Macroeconomic Challenges Facing
Low-Income Countries
New Perspectives

January 30–31, 2014

Output Data Revisions in Low-Income Countries

Eduardo Ley
The World Bank

Florian Misch
Centre for European Economic Research

Paper presented at the joint RES-SPR Conference on "Macroeconomic Challenges Facing Low-Income Countries"
Hosted by the International Monetary Fund
With support from the UK Department of International Development (DFID)
Washington, DC—January 30–31, 2014

Preliminary

Output Data Revisions in Low-Income Countries

Eduardo Ley
The World Bank, Washington DC, USA

Florian Misch
Centre for European Economic Research, Mannheim, Germany

Version: January 17, 2014 — 7:05 P.M.

Abstract: Reliable information on changes and the level of economic activity are central for macroeconomic analysis, but preliminary output data available when policy decisions are made are typically revised as more and better information become available. There is a large literature on developed countries documenting such revisions and discussing its implications, but evidence from developing, notably low-income economies, is scarce. The objective of this paper is therefore to analyze the nature, causes and policy implications of output data revisions, $i.e.$, the discrepancy between forecasts, nowcasts as well as 'early' backcasts of real GDP growth and the level of nominal GDP in low-income countries. First, using data from the IMF's World Economic Outlook, we show that output data revisions in low-income countries are, on average, larger than in other countries, and that they are much more optimistic. Second, we analyze the determinants of output data revisions and show that economic and technical factors drive revisions, rather than factors related to the political economy of the IMF itself as commonly assumed.

Keywords: GDP revisions, output gap revisions, growth revisions

JEL Codes: E01, E27

This working paper is part of a research project on macroeconomic policy in low-income countries supported by the U.K.s Department for International Development (DFID). I thank Laurence Ball, Andrew Berg, Raphael Espinoza, Camelia Minoiu, and Chris Papageorgiou, as well as workshop participants of a workshop at the IMF in July 2013 for excellent comments. We also thank Toh Kuan for facilitating access to the data. Financial support from the IMF and DFID is gratefully acknowledged. Eduardo Ley passed away on July 1st, 2013, before the completion of the paper. I remember him as an excellent researcher, manager and a friend.

1. Introduction

Reliable and timely measures of the level and the change of economic activity are central for macroeconomic policy in low-income countries, much like in any other countries. However, an important problem of output data relates to revisions over time, *i.e.*, that real-time estimates of output growth and output levels get revised as better and more information becomes available. As macroeconomic policy decisions are inevitably based on real-time output data, such output data revisions are problematic. The objective of this paper is to examine output data revisions in low-income countries relative to other countries and their determinants.

Correctly measuring output levels and output growth in low-income countries is challenging. One problem in compiling reliable GDP statistics relates to informal and therefore hidden economic activity. This is a problem in all countries, although the size of informal activity is likely to be larger in low-income countries. Henderson et al. (2012) for instance therefore propose to use luminosity as measured from outer space as an indicator of 'true' economic activity. In addition, it may be argued that GDP which is the most widely used measure of economic activity is incomplete, notably in resource-rich, low-income countries, irrespective of criticism that it poorly reflects social well-being. Hamilton and Ley (2010) argue that in countries with significant natural resources, GDP should be adjusted for net foreign factor income, depreciation of capital and notably for depletion of natural capital yielding the adjusted net national income as a complementary measure of economic progress. Jerven (2012) suggests that GDP data in Sub-Saharan Africa are not reliable based on anecdotal evidence. One recent case of GDP level revisions that has attracted much public attention is the rebasing in Ghana, shifting it to middle-income status overnight.

In this paper, we focus on another aspect that undermines the reliability of output data, at least in real time, namely revisions to output growth and its level; *i.e.*, the discrepancy of output figures for particular years and countries across different data releases or vintages. To provide an example of such revisions in developing countries, consider the case of Malawi: according to estimates made by the IMF in spring 2002, growth was projected to be over 3% in same year, but this figure was subsequently downward revised by 1.3 percentage points.

Given that such revisions are of economic significance, their analysis of revisions have become a standard element of macroeconomic research on developed countries; see Croushore (2011) for a summary of the literature. Output data revisions may entail adverse, tangible economic consequences in all countries. Fiscal and monetary policy decisions are inevitably based on preliminary output data so that policy mistakes occur if revisions are significant. This may, for instance, result in unplanned and significant public debt accumulation as suggested by Easterly (2012) and Ley and Misch (2013).

However, much of the existing literature focuses on mature economies, and evidence is scarce for LICs, even though it seems plausible that in LICs, output growth and level revisions are larger. On the one hand, overall data quality is often poor which is likely to permeate to the reliability output data in real time as well, due to, for example, weak statistical capacity in public administration. On the other hand, some other, inherent features of LICs including in particular greater vulnerability to various types of shocks make output data revisions more likely. Output data revisions are not only likely to be larger in LICs; their consequences may also be more severe. For instance, given that social safety nets are largely absent, overoptimism in growth forecasts may be particularly harmful as remedial measures, for instance to protect the poor, cannot be taken even if sufficient public resources were available. Even growth surprises leading to a boost in public revenue may be more difficult to deal with, given that fiscal management capacities of LICs tend to be weaker compared to other countries.

Our data on output data revisions come from the IMF's World Economic Outlook comprising over 20 different vintages and a large number of countries and information for real output growth and nominal GDP in levels. Apart from the cross-country comparability of these data, their central advantage is that it is likely to be one of the best source of information, and that compared to other data from national sources, political factors resulting biases are likely to be less important compared to government GDP forecasts for instance. Another advantage of WEO is that it is of key importance for macroeconomic policy making in LICs contrary to, for example, the Penn World Tables (PWT) which are a primary data source for economic research.[1] Finally, WEO data allows us to study the revisions of forecasts, nowcasts and backcasts which are likely to differ in terms of their reliability and which are needed for different purposes in the context of policy making.

This paper makes two contribution. First, we show that LICs perform worse across most dimensions of output data revisions. The mean absolute error is more than twice as high as in OECD countries, overoptimism in preliminary growth figures is more severe than in other countries, and extreme revisions occur much more frequently. These results also hold across most release dates and apply to forecasts, nowcasts and backcasts, with only few exceptions. Contrary to existing papers, we also document revisions to same-year estimates of GDP in levels which can be seen as the accumulated effects of revisions to growth and inflation over several years.

Second, we examine what drives output data revisions, both across all countries and within LICs only. We consider three different sets of possible explanations: economic factors mostly related to shocks and growth volatility, the capacity of the national statistical services and political economy-related factors. While unsurprisingly, economic factors contribute significantly to output data revisions, we also show that upgrading the capacity of the national statistical services and increased IMF surveillance lower the extent of revisions. By contrast, our evidence does not robustly support various existing hypotheses that the IMF purposely manipulates preliminary growth figures for political reasons. Finally, when we control for these other factors, differences in terms of the bias in preliminary output data between low-income countries and other countries are mostly not significant anymore.

[1] Revisions of GDP data for a large number of countries have also been documented in the PWT by Johnson et al. (2013) and Ponomareva and Katayama (2010). The former paper reports that some well known results from the cross-country growth literature are not robust to changes in the vintage of the underlying PWT data. They attribute revisions of output data for a given country and year to updated information from the underlying national income accounts data (which also drives revisions to data in WEO), but also to other factors inherently related to the way PWT data are compiled. This means that contrary to WEO, a priori, there is ambiguity about whether PWT data from newer releases are better. Ponomareva and Katayama (2010) show that revisions in the PWT tend to be smaller in high-income OECD (HIC-OECD) countries.

We extend and update the existing literature. Timmermann (2007) is most closely related to the first part of our paper. He reports few (though to some extent different) statistics on growth revisions and their differences across geographical regions using WEO data as well.[2] The second contribution extends the results by Aldenhoff (2007) and Dreher et al. (2008). Both papers use WEO data to test several alternative political economy-related hypotheses about the determinants of forecast errors, but they ignore any other reasons about why GDP estimates may be incorrect.

Our results have important policy implications. First, preliminary output figures should be treated with caution, and when they are used for policy purposes in LICs, significant error margins should be taken into account depending on the characteristics of the country. Second, output data revisions are not exogenously given but, at least to some extent are also dependent on the statistical capacities of the national services in in question. This implies that investments in statistical capacity, in particular to meet certain international data dissemination standards, may pay off in the sense of lowering the magnitude of output data revisions.

The paper is organized as follows. The next section briefly discusses the concept of revisions to output growth and the level of nominal GDP. Section 3 presents the data used. Section 4 provides descriptive statistics on output data revisions, whereas Section 5 presents evidence on their determinants. Section 6 concludes.

2. Conceptual Framework

While many macroeconomic variables get revised over time, in this paper, we focus on revisions of real output growth and output levels in nominal terms, *i.e.*, the discrepancy between figures according to the latest releases and preliminary figures released in real time. Let $x_{i,t,s}$ denote output for country i, year t, as seen from vintage / according to

[2] Preceding Timmermann (2007), there are several older papers mostly from the 1990s and 1980s performing similar exercises using WEO data; see for instance Artis (1988, 1997) and Barrionuevo (1993). Some other, more recent papers which study revisions in WEO data do not consider developing countries and are hence not relevant for this paper; see for instance Batchelor (2000), Beach (1999), Loungani (2000) and Pons (2000).

the data release date s. The law of motion identity is

$$x_{i,t,s} = (1 + \gamma_{i,t,s}) \cdot x_{i,t-1,s} \tag{1}$$

where $\gamma_{i,t,s}$ denotes the rate of output growth as estimated at release date s. Output in nominal terms, $v_{i,t,s}$, grows at

$$v_{i,t,s} = (1 + \gamma_{i,t,s}) \cdot (1 + \pi_{i,t,s}) \cdot p_{i,t-1,s} \cdot x_{i,t-1,s} \tag{2}$$

where $\pi_{i,t,s}$ denotes the rate of inflation in t as estimated at release date s where p denotes the price level.

We measure revisions of output growth, real output levels and nominal output levels, respectively, as

$$\varepsilon^{\gamma}_{i,t,s} = (\gamma_{i,t,s^*} - \gamma_{i,t,s}) = \text{actual} - \text{predicted}$$

,

$$\varepsilon^{x}_{i,t,s} = \frac{(x_{i,t,s^*} - x_{i,t,s})}{x_{i,t,s}} = \frac{\text{actual} - \text{predicted}}{\text{predicted}}$$

and

$$\varepsilon^{v}_{i,t,s} = \frac{(v_{i,t,s^*} - v_{i,t,s})}{v_{i,t,s}} = \frac{\text{actual} - \text{predicted}}{\text{predicted}}$$

where s^* is the latest release date and where $s^* - t \geq 5$. This reflects our assumption that final data become available 5 years later which we explain in the next section in greater detail. We limit our analysis to s ranging from spring in $t-1$ to fall in $t+3$.

We distinguish forecasts, nowcasts and backcasts. Forecasts of GDP in t are made prior to t, nowcasts are released in the year they refer to, and backcasts are released in years after t. Forecasts prior to year t and nowcasts, in particular those made in spring of year t, are probably most relevant for macroeconomic policy. If the level of economic activity could be reliably assessed from this perspective of the first quarter, then corrective measures in the area of macroeconomic policy could perhaps be implemented to re-direct fiscal policy for the remaining year. By contrast, revisions of backcasts are mainly important for ex-post evaluation of economic policy and economic research; for instance, if ex-post evaluation uses backcasts subject to revisions, the conclusions may be misleading if actual GDP figures significantly deviate from preliminary ones.

Data releases may also be distinguished in terms of their economic content, *i.e.*, whether they are based on estimates or on actual data. Revisions to GDP growth estimates obviously occur as new information becomes available, but methodological changes, at least in principle, also play a role. Forecasts are necessarily based on estimates only and may obviously revised due to unforeseen events or shocks; nowcasts are also typically estimates, but which may also take into account some additional information. By contrast, at the time when backcasts are made, all relevant information could, at least in principle, be already available. Under this view, forecasts and nowcasts should be generally less reliable than backcasts. In practice however, 'early' backcasts made in $t+1$ are often still estimates implying that they are also highly preliminary, especially in the case of low-income countries. Our stylized facts which we present below confirm this: while later backcasts generally improve in terms of accuracy, revisions may still be economically significant. However, revisions of backcasts may also be driven by methodological changes such as rebasing.[3]

This simple framework implies that revisions to the level of output are driven by revisions to nowcasts and backcasts of growth. In other words, considering revisions to GDP levels is informative as they reflect the accumulated effects of growth revisions. Revisions to nominal GDP are also driven by revisions to inflations nowcasts and backcasts. In practice, an important source of revisions comes from rebasing which is not explicitly captured by this framework. Revisions to nominal GDP matter in practice as fiscal aggregates are often measured in terms of GDP; such ratios are affected by changes to the denominator.

3. Data

[3] Rebasing may only affect nominal GDP if the level is lifted 'upwards' with the relative change left constant. Every vintage for a given country from WEO is internally consistent in the sense that for instance, figures of all years within a given vintage use the same currency and the same base year.

3.1. Real-time and Final WEO Data

Our data on output data revisions come from the IMF World Economic Outlook; they contain real growth rates and nominal GDP in levels released in spring and fall from 1990 to 2012—*i.e.*, real-time output figures from 24 different vintages where we consider the one from fall 2012 as the one that contains final figures, with some exceptions (as for some countries, there were no output figures released in fall 2012). Every vintage contains GDP growth rates and nominal GDP level figures for past years, the concurrent year and for up to five years in to the future. The revisions/errors are then calculated as the difference between actual outcomes as measured by the most recent vintage (the fall 2012 estimate) and the real-time figures. We assume that final estimates are available five years later, so the 1990–2012 data allow us to study the reliability of the 1990–2007 concurrent estimates. Using data on exchange rates from the WEO, we correct revisions of GDP levels that occur due to currency reforms.

There are obviously other sources of GDP data including in particular the Penn World Tables and the World Bank's World Development Indicators. While according to Johnson et al. (2013), these latter two sources are used more frequently in empirical cross-country research, here we argue that for most LICs, WEO data is probably the highest-quality data source available for policy making, and more timely than other sources. Information from WEO is the result of a comprehensive and systematic procedure. The country desks, in consultation with country governments and other observers, submit GDP figures to the WEO division. The WEO division makes sure that 'the pieces fit in', checking the compatibility of the GDP figures between countries that have significant trade, or share significant trade partners. Several iterations with individual desks may occur before it is settled on the published WEO (spring and fall). In addition, compared to government GDP forecasts, WEO data are likely to be less affected by political interference.[4] Finally,

[4] There are no systematic differences between WEO data and other sources that are deemed reliable; Timmermann (2007) and Abreu (2012) for instance suggest that the quality of consensus forecasts and WEO forecasts is similar. Irrespective of the relatively high quality of WEO data, Aldenhoff (2007) and Dreher *et al.* (2008) still find evidence of political interference, and recently, Blanchard and Leigh (2013) show that during the global financial crisis, WEO forecasters underestimated the magnitude of fiscal multipliers.

release dates are identical across countries which enables cross-country comparisons of GDP data revisions, contrary to GDP data from national sources where release dates and therefore the informational content of given vintages may differ.

Our data cover 175 countries. While we do not discard any outliers to understand the full spectrum of output data revisions, we omit three countries where output figures have not been revised at all including Afghanistan, Liberia and Somalia, probably because original figures were never re-assessed. While obviously, each vintage also contains output figures reaching back up to the 1960s, we ignore all output figures for the time prior to 1990 as our data do not contain and data releases from this period (and thereby would not allow us to study revisions). Table 1 summarizes the number of observations for each release date, the share of observations referring to LICs and the notational and economic categorization of different data releases.

Table 1. Vintage classification and Sample

(175 countries: 1990-2007; $N = 28,303$)

Release of data		Notational term	Economic content	no. of	% of LICs
Year	Season	of release	of release	obs.	
t-1	spring	forecast	estimates	2,865	20
t-1	fall	forecast		2,867	20
t	spring	nowcast	estimates	3,040	20
t	fall	nowcast		3,042	20
t+1	spring	backcast		2,894	20
t+1	fall	backcast		2,896	20
t+2	spring	backcast	historical / actual	2,748	20
t+2	fall	backcast	data	2,749	20
t+3	spring	backcast		2,600	20
t+3	spring	backcast		2,602	20
Total number of observations				28,303	20

Year t is the year the output figures refer to. For many countries, backcasts may still be estimates.

3.2. Exogenous Variables

In order to examine the determinants of output data revisions, we collect information on various control variables which may determine the magnitude and nature of output data revisions. We group them into four categories, economic factors, political factors, technical capacity, and other dummy variables reflecting the decade and whether the country classifies as a LICs. Table 2 provides an overview of the control variables, including a short description.

First, we include in our dataset several variables indicating whether a particular country has suffered, or was at least susceptible, from adverse and unforeseen shocks emanating from conflicts, natural disasters or, at least potentially, natural resource price movements. In addition, we include indicators about the ease to which growth in a particular country can be estimated; we hypothesize that this is affected by both growth volatility in a country over the period under consideration and whether there as a cyclical turning point in a particular year and country.

Second, the statistical capacity of countries may determine their ability to provide accurate information in a timely manner to the IMF. We therefore include information on whether and since when a particular country in a given year adhered to one of two data dissemination standards defined be the IMF, the General Data Dissemination System (GDDS) or to the Special Data Dissemination Standard (SDDS), for publically provided statistics and data. We expect that both, adherence to the SDDS and to the GDDS, imply that the statistical services of the countries are more apt to estimate GDP in real time; as WEO data is obviously also based on information from national authorities, we expect that adherence reduces revisions as well. The difference between both standards is that requirements of the SDDS are more demanding so that, generally speaking, it is almost only followed by more advanced countries. Due to data limitations which we explain in more detail in the Appendix, we are unable more specific indicators of national accounts data periodicity and timeliness of a given country.

As an additional indicator of statistical capacity, we use the log of the population as an indicator of the size of the country. We argue that due to economies of scale in the provision of national statistics, data dissemination is relatively less costly for larger countries. Dreher et al. (2008) include log GDP in their regressions as an indicator of country size, albeit for other reasons. They argue that governments tend to prefer optimistic forecasts, and they argue that the level of GDP correlates with the probability that WEO projections are manipulated in the favor of the member country government. Johnsons et al. (2013) also find that the smaller the country, the larger are the revisions. However, as noted in their paper, in the case of the PWT, this result reflects a particular aspect of the PWT methodology for calculating PPP prices and therefore does not relate to our analysis.

Third, political factors may affect the accuracy of real-time output data. While the IMF is much more technocratic than say national governments, it is subject to various political constraints as well. For instance, there may be an optimism bias if countries are in recession because it there may frequently the believe that the recession will soon be over. Alternatively, if a country has currently an IMF program, there may also be overoptimism as obviously, the IMF has to believe in the benefits from its 'own medicine'. Along the same lines, when an IMF program is agreed, WEO forecasts may be overly pessimistic as there needs to be a rationale for starting a program. Similarly, Aldenhoff (2007) and Dreher et al. (2008) hypothesize that the IMF purposely biases its growth estimates as a means to as a means to defend or justify its lending. We therefore include in our dataset information on whether coinciding with the release year of a particular vintage of WEO data, an IMF program was in effect for at least 5 months, or whether an IMF program has been agreed on at that time. In addition, we include a dummy indicating whether a country was in recession according to final GDP data (*i.e.*, suffered from negative growth). This may for instance affect the overall 'sentiment' of GDP forecasters.

Fourth, we generate several dummy variables controlling for the vintage (*i.e.*, the release date of the data), a '90s' dummy controlling for whether there was a change over time in the quality of preliminary output figures, and a dummy indicating whether the country in question is a low-income country

4. Stylized Facts of Output Data Revisions

This section documents revisions to real GDP growth and to the level of nominal GDP. Figure 1 plots the spring nowcast of growth, *i.e.*, concurrent growth estimates versus final estimates across all countries. It shows that in general, deviations from the diagonal are frequent suggesting that real-time and final estimates often diverge. Table 3 provides more

Table 2. Control variables

Category	Variable	Description
Economic	conflict	intra-/or inter-state conflict (dummy)
	resources	resource rents > 20% of GDP (dummy)
	volatile	Growth volatility (std. dev. of final growth)
	tphp	Turning point (dummy, based on HP filtering)
	disaster	Occurance of major disaster (dummy)
Political	imf	IMF program agreed (dummy)
	imf5	IMF program in effect (at least 5 months; dummy)
	recession	estimated neg. growth at time of data release (dummy)
Technical capacity	gdds	Subscription to GDDS (dummy)
	sdds	Subscription to SDDS (dummy)
	quarterly	Quarterly GDP series (dummy)
	population	Log of population (in millions)
Other dummies	p1990	1990s (dummy)
	lic	Low-income countries (dummy)

details and considers the difference between final and preliminary growth rates released in different vintages in percentage points across different country income groups. The first panel refers to preliminary estimates for t made in spring of $t-1$, the second to estimates made in spring of year t and the third panel to spring of year $t+1$ (for reasons related to space, we omit the fall estimates). Going from left to right in each row, we first present the percentiles, we then present the mean (which coincides with the bias in the preliminary growth figures including whether it is significant), the standard deviation, and the mean absolute revision. The latter is informative as positive and negative revisions may, at least to some extent, cancel each other out, lowering the mean.

Overall, preliminary growth figures generally improve with time, *i.e.*, later figures are more accurate than earlier ones for a given year. However, there is significant heterogeneity across country income groups. Absolute revisions are, on average, largest in high-income countries which are not members of the OECD (see last column). One reason is that these countries are either large natural resource exporters and therefore vulnerable to price movements and other countries of this groups recovered from conflicts. However, this result is driven by overpessimism, rather than overoptimism, which is arguable less harmful. By contrast, in LICs, mean absolute revisions are also relatively high on average, but here, there is a large optimism bias for forecasts and nowcasts contrary to all other country income groups.

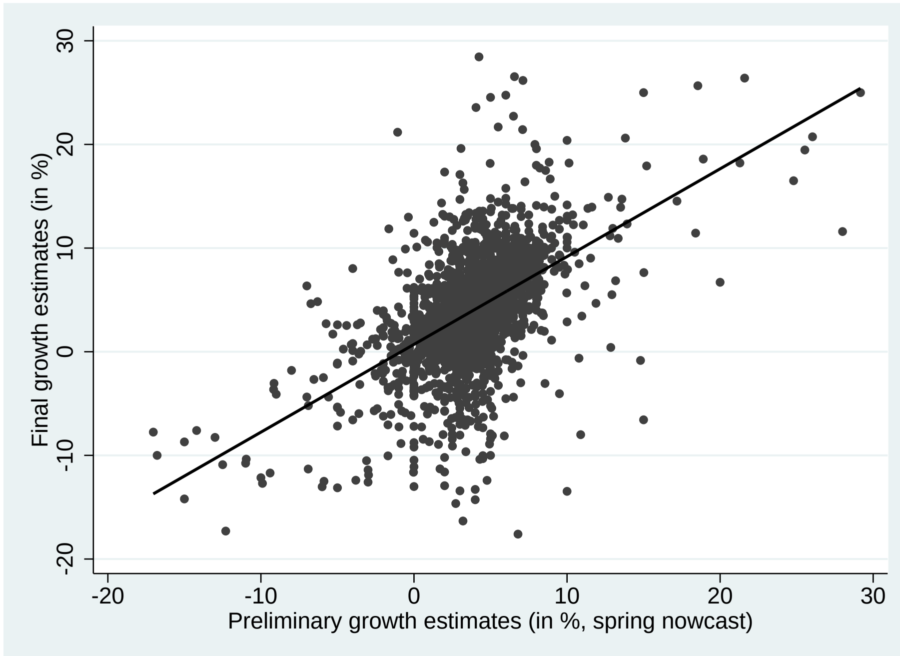

Fig. 1. Scatter Plot of the Growth Rates: Concurrent vs Final (169 countries: 1990-2010)

Looking at the percentiles, Table 3 implies that in LICs, in 10% of the cases, the actual growth outcome was almost 8% below the forecasted one (compared to just over 2.1% in HIC-OECD countries), and the mean absolute error (MAE) of LICs exceeds the MAE of HIC-OECD countries by a factor of 2 to 3 depending on the release date considered.

Specific country examples are useful to illustrate the magnitude of output data revisions and their differences across country income groups. As mentioned above, Malawi's growth estimate for 2002 made in spring of the same year was downward revised by 1.3% from around 3% to just below 1.7%. By contrast, in the Comoros, growth h was revised upwards by 1.3 percentage points, from around 3%. Such downward and upward revisions of equal or greater magnitude occur frequently in LICs, namely each in more than 25% of the cases (see values for 25th and 75th percentiles in LICs for spring nowcasts). While in HIC-OECD countries, chances that upward revisions of at least this magnitude occur are only slightly lower (as an indication, see value for 75th percentile in HIC-OECD countries of spring nowcast), downward revisions of 1.3 % are much rarer, and only occur in around 10% of the cases (not shown in Table 3).

Table 4 displays correlation coefficients between preliminary and final growth rates

Table 3. Growth Revisions: LICs vs. Other Countries

(175 countries: 1990-2007)

Release	Country group	Percentiles					Moments		Mean abs.
		10	20	50	75	90	Mean	StdDev	error (MAE)
fall of t-1	High income: OECD	-2.12	-0.95	0.19	1.23	2.32	0.08**	2.17	1.51
	High income: nonOECD	-4.60	-1.62	0.67	2.86	5.62	1.25*	11.63	4.58
	Upper middle income	-5.09	-2.27	0.34	2.57	4.80	0.04	4.59	3.25
	Lower middle income	-4.53	-1.80	0.08	1.90	4.48	-0.21**	6.90	3.29
	Low income	-7.95	-3.24	-0.56	1.33	4.32	-1.36***	5.77	3.76
	All countries	-4.78	-1.83	0.07	1.86	4.18	-0.18**	6.35	3.20
fall of t	High income: OECD	-0.93	-0.26	0.41	1.07	1.89	0.46***	1.38	1.02
	High income: nonOECD	-2.81	-0.92	0.78	2.36	6.25	0.90**	10.26	3.87
	Upper middle income	-3.35	-1.27	0.61	2.32	4.42	0.60***	3.62	2.54
	Lower middle income	-3.09	-1.13	0.40	1.85	3.74	0.36**	4.64	2.57
	Low income	-5.04	-1.92	0.00	1.45	4.05	-0.52**	5.07	3.02
	All countries	-3.12	-0.98	0.38	1.74	3.72	0.31***	5.07	2.51
fall of t+1	High income: OECD	-0.62	-0.13	0.30	0.76	1.20	0.31***	1.02	0.71
	High income: nonOECD	-1.86	-0.55	0.35	1.76	5.09	1.09**	7.20	2.79
	Upper middle income	-1.96	-0.41	0.28	1.36	2.97	0.45***	2.89	1.69
	Lower middle income	-1.79	-0.30	0.20	1.23	2.79	0.53***	3.54	1.68
	Low income	-2.46	-0.71	0.10	1.34	3.51	0.28*	3.51	1.98
	All countries	-1.73	-0.35	0.24	1.14	2.90	0.48***	3.68	1.69

Source: WEO data and own compilation. Negative forecast errors denote over-optimism; positive ones over-pessimism

* / ** / *** denotes significance at the 1% / 5% / 10% level

across different release dates and mirrors these differences between OECD countries and LICs. The coefficients of OECD countries exceed those of LICs by 0.1 to 0.2 depending on the release date. However, there appear to be no systematic patterns in terms of differences of the correlation coefficients across the remaining country income groups, although LICs always rank lowest or second lowest except for 'late' backcasts release in winter of $t + 3$.

Table 4. Correlations between preliminary and final growth

(175 countries: 1990-2007)

Country group	t-1		t		t+1		t+2		t+3	
	spring	winter	spring	winter	spring	winter	spring	winter	spring	winter
High income: OECD	0.43	0.51	0.69	0.84	0.90	0.92	0.92	0.94	0.94	0.94
High income: nonOECD	0.35	0.38	0.56	0.68	0.69	0.82	0.88	0.85	0.84	0.87
Upper middle income	0.34	0.42	0.60	0.73	0.82	0.85	0.86	0.88	0.88	0.89
Lower middle income	0.23	0.20	0.42	0.67	0.78	0.83	0.84	0.85	0.85	0.85
Low income	0.28	0.34	0.43	0.54	0.77	0.82	0.84	0.87	0.88	0.89
All countries	0.31	0.32	0.49	0.65	0.75	0.83	0.86	0.86	0.86	0.87

Source: WEO data and own compilation. Negative forecast errors denote over-optimism; positive ones over-pessimism

In Table 5, we then consider the occurrence of overoptimism and of extreme downward revisions of preliminary growth rates defined as revisions involving changes from positive to negative growth, *i.e.*, the sign of preliminary and final growth rates differ, and downward growth revisions exceeding 5% and 10%, respectively. Across all release dates which we consider, sign changes occur most frequently in LICs, and they occur more than twice more often than in OECD countries. In fact, the sign of forecasts is revised in 16% of the cases, and these revisions, amount, on average, to almost 10 percentage points implying that such revisions are not negligable in absolute terms either. Along the same lines, in LICs, downward revisions by more than 5 percentage points occur in more 1 out of 10 cases for nowcasts. This is twice as often as in the case of the remaining country income groups.

Table 5. Extent of Overoptimism

(175 countries: 1990-2007)

Vintage	Country group	Overoptimism (% of instances)			Sign change (⊕ to ⊖)	
		total	<-5%	<-10%	%	mean
Release in fall of t-1	High income: OECD	46.64	1.78	0.40	7.51	-4.15
	High income: nonOECD	42.18	8.16	2.04	10.20	-6.92
	Upper middle income	45.16	9.99	2.98	13.41	-7.17
	Lower middle income	48.21	8.63	2.47	9.74	-8.99
	Low income	58.80	16.41	6.84	16.07	-9.99
	All countries	48.76	9.28	3.07	11.55	-8.04
Release in fall of t	High income: OECD	31.66	0.56	0.00	2.98	-1.74
	High income: nonOECD	37.18	5.45	1.60	7.69	-5.02
	Upper middle income	38.76	4.21	0.70	7.44	-4.63
	Lower middle income	41.11	4.99	1.51	7.32	-6.13
	Low income	49.84	10.16	4.03	12.42	-8.24
	All countries	40.27	5.13	1.58	7.66	-6.07
Release in fall of t+1	High income: OECD	31.84	0.39	0.00	1.37	-2.02
	High income: nonOECD	36.82	2.03	0.68	5.07	-4.37
	Upper middle income	39.03	2.65	0.29	4.71	-4.77
	Lower middle income	36.74	2.43	0.61	3.77	-4.79
	Low income	41.23	4.09	0.68	6.13	-5.93
	All countries	37.33	2.42	0.45	4.18	-4.91

Source: WEO data and own compilation.

In Tables 6 and 7, we finally examine revisions of the level of real and nominal GDP of backcasts released in spring of year $t+1$. These revisions matter for policy as well as for instance fiscal aggregates are often expressed relative to the size of the economy. In addition, they embody the accumulated effects of growth revisions over previous years and the effects of inflation revisions (in the case of nominal GDP only).

In the case of revisions to real GDP (Table 6), the picture is somewhat more mixed than previously. While the chances of for example a 14% downward revisions amount to 10% and are much greater than in all other country income groups, there does not seem to be a bias, and the mean absolute error is smaller than in lower middle income economies and non-OECD high-income countries. In the case of nominal GDP revisions, the range as measured by the distance between the 10th and 90th percentile is greater in LICs than in all other income countries, but across all other dimensions, LICs do not rank last, but their performance is always worse than in the case of high-income OECD countries. In general, revisions to nominal GDP are very large, but their economic interpretation is not obvious, because these revisions are, at least in part, explained by the combined effect of growth and inflation revisions.[5]

Table 6. Revisions of Real GDP (in %)

(147 countries: 1990-2007)

Country group	Percentiles			Moments		Mean absolute
	10	50	90	Mean	StDev	error (MAE)
High income: OECD	-5.96	1.61	10.85	2.28	7.39	5.34
High income: nonOECD	-8.63	2.03	26.64	9.54	35.26	15.08
Upper middle income	-8.05	1.97	13.36	2.48	10.89	7.23
Lower middle income	-8.64	1.23	19.64	5.36	24.27	11.45
Low income	-14.46	-0.33	14.82	-0.02	13.21	8.97
All countries	-8.96	1.23	14.99	3.44	19.26	8.97

Source: WEO data and own compilation. Preliminary estimates from spring in t+1.

[5] In Table 6, we are forced to exclude shorter vintages to ensure comparability across different vintages. Note that in Table 7, to ensure comparability across different release dates, we only include countries which did not implement monetary reforms; to limit the percentage of observations we loose due to this exclusion criterion, we also only consider revisions from the 2000s.

Table 7. Revisions of Nominal GDP (in %)

(152 countries: 2000-2007)

Country group	Percentiles			Moments		Mean absolute
	10	50	90	Mean	StDev	error (MAE)
High income: OECD	-0.65	3.73	11.87	0.94	21.17	9.42
High income: nonOECD	-1.59	12.85	43.77	17.53	19.85	18.91
Upper middle income	-0.91	9.22	31.83	12.75	12.31	13.59
Lower middle income	-4.29	6.08	34.03	10.85	18.49	13.08
Low income	-3.68	6.29	42.91	14.42	22.26	16.66
All countries	-2.27	6.67	33.71	10.81	19.38	16.66

Source: WEO data and own compilation. Preliminary estimates from spring in t+1.

5. Econometric Results

In this section, we present evidence on the determinants of growth revisions. We first turn to the determinants of growth revisions in absolute (*i.e.*, the absolute error) which is the endogenous variable in the regressions presented in Table 8. Whereas specification (1) includes all release dates and, to control for differences between vintages, vintage fixed effects, specifications (2), (3) and (4) only include forecasts, nowcasts and backcasts of $t+1$, respectively (in each case, we do not differentiate between spring and fall releases to increase the number of observations). Specification (5) is identical to specification (3) except that we only include LICs. Throughout this section, we use heteroscedasticity-consistent standard errors.

Across all specifications, the 'economic' variables are highly significant in almost all cases and all have mostly the expected signs. Conflicts, growth volatility and cyclical turning points increase revisions. By contrast, the coefficient of 'resources' which we include as a measure of susceptibility to changes in natural resource prices is insignificant. This means that either natural resource price movements are well predicted, or that they do not have effects that are large enough over the period of consideration. In addition, disasters are consistently positive but not always significant possibly pointing to heterogeneity of the effects of disasters on GDP which we do not capture in the coding of the variable.

Technical capacity is also critical. Adhering to the data standards prescribed by the IMF significantly and robustly lowers the extent of revisions, but for forecasts, the effects are not significant. Larger countries tend to experience lower revisions; the coefficient is negative and significant in most specifications. Agreeing to an IMF program does not have a robust effect, but interestingly, undergoing an IMF program at the time when the preliminary growth figures are released lowers growth revisions. While the reason may relate to closer surveillance through the IMF, this effect is not always significant.

Finally, the LIC dummy is positive and significant throughout. This implies that the control variables do not capture the factors that lead to larger output data revision. There also appears to improvement in the accuracy of preliminary growth figures over time as the coefficient of the 1990s dummy is significant, except in the last specification (5) which only focuses on LICs.

In Table 9, we analyze the determinants of growth revisions (not in absolute terms). The results from these regressions shed more light on the underlying factors that are correlated with a positive or negative bias in preliminary growth figures for LICs, and what could, at least potentially, correct for this. The coefficients measuring the effects of turning points and of growth volatility are not robust or not significant. This is not surprising, given that their effects could both be positive or negative (*i.e.*, either lead to overoptimism or pessimism), depending on the circumstances. By contrast, conflicts and, to a lesser extent, disasters, have always negative effects on output and therefore result in an optimism bias in preliminary GDP estimates. The indicators on statistical capacity are not robust and/or mostly not significant implying that they have neither negative or positive effects (but they lower growth revisions in absolute terms as shown above).

The picture of the effects of IMF programs is however more nuanced. A larger population tends to lead to a negative bias. By contrast, the presence of IMF programs causes an overoptimism bias according to at least to those specifications including all countries. However, when only considering LICs in the sample - those countries where IMF programs are more frequent than in other countries - the effect is not significant. There is also some indication that in years when IMF programs are agreed, preliminary growth figures are too pessimistic which possibly serves as a justification for the program itself, but this effect is not robust across all specifications in Table 9. The LIC dummy ceases to be significant in specification (3) implying again that the control variables capture important factors resulting in differences in terms of growth revisions between LICs and other countries. The coefficient of the 1990 period dummy is negative implying that overoptimism was worse in the 1990s.

In specifications 1 and 2 of Table 10, we again only include LICs, but now we examine the factors leading to biases in forecasts and backcasts. The coefficients of the 'economic' and the 'technical capacity' variables are similar. However, the coefficients of the variables relating to IMF programs cease to be significant. Finally, in the last three specifications, we examine whether some of the economic factors alone explain the optimism bias in preliminary growth figures of LICs. We therefore only choose to include the economic variables. Given that the constant is not negative and/or significant, one may argue that the overoptimism bias of preliminary growth figures in LICs is caused by vulnerability to shocks alone and not by any other factors.

6. Conclusions

Output data revisions may be large and significant in economic terms in all countries, and they may have adverse economic consequences. This paper presented stylized facts which suggest that in LICs, growth revisions are larger, the bias is (more) negative suggesting that preliminary growth figures are more often too optimistic, and extreme downward revisions occur more frequently compared to other countries. These results are robust across different release dates. The second part of the paper examined whether these differences remain significant once other differences between LICs and non-LICs are controlled for. Our results suggest that at least the overoptimism bias in LICs disappears once vulnerability to economic shocks is controlled for.

The second part also analyzed the determinants of absolute growth revisions and factors biasing growth revisions. With respect to absolute growth revisions, we showed that statistical capacity and increased IMF surveillance may be suitable tools to lower them. With respect to reducing the bias, the picture is more mixed, and we find some (but very limited) evidence that IMF programs cause specific biases. However, we also find some evidence that countries with larger populations tend to suffer less from overoptimism bias and have, on average, larger absolute growth revisions. Finally, we find at best very limited support for hypotheses that the IMF purposely biases preliminary growth figures to justify its lending activities, but the effects are not robust.

However, whether the correlations we report can be interpreted as causal is of course subject to debate. While not all of the effects remain robust to the inclusion of country and time effects, here one problem is that there is very little within-country variation in particular for the variables relating to technical capacity. Nevertheless, the coefficients of variables relating to the vulnerability to economic shocks are also robust to the inclusion of country and time effects. which suggests that this result is robust as well.

Of course, the extent to which output data revisions occur is only one dimension of output data quality, and other indicators are also important. In addition, even when output data revisions are small, there is no guarantee that final figures are more accurate than in cases when revisions are large. Nevertheless, output data revisions remain still an important challenge for macroeconomic policy, and this dimension of data quality is easily quantifiable and is suitable to evaluate the effects of, for instance, adherence to international standards.

7. References

Abreu, I., 2011, "International organisations' vs. private analysts' growth forecasts: an evaluation" *Economic Bulletin*, vol. 17, no. 2, pp. 23-44.

Aguiar, M., and Gopinath, G. (2007),"Emerging Market Business Cycles: The Cycle Is the Trend" *Journal of Political Economy*, pp.69 - 102.

Aldenhoff, F.-O., 2007, "Are Economic Forecasts of the International Monetary Fund Politically Biased? A Public Choice Analysis." *The Review of International Organizations*, vol. 2, no. 3, pp. 239-260.

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, pp. 1-49.

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.

Barrionuevo, J.M., 1993, "How Accurate Are the World Economic Outlook Projections? Staff Studies for the World Economic Outlook (Washington, International Monetary Fund.

Batchelor, R. (2000), "The IMF and OECD versus consensus forecasts. City University Business School, London, August 2000.

Beach, W. W., Schavey, A. B., and Isidro, I. M. (1999), "How reliable are IMF economic forecasts?," Heritage Foundation CDA 99-05.

Blanchard, O., and D. Leigh, 2013, "Growth Forecasts and Fiscal Multipliers." IMF Working Paper, no. WP/13/1.

Croushore, D., 2011, "Frontiers of Real-Time Data Analysis." *Journal of Economic Literature*, vol. 49, no. 1, p. 29.

Dreher, Axel, 2006, "IMF and Economic Growth: The Effects of Programs, Loans, and

Compliance with Conditionality.", *World Development*, vol. 34, no. 5, pp. 769-788.

Dreher, A., S. Marchesi, and J. Vreeland, 2008, "The politics of IMF forecasts." *Public Choice*, vol. 137, pp. 145-171.

Easterly, W., 2012, "The Role of Growth Slowdowns and Forecast Errors in Public Debt Crises." in: "Fiscal Policy After the Financial Crisis.", eds. A. Alesina and F. Giavazzi, forthcoming.

Hamilton, K., Ley, E., 2010, "Measuring National Income and Growth in Resource-Rich, Income-Poor Countries." Economic Premise, The World Bank.

Henderson, J.V., Storeygard, A., Weil, D.N., 2012, "Measuring Economic Growth from Outer Space." American Economic Review 102, 994-1028.

Ince, O., Papell, D.H., 2011, "The (Un)Reliability of Real-Time Output Gap Estimates with Revised Data." unpublished manuscript.

Jerven, M., 2013, "Poor Numbers: How We Are Misled by African Development Statistics and What to Do About It." Cornell University Press.

Ley, E., and F. Misch, 2013b, "Cross-Country Real-Time Estimates of the Output Gap." unpublished manuscript.

Ley, E., Misch, F., 2013a, "Real-time macro monitoring and fiscal policy." World Bank Policy Research Working Paper.

Loungani, P. (2000), "How accurate are private sector forecasts? Cross-country evidence from consensus forecasts of output growth," *IMF working paper*, 00/77. International Monetary Fund, Washington, DC.

Pons, J. (2000), "The accuracy of IMF and OECD forecasts for G7 countries," *Journal of Forecasting*, 19(1), 53-63.

Timmermann, A., 2007, "An evaluation of the World Economic Outlook forecasts." *IMF Staff Papers*, vol. 54, pp. 1-33.

8. Appendix: Definition of Output Gaps and Turning Points

In order to examine the effects of the cyclical position of the economy on the reliability of output figures, it is useful to define the output gap within this framework; the output gap is the result of fluctuations of actual output x_t around potential output represented by its trend, $\bar{x}_t$. Analytically, it is:

$$z_t \equiv \log\left(\frac{x_t}{\bar{x}_t}\right) = [\log(x_t) - \log(\bar{x}_t)] = (y_t - \bar{y}_t) \tag{3}$$

where $y_t = \log(x_t)$. The output gap, z_t, defined in (3), reflects the cyclical position of the economy: when it is positive then economic activity is above potential which can loosely be defined as an upswing, whereas when it is negative then output is below potential which can loosely be interpreted as a recession. Cyclical turning points / points of inflection can be identified as follows algebraically: define the slopes of the components:[6]

$$\begin{aligned} \bar{g}_t &= (\bar{y}_t - \bar{y}_{t-1})/\bar{y}_{t-1} \\ \tilde{g}_t &= (z_t - z_{t-1})/z_{t-1} \end{aligned} \tag{4}$$

There is a *turning point* at t (*i.e.*, the ordering of the size of the slopes changes):

$$\begin{array}{ccc} \tilde{g}_{t-1} > \bar{g}_{t-1} & \text{and} & \tilde{g}_t < \bar{g}_t \\ & \text{or} & \\ \tilde{g}_{t-1} < \bar{g}_{t-1} & \text{and} & \tilde{g}_t > \bar{g}_t \end{array} \tag{5}$$

which can also be expressed as:

$$\text{sign}(\tilde{g}_{t-1} - \bar{g}_{t-1}) \neq \text{sign}(\tilde{g}_t - \bar{g}_t) \tag{6}$$

[6] Draw a $\cos(y)$ wave and the standard y axis. Turning points occur at every $f\pi$, $f \in \{1, 2, 3, \ldots\}$, at which point the slope of the cosine wave changes sign. Now tilt the x axis at a 45 degree angle; the turning points are the same, the slope of the cycle goes from steeper/flatter than 45^o to flatter/steeper.

Table 8. The Determinants of Growth Revisions in Absolute Terms

VARIABLES	(1) abs. revision	(2) abs. revision	(3) abs. revision	(4) abs. revision	(5) abs. revision
tphp	0.326***	0.408***	0.415***	0.329***	0.999***
volatile	0.243***	0.357***	0.310***	0.186***	0.354***
resources	-0.0139	0.115	-0.0478	-0.0487	-0.485
conflict	0.600***	1.495***	0.780	0.228	0.976
disaster	0.527***	0.927***	0.527**	0.286	0.408
gdds	-0.187***	-0.181	-0.209*	-0.193**	-0.673**
sdds	-0.395***	-0.198	-0.470***	-0.536***	
population	-0.160***	-0.139***	-0.188***	-0.173***	-0.221
imf	0.0327	0.0751	0.131	-0.239**	-0.352
imf5	-0.270***	-0.117	-0.134	-0.403***	-0.584**
recession	0.650***	2.307***	0.699**	0.243	0.512
lic	0.292***	0.446**	0.318**	0.471***	
p1990	0.315***	0.372***	0.260**	0.235**	0.229
Constant	1.943***	1.027***	1.079***	1.018***	1.547***
Observations	17,816	3,964	3,964	3,630	804
R-squared	0.166	0.173	0.153	0.113	0.159

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
(1) All vintages included with vintage fixed effects
(2) Forecasts only
(3) Nowcasts only
(4) t+1 Backcasts only
(5) Nowcasts & LICs only

Table 9. The Determinants of Growth Revisions

VARIABLES	(1) revision	(2) revision	(3) revision	(4) revision	(5) revision
tphp	-0.0518	-0.134	0.113	-1.326***	-1.323***
volatile	0.205***	0.153***	0.106**	-0.00159	
resources	0.230	0.0442	-0.0604	-0.534	
conflict	-3.551***	-2.687***	-1.004**	-4.015***	-3.964***
disaster	-0.175	-0.245	0.0861	-2.112***	-2.007***
gdds	-0.0964	0.140	-0.0665	-0.638*	-0.741**
sdds	-0.272	-0.166	-0.365***		
population	0.0731	0.123***	0.0458	1.038***	1.020***
imf	-0.146	-0.242	0.101	0.854**	0.781**
imf5	-0.537***	-0.503***	-0.467***	-0.484	
recession	-4.213***	1.598***	0.866***	3.125***	3.226***
lic	-1.172***	-0.792***	-0.198		
p1990	-0.855***	-0.499***	-0.0967	-1.480***	-1.540***
Constant	0.264	0.173	0.215	-0.749	-0.973**
Observations	3,964	3,964	3,630	804	804
R-squared	0.107	0.057	0.027	0.167	0.164

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

(1) Forecasts only

(2) Nowcasts only

(3) t+1 Backcasts only

(4) Nowcasts & LICs only

(5) Nowcasts & LICs only

Table 10. The Determinants of Growth Revisions

VARIABLES	(1) revision	(2) revision	(3) revision	(4) revision	(5) revision
tphp	-1.206***	-0.476*			
volatile	0.0976	0.00323	-0.140*	-0.180**	-0.0349
resources	-5.736***	0.603	-4.670**	0.508	0.407
conflict	-4.884***	-1.758**	-5.854***	-4.016***	-0.845
disaster	-2.326**	-1.178***			
gdds	-1.486***	-0.557**			
sdds	-0.810	0.198			
population	1.060***	0.617***			
imf	-0.0823	0.208			
imf5	-0.296	-0.261			
recession	-3.886***	2.379***			
p1990	-2.356***	-1.052***			
Constant	-0.679	-0.129	-0.170	0.432	0.419*
Observations	804	736	1,170	1,240	1,170
R-squared	0.210	0.082	0.106	0.067	0.006

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
(1) Forecasts only & LICs only
(2) t+1 Backcasts & LICs only
(3) Forecasts only & LICs only
(4) Nowcasts & LICs only
(5) t+1 Backcasts & LICs only