIS	0.06	0.11	-0.04	-1.29		0.0	6.7	0.0	0.0		3.3	0.0	10.0	60.0	
SSA	0.10	0.19	0.07	-0.32		11.6	18.6	25.6	9.3		7.0	7.0	4.7	20.9	
Program	0.08	0.08	-0.05	-0.57		10.3	13.4	10.3	5.2		3.1	9.3	21.6	24.7	
Correlation to Euro Area Growth Forecast															
World	0.08	0.08	1.12	1.97		7.1	2.6	27.6	32.7		4.5	7.1	3.2	4.5	
Advanced Economies	0.15	-0.35	0.13	0.17		5.9	0.0	11.8	11.8		11.8	5.9	0.0	0.0	
Emerging Market Economies	0.18	0.19	1.79	2.58		6.0	2.4	38.6	39.8		1.2	2.4	0.0	2.4	
Low-Income Economies	-0.01	-0.11	0.69	2.05		8.9	3.6	16.1	28.6		7.1	14.3	8.9	8.9	
EMDE Fuel Exporters	0.00	0.34	2.34	6.97		4.2	0.0	50.0	58.3		4.2	4.2	0.0	0.0	
EMDE Fuel Importers	0.08	0.08	1.11	1.91		7.8	3.5	25.2	30.4		3.5	7.8	4.3	6.1	
DASIA	0.15	0.10	1.19	1.96		16.0	4.0	24.0	48.0		0.0	8.0	0.0	4.0	
EEUR	0.11	-0.05	0.82	-0.10		0.0	0.0	0.0	0.0		0.0	0.0	0.0	11.1	
LAC	0.21	0.16	1.52	2.43		6.3	3.1	31.3	40.6		0.0	3.1	3.1	3.1	
MENAP_CIS	0.09	0.50	2.43	5.06		6.7	6.7	60.0	46.7		6.7	3.3	0.0	0.0	
SSA	-0.06	-0.33	0.56	1.98		4.7	0.0	16.3	23.3		7.0	14.0	9.3	9.3	
Program	0.07	0.07	0.96	1.97		8.7	3.3	21.7	26.1		3.3	6.5	4.3	5.4	

Note: The calculations are based on country-by-country regressions of growth forecast errors on the US / China / Euro Area growth forecast for the same horizon: EMDE = Emerging Market and Developing Economies; DASIA = Developing Asia; EEUR = Emerging Europe; LAC = Latin America and the Caribbean; MENAP_CIS = Middle East, North Africa, Afghanistan, Pakistan, and Commonwealth of Independent States; SSA = Sub-Saharan Africa; Program = IMF program countries.

Table 6. Correlations between Growth Forecast Errors and Terms-of-Trade Forecasts

	Median coefficient on terms-of-trade forecast				Percent of countries with statistically significant positive coefficients				Percent of countries with statistically significant negative coefficients			
	$(h=0, F)$	$(h=1, F)$	$(h=2, F)$	$(h=5, F)$	$(h=0, F)$	$(h=1, F)$	$(h=2, F)$	$(h=5, F)$	$(h=0, F)$	$(h=1, F)$	$(h=2, F)$	$(h=5, F)$
World	-0.04	0.09	-0.31	-1.21	4.7	7.1	10.6	8.8	6.5	6.5	12.9	22.4
Advanced Economies	1.51	1.85	0.52	-2.16	11.4	5.7	11.4	11.4	5.7	0.0	5.7	17.1
Emerging Market Economies	-0.41	-0.82	-0.72	-2.15	1.3	6.3	10.0	7.5	10.0	11.3	18.8	27.5
Low-Income Economies	-0.04	0.09	-0.03	-0.30	5.5	9.1	10.9	9.1	1.8	3.6	9.1	18.2
EMDE Fuel Exporters	-0.91	-0.54	-3.17	-3.66	4.3	8.7	4.3	4.3	0.0	8.7	17.4	39.1
EMDE Fuel Importers	-0.05	-0.06	0.00	-0.66	2.7	7.1	11.6	8.9	8.0	8.0	14.3	20.5
DASIA	-0.11	-0.39	0.08	-1.69	4.5	4.5	9.1	22.7	0.0	9.1	9.1	13.6
EEUR	0.92	0.11	1.24	-1.31	0.0	0.0	12.5	0.0	0.0	0.0	0.0	25.0
LAC	-0.16	0.05	-0.38	-0.44	0.0	9.4	12.5	0.0	9.4	9.4	15.6	25.0
MENAP_CIS	-0.92	-0.89	-2.14	-2.22	6.7	6.7	13.3	10.0	13.3	13.3	30.0	40.0
SSA	-0.04	0.03	-0.36	-0.64	2.3	9.3	7.0	7.0	4.7	4.7	9.3	16.3
Program	-0.04	0.29	-0.17	-0.49	3.2	10.8	11.8	6.5	6.5	8.6	15.1	20.4

Note: The calculations are based on regressions of growth forecast errors on the terms-of-trade forecast for the same country and forecast horizon. EMDE = Emerging Market and Developing Economies; DASIA = Developing Asia; EEUR = Emerging Europe; LAC = Latin America and the Caribbean; MENAP_CIS = Middle East, North Africa, Afghanistan, Pakistan, and Commonwealth of Independent States; SSA = Sub-Saharan Africa; Program = IMF program countries.

Table 7. Comparison of Accuracy of Consensus Economics and WEO Forecasts

	March & September				April & October			
	(h=0, F)	(h=0, S)	(h=1, F)	(h=1, S)	H=F, 0	H=S, 0	H=F, 1	H=S, 1
Median	-0.03	-0.05	-0.01	0.03	0.08	0.02	0.03	0.05
25th percentile	-0.12	-0.12	-0.08	-0.03	-0.01	-0.08	-0.05	-0.03
75th percentile	0.06	0.04	0.04	0.18	0.22	0.12	0.12	0.21
Percent of countries where the WEO forecast is more accurate	61.2%	65.9%	56.5%	40.0%	27.1%	47.1%	38.8%	34.1%
Albania	-0.08	-0.15	0.32	1.27	-0.15	-0.15	0.13	1.08
Argentina	0.04	-1.02	-0.88	-0.97	-0.41	-2.20 *	-2.71 *	-2.12
Armenia	2.18	-0.06	0.07	0.91	0.87	0.04	-0.63	1.05
Australia	0.00	-0.14	-0.03	-0.37	0.02	-0.17	0.00	-0.36
Austria	0.04	-0.03	0.05	-0.24 **	-0.03	-0.10	0.09	-0.13
Azerbaijan	-0.72	3.46	1.38	2.03	-1.60	4.13	-0.54	3.98
Bangladesh	0.14	0.15	0.17	0.43 **	0.12	0.11	0.16	0.39 **
Belarus	0.11	1.86	6.27	0.31	0.27	0.34	6.28	0.77
Belgium	0.05	0.05	0.24 **	-0.10	0.00	0.02	0.04	0.05
Bolivia	0.36 **	0.50 **	0.62	0.57	0.31	0.53 **	0.67	0.64
Bosnia and Herzegovina	-0.25 **	0.34	-0.68	0.22	-0.37 **	0.17	-0.91	-0.38
Brazil	-0.02	0.35	-0.10	-0.52	-0.09	-0.05	-0.68	-0.90
Bulgaria	-0.09	-0.32	-0.99	0.17	-0.16	-0.38 *	-1.13	0.03
Canada	-0.01	-0.01	0.06	-0.24	-0.01	-0.06	-0.03	-0.25
Chile	-0.13 **	0.03	-0.15	0.43	-0.22 **	-0.08	-0.25	0.28
China	-0.02	0.36 **	0.20	-0.14	-0.06	0.19	-0.01	-0.16
Colombia	0.16	0.35	0.20	0.10	-0.30 *	0.14	-0.11	-0.15
Costa Rica	1.03	0.10	-0.30	-1.82 **	0.76	0.08	-0.35	-1.79 ***
Croatia	-0.11	0.22	-0.44	0.76	-0.20 *	0.01	-1.11 *	0.27
Cyprus	0.44	-0.07	3.46	3.66	0.27	0.27	2.72	2.58
Czech Republic	-0.03	0.15	0.04	-1.01	-0.11	0.07	-0.11	-1.11 *
Denmark	-0.01	0.14	0.18	-0.11	-0.16 *	0.07	-0.07	-0.11
Dominican Republic	-0.28	-1.42	0.35	0.67	-0.10	-1.87 *	0.63	1.70
Ecuador	1.08	-0.24	-0.64	-3.51	-0.10	-0.61	-1.23	-3.29
Egypt	0.09 *	0.11	0.17	0.89	0.03	0.06	0.13	0.31
El Salvador	0.01	-0.07	-0.17	-0.27 **	-0.04	-0.05	-0.17	-0.25 *
Estonia	0.16	-0.05	-1.67	0.01	-0.07	-0.81	-1.86	0.27
Finland	0.50	0.36 *	0.13	-1.40	0.03	0.27	-0.18	-0.89
France	0.00	0.01	-0.04	-0.26 **	-0.02	0.01	-0.12	-0.22 *
Georgia	0.79	0.42	0.78	0.38	0.76	-0.20	0.39	1.19
Germany	-0.01	-0.19	-0.31	-0.87	-0.04	-0.20	-0.42	-0.79
Greece	0.02	-0.20	1.05	0.41	-0.28	-0.94	0.03	-1.15
Guatemala	-0.02	0.04	-0.01	-0.04	0.01	0.05	-0.14	-0.04
Honduras	0.11	0.00	0.05	-0.13	0.09 *	0.03	-0.01	-0.08
Hong Kong SAR	0.32	0.12	-1.01	-1.54	0.12	-0.14	-1.35	-1.59
Hungary	-0.24 **	-0.28	-0.49	-0.28	-0.23 **	-0.36 *	-0.96	-0.66
India	0.69	1.12	-0.40	0.42	0.51	0.42	-0.54	-0.43
Indonesia	-0.01	0.03	-0.03	-0.40	0.00	0.03	-0.11	-0.29
Ireland	0.72	1.19	1.54	0.22	0.06	0.92	1.17	0.28
Israel	0.41 **	0.06	0.47	-1.26	0.20	0.02	0.18	-1.10
Italy	0.05	0.19	0.04	-0.23	0.00	0.11	-0.14	-0.20
Japan	0.09	-0.09	0.47 **	0.05	-0.02	-0.23 **	0.39 **	0.05
Kazakhstan	0.07	-0.60	0.28	-1.69	-0.05	-0.95 *	-0.74	-1.32
Korea	0.10	-0.05	0.01	-0.84	0.06	-0.08	-0.24	-0.70
Latvia	0.37	-0.61	0.24	0.40	0.23	-0.81	0.54	1.03
Lithuania	0.19	0.98	-0.37	0.65	0.12	0.33	-0.06	0.79
Macedonia, FYR	0.42	0.34	0.14	-0.11	0.09	0.45	-0.16	-0.32
Malaysia	-0.08	-0.15	-0.90	-1.53	-0.12	-0.27	-0.92	-1.33
Mexico	0.06	0.17	0.50 **	-0.40	-0.02	0.04	0.45 **	-0.18
Moldova	0.02	2.64	1.60	-4.91	-1.53	1.55	0.12	-4.37
Netherlands	0.00	0.06	0.03	0.05	0.01	0.04	-0.06	0.14
New Zealand	-0.12	-0.17	0.03	-0.10	-0.23 *	-0.23	-0.06	-0.11
Nicaragua	0.16	0.38	-0.51	-0.57	-0.01	0.28	-0.41	-0.62
Nigeria	-0.09	-1.31	1.29	0.05	-0.71	1.64	-0.44	0.18
Norway	0.53 **	0.15	0.30	0.39 *	0.52 **	0.14	0.26	0.40 **
Pakistan	-0.39	0.28	-0.31	0.12	-0.41	0.23	-0.26	0.31
Panama	0.81	1.32	3.40 **	2.29 **	-0.11	0.96	2.69 **	2.07 *
Paraguay	3.82	9.05	9.51	3.21	2.97	8.44	8.03	3.48
Peru	0.06	0.52	0.81	0.86	-0.23	-0.14	0.53	0.03
Philippines	0.04	0.04	0.01	-2.19	-0.14	0.02	-0.13	-2.30
Poland	0.03	0.07	0.20	-0.28	-0.02	0.04	0.10	-0.26
Portugal	0.13	0.11	0.27	0.20	0.04	0.15	0.11	-0.01
Romania	-0.07	0.45	0.18	-0.84	-0.12	0.20	0.28	-1.15
Russia	0.17	-0.20	-0.57	-1.86	0.03	-0.34	-1.15	-2.04
Saudi Arabia	0.03	-0.28	0.15	-0.12	-0.06	-0.68	0.13	-0.28
Serbia	0.17	-0.01	1.03	-0.19	0.01	-0.13	0.22	-0.46
Singapore	0.22	-0.48	-1.23	-7.78	-0.48	-1.06	-2.08	-7.43
Slovak Republic	-0.03	0.36	0.29	-0.80 *	-0.04	0.26	0.11	-0.42
Slovenia	0.34	0.82	-1.68	0.58	-0.09	0.45	-2.02	0.81
South Africa	-0.03	0.17	0.51	0.36	-0.13 **	0.02	0.15	0.29
Spain	-0.01	0.14	-0.21	-0.36	-0.02 *	0.09	-0.32	-0.40
Sri Lanka	-0.01	0.15	-0.11	-0.49	-0.14	-0.12	-0.33	-0.46
Sweden	0.11	-0.30	-0.85	-0.91	-0.08	-0.41	-1.13	-0.92
Switzerland	-0.12 *	0.02	0.01	-0.24	-0.10	-0.06	-0.09	-0.17
Taiwan Province of China	0.46	1.41	-0.38	-4.43	0.01	1.06	-0.92	-4.40
Thailand	0.23	0.14	-0.91	-4.04	0.09	-0.14	-1.01	-3.61
Turkey	0.84	0.65	-0.09	-3.00	0.12	-0.79	-0.49	-2.97
Turkmenistan	-1.40	-3.96 *	-2.21	-2.01	-0.54	-4.02 *	-4.54	-1.40
Ukraine	-0.61	1.01	0.24	0.99	-0.95	0.10	-0.69	0.57
United Kingdom	0.02	0.02	0.05	-0.37 *	-0.01	0.02	-0.13	-0.32
United States	0.02	0.08	-0.12	-0.57	-0.04	-0.01	-0.15	-0.63
Uruguay	0.29	4.01	1.84	0.71	0.12	2.53	2.37	0.79
Uzbekistan	-0.13	0.21	-0.11	-0.02	-0.13	0.22	-0.43	0.32
Venezuela	1.62	5.60	7.83	20.51	-0.72	6.03	6.23	18.32
Vietnam	-0.03	0.25	-0.07	-0.52	-0.04	-0.02	-0.25	-0.53

Note: The first four rows show the quartiles of the U -statistic and percent of negative U -statistic for different forecast horizons. The remaining rows show Diebold and Mariano (1995) test statistics.