Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Forecast Bias in Latin American Medium-Term Fiscal Frameworks: Evidence from Chile, Colombia, Peru, and Uruguay

Javier Torres

j.torresgomez@up.edu.pe

University of the Pacific

Manuel Arriola-Montenegro

University of the Pacific

Research Article

Keywords: Forecast bias, Medium-Term Fiscal Frameworks, GDP growth projections, Latin America, macroeconomic forecasting, unbiasedness tests

Posted Date: October 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10935615/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Forecast Bias in Latin American Medium-Term Fiscal Frameworks: Evidence from Chile, Colombia, Peru, and Uruguay

Javier Torres[1*] and Manuel Arriola-Montenegro[1]

[1*]Department of Economics, Universidad del Pacífico, Av. Salaverry 2020, Jesús María, Lima, 15072, Peru.

*Corresponding author(s). E-mail(s): j.torresgomez@up.edu.pe;
Contributing authors: ma.arriolam@up.edu.pe;

Abstract

Medium-term fiscal frameworks (MTFs) anchor the annual budget within a multi-year macroeconomic outlook and are a central institutional device for fiscal-policy credibility in Latin America. This paper evaluates whether the real gross domestic product (GDP) growth forecasts published by the MTFs of Chile, Colombia, Peru, and Uruguay are unbiased. Using a unified database of multi-year forecast vintages from each country, we apply the Holden and Peel (1990) test to country-specific subsamples, and we find a recurrent two-regime structure. In every country the unbiasedness null is rejected in the second subsample, with overestimation biases ranging between 1.4 and 2.4 percentage points per year, approximately. Because macroeconomic projections enter medium-term budget planning and, in different ways across countries, interact with the operation and assessment of fiscal rules, persistent growth overestimation may weaken the reliability of the fiscal baselines on which policy decisions are made. How that bias maps into compliance with a numerical rule differs across the four countries, because their rules rest on different structural, cyclical and debt concepts.

Keywords: Forecast bias, Medium-Term Fiscal Frameworks, GDP growth projections, Latin America, macroeconomic forecasting, unbiasedness tests

JEL Classification: E37 , C53 , E17 , H68

ORCID: Javier Torres 0000-0001-6850-1395; Manuel Arriola-Montenegro 0009-0005-5733-9413

Manuscript statistics: Main text 5,587 words (excluding title page; abstract; tables; figures; references); 2 figures and 3 tables in the main text.

1 Introduction

Medium-term fiscal frameworks (MTFs) are official documents in which a government's economic authority lays out a multi-year macroeconomic outlook (typically three to four years ahead) together with the fiscal projections that follow from it. They serve a dual purpose: to anchor the annual budget within a longer-term view of public finances, and to communicate to citizens, markets, and supervisory bodies the assumptions under which fiscal policy is being conducted. Filc and Scartascini (2011) trace the modern MTF back to fiscal-reform episodes in the United Kingdom and Germany during the 1960s, and document the wave of Latin American adoptions during the 2000s, which sought to lock in fiscal discipline after the volatile decades that preceded it.[1]

Because MTFs are reviewed by markets and by domestic political actors, the quality of the macroeconomic forecasts they contain is a first-order policy issue. The link runs through fiscal planning itself. A deficit ceiling, a debt anchor or a structural balance target cannot be applied to an observed outcome at the moment the budget is approved: compliance has to be assessed against a projected macroeconomic path, and in these four countries the projections that inform that path are produced and published by the same authority whose discretion the rule is meant to restrain. An optimistic projection can therefore relax the effective constraint without any ceiling being breached. If the projections embedded in the document are systematically biased, the fiscal rules and spending plans that rest on them will be miscalibrated, the legislative debate around the budget will be informed by an inaccurate baseline, and the credibility benefit that motivated the framework in the first place will be eroded. Evaluating whether official multi-year forecasts are unbiased is therefore a precondition for assessing the broader functioning of MTFs as a fiscal-policy institution.

This paper conducts that evaluation for four Latin American economies (Chile, Colombia, Peru, and Uruguay) that have published medium-term fiscal frameworks for at least two decades. The four documents differ in institutional setting and update frequency, yet they share a common feature. Each one issues a real gross domestic product (GDP) growth forecast for the current year and for up to three years ahead, although the number of horizons actually reported varies across documents and years. Specifically, the documents analyzed in this paper are:

- **Chile.** The *Informe de Finanzas Publicas* (IFP), prepared by the Direccion de Presupuestos (DIPRES) of the Ministry of Finance. The IFP is issued quarterly, with the third-quarter release containing the macroeconomic baseline that accompanies the budget bill sent to Congress. Earlier vintages of the same exercise appeared as an annex of macroeconomic assumptions inside the budget itself (Dirección de Presupuestos, 2026). The submission of an *Informe de Finanzas Publicas* prior to the congressional consideration of the budget is mandated by Law 19896 of 2003 (Congreso Nacional de Chile, 2003).
- **Colombia.** The *Marco Fiscal de Mediano Plazo* (MFMP), published every June by the Ministerio de Hacienda y Credito Publico. The MFMP was established by

[1] Schick (2009) and Brumby and Hemming (2013) review the role of MTFs as institutional devices for credibility and the cross-country variation in their design.

Law 819 of 2003 (Congreso de la República de Colombia, 2003), and the first issue appeared in June 2004; it provides a ten-year horizon of macroeconomic and fiscal projections (Ministerio de Hacienda y Crédito Público, 2026).

- **Peru.** The *Marco Macroeconomico Multianual* (MMM), prepared by the Ministerio de Economia y Finanzas (MEF) under Legislative Decree 1276 (Poder Ejecutivo del Perú, 2016). The MMM has been published since 2000 and typically appears twice a year (an April-May release and a third-quarter revision) with a four-year horizon (Ministerio de Economía y Finanzas del Perú, 2026).
- **Uruguay.** The macroeconomic scenario reported in the *Presupuesto Nacional* (issued once per five-year administration) and in the annual *Rendicion de Cuentas* updates filed with the legislature; both are produced by the Ministerio de Economia y Finanzas (Ministerio de Economía y Finanzas del Uruguay, 2026). Both documents are constitutionally mandated: Article 214 of the Constitution requires each incoming administration to submit a five-year Presupuesto Nacional and the Executive to file an annual Rendición de Cuentas within six months of the close of each fiscal year (República Oriental del Uruguay, 1996).

Each of these documents also sits alongside a numerical fiscal rule, and the macroeconomic scenario it publishes is part of the baseline against which fiscal plans are drawn. How directly the growth projection enters the rule differs across the four countries. Peru offers the most direct link, since the fiscal rules are set out in the same legal framework that governs the MMM. The Colombian rule combines a debt anchor with a structural primary balance and cyclical and commodity adjustments, and the Chilean rule is assessed on a cyclically adjusted balance built on trend output estimated with the participation of expert committees, so in neither case does the published growth forecast map one for one into fiscal space. Uruguay requires a further caveat: its current rule was introduced only in 2020 and therefore postdates most of the sample studied here.

The four frameworks are heterogeneous along several dimensions. Chile and Peru update their projections two or more times a year, whereas Colombia produces a single annual document and Uruguay updates only within the budgetary cycle. The Chilean and Peruvian frameworks have been in continuous operation since the early 2000s, while Colombia's first MFMP dates from 2004 and Uruguay's digitally available macroeconomic projections begin only in the second half of the 2000s. Whatever the differences in form, all four share the property of producing official, on-the-record forecasts of real GDP growth over a medium-term horizon of up to three years, which makes them directly comparable through the lens of forecast evaluation.

The methodology builds on the literature on statistical forecast evaluation. Mincer and Zarnowitz (1969) proposed the canonical regression test of rationality, but Mankiw and Shapiro (1986) showed that its small-sample distribution can over-reject the null with highly autoregressive series, a concern that is acute given our short samples. We therefore follow Holden and Peel (1990) and test the necessary moment condition for unbiasedness, $\mathbb{E}(Y_{t+j} - F_{t|t+j}) = 0$, directly.

Forecast bias has been documented in a variety of official forecasting bodies. Frankel (2011) finds particularly pronounced optimism among countries subject to the

Stability and Growth Pact, while Frankel and Schreger (2013) show that the interaction between forecast bias and fiscal institutions depends importantly on institutional design. Jonung and Larch (2006) argue that the incentive to over-project is structural rather than incidental and propose independent forecasts as the remedy, and Beetsma, Giuliodori, and Wierts (2009) show that the multi-annual plans submitted under the European fiscal framework are systematically not delivered. Eicher and Kawai (2024) document persistent bias in International Monetary Fund (IMF) reserve and debt forecasts for balance-of-payments program countries, and Aldenhoff (2007) provides a public-choice rationalization for biased IMF growth forecasts. For revenue and budget forecasts at the country level, Auerbach (1999) provides early benchmark evidence for the United States. The literature therefore documents systematic bias in a number of official forecasting settings, although the evidence is not universal, and the direction, magnitude, and timing of the bias appear to be context-specific. To our knowledge, the four Latin American MTFs studied here have not been jointly evaluated using a common methodology.

The remainder of the paper is organized as follows. Section 2 describes the dataset, the variable of interest, and the period covered, and presents the descriptive analysis country by country with the help of the error-series plots. Section 3 describes the methodology used to identify bias in the breaks visually identified in section 2. Section 4 presents the main results and section 5 concludes.

2 Database and Descriptive Analysis

2.1 Data sources and variable of interest

The dataset is constructed from the medium-term fiscal frameworks produced by each country, complemented with realized GDP growth from the World Bank's World Development Indicators (World Bank, 2026). Each MTF report contains an official real GDP growth projection for the year in which it is published (t) and for up to three years ahead ($t+1$, $t+2$, $t+3$). We extracted, for each country and each publication year, the medium-term forecasts $F_{t|t+1}$, $F_{t|t+2}$, and $F_{t|t+3}$ reported in the official document, together with the realized growth rate Y_{t+j} for each target year.

For every triplet (country, publication year, horizon) the forecast error is constructed as

$$e_{t|t+j} \;=\; Y_{t+j} \;-\; F_{t|t+j}, \qquad j \in \{1, 2, 3\}. \tag{1}$$

By construction, $e_{t|t+j} > 0$ indicates that realized growth exceeded the projection (the MTF underestimated growth), whereas $e_{t|t+j} < 0$ indicates that realized growth fell short of the projection (the MTF overestimated growth). Under unbiasedness, $\mathbb{E}(e_{t|t+j}) = 0$ for each horizon j.

The period covered varies by country, reflecting the date at which each MTF began to publish multi-year GDP projections. Chile's series spans publication years 2000 to 2022, with realized growth available through 2024; Colombia begins in 2004 (the first issue of the MFMP, following Law 819 of 2003); Peru spans 2002 to 2022; and Uruguay spans 2006 to 2022 with a few internal gaps in years for which the budget exercise or accountability report was not digitally available.

The rest of this section uses the per-country error-series plots collected at the end of the paper to motivate the cutoff years that will partition each sample in the empirical analysis. For each country we look at the average forecast error in year t, defined as the simple mean of $e_{t|t+1}$, $e_{t|t+2}$, $e_{t|t+3}$, and we overlay two centered moving averages, a short one with a three-year window and a long one with a five-year window. The cutoff selected for each country is the year around which the long moving average changes its qualitative behavior in a way that is consistent with a regime change in the sign and magnitude of the bias.

2.2 Chile

Figure 1(a) shows the average annual forecast error of the IFP. From 2005 onward both moving averages remain in negative territory for essentially every year of the sample, which signals a sustained regime of overestimation. The short moving average stays in the range $[-3, -1]$ percentage points for most of the period, while the long moving average behaves as a smooth envelope that does not return to zero until the very end of the sample.

The only departures from this regime come from atypical years. The bar for 2009, which records the impact of the global financial crisis on the realized series, lies visibly below the moving averages and reflects a shock that no IFP vintage anticipated. The bars for 2020 and 2022 introduce additional volatility associated with the COVID-19 contraction and the subsequent rebound, but they do not alter the qualitative reading of the moving averages, which remain below zero. The conclusion that emerges from the figure is that the Chilean forecasts have been systematically optimistic since 2005, with sporadic outliers concentrated in years of pronounced macroeconomic stress.

The empirical implication is direct. The natural cutoff for the Chilean sample is the 2004/2005 boundary: the period up to and including 2004 captures the few years that precede the onset of the overestimation regime, and the period from 2005 onward captures the regime itself. The pre-cutoff window should therefore display a small and statistically insignificant mean forecast error, while the post-cutoff window should display a clearly negative and significant one. Both subsamples, together with the full sample, are reported in the Results section.

2.3 Colombia

Figure 1(b) reports the average annual forecast error of the MFMP. The plot can be naturally split into two phases. For the period 2004 to 2011 the long moving average sits very close to zero, oscillating within a narrow band that contains the horizontal axis, while the short moving average swings around zero with no clear sign. This is the visual signature of a regime in which the MFMP forecasts are, on average, broadly consistent with realized growth and do not display a systematic direction of error.

From 2012 onward the picture changes qualitatively. The long moving average leaves the neighborhood of zero and remains entirely negative for the rest of the sample, indicating that the average error becomes systematically negative once that point is passed. The short moving average follows the same descent and stays below

zero for almost the entire post-2011 window, with only minor positive blips in the post-pandemic observations. The combination of an essentially zero long moving average up to 2011 and an entirely negative long moving average from 2012 onward justifies treating 2011 as the last year of an unbiased regime and 2012 as the first year of an overestimation regime. The cutoff used in the Results section therefore partitions the Colombian sample at the 2011/2012 boundary.

2.4 Peru

Figure 1(c) shows the cleanest pattern in the four-country panel. Both moving averages, short and long, display a clearly descending trend over the whole sample: starting from positive values in the early 2000s, they decline through the second half of the decade and end up well below zero in the 2010s. The long moving average crosses zero only once, around 2010, and from that point onward the average forecast error is systematically negative.

The visual evidence indicates that the marked overestimation regime begins in 2011. The bars from 2011 to the end of the sample are predominantly negative, with magnitudes that are large in absolute value during the second half of the 2010s. The two pandemic-related bars (2020 and 2021) introduce large fluctuations, but they do not pull the moving averages back into positive territory: the long moving average remains clearly negative through 2022. The cutoff for Peru is therefore set at 2010, with the regression splitting the sample into the years up to and including 2010 and the years from 2011 onward.

2.5 Uruguay

Figure 1(d) shows a pattern that is qualitatively similar to Peru's. In the late 2000s the moving averages sit above zero, consistent with a regime in which Uruguayan forecasts tend to underestimate realized growth, partly because of the strong post-recession rebound of 2010. Around 2011 the moving averages begin to descend and they cross zero shortly thereafter, settling firmly into negative territory from 2012 onward. The long moving average reaches its trough in the second half of the 2010s, when the average error is the most negative of any country in the panel.[2]

The marked overestimation regime for Uruguay therefore begins in 2012, one year later than the Peruvian shift. As in Peru, the post-pandemic observations introduce a partial reversal towards positive average errors in 2020 and 2021, but these are too few to alter the medium-term reading of the moving averages. The cutoff for Uruguay is consequently set at 2011, with the regression splitting the sample into the years up to and including 2011 and the years from 2012 onward.

[2] The Uruguayan series is irregular relative to the other three countries in the panel. No publicly accessible digital projection was located for fiscal year 2009. In addition, in several years, fewer than the three forecast horizons used by the other countries. Specifically, the documents published in 2007, 2014, 2017 and 2018 contain only two forecast horizons (one- and two-year-ahead), and the documents published in 2008 and 2019 contain a single one-year-ahead forecast. The dataset preserves the original document structure, so observations with missing horizons are treated as missing rather than imputed; this is the reason behind the smaller effective sample size for Uruguay in the regressions reported in the following section.

3 Methodology

Following the approach developed by Holden and Peel (1990), this study examines whether the real GDP growth projections reported in each country's MTF display a systematic bias. Under an unbiased forecasting scheme, the projections should, on average, match the actual observed GDP growth figures. Specifically, the forecast error defined in equation (1) should have an expected value of zero and should show no systematic patterns across time.

3.1 Testing for unbiasedness

The Mincer-Zarnowitz regression (Mincer & Zarnowitz, 1969) is the canonical test of forecast rationality. It regresses the realized value of a variable on the corresponding forecast and tests the joint null hypothesis that the intercept is zero and the slope is one. Holden and Peel (1990) show, however, that this joint condition is sufficient but not necessary for unbiasedness. A forecast can be unbiased and still reject the Mincer-Zarnowitz null because of correlation between the forecast and the error term. They propose, as an alternative, to test the necessary moment condition directly, namely that the average forecast error is equal to zero. This concern is particularly relevant in short samples when the variables involved are highly persistent (Mankiw & Shapiro, 1986), which is the setting here given the limited number of MTF publications available for each country.

For each country and each subsample considered, we estimate the following intercept-only regression:

$$e_{t|t+j} \;=\; \mu \;+\; \varepsilon_{t,j} \tag{2}$$

where $e_{t|t+j}$ is the forecast error defined in equation (1), pooled across the three forecast horizons $j \in \{1, 2, 3\}$. The parameter μ captures the average forecast bias in the corresponding subsample. Pooling the horizons increases precision in the relatively short country samples and estimates the average bias across the medium-term projection profile rather than at a single horizon. To accommodate the heterogeneity in forecast variance across years and horizons, the standard errors used to construct the t-statistic are heteroskedasticity-robust (Huber-White) (White, 1980) throughout.

The null hypothesis of unbiasedness is

$$H_0 \colon \mu \;=\; 0\,, \tag{3}$$

which is tested using a standard t-test. A rejection of H_0 in favor of $\mu < 0$ indicates systematic overestimation of realized growth, while a rejection in favor of $\mu > 0$ indicates systematic underestimation.

3.2 Sample partitions

Section 2 motivates a country-specific sample cutoff. For each of the four countries, the average forecast error series suggests a year that separates an early phase in which the forecasts are close to unbiased from a later phase in which the forecasts systematically overestimate realized growth. The empirical design therefore reports, for each country,

three regressions of equation (2): one estimated on the full sample, one estimated on a restricted "pre" window, and one estimated on a restricted "post" window. The partitions are summarized in Table 1.

Table 1 Sample partitions used in the empirical analysis

Country	Full sample	Pre window	Post window
Peru	2002–2022	≤ 2010	> 2010
Colombia	2004–2022	≤ 2011	> 2011
Uruguay	2006–2022	≤ 2011	> 2011
Chile	2000–2022	≤ 2004	> 2004

The cutoffs are complementary. The pre window ends in one year and the post window begins in the next. The Chilean pre window is the shortest in the panel (the years 2000 to 2004), which reflects the early date at which the overestimation regime begins in that country. Because these cutoffs are motivated by the same data used in the regressions that follow, the conventional pre/post significance levels should be read descriptively; the Appendix addresses breakpoint uncertainty formally.

4 Results

This section reports the estimates of the Holden-Peel regression of equation (2) for each of the four countries. Table 2 presents three panels, the full sample, a restricted "pre" subsample and a restricted "post" subsample, with one column per country, with the country-specific cutoffs motivated in Section 2 and summarized in Table 1. The reported coefficient is the mean forecast error in the corresponding window, with heteroskedasticity-robust standard errors in parentheses. Statistical significance is reported at the standard 1%, 5% and 10% levels.

4.1 Peru

Column (3) of Table 2 reports the estimated mean forecast error of the MMM for the full sample, the pre-2010 window, and the post-2010 window.

The full-sample estimate is -0.95 percentage points, marginally significant at the 10% level. This figure suggests a mild tendency for the MMM forecasts to overestimate realized growth on average, but the marginal significance of the full-sample coefficient is consistent with the visual reading of Section 2: the average error series mixes a phase of underestimation with a phase of overestimation, so the full-sample mean is close to zero.

The picture changes substantially when the sample is partitioned around the 2010 cutoff. In the pre-2010 window the average forecast error is $+0.97$ percentage points, significant at the 10% level. The MMM forecasts published between 2002 and 2010 therefore underestimated realized growth by approximately one percentage point per year, a pattern that is consistent with the commodity boom outperforming the conservative projections of the early years.

Table 2 Mean forecast error
$e_{t|t+j} = Y_{t+j} - F_{t|t+j}$, pooled across $j \in \{1, 2, 3\}$

VARIABLES	(1) Chile	(2) Colombia	(3) Peru	(4) Uruguay
(a) Full sample				
Constant	-1.14***	-0.72	-0.95*	-1.04**
	(0.38)	(0.47)	(0.56)	(0.46)
Observations	68	56	62	39
(b) Pre-cutoff window				
Constant	-0.34	0.38	0.97*	1.07**
	(0.43)	(0.44)	(0.50)	(0.45)
Observations	15	24	27	12
(c) Post-cutoff window				
Constant	-1.37***	-1.54**	-2.43***	-1.98***
	(0.47)	(0.73)	(0.84)	(0.55)
Observations	53	32	35	27

Coefficients in percentage points
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In the post-2010 window the average forecast error is -2.43 percentage points, significant at the 1% level. The magnitude is much larger than the mild bias of the full sample and confirms that, from 2011 onward, the MMM has systematically overestimated GDP growth by close to two and a half percentage points per year. The contrast between the two subsamples, an underestimation regime of roughly one percentage point and an overestimation regime of roughly two and a half percentage points, supports the regime-change interpretation of the visual evidence and supports the choice of 2010 as a natural cutoff for the Peruvian series.

4.2 Colombia

Column (2) of Table 2 reports the estimates for Colombia under the 2011/2012 partition.

The full-sample coefficient is -0.72 percentage points, not statistically distinguishable from zero. As in the Peruvian case, the lack of significance at the aggregate level reflects the offsetting effect of two regimes: a first phase in which the average error is close to zero and a second phase in which it is systematically negative.

The pre-2011 window confirms this reading. The mean forecast error from 2004 to 2011 is $+0.39$ percentage points and is not statistically different from zero. The Colombian MFMP issued during the first decade of the framework therefore behaves approximately as an unbiased forecasting sequence, which is consistent with the long moving average sitting close to zero throughout that period in Figure 1(b).

The post-2011 window tells a different story. The mean forecast error from 2012 onward is -1.54 percentage points, significant at the 5% level. The MFMP forecasts published from 2012 onward have therefore overestimated realized growth by

roughly one and a half percentage points per year. The combination of an insignificant pre-period coefficient and a significantly negative post-period coefficient supports the interpretation that the Colombian MFMP transitioned from an unbiased regime to an overestimation regime around the 2011/2012 boundary, exactly as suggested by the change in the long moving average documented in Section 2.

4.3 Uruguay

Column (4) of Table 2 reports the estimates for Uruguay under the 2011/2012 partition.

The Uruguayan results are the most clear-cut of the three two-regime countries. The full-sample coefficient is -1.04 percentage points, significant at the 5% level, which already suggests a systematic overestimation on average.

The pre-2011 window yields a positive and statistically significant mean forecast error of $+1.07$ percentage points, significant at the 5% level. Despite the irregular structure of the Uruguayan data, with several years reporting fewer than three forecast horizons, the available evidence indicates that the macroeconomic scenario in the budget and accountability documents underestimated realized growth in the second half of the 2000s, a period that includes the strong post-recession rebound of 2010.

The post-2011 window confirms the regime change with the largest magnitude observed in the panel. The mean forecast error from 2012 onward is -1.98 percentage points, significant at the 1% level. The Uruguayan forecasts have therefore overestimated GDP growth by close to two percentage points per year since 2012, with a level of statistical significance that is unambiguous. As in the Peruvian case, the two-regime structure is statistically supported on both sides of the cutoff, which strengthens the interpretation of 2012 as the year in which the systematic overestimation regime begins.

4.4 Chile

Column (1) of Table 2 reports the estimates for Chile under the 2004/2005 partition motivated in Section 2.

The full-sample estimate is -1.14 percentage points, significant at the 1% level. The Chilean IFP has therefore overestimated realized growth by roughly one and a fifth percentage points per year, averaged over the entire sample. The full-sample significance is consistent with the visual evidence that the bias appears very early in the Chilean series.

The pre-2004 window yields an estimated mean forecast error of -0.34 percentage points, not statistically distinguishable from zero. The small magnitude and the lack of significance are consistent with the visual reading of Figure 1(a), in which the moving averages oscillate around zero during the very first years of the sample. The Chilean IFP forecasts published between 2000 and 2004 are therefore broadly consistent with an unbiased forecasting regime, even though the sample of 15 observations is small.

The post-2004 window yields an estimated mean forecast error of -1.37 percentage points, significant at the 1% level. The Chilean forecasts published from 2005 onward have therefore overestimated realized growth by close to one and a half percentage

points per year. The contrast with the pre-2004 window is sharp: the regime moves from an insignificant near-zero bias to a clearly negative and significant overestimation, exactly the pattern that the visual evidence motivates. The two-regime interpretation of the Chilean series is therefore supported by the formal test, with 2004/2005 acting as the cutoff between an early unbiased phase and the sustained overestimation regime that follows.

4.5 Summary: results and limitations

Across the four countries, the Holden-Peel test rejects the null of unbiasedness in the post-cutoff window with statistical significance, with coefficients ranging from -1.37 to -2.43 percentage points. The interpretation is unambiguous: every country in the panel has overestimated realized growth in the recent regime. In the pre-cutoff window the pattern is more heterogeneous: Peru and Uruguay exhibit a positive and significant mean error, indicating that the early phase of each framework was characterized by mild underestimation; Colombia and Chile fail to reject unbiasedness, consistent with an early phase that behaves approximately as an unbiased forecasting sequence. Two stylized facts therefore emerge from the table: the four countries share a recent regime of systematic overestimation, and the year that separates the early phase from the recent regime falls between 2004 (for Chile) and 2011 (for Colombia and Uruguay), with Peru in between.

This study has limitations that point to directions for future research. The methodology employed here does not identify the structural causes of the documented biases, so future work could investigate the role of political-economy incentives, of the timing of methodological updates inside each ministry, or of the response to global commodity-price shocks in explaining the cross-country differences in the magnitude and the timing of the bias.

5 Conclusions

This paper evaluates the unbiasedness of the real GDP growth forecasts of four Latin American medium-term fiscal frameworks: Chile, Colombia, Peru, and Uruguay. The evaluation combines a unified database of forecast vintages with realized GDP growth from the World Bank.

The empirical estimates confirm a regime change interpretation of the MTF's forecasts. In the post-cutoff window, the average forecast error is negative and statistically significant at the 5% level at least in all four countries, with magnitudes ranging from -1.37 percentage points for Chile (from 2005 onward) to -2.43 percentage points for Peru (from 2010 onward).

The Holden-Peel test rejects the unbiasedness null in every post-cutoff subsample. In the pre-cutoff window the picture is more nuanced: for Peru and Uruguay the average error is positive and statistically significant, indicating an early regime of mild underestimation; for Colombia and Chile the average error is not statistically distinguishable from zero.

The magnitudes of the post-cutoff biases are economically meaningful. The four countries overestimate realized growth by between around 1.4 and 2.4 percentage

points per year in the post-cutoff window. Chile stands out by having its 2nd window around 2004-2005, while the rest around 2010-2012. Given that the historical average of GDP growth in these economies during the analyzed years lies in the three to five percent range, an annual bias of two percentage points represents a substantial fraction of the realized variation, with direct implications for the credibility of the fiscal projections that depend on these growth assumptions.

The cross-country comparison points to a shared inflection period that coincides with the end of the commodity boom and the beginning of the deceleration phase of the early 2010s. The empirical evidence is consistent with the hypothesis that three of the four MTFs reacted with a lag to this regional macroeconomic transition, and that the systematic overestimation observed since the early 2010s reflects, at least in part, a delayed update of the underlying medium-term growth assumptions. The present analysis does not identify this mechanism causally.

These findings have clear implications for the institutional governance of medium-term fiscal frameworks. If the forecasts that anchor the budget process display a systematic and statistically significant overestimation of GDP growth, the fiscal rules calibrated against those projections will deliver looser actual fiscal stances than intended, and the credibility benefit that motivated the adoption of the framework in the first place will be eroded. The order of magnitude follows from the coefficients themselves: if each of the three medium-term projections of a vintage overstates growth by the post-cutoff mean, the level of output three years out is overstated by roughly four to seven percent. Overestimation of that size distorts the macroeconomic baseline on which medium-term fiscal planning rests. How it maps into compliance with a fiscal rule differs across the four countries, because their rules rely on different structural, cyclical, debt and expenditure concepts. A periodic external evaluation of the bias of the forecasts, along the lines of the Holden-Peel test applied here, is therefore a desirable complement to the publication of the frameworks themselves.

Figures

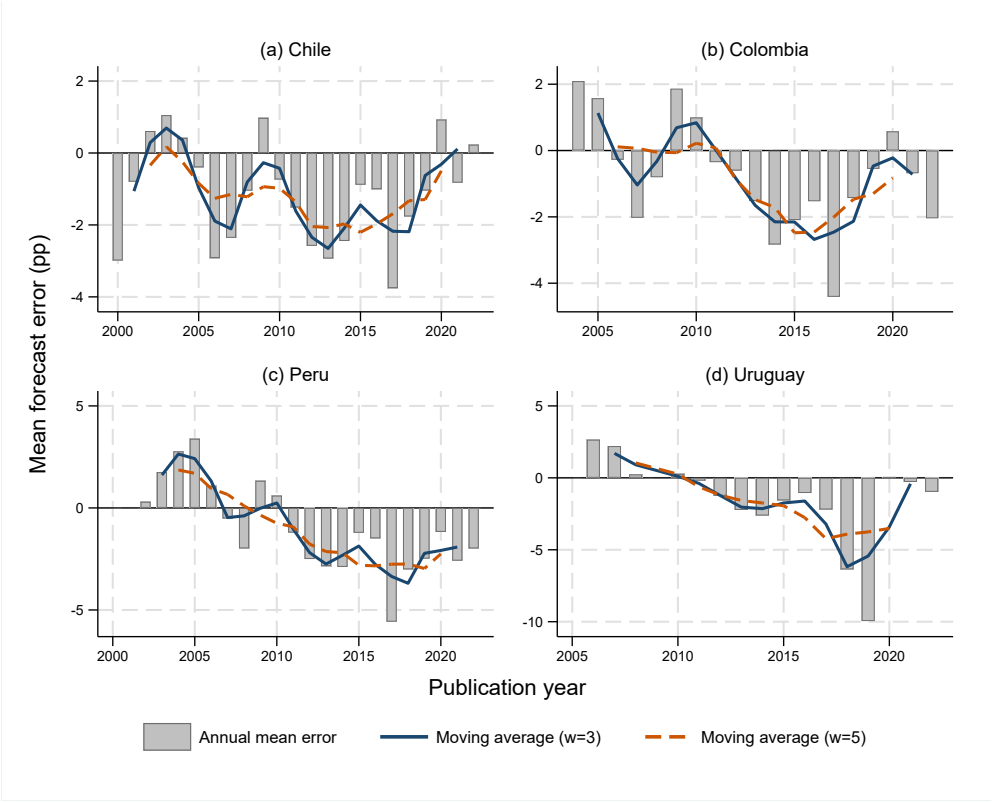

Fig. 1 Average forecast error by year of publication. Panels: (a) Chile, (b) Colombia, (c) Peru, (d) Uruguay

Notes: Each bar is the mean of $e_{t|t+1}$, $e_{t|t+2}$, $e_{t|t+3}$ for the corresponding publication year, in percentage points. The two superimposed lines are centered moving averages with short and long windows, of three and five years respectively; both are undefined at the ends of each sample. Error = realized − forecast, so positive values indicate underestimation and negative values overestimation. The Uruguayan panel has no bar for 2009, a year for which no projection is publicly available.

Acknowledgements. We thank Jason Cruz for methodological insights and participants at the July 2026 Annual Meeting of the Peruvian Economic Association for their comments.

Statements and Declarations

Funding. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interests. The authors have no competing interests to declare that are relevant to the content of this article.

Data availability. The dataset is transcribed from public documents that are cited individually in the reference list. The compiled dataset and the code that reproduces every table and figure are available from the authors on request.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process. During the preparation of this work the authors used Claude Code for data processing and to improve the writing style of the manuscript. The authors reviewed and edited the output as needed and take full responsibility for the content of the published article.

Author contributions. Both authors contributed to the study conception and design, data collection, analysis and writing. Both authors read and approved the final manuscript.

Appendix

Endogenous identification of the cutoff

The cutoffs used in the previous subsection are read off the average forecast error series in Section 2. Selecting the break point by visual inspection of the same series that is later tested is a source of concern: the location of the cutoff is then endogenous to the data, and the nominal significance of a pre/post comparison estimated at a data-chosen date overstates the true significance, because the break date is a nuisance parameter that is not identified under the null of no break (Andrews, 1993; Davies, 1977). To guard against this, we complement the visual analysis with a formal, data-driven procedure that estimates the cutoff and tests for the presence of a break.

For a candidate cutoff year τ, define the indicator $D_t(\tau) = \mathbf{1}\{t \le \tau\}$ and augment the intercept-only regression (2) with a shift in the mean error at τ:

$$e_{t|t+j} \;=\; \mu \;+\; \delta\, D_t(\tau) \;+\; \varepsilon_{t,j}\,. \tag{4}$$

The null of no break at τ is $H_0 : \delta = 0$, a single-restriction Chow test (Chow, 1960) of a shift in the mean forecast error. Because the model contains only a constant, the Chow statistic is

$$F(\tau) \;=\; \frac{[\,\mathrm{SSR}_R - \mathrm{SSR}_U(\tau)\,]/1}{\mathrm{SSR}_U(\tau)/(N-2)}\,, \tag{5}$$

where SSR_R is the residual sum of squares of the constant-only model (2) and $\mathrm{SSR}_U(\tau)$ that of the two-regime model (4); with a single restriction $F(\tau)$ equals the square of the t-statistic on δ and is equivalent to a two-sample comparison of the mean error before and after τ.

Since the break date is unknown, we estimate it as the value that maximizes the statistic over a set of candidate years,

$$\hat{\tau} \;=\; \arg\max_{\tau \in [\underline{\tau},\,\overline{\tau}]} F(\tau)\,, \tag{6}$$

This maximization over an unknown break date follows the search procedure introduced by Quandt (1960) and yields a sup-F statistic. To keep each regime adequately populated, the candidate set $[\underline{\tau}, \overline{\tau}]$ trims 15 percent of the span of publication years at each end, the symmetric trimming recommended by Andrews (1993). Because $\hat{\tau}$ is estimated by maximization, $\sup_\tau F(\tau)$ does not follow the usual F (or χ^2) distribution; we therefore compare it with the asymptotic critical values tabulated by Andrews (1993) for a single tested parameter and 15% trimming, whose 5% value is approximately 8.85 (against 3.84 for the ordinary test with a known date). A survey of this class of tests is provided by Hansen (2001). Throughout we report the homoskedastic version of the statistic; heteroskedasticity-robust and target-year-clustered variants leave the estimated dates $\hat{\tau}$ unchanged and are discussed with the results.

This procedure endogenizes the cutoff that Section 2 selects visually. A close match between $\hat{\tau}$ and the visually chosen cutoff provides formal support for the descriptive reading, while the value of $\sup_\tau F(\tau)$ relative to the Andrews threshold indicates whether the transition is a sharp, datable break or a smoother change of regime.

The country-specific cutoffs used in Section 4 above were motivated visually in Section 2. We now subject them to the formal, data-driven procedure of this appendix, which estimates the break date by maximizing the Chow statistic (5) over candidate years. Figure 2 plots the profile $F(\tau)$ for each country, and Table 3 collects the estimated break $\hat{\tau}$, the value of the statistic at the peak, and the mean forecast error on each side of $\hat{\tau}$.

Table 3 Endogenous break: sup-F estimates versus the visual cutoffs

Country	Visual cutoff	$\hat{\tau}$ (estimated)	sup-F	Sharp break (Andrews 5%)	$\hat{\mu}_{\text{pre}}$ (pp)	$\hat{\mu}_{\text{post}}$ (pp)
Peru	2010	2010	10.43	Yes	+0.97*	−2.43***
Uruguay	2011	2011	11.95	Yes	+1.07**	−1.98***
Colombia	2011	2012	4.36	No	+0.27	−1.64*
Chile	2004	2005	1.59	No	−0.35	−1.43***

$\hat{\tau} = \arg\max_{\tau} F(\tau)$; candidate years trimmed by 15 percent at each end.
A sharp break means that sup-F exceeds the Andrews (1993) sup-Wald 5% critical value (≈ 8.85).
$\hat{\mu}_{\text{pre}}$ and $\hat{\mu}_{\text{post}}$: mean forecast error (pp) on each side of $\hat{\tau}$; stars from
the Holden-Peel robust t-test of a zero mean in each subsample (*** p<0.01, ** p<0.05, * p<0.1).

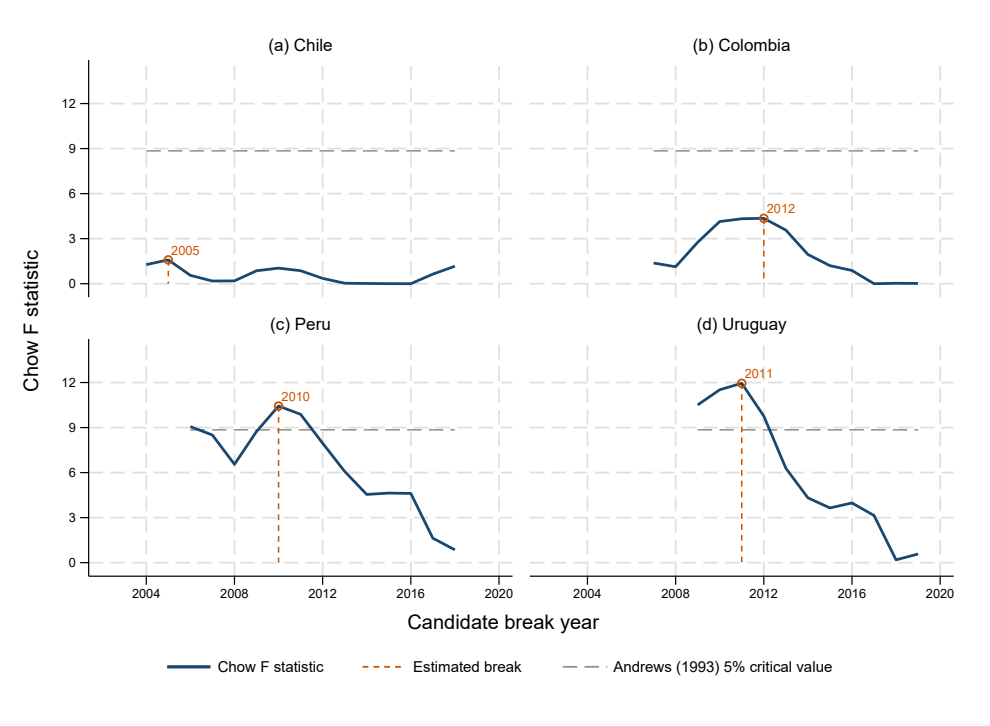

Fig. 2 Endogenous identification of the break. Panels: (a) Chile, (b) Colombia, (c) Peru, (d) Uruguay

Notes: For each country, the solid line plots the homoskedastic Chow statistic $F(\tau)$ testing a shift in the mean forecast error at candidate cutoff year τ (with $D_t(\tau) = \mathbf{1}\{t \leq \tau\}$), pooling the three forecast horizons. The marker and the vertical spike locate the estimated break $\hat{\tau} = \arg\max_\tau F(\tau)$, which is labelled with its year; the dashed horizontal line is the Andrews (1993) sup-Wald 5% critical value (≈ 8.85). Candidate years are trimmed by 15 percent at each end of the sample.

The estimated dates line up closely with the visual cutoffs. For Peru and Uruguay the sup-F date coincides exactly with the cutoff read off the moving averages ($\hat{\tau} = 2010$ and $\hat{\tau} = 2011$, respectively), and the mean errors on each side of $\hat{\tau}$ reproduce the Holden-Peel coefficients reported in Table 2. For Colombia and Chile the estimated date falls one year after the visual cutoff (2012 instead of 2011, and 2005 instead of 2004), a difference that leaves the estimated magnitudes essentially unchanged. The data-driven procedure therefore confirms the partition used in the empirical analysis rather than overturning it.

The height of the profile relative to the Andrews threshold sharpens the interpretation. For Peru and Uruguay the peak of $F(\tau)$ clears the 5% sup-Wald critical value (≈ 8.85), so the null of no break is rejected even after accounting for the estimation of the date: these two countries display a sharp, datable transition from underestimation to overestimation. For Chile and Colombia the profile is flat and remains below the threshold, providing no evidence of a statistically sharp, datable break at conventional levels and remaining consistent with a more gradual transition; this is consistent with

the Chilean bias appearing very early and persistently, and with the milder Colombian shift. The gradual nature of the transition does not weaken the central finding: in both countries the post-cutoff mean error remains negative and significant under the Holden-Peel test (Table 2).[3]

[3]The estimated break dates are unchanged when the statistic is computed with heteroskedasticity-robust or target-year-clustered standard errors. Under those variants the sharp-break conclusion for Peru and Uruguay is preserved, while the statistics for Chile and Colombia remain below conventional thresholds.

References

Aldenhoff, F.-O. (2007). Are economic forecasts of the International Monetary Fund politically biased? A public choice analysis. *The Review of International Organizations*, *2*(3), 239–260, https://doi.org/10.1007/s11558-006-9010-x

Andrews, D.W.K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, *61*(4), 821–856, https://doi.org/10.2307/2951764

Auerbach, A.J. (1999). On the performance and use of government revenue forecasts. *National Tax Journal*, *52*(4), 767–782, https://doi.org/10.1086/ntj41789430

Beetsma, R., Giuliodori, M., Wierts, P. (2009). Planning to cheat: EU fiscal policy in real time. *Economic Policy*, *24*(60), 753–804, https://doi.org/10.1111/j.1468-0327.2009.00230.x

Brumby, J., & Hemming, R. (2013). Medium-term expenditure frameworks. R. Allen, R. Hemming, & B.H. Potter (Eds.), *The international handbook of public financial management* (pp. 219–236). London: Palgrave Macmillan.

Chow, G.C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, *28*(3), 591–605, https://doi.org/10.2307/1910133

Congreso de la República de Colombia (2003). *Ley 819 de 2003: Normas orgánicas en materia de presupuesto, responsabilidad y transparencia fiscal.* (Diario Oficial 45.243, July 2003. https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=13712. Accessed 27 July 2026)

Congreso Nacional de Chile (2003). *Ley 19.896: Introduce modificaciones al Decreto Ley 1.263 de 1975, Orgánico de Administración Financiera del Estado.* (Diario Oficial, September 2003. https://www.bcn.cl/leychile/navegar?idNorma=214205. Accessed 27 July 2026)

Davies, R.B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, *64*(2), 247–254, https://doi.org/10.2307/2335690

Dirección de Presupuestos (2026). *Informe de Finanzas Públicas.* Ministerio de Hacienda de Chile. (https://www.dipres.gob.cl/598/w3-propertyvalue-24862.html. Accessed 27 July 2026)

Eicher, T.S., & Kawai, R. (2024). Systemic bias of IMF reserve and debt forecasts for program countries. *International Journal of Forecasting*, *40*(3), 985–1001, https://doi.org/10.1016/j.ijforecast.2023.08.007

Filc, G., & Scartascini, C. (2011). *Is Latin America on the right track? An analysis of medium-term frameworks and the budget process* (IDB Working Paper). Washington, DC: Inter-American Development Bank.

Frankel, J.A. (2011). Over-optimism in forecasts by official budget agencies and its implications. *Oxford Review of Economic Policy*, *27*(4), 536–562, https://doi.org/10.1093/oxrep/grr025

Frankel, J.A., & Schreger, J. (2013). Over-optimistic official forecasts and fiscal rules in the eurozone. *Review of World Economics*, *149*(2), 247–272, https://doi.org/10.1007/s10290-013-0150-9

Hansen, B.E. (2001). The new econometrics of structural change: Dating breaks in U.S. labor productivity. *Journal of Economic Perspectives*, *15*(4), 117–128, https://doi.org/10.1257/jep.15.4.117

Holden, K., & Peel, D.A. (1990). On testing for unbiasedness and efficiency of forecasts. *The Manchester School*, *58*(2), 120–127, https://doi.org/10.1111/j.1467-9957.1990.tb00413.x

Jonung, L., & Larch, M. (2006). Improving fiscal policy in the EU: The case for independent forecasts. *Economic Policy*, *21*(47), 492–534, https://doi.org/10.1111/j.1468-0327.2006.00162.x

Mankiw, N.G., & Shapiro, M.D. (1986). Do we reject too often? Small sample properties of tests of rational expectations models. *Economics Letters*, *20*(2), 139–145, https://doi.org/10.1016/0165-1765(86)90161-8

Mincer, J., & Zarnowitz, V. (1969). The evaluation of economic forecasts. J. Mincer (Ed.), *Economic forecasts and expectations: Analysis of forecasting behavior and performance* (pp. 3–46). New York: National Bureau of Economic Research.

Ministerio de Economía y Finanzas del Perú (2026). *Marco Macroeconómico Multianual.* República del Perú. (https://www.mef.gob.pe/es/marco-macroeconomico/marco-macroeconomico-multianualmmm. Accessed 27 July 2026)

Ministerio de Economía y Finanzas del Uruguay (2026). *Rendición de Cuentas y Presupuesto Nacional.* República Oriental del Uruguay. (https://www.gub.uy/ministerio-economia-finanzas/rendicion-cuentas. Accessed 27 July 2026)

Ministerio de Hacienda y Crédito Público (2026). *Marco Fiscal de Mediano Plazo.* República de Colombia. (https://www.minhacienda.gov.co/politica-fiscal/documentos-planeacion-financiera/marco-fiscal-mediano-plazo. Accessed 27 July 2026)

Poder Ejecutivo del Perú (2016). *Decreto Legislativo 1276: Marco de la Responsabilidad y Transparencia Fiscal del Sector Público No Financiero.* (Diario Oficial El Peruano, December 2016. https://www.gob.pe/institucion/mef/normas-legales/166907. Accessed 27 July 2026)

Quandt, R.E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association*, *55*(290), 324–330, https://doi.org/10.2307/2281745

República Oriental del Uruguay (1996). *Constitución de la República, artículo 214.* (1967 text with the 1996 constitutional reform. https://www.impo.com.uy/bases/constitucion/1967-1967/214. Accessed 27 July 2026)

Schick, A. (2009). *Evolutions in budgetary practice: Allen Schick and the OECD Senior Budget Officials.* Paris: OECD Publishing.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*(4), 817–838, https://doi.org/10.2307/1912934

World Bank (2026). *World Development Indicators: GDP growth (annual %).* The World Bank Group. (Indicator NY.GDP.MKTP.KD.ZG. https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG. Accessed 27 July 2026)