International Journal of Forecasting 39 (2023) 431–449

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

The accuracy of IMF crises nowcasts☆

Theo S. Eicher [a,*], Yuan Gao Rollinson [b]

[a] University of Washington, United States of America
[b] International Monetary Fund, United States of America

ARTICLE INFO

Keywords:
Forecast evaluation
Economic crisis
Economic growth
IMF programs
Conditionality

ABSTRACT

The International Monetary Fund (IMF) provides loans to countries in economic crises as a lender of last resort. IMF loan approvals are tied to policy reforms and quantitative targets that reflect the IMF's crisis assessment. An extensive literature scrutinizes the efficacy of IMF loan programs, instead, we examine the accuracy of the IMF's crisis assessments (nowcasts) that predicate program designs. Analyzing an unprecedented 602 IMF loan programs from 1992 to 2019, we contradict previous findings that IMF nowcasts are generally optimistic. Disentangling the structure of the IMF's nowcast bias, we find the IMF systematically overestimates high-growth recoveries GDPs, while low-growth recoveries for low-income countries (LICs) are underestimated. In contrast, non-LICs' nowcasts exhibit no statistically significant optimistic and pessimistic bias. Interestingly, shorter nowcast horizons do not improve accuracy, and GDP growth nowcasts improved substantially since 2013, while inflation nowcasts remain inefficient. We also isolate the sources of IMF nowcast inefficiencies according to ((*i*) program objectives, ((*ii*) program conditionality type, ((*iii*) geographic regions, ((*iv*) global crises, and ((*v*) geopolitics (elections, conflicts, and disasters).

Most forecasts are wrong, but some are useful[1]

1. Introduction

The core mission of the International Monetary Fund (IMF) is to ensure the stability of the global economy through surveillance of member economies and by lending to crisis countries (IMF, 2020e). As global lender of last resort, IMF crisis loan programs stipulate policy conditions and quantitative economic targets, which must be implemented before loan tranches are disbursed (Fischer, 1999). IMF program design and conditionality can be contentious, as crisis countries often require substantial and often unpopular policy changes to ameliorate the economic problems that IMF nowcasts establish at the time of crisis.[2] IMF nowcasts thus constitute a unique dataset to evaluate the accuracy of the assumptions that constitute the basis for IMF program conditions.[3] Subsequent program performance evaluations and loan tranche

☆ We thank the IMF for access to the Monitoring of Fund Arrangement (MONA) database and to archived executive board loan program documents. Chris Papageorgiou, Charis Christofides, David Kuenzel, Justene Maldonado, Eric Maldonado, and two referees provided helpful suggestions. Chengjun Zhang, Liyuan Zhang, Reina Kawai, Elena Zhu, Ziqing Wang, and Xuanming Da provided excellent research assistance.
* Corresponding author.
E-mail address: te@uw.edu (T.S. Eicher).

[1] Variant of Box (1976) statistical aphorism.

[2] See e.g., Feldstein (1998) and Stiglitz (2002) who attribute slow crisis recoveries to IMF conditionality.

[3] IMF program designs assess current crisis conditions and forecast future outcomes. Program approvals require one-off forecasts that must rely on interim data in lieu of regular data vintages; these forecasts can be described as "nowcasts" following the definitions provided by *The Handbook of Forecasting* (Elliott & Timmermann, 2013) where a nowcast is defined as, an "*estimate of the current state of the economy. It constitutes a first important step in any forecasting exercise because macroeconomic data only become available with some time lag*" (p 273) and as "*the prediction of the present, the very near future and the very recent past. The term is a contraction for now and forecasting and has*

https://doi.org/10.1016/j.ijforecast.2021.12.007
0169-2070/

disbursements also depend on the benchmarks established by the original nowcasts. Thus, nowcast bias and/or inefficiency influence not only program scope and design but also the apparent magnitude and speed of recoveries (Park, 2006, Ch9).

Nowcasts are unbiased and efficient when they incorporate all relevant information available at the time that nowcasts are established (Nordhaus, 1987). There have been a number of evaluations of IMF forecasts, but previous evaluations focus largely on official final data from the IMF's World Economic Outlook (WEO) database, not the actual nowcast data established at the time of crisis.[4] Our dataset predates any WEO entries and represents a unique, long confidential snapshot of the IMF's understanding of economies in crises.

Eicher et al. (2019, EKPC from here forward) also formally examined the accuracy of IMF nowcasts using actual crisis data, not WEO data. They did not explore the possible sources of bias and inefficiencies, as we do below. EKPC also examined only 110 out of 602 available programs between 2002 and 2016, but they excluded a large share of observations due to concerns about outliers and database errors. We audit the IMF's *Monitoring of Fund Arrangements* (MONA) database and verify that about one-third of EKPC's excluded programs should have been included. Large outliers are simply a defining characteristic of countries requiring IMF crisis assistance. Our database audit allowed us to include an additional 492 programs and an additional 14 years of nowcast data.[5] To evaluate nowcasts, EKPC used estimates of final data, not official final data which differ substantially (see Fig. 1). Only 57% of the official final GDP growth data lie within 1% of the estimated GDP growth data, which implies EKPC's nowcast assessments are contaminated by errors in the final data.

EKPC found inflation nowcasts to be unbiased and efficient. Using the additional 492 programs and official (not estimated) final data, we find IMF inflation nowcasts are actually biased and inefficient. EKPC also found nominal GDP growth nowcasts unbiased and efficient, but our expanded dataset shows both nominal and real GDP growth nowcasts to be inefficient, systematically overestimating high-growth crisis recoveries and underestimating low-growth recoveries. Luna (2014) also studied IMF crisis forecasts in a sample of 94 program countries and found the optimism in IMF crisis forecasts was driven by countries with large loans. Our results in the substantially larger and longer sample do not confirm that either approved or drawn loan size affects nowcast bias or inefficiency. Luna (2014) also suggested that apparent optimism in IMF forecasts may simply reflect inadequate program execution on the part of the countries. We find no evidence that canceled programs affect nowcast efficiency.[6]

It is generally expected that forecasts improve as time horizons shorten and IMF-IEO (2014) established the same for WEO forecasts. In contrast, we find that this is not the case for IMF crisis countries nowcasts that are equally likely to be inefficient and/or biased for short or long-time horizons. Exploiting our 28-year time series, we can document that recent GDP growth nowcast accuracy has improved in recent years, while inflation nowcasts continue to struggle with inefficiency.

We also explore the sources of bias and inefficiency in IMF nowcasts, focusing on information IMF economists possessed at the time nowcasts were established. Several categories of covariates are examined to see if their effects are properly accounted for by IMF nowcasts. These categories cover regressors that relate to (*i*) international crises, (*ii*) geographic regions, (*iii*) conditionality (quantitative performance criteria (QPC) and structural performance criteria (SPC)), (*iv*) program objectives, (v) loan size and loan cancellation, and (v*i*) geopolitics (elections, conflicts, and natural disasters). Inflation, nominal, and real GDP growth nowcasts are shown to be associated with different sources of bias and inefficiencies. For inflation, we show that the effects of conditionality relating to ceilings on government and central bank credit are underestimated in IMF nowcasts. For real GDP, the effects of ceilings on government credit were overestimated. Nominal GDP nowcasts feature the largest number of conditionality effects improperly integrated into nowcasts, and these include ceilings on external debt/arrears, ceilings on government deficit/debt, and reforms of the current/capital accounts.

A voluminous literature describes a common theme of IMF forecasts being optimistic.[7] For crisis countries, we show that a focus on small samples and average forecast errors, without formal statistical tests, may provide misleading results. Our approach is novel in that our

been used for a long-time in meteorology and recently also in economics. Now-casting is relevant in economics because key statistics on the present state of the economy are available with a significant delay" (p 196). *The Oxford Handbook of Forecasting* (Clemens & Hendry, 2011) has the same definition but adds also "*Nowcasting is particularly relevant for those key macroeconomic variables which are collected at low frequency, typically on a quarterly basis, and released with a substantial lag*" (p.193), which is particularly relevant for IMF crisis nowcasts.

4 See among others: Aldenhoff (2007), Artis (1988, 1997), Barrionuevo (1993), Batchelor (2001), Beach and Schavey (1999), Loungani (2001), Pons (2000) and Timmermann (2007). Similar to Batchelor, Beach et al. find IMF WEO forecasts for developing countries overestimate real GDP growth and underestimate inflation, while forecasts for industrialized nations are unbiased and efficient. Artis and Loungani compare IMF forecasts to consensus forecasts and find no substantial differences. Neither of these papers focuses on crisis countries.

5 Appendix A1 documents our audit of the database, detailing 11 different types of errors and corrections.

6 Luna (2014) also suggests that IMF nowcasts may be overly optimistic because country authorities provided faulty data to the IMF; we have no data to test this hypothesis. Another approach to explaining optimistic IMF forecasts bias is based on institutional factors that we could not explore for lack of data: Genberg and Martinez (2014) find that the IMF desk economists with more experience produce smaller forecast errors. Beaudry and Willems (2022) find IMF mission chiefs' optimism systematically influences IMF forecasts.

7 Baqir et al. (2003) examine 29 MONA program countries to find positive bias, as did (Baqir et al., 2005) in a sample of 94 MONA countries. Atoyan and Conway (2011) examine fiscal and current account balances for 291 MONA program countries to find positive forecast errors. Luna (2014) examines real GDP growth and inflation forecasts for 103 MONA program countries and finds positive forecast bias for countries with "exceptional access to IMF resources". Neither paper employs formal tests for bias/efficiency. Genberg and Martinez (2014) and Timmermann (2007) find optimistic forecast bias using WEO data.

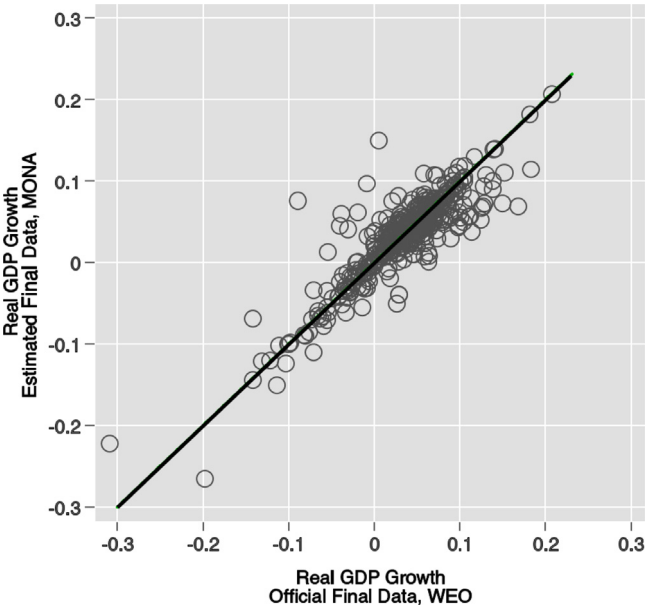

Fig. 1. Official final data (WEO) vs. estimates of final data (MONA).

large sample allows formal tests that decompose the perceived optimism to lay bare the structure of the nowcast inefficiency. The particular manner in which optimistic and pessimistic nowcasts average out for high and low-growth crisis recoveries is shown to contain important information. We show that IMF optimism is only a feature for high-growth recoveries in low-income countries (LICs), while slow recovering program countries actually suffer from excessively pessimistic IMF nowcasts. Once LICs are purged from the sample, we produce the important result that the remaining share of nearly 250 nowcasts is unbiased and efficient. For a full cross-country WEO sample that includes non-crisis countries, Loungani (2001) also reports that forecast accuracy differs by country income levels. Interestingly, he reports upward-biased forecast errors, indicating exactly the opposite result as we found in the crisis sample, in that he finds forecasts to be excessively pessimistic.

Overly optimistic nowcasts for high-growth recoveries may lead to an underestimate of the required financial support and quantitative/reform adjustments. Also, overly optimistic nowcasts for high-growth recoveries may translate into overly optimistic program targets that are too difficult to reach. LICs' nowcasts are shown to exhibit the greatest inefficiencies and by far the most optimistic nowcasts for high-growth recoveries. Most unsettling, Beaudry and Willems (2022) show that optimistic bias in IMF GDP growth forecasts can induce subsequent economic contractions through the greater accumulation of public and private debt. As the COVID-19 pandemic has raised the demand for IMF programs, improved nowcasts, and an understanding of the particular areas that produced nowcast inaccuracies are more important than ever. The key policy take-away from our results is clearly that LICs' nowcasts are highly problematic, suffering from profound inefficiency and bias that is at times strong enough to dominate the results for the full sample. Improved nowcasts procedures for this subset of countries are particularly important as these are also the countries in greatest need of COVID-19 financing (IMF, 2020g).

The remainder of the paper is organized as follows. Section 2 lays out the methodology, discusses the data, and provides the first results. Section 3 establishes nowcast bias and/or inefficiency across relevant subsamples. Section 4 examines why nowcasts have been inefficient and/or biased, and Section 5 investigates if nowcast efficiency changed. Section 6 checks if time horizons affected nowcast accuracy, and Section 7 concludes.

2. Evaluating nowcasts: Data and methodology

2.1. Data

Our nowcast data originates with the IMF's *MONA* database (IMF, 2020d). The database reports data that the IMF establishes at the time of crisis to design the loan program. When we last accessed the database, it covered 602 programs launched from 1992 to 2019. Each program is identified by program type, approval date, and loan size (approved and drawn loan amounts).[8] The database also reports macroeconomic indicators for program countries, including three years of historical data that predates the crisis year, t, the nowcast for t, and four additional years

[8] IMF (2020a) provides a full description of each of the 13 program types. These program types include Extended Credit Facility (ECF), Extended Fund Facility (EFF), Enhanced Structural Adjustment Facility (ESAF), Exogenous Shock Facility (ESF), Flexible Credit Line (FCL), Structural Adjustment Facility (SAF), Stand-By Agreements (SBA), Standby Credit Facility (SCF), Policy Coordination Instrument (PCI), Precautionary Credit Line (PCL), Precautionary Liquidity Line (PLL), Poverty Reduction and Growth Trust (PRGT), and Policy Reform Instrument (PSI).

of forecasts. Also reported are program conditionalities or performance criteria, which are grouped into 9 "quantitative" and 11 "structural" categories.[9] QPC refer to numeric benchmarks (e.g., fiscal deficit targets), SPC list policy, and institutional reforms (e.g., income tax reform) that must be implemented by each program country. We examine the MONA database's nowcasts established in crisis year t for program year t. Since IMF projections are based on growth rates, we examine nowcasts for real/nominal GDP growth and CPI inflation from $t-1$ to t.

For a number of reasons, nominal GDP, real GDP, and CPI are not directly dependent; as the GDP deflator (which is not consistently reported in MONA) and CPI inflation differ substantially. T-tests reject identical means at the $Pr(|T| > |t|) = 0.0002$ level, with similar levels of rejection for quantile means tests of the respective distributions. This is due to the IMF's Financial Programming approach (Barth et al., 2000), which combines real and nominal quantities in different ways, some based on CPI, and some based on GDP deflator. GDP, for example, is modeled mostly in real terms, but fiscal variables are in nominal terms (as CPI/wages matter). Also, GDP deflator forecasts are calculated based on the forecasted ratio of nominal GDP to real GDP, while the CPI is modeled as a function of completely different factors (e.g., past CPI, reserves, money growth, international food commodity prices, and exchange rates), without taking into account nominal/real GDP forecasts. GDP deflator forecasts are thus (indirectly) aggregating the deflators in consumption, investment, exports, and imports, while the CPI captures only price changes in consumption goods. In addition, Easterly (2006) notes that the nature of the IMF Financial Programming model implies that its equation and even identities "produce large statistical discrepancies, which weakens the case for them as consistency checks". Easterly notes that two related quantities may diverge substantially because some equations contain variables that take policy as endogenous along with other variables that assume policy as exogenous. Our approach to examining real GDP, nominal GDP, and inflation forecasts accuracy has a long history in this literature that dates back to the earliest studies in this literature (Kenen & Schwartz, 1986; and Artis, 1988) all the way to the most recent study (EKPC, 2019).

The MONA database also associates each program with IMF Executive Board documents which became available when IMF archives opened in 2009. We review these documents in our MONA audit and provide a surprisingly extensive list of database errors in Appendix A. After the start of each program, IMF economists review country performance at regular intervals (monthly, quarterly, and semi-annually) and enter estimates of realized data. EKPC used data from these subsequent reviews to obtain estimates of final data for t. There are four reasons why this approach is unnecessarily restrictive and inaccurate. First, 48 programs do not have estimates of final data in MONA and were thus excluded by EKPC analysis, although official final data exists. Second, the EKPC approach constrained the analysis to programs that lasted at least 18 months, omitting another 14 programs. Third, in fear of database errors, EKPC excluded another 106 programs when data were unbalanced or when observations exceeded four standard deviations from the mean. Fourth, MONA's estimates of final data can differ substantially from the IMF's official final data, which is reported in the IMF's WEO database (IMF, 2020f), as illustrated in Fig. 1. These differences between MONA's estimated final data and WEO's official final data contaminate EKPC's assessment of nowcast accuracy, as errors in MONA's estimated final data are attributed to nowcast inaccuracy. Therefore, we evaluate IMF nowcast accuracy based on official final data obtained from the IMF's WEO.

2.2. A methodology to evaluate nowcast accuracy

Forecast evaluation exercises often focus on a range of forecast-error summary statistics that compare the performance of different forecasts, for example, the mean absolute error (MAE) and the root mean square error (RMSE).[10] Such statistics are informative only when two or more forecasts are being compared. The forecast evaluation in this paper involves, however, only a single forecast, since only the IMF has access to country crisis data; and the loan documents remained largely confidential until 2009. Hence, we cannot compare crisis nowcasts from different sources and instead focus on understanding the accuracy of the only available nowcast. The importance of these IMF nowcasts is thus also derived from the fact that IMF nowcasters hold a unique position as they enjoy exclusive access to confidential country data in times of crises.[11]

The literature on forecast accuracy evaluation dates back to Mincer and Zarnowitz (1969), who based their analysis on the seminal work of Theil (1961). Theil introduced the concept of a "*prediction-realization diagram*" (Fig. 2), which displays IMF nowcasts, F_t, on the horizontal axis and official final data, A_t, on the vertical axis. IMF nowcasts are established at the time of crisis when the program is designed and approved, t, for the year of the crisis. We examine real/nominal GDP growth and inflation nowcasts. Mincer and Zarnowitz label the solid 45-degree line in the prediction-realization diagram the "*line of perfect forecasts*" as it represents coordinates where nowcasts equal the official final data, $F_t = A_t$. For real GDP growth, Fig. 2 indicates that IMF nowcasts overestimate high-growth recoveries and underestimate

[9] See Table 1 for conditionality categories.

[10] A review of these statistics is provided by Hyndman and Athanasopoulos (2018).

[11] IMF programs can include up to 4-year forecasts; we examine only nowcasts produced for the crisis year, t, at the time of crisis in year t. IMF crisis nowcasts are conditional on the assumption that IMF conditionality is fulfilled. Rational expectations forecasts would also account for the probability that conditionality is not fulfilled. Since IMF loans are contracts (or "arrangements" in IMF parlance) between governments and the IMF, and since IMF conditionality is explicitly outlined in these contacts, and since the IMF is the lender of last resort, one might expect implementation of conditionality to be generally high. Some programs do get cancelled due to non-performance; we reran all tables without cancelled programs and find cancelled programs do not explain systematic nowcast bias or inefficiency. Result available upon request.

Table 1
Variable list, sources and descriptions.

Variable	Data Source	Description
IMF Nowcasts	IMF (2020d) MONA Database. See Appendix A for details	"t-1" to "t" peroid growth rates for real GDP levels (RGDPC/NGDP_R), nominal GDP levels (NGDP), end-of-period inflation (PCPIC/PCPIE)
Final Realized Data	IMF (2020f) WEO Database	"t-1" to "t" peroid growth rates for real GDP levels (NGDP_R), nominal GDP levels (NGDP), end of period inflation (PCPIE)
2008 Crisis		Dummy variable for the 2008 Global Financial Crisis. Program received a "1" when program commenced between 9/15/2018 and 9/15/2019 , where the start date is the Lehman Brothers' bankruptcy filing date.
2007 Crisis		Dummy variable of the 1997 Asian Crisis. Program received a "1" for programs that commenced between 7/2/1997 and 1/1/1999 ; and (ii) the country was also identified as an Asian Crisis Country by Kaminsky et al. (2003). In July 1997, Thailand was forced to float exchange rate, which is generally seen as the start of the crisis.
Regions		Dummy variables for Africa, Americas, Asia, and Europe.
Quantitative Conditionality: QPCs	IMF (2020b) MONA Database Glossary	Quantitative conditionality. Dummy variables given by MONA's "Main Criteria" defined by MONA Glossery.
Structural Conditionality: SPCs	IMF (2020b) MONA Database Glossary	Structural conditionality. Dummy variables given by MONA's "Indicative Targets" defined by MONA Glossery. We grouped them as: Gen_Gov't_Reform (1.1-1.9), CB_Stats_Regs_Indep.(2.1-2.2), Civil_Service_Wage/Empl. (3), Pension_Reform (4.1-4.2), Gov't_Enterprise_Pricing (5.1-5.3), Financial_Sector_Reform (6.1-6.2), Open_Current&Capital_Account (7), Reduce_Trade_Tariff/Quota (8), Labor_Mkt_Wage/Empl.(9), Improve_Econ_Statistics (10), Legal/Market_Reforms (11.1-11.4)
Program Objectives	IMF (2020a) Crisis Lending Fact Sheet	We grouped 13 IMF programs types into 5 program objectives: (i) BOP_Stablization: Stand-By Agreements (SBA); (ii) BOP_shocks_precautionary: Exogenous Shock Facility (ESF), Standby Credit Facility (SCF), Flexible Credit Line (FCL), Precautionary Credit Line (PCL), Precautionary Liquidity Line (PLL); (iii) Struct_Adj_Poverty_Growth: Structural Adjustment Facility (SAF), Enhanced Structural Adjustment Facility(ESAF), Poverty Reduction and Growth Trust (PRGT); (iv) long-term BOP and structural reform assistance: Extended Credit Facility (ECF), Extended Fund Facility (EFF); (v) Non_Financial_Reforms: Policy Reform Instrument (PSI), Policy Coordination Instrument (PCI).
LoanAmount	IMF (2020d) MONA Database for loan amount drawn, IMF (2020c) IFS Database for quota	Loan Amount is the ratio of drawn loan size to IMF country annual quota in the program start year.
Elections	Beck et al. (2001) for data before 1998; IFES (2020) for post 1998 data	Election dummy cover two types of national elections: (i) head of state or government election; (ii) legislative election. Program received a "1" if an election occurred up to 1 year prior to the program start date.
Conflicts	UCDP/PRIO Armed Conflict Dataset by Harbom et al. (2009)	Conflict dummy covers intra-state and inter-state conflicts. Program received a "1" if program country experienced a conflict up to one year prior of program start date.
Disasters	EM-DAT (2020)	Disaster dummy covers natural disasters. Program received a "1" if a disaster occurred up to 1 year prior to the program start date.

low-growth recoveries. A similar pattern exists for nominal GDP growth, but we suspect the finding is obscured by influential observations, which we examine further below. The white shaded area around the dashed regression line represents the 95% confidence interval.

Our test of nowcast bias and efficiency is based on Mincer and Zarnowitz regressions, which have been frequently applied: see e.g., Romer and Romer (2000), Rossi and Sekhposyan (2011), or Granger and Newbold (2014). The identity $A_t \equiv F_t + \mu_t$, which includes the forecast error, μ_t, produces the regression

$$A_t = \alpha + \beta F_t + \varepsilon_t. \quad (1)$$

Nowcasts are conventionally chosen as the "independent" variable because they are available before the official final data is published. Mincer and Zarnowitz (1969) point out that the regression slope, β, equals unity only when the forecast error, μ_t, is uncorrelated with the forecast values,

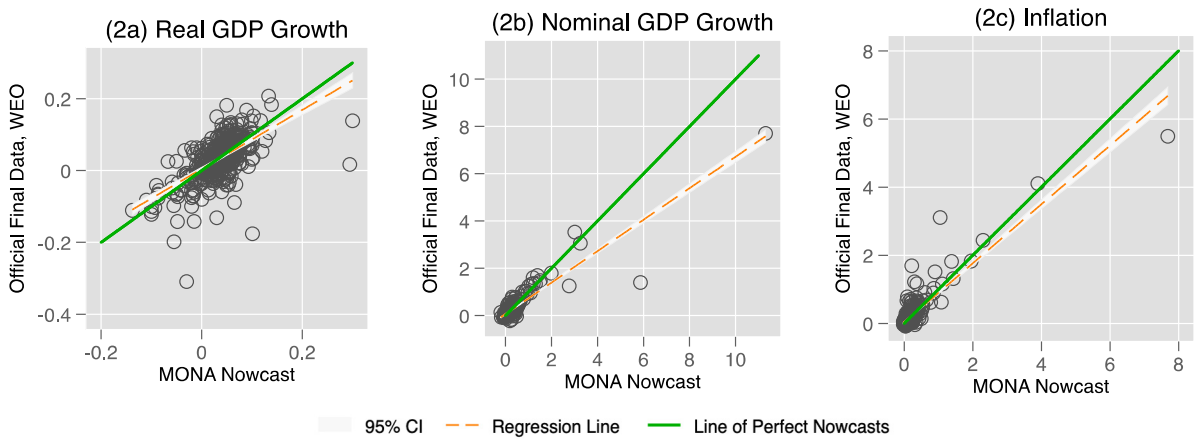

Fig. 2. Prediction-realization diagrams (full sample).

F_t which is when the residual variance of the regression, $\sigma(\varepsilon)$, equals the variance of the forecast error, $\sigma(\mu)$. In this case, the forecast is efficient. If we also have $\alpha = 0$, the forecast is unbiased. To test the accuracy of forecasts, Mincer and Zarnowitz suggest the joint null hypothesis $\alpha = 0$ & $\beta = 1$ for unbiased and efficient nowcasts. Since estimates of α and β are generally correlated, individual T statistics are insufficient, and the joint test is required (Wallis, 1989).

If the Mincer and Zarnowitz null of $\alpha = 0$ & $\beta = 1$ is rejected, nowcasts are inefficient but they may or may not be biased. Holden and Peel (1990) derived a necessary and sufficient condition for unbiased nowcasts, which simply tests whether the regression line intersects the *Line of Perfect Nowcasts* at the respective expected values, $E(A_t) = E(F_t)$. When the Holden and Peel test of $A_t - F_t = \gamma + \nu_t$ rejects the null of $\gamma = 0$, the nowcast is then said to be biased. Here, it is important to note that inefficiency may be more informative than bias since a nowcast exhibits no bias at all when it is half of the time 20% higher and half of the time 20% lower than the official final data. This is why Mincer and Zarnowitz stress that unbiasedness may be desirable, but not by itself informative about forecast accuracy. Of course, other things being equal, the smaller the bias, the greater the accuracy of the forecast; "other things" here being the distances between the points in the prediction-realization diagram and the Line of Perfect Nowcasts. These distances can be expressed by the variance of the forecast error around its mean, which Mincer and Zarnowitz introduce as an inverse measure of forecast efficiency. Rotations of regression lines to better match lines of perfect forecasts reduce this variance to increase efficiency.

Nowcasts are inefficient when they do not incorporate all available information. That is why (Nordhaus, 1987) noted that forecast efficiency shares similarity with stock market efficiency as both concepts imply that efficiency exists only when all relevant and available information was considered. Even if the Holden and Peel test indicates unbiased forecasts, the slope coefficient, β, in (1) may indicate a statistically significant deviation from unity to suggest inefficiency. Tables 2 and 3 present the results of our Mincer and Zarnowitz regressions. Diagnostic tests for all regressions show no concerns regarding heteroscedasticity or serial correlation in residuals, and we address potentially non-normal residuals by using robust standard errors.

Regressions (1a)–(4c) in Table 2 are Mincer and Zarnowitz regressions associated with Fig. 2 for real/nominal GDP growth and inflation. In regressions (1a)–(1c), we revisit EKPC results based on their sample of 110 out of 602 available programs, estimated final data, and excluded programs due to program duration and data variation. The regressions show that EKPC's sample rejects efficiency only for real GDP growth, while all nowcasts are indicated as unbiased.[12] In regressions (2a)–(2c), we rerun the EKPC sample with our corrected data and with actual final values rather than estimated final values. Results do not change substantially, although we start to see the first signs of inefficiency (for nominal GDP growth) and bias (marginally significant, for real GDP growth and inflation at 10% level). Regressions (3a)–(3c) represent all available data for the EKPC time period, including audited data, and programs EKPC had omitted due to program duration or because program data exceeded four standard deviations from the mean. All nowcasts are unbiased, and only nominal GDP is inefficient at the 1% significance level. In summary, EKPC results and our results for EKPC's time period (with audited, verified, and final outcome data) are different for real and nominal GDP, but similar for inflation.

Regressions (4a)–(4c) include all programs with all available, audited data, for all years (1992–2019), with the official (not estimated) final data, and without eliminating programs due to program duration or data variance. Real GDP growth nowcasts are now shown to be inefficient but unbiased with a slope coefficient that is significantly below unity. This suggests low (high) growth recoveries are under (over) estimated in IMF nowcasts.

[12] In contrast to EKPC, we do not include regional or crisis dummies in the benchmark; the inclusion of such dummies invalidates the Mincer and Zarnowitz null hypothesis. The question of whether IMF nowcasts do or do not properly integrate regional or global crisis information is examined in Section 4 using a different test.

Table 2
Bias and inefficiency of IMF nowcasts.

Dependent variable:	Actual Real GDP Growth				Actual Nominal GDP Growth				Actual Inflation			
	(1a)	(2a)	(3a)	(4a)	(1b)	(2b)	(3b)	(4b)	(1c)	(2c)	(3c)	(4c)
	EKPC (2019)	Our data, EKPC sample	Our data, EKPC time period	All available, audited data	EKPC (2019)	Our data, EKPC sample	Our data, EKPC time period	All available, audited data	EKPC (2019)	Our data, EKPC sample	Our data, EKPC time period	All available, audited data
Constant, α	0.016**	0.014**	0.010*	0.004	0.014*	0.025**	0.020***	0.062***	0.002	0.006	0.006	0.038***
p-value ($\alpha = 0$)	0.013	0.012	0.098	0.206	0.071	0.011	0.002	0.000	0.875	0.663	0.420	0.003
IMF Nowcast, β	0.621***	0.783**	0.753	0.821**	0.926	0.832**	0.864**	0.666***	1.091	1.042	0.982	0.864
p-value ($\beta = 1$)	0.010	0.045	0.136	0.044	0.193	0.032	0.018	0.000	0.702	0.863	0.890	0.185
Observations	110	110	269	597	110	110	270	596	100	110	268	595
Adjusted R-square	0.402	0.482	0.403	0.404	0.742	0.655	0.638	0.836	0.545	0.553	0.566	0.810
MZ F-test ($\alpha = 0$, $\beta = 1$)	3.460**	3.410**	1.414	2.731*	1.740	3.356**	4.820***	13.93***	1.021	1.827	1.813	8.258***
p-value ($\alpha = 0$, $\beta = 1$)	0.035	0.037	0.245	0.066	0.180	0.039	0.009	0.000	0.364	0.166	0.165	0.000
HP T-test ($\gamma = 0$)	−0.111	1.816*	0.227	−1.464	1.180	1.100	1.281	−0.946	1.408	1.681*	1.347	2.595**
p-value ($\gamma = 0$)	0.912	0.072	0.820	0.144	0.240	0.274	0.201	0.344	0.162	0.096	0.179	0.010
Diagnostic Tests												
H0: Heteroskedasticity	11.18***	5.005*	145.5***	62.170***	0.362	4.106	8.394**	240.740***	28.52***	57.53***	109.8***	77.020***
H0: No serial correlation in residuals	0.687	1.595	0.289	0.323	6.621**	2.853*	2.297	0.002	2.219	7.381***	3.070*	0.029
H0: Normal residuals	0.848***	0.960***	0.890***	0.859***	0.936***	0.966***	0.918***	0.529***	0.877***	0.879***	0.838***	0.496***

Notes: Robust standard errors in parentheses unless otherwise indicated; $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Mincer and Zarnowitz (MZ) null: nowcast is unbiased and efficient.
Holden and Peel (HP) null: nowcast is unbiased.
Heteroskedasticity is assessed using the Cameron-Trivedi test; serial correlation is assessed using the Breusch–Godfrey test; normality is assessed using the Shapiro–Wilk W test. EKPC regressions in 1a-c included region dummies. MAE (mean absolute error) and RMSE (root mean square error) are scaled by 100.

Table 3
Bias and inefficiency of IMF nowcasts by subsample.

Dependent variable:	Real GDP Growth					Nominal GDP Growth					Inflation				
	(1a)	(2a)	(3a)	(4a)	(5a)	(1b)	(2b)	(3b)	(4b)	(5b)	(1c)	(2c)	(3c)	(4c)	(5c)
	All	Non-Hyper	Non-Hyper LICs	Non-Hyper Non-LICs	Hyper Inflation	All	Non-Hyper	Non-Hyper LICs	Non-Hyper Non-LICs	Hyper Inflation	All	Non-Hyper	Non-Hyper LICs	Non-Hyper Non-LICs	Hyper Inflation
Constant, α	0.004	0.008**	0.020***	0.003	−0.008	0.062***	0.003	0.011	−0.004	0.188***	0.038***	−0.004	0.002	−0.010	0.197***
p-value ($\alpha = 0$)	0.206	0.010	0.001	0.390	0.234	0.000	0.704	0.520	0.665	0.000	0.003	0.658	0.929	0.168	0.001
IMF Nowcast, β	0.821**	0.775***	0.601***	0.835	0.801	0.666***	1.020	0.987	1.052	0.624***	0.864	1.242	1.262	1.210	0.792**
p-value ($\beta = 1$)	0.044	0.002	0.001	0.145	0.437	0.000	0.798	0.919	0.551	0.000	0.185	0.156	0.407	0.129	0.023
Observations	597	526	276	250	71	596	525	276	249	71	595	526	275	251	69
Adjusted R-square	0.404	0.353	0.155	0.486	0.383	0.836	0.665	0.625	0.699	0.834	0.810	0.309	0.207	0.595	0.822
MZ F-test ($\alpha = 0 \& \beta = 1$)	2.731*	4.841***	5.685***	1.362	0.988	13.93***	1.685	2.039	0.202	11.26***	8.258***	6.062***	6.222***	1.190	5.8***
p-value ($\alpha = 0 \& \beta = 1$)	0.066	0.008	0.004	0.258	0.377	0.000	0.186	0.132	0.818	0.000	0.000	0.003	0.002	0.306	0.005
HP T-test ($\gamma = 0$)	−1.464	−0.882	−0.345	−1.001	−1.334	−0.946	1.757*	1.834*	0.508	−1.525	2.595**	2.997***	2.692***	1.345	1.135
p-value ($\gamma = 0$)	0.144	0.378	0.730	0.318	0.187	0.344	0.079	0.068	0.612	0.132	0.010	0.003	0.008	0.180	0.260
Diagnostic Tests															
H0: Heteroskedasticity	62.17***	13.86***	7.636**	43.06***	12.78***	240.7***	46.52***	35.18***	18.10***	34.15***	77.02***	26.30***	22.86***	74.21***	5.072*
H0: No serial correlation in residuals	0.323	0.002	0.413	0.319	1.764	0.002	0.200	1.413	1.363	0.400	0.029	0.293	0.016	0.010	0.041
H0: Normal residuals	0.859***	0.902***	0.860***	0.951***	0.838***	0.529***	0.821***	0.842***	0.779***	0.729***	0.496***	0.569***	0.567***	0.785***	0.750***
St Dev: Actual Data	0.050	0.043	0.040	0.042	0.077	0.426	0.137	0.141	0.129	1.051	0.382	0.122	0.145	0.089	0.893
St Dev: Nowcast Error	0.039	0.035	0.038	0.031	0.062	0.261	0.079	0.086	0.071	0.718	0.175	0.102	0.130	0.058	0.430
MAE	2.457	2.264	2.406	2.108	3.882	7.202	4.747	5.285	4.150	25.355	6.482	4.512	5.789	3.111	21.506
RMSE	3.914	3.496	3.838	3.075	6.186	26.089	7.953	8.660	7.088	72.427	17.579	10.286	13.107	5.787	43.108

Notes: Robust standard errors in parentheses unless otherwise indicated; $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Mincer and Zarnowitz (MZ) null: nowcast is unbiased and efficient.
Holden and Peel (HP) null: nowcast is unbiased.
Heteroskedasticity is assessed using the Cameron–Trivedi test; serial correlation is assessed using the Breusch–Godfrey test; normality is assessed using the Shapiro–Wilk W test. MAE (mean absolute error) and RMSE (root mean square error) are scaled by 100.

Nominal GDP growth, using the full dataset, is now also inefficient but unbiased with a substantially smaller, significant slope coefficient that is far below unity (0.666). This again suggests nominal GDP nowcasts underestimate (overestimate) low (high) growth recoveries.

Inflation shows the greatest divergence from previous results. EKPC found near-perfect, unbiased, and efficient IMF inflation nowcasts with near unitary slope and zero intercept, suggesting that nowcasts closely match the Line of Perfect Nowcasts. When we include the previously omitted programs, as well as official (not estimated) final data, along with additional years of available data, we find IMF inflation nowcasts exhibit statistically significant bias and inefficiency. As was the case for GDP growth, IMF nowcasts for inflation under (over) estimate low (high) inflation recoveries.

The fact that high GDP growth crisis recoveries are overly optimistic nowcast in the full sample presents major problems for IMF crisis program countries. First, overly optimistic nowcasts for high-growth recoveries imply underestimated loan requirements and mismatched quantitative and structural adjustment targets. On the other hand, overly pessimistic nowcasts for the low-growth recoveries overstate programs' financial needs, which may lead to misallocation of resources. Second, overly optimistic nowcasts can translate into overly optimistic program targets and performance criteria (e.g., government revenues or import volumes) that may be difficult or impossible to reach. Third, overly optimistic nowcasts for high-growth recoveries affect program evaluations, as seemingly below-par performances may in fact be due to excessively optimistic nowcasts.

3. Does nowcast bias and efficiency vary by subsample?

Fig. 2(b)–(c) suggest that the assessment of IMF nowcast accuracy is impacted by influential observations, likely relating to nowcast inaccuracies in high inflation countries. Below we examine whether IMF nowcast accuracy differs by subsamples, specifically subsamples that differ by income levels and inflation. For country income groupings, we use the World Bank's time-variant LIC demarcation,[13] and for inflation, we use Dornbusch and Fischer's (1986) threshold of inflation > 25% to identify hyperinflation crises. Note that our hyperinflation sample is best described as *anticipated hyperinflation countries* since we code countries for hyperinflation only if IMF nowcasters predicted hyperinflation. Hence, we evaluate only how accurate IMF nowcasts are for countries and crises for which the IMF expected hyperinflation.[14]

Separating the samples by hyperinflation countries and LICs, we obtain the subsample prediction-realization diagrams in Figs. 3.1–3.3, which shows a substantially improved fit and nowcasts for the non-hyperinflation-non-LIC subsamples (as compared to Fig. 2). Table 3 reports Mincer and Zarnowitz regressions for the subsamples. For real GDP growth, nowcast accuracy differs substantially across subsamples. The full sample, the non-hyperinflation sample, and the non-hyperinflation LIC sample are all inefficient. Efficiency arises only once we purge the full sample of LICs and hyperinflation nowcasts (see regression (4a)). This implies that the inefficiency in the full sample is driven by nowcasts for hyperinflation countries and LICs.

Slope coefficients are below unity for all subsamples, indicating overly optimistic nowcasts for high real GDP growth recoveries and overly pessimistic estimates for low-growth recoveries. This effect is most pronounced for non-hyperinflation LICs with an astonishing low β of 0.6, indicating enormous improvement potential for IMF nowcasts for this subsample of 276 programs. By calculating the intersection between the regression line in (1a) and the Line of Perfect Nowcasts, we find that fragile non-hyperinflation LICs (LICs with below 5% real GDP growth recoveries), may have been negatively impacted by conditionality and performance evaluations that were based on excessively pessimistic IMF crisis nowcasts.

Surprisingly, IMF real GDP growth nowcasts for hyperinflation countries are not statistically significantly biased or inefficient. This is likely an artifact of the high variance in the official final data for hyperinflation countries (see statistics for (5a) in Table 3), which is twice the magnitude of other subsamples variances. High variance outcomes are harder to predict, resulting in substantially larger nowcast errors, as reported in Table 3. The large standard errors in the hyperinflation subsample then widen the confidence bands to the point where the Mincer and Zarnowitz F-test cannot rule out that intercept and slope coefficients are consistent with the Line of Perfect Nowcasts (see Fig. 3.3a).

It is important to note that the majority (55%) of hyperinflation events occurred before 1997. In that sense, the noise introduced by hyperinflation nowcast errors is not representative of ongoing nowcast dynamics over the years in the full sample. This contrasts the effects of LICs' nowcast errors, whose contribution to the inefficiency of the full sample inefficiency is steady over the entire 28-year time period.

For nominal GDP growth nowcasts in hyperinflation countries, we find a pattern similar to what we observed for real GDP growth. Table 3 shows that nowcast inefficiency of nominal GDP growth in the full sample is driven entirely by nowcast inaccuracies of hyperinflation countries. Once hyperinflation countries are excluded the nowcasts are unbiased and efficient. Both the hyperinflation and nowcast slope coefficients are highly significant, but the hyperinflation coefficient is even lower (at 0.624) than the full sample's (0.666). This indicates again that IMF nowcasts substantially overestimate nominal GDP growth in "high hyperinflation" recoveries and underestimate nominal GDP growth in "low hyperinflation" recoveries. Once we purge the full sample of hyperinflation

[13] Low-income country classification is a time-varying measure based on World Bank's data of GNI per capita in each year. See http://databank.worldbank.org/data/download/site-content/OGHIST.xls.

[14] If hyperinflations or policies leading up to hyperinflations were a surprise to IMF nowcasters (and hence not part of their information set), we would not want to link such surprises to IMF nowcast bias and inefficiency.

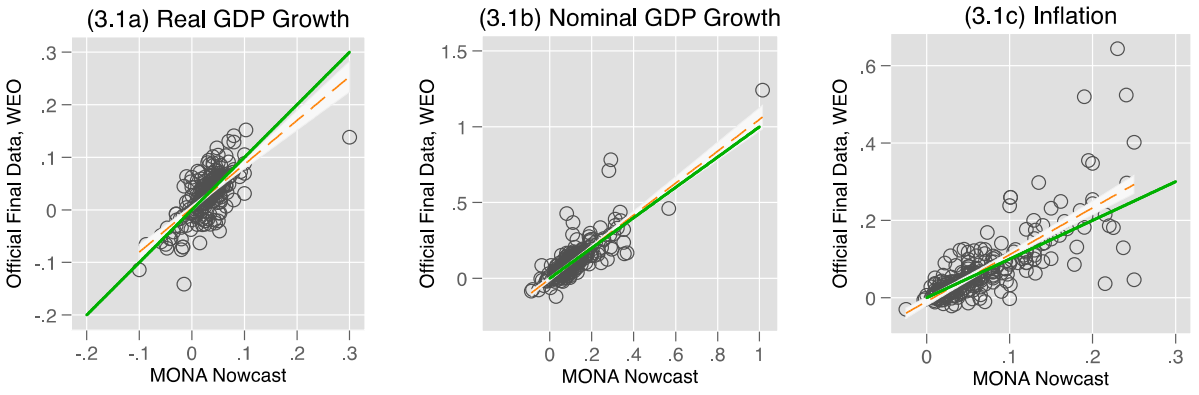

Fig. 3.1. Prediction-realization diagrams. Non-hyperinflation-non-low-income countries.

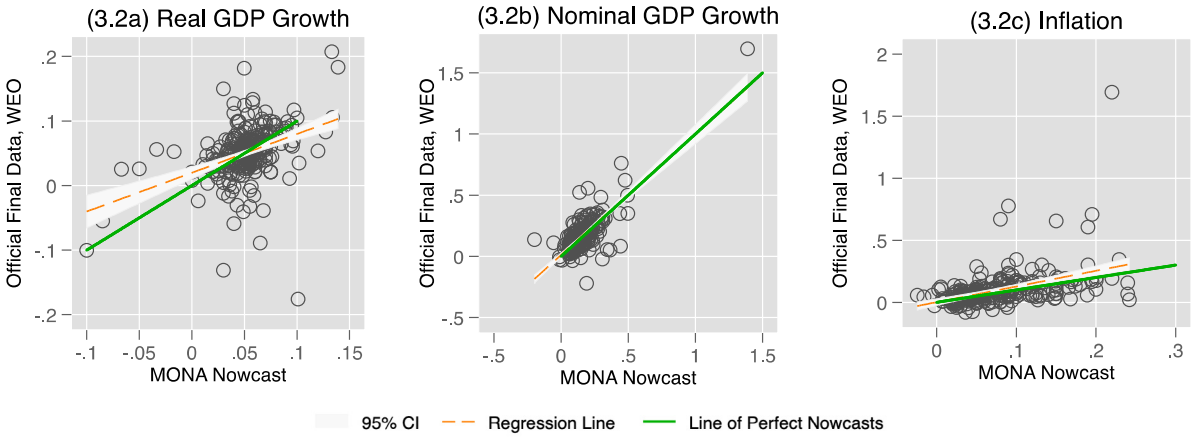

Fig. 3.2. Prediction-realization diagrams. Non-hyperinflation-low-income countries.

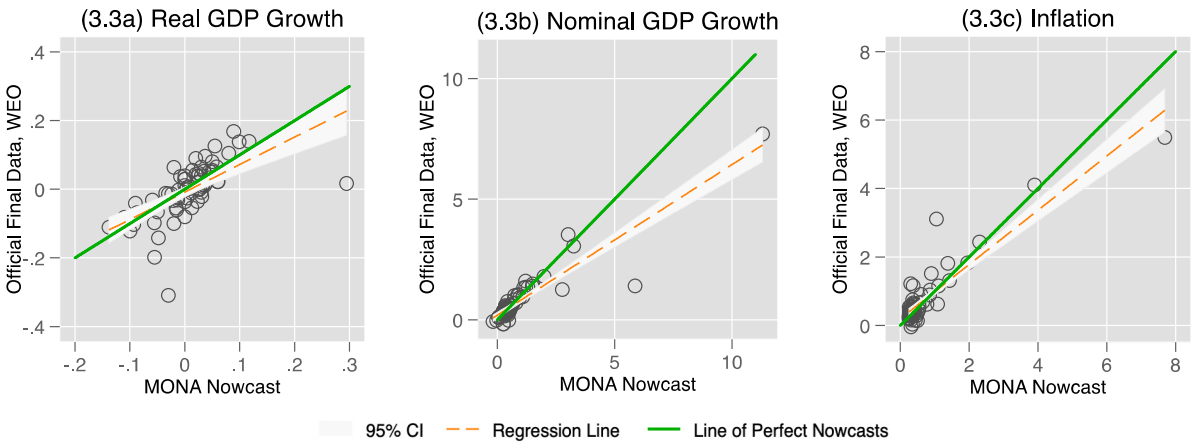

Fig. 3.3. Prediction-realization diagrams. Hyperinflation countries.

countries, we find slope coefficients near unity across subsamples (ranging from 0.987 to 1.052), indicating efficient nowcasts.

Interestingly, both the non-hyperinflation and the non-hyperinflation LIC subsamples exhibit biased nominal GDP growth nowcasts (regression (2b) and (3b)). This is because the positive intercept together with high slope coefficients imply that optimistic bias program nowcasts and pessimistic bias program nowcasts cannot average out. Here, it is important to note that the bias is driven entirely by LICs. Once LICs are removed from the non-hyperinflation sample, the slope is just about unity (1.052)

and nowcasts are unbiased and efficient (see regression(4b) in Table 3). In summary, for nominal GDP growth, we find that the full sample must exclude (*i*) hyperinflation countries to achieve efficiency and (*ii*) LICs to achieve unbiased nowcasts.

Inflation nowcasts produce the largest accuracy variations across subsamples. Biased and inefficient nowcasts are observed for the full sample, the non-hyperinflation sample, and the non-hyperinflation LIC samples. As in the case of nominal and real GDP growth, we must purge hyperinflation and LIC programs from the full sample to obtain efficient and unbiased nowcasts. Slope coefficients are high and even exceed unity for non-hyperinflation samples (LICs or Non-LICs), implying that IMF nowcasts actually under-predict inflation for high inflation recoveries (greater than 1.7%) and over predict inflation for low inflation events. Only the hyperinflation subsample exhibits a low slope coefficient (0.792), indicating that IMF nowcasts decisively overestimate "high hyperinflation" recoveries and underestimate "low hyperinflations" recoveries from economic crises.

The important finding of our subsample analysis is that non-hyperinflation-non-LIC programs (about 42% of the sample) exhibit unbiased and efficient nowcasts for real and nominal GDP growth and for inflation. Hyperinflation countries contribute to bias and inefficiency, but given that their numbers are concentrated in a few early years in the 28-year sample, the real drivers of bias and inefficiency in the full sample are non-hyperinflation LICs. The results raise the question as to the drivers of inefficiency, which we explore in the next section.

4. Why are IMF nowcasts inefficient?

Sinclair et al. (2010, SJS from here forward) and Sinclair et al. (2012) propose a methodology to investigate potential sources of forecast inefficiencies. They suggest including additional covariates in (1) that represent information available to forecasters.

$$A_t = \alpha + \beta F_t + \delta X_t + \varepsilon_t, \tag{2}$$

where X_t are additional candidate covariates known to forecasters at the time of the forecast. SJS propose the joint null hypothesis of $\beta = 1$ & $\alpha = \delta = 0$ to identify whether the information content of the additional covariates is properly included in the nowcast. If the null is rejected, SJS note that the information contained in X was not fully integrated into the nowcast, which then identifies possible sources of inefficiency.

The approach is helpful here because it allows us to understand whether IMF nowcasts (fully) incorporate the effects of key pieces of information that are known to nowcasters at the time the program is designed. How quickly and accurately the news of global contagion or specific program conditionality is integrated into the nowcasts is an important determinant of the bias and inefficiency generated by deviations from final outcomes. Gultekin et al. (2006) study forecast revisions to highlight how inadequate consideration of news drives forecast inefficiency. While we do not examine forecast revisions in our short time horizon, we do want to understand if particular crisis dimensions contribute to nowcast bias and inefficiency.

Both EKPC (2019) and Luna (2014) include examinations along similar lines. EKPC use the national income identity to attribute the GDP forecast error to forecast errors in the identity's subcomponents. Luna (2014) examined forecasts for government balance and current account balance variables in addition to GDP and inflation. He suggests that predictions might be inaccurate not because of faulty forecasting—but because the country did not fulfill all conditionality. A simple way to test this is to include a dummy for canceled programs in (2). More interestingly, it would be to examine deviations from the agreed and achieved conditionality. However, there is no complete database on IMF conditionality, although the IMF does provide a list of detailed policy areas in which a country received conditionality. We use these data below.

We examine six groups of candidate covariates to test whether they were properly accounted for in IMF nowcasts, specifically, (*i*) international crises, (*ii*) regions, (*iii*) conditionality (QPC and SPC), (*iv*) program objectives, (v) loan size (approved and drawn amount) and loan cancellation, and (*vi*) geopolitics (elections, conflicts, and natural disasters). Two major international crises occurred during the years covered by our sample (1997 Asian crisis, 2008 global financial crisis), and one might suspect that the effects of global contagion were difficult to integrate into IMF country nowcasts at the time. Genberg and Martinez (2014) and IMF-IEO (2014) find that IMF WEO forecasts tend to be consistently over-optimistic in times of regional and global recessions. Here, it is important to emphasize that we are examining only sources of inefficiencies based on information available to IMF nowcasters at the time of the nowcast. Hence, our crisis dummy is set to one only for countries whose programs started *after* the 1997 and 2008 crises commenced.[15]

Regional effects for Africa, Asia, and Latin American dummies commonly hold explanatory power in growth regressions; hence, we investigate if nowcasts fully account for these.[16] Another set of regressors we investigate as possible sources of inefficiencies relates to IMF conditionality. Conditionality is specifically designed to affect program countries' recoveries; hence, nowcasts must exercise particular caution to integrate their effects. For example, conditionality often relates to fiscal discipline and credit targets that directly affect GDP growth and/or inflation. The IMF classifies conditionality as "QPC" (e.g., "dollar ceiling on external debt") and "SPC" (e.g., policy reforms). Table 1 reports how MONA groups conditionality into 9 categories of QPCs and 11 categories of SPCs, which we can add to regression (1).

[15] The 1997 Asian crisis dummy received a "1" indicator for programs that (*i*) commenced between 7/2/1997 and 1/1/1999 and (*ii*) that were identified as affected Asian crisis countries by Kaminsky et al. (2003). In July 1997, Thailand was forced to float its exchange rate, which is generally seen as the start of the crisis. The 2008 global financial crisis dummy received a "1" indicator for programs that commenced between 9/15/2008 and 9/15/2009, where the start date is the Lehman Brothers' bankruptcy filing date.

[16] See, e.g., Barro (1991) and Masanjala and Papageorgiou (2008) for Africa and Fernandez et al. (2001) for Asia and Latin America.

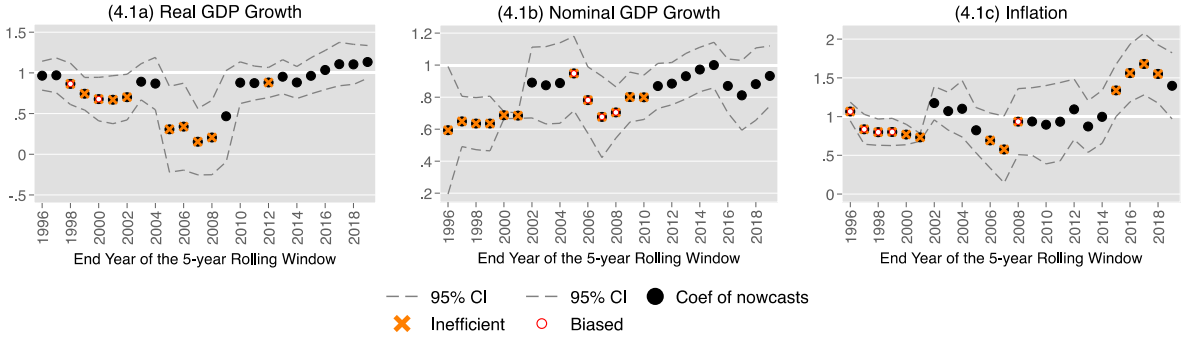

Fig. 4.1. Nowcast bias and efficiency over time (full sample). (Rolling 5-year averages.)
Source: Figures based on regressions in Appendix B2.

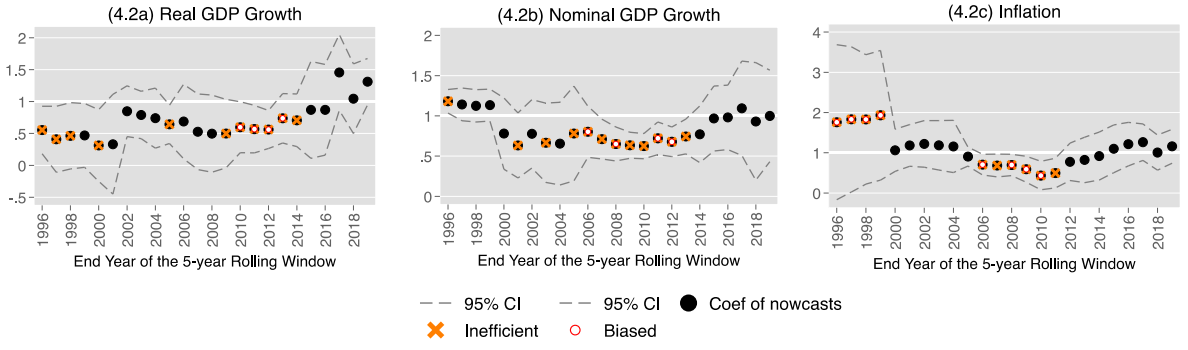

Fig. 4.2. Nowcast bias and efficiency over time (non-hyperinflation LICs). (Rolling 5-year averages.)
Source: Figures based on regressions in Appendix B2.

The IMF also offers countries a menu of program types to focus recoveries on different aspects of an economy. Different program types address issues from external financing difficulties to non-financial reforms and poverty/growth.[17] We examine if the policy focuses of IMF programs were properly accounted for in IMF nowcasts. The previous literature on IMF program performance also notes the importance of program-loan size (e.g., Dreher, 2006). Luna (2014) also notes that greater access to IMF lending is correlated with optimistic IMF forecast bias. Hence, we examine whether the program loan size relative to IMF quota is fully accounted for in IMF nowcasts. We obtain loan data from MONA and quota data from the *International Financial Statistics* database (IFS, IMF, 2020c). Results for drawn vs. approved loan sizes are identical; hence, we only report the latter.[18]

Finally, we examine a block of non-economic regressors that may exert effects on the economy, including elections, conflicts (civil and international), and natural disasters. In its review of program design and conditionality (IMF, 2019) noted that forecast errors of program countries are impacted by political transitions, conflicts, and natural disasters. We are asking again whether knowledge of such non-economic factors was properly integrated into the nowcasts. The non-economic dummies exhibit a one only if the event occurred up to one year before the program was approved. IMF nowcasters were thus well aware of the election results, civil wars, or natural disasters and we can test if the effect was properly accounted for in their nowcasts.[19]

[17] We grouped 13 IMF programs types into 5 program objectives: (i) BOP Stabilization: *Stand-By Agreements (SBA)*; (ii) BOP Shocks (precautionary): *Exogenous Shock Facility (ESF), Standby Credit Facility (SCF), Flexible Credit Line (FCL), Precautionary Credit Line (PCL), Precautionary Liquidity Line (PLL);* (iii) Structural Adjustment Poverty Reduction and Growth: *Structural Adjustment Facility (SAF), Enhanced Structural Adjustment Facility(ESAF), Poverty Reduction and Growth Trust (PRGT);* (iv) Long-Term BOP Reforms*: Extended Credit Facility (ECF), Extended Fund Facility (EFF);* (v) Non-Financial Reforms*: Policy Reform Instrument (PSI), Policy Coordination Instrument (PCI).*

[18] Atoyan and Conway (2011), Luna (2014), and IMF (2019) also point out that IMF program forecasts are conditional on the assumption that program targets are successfully implemented so that implementation failures on the part of the country could explain an optimistic IMF forecast bias. Cancellations are by definition not part of the nowcasters' information sets and thus represent orthogonal errors to the IMF nowcasts. We examined whether cancelled programs explain systematic bias and efficiency and found no evidence (results available upon request).

[19] Table 1 provides details on elections, conflicts, and disasters data. The election dummy covers head of state/government elections and legislative elections. Programs received a "1" indicator if elections occurred up to one year prior to program start, based on Beck et al. (2001) and IFES (2020). The conflict dummy covers intra/inter-state conflicts. Programs received a "1" indicator if (civil) wars occurred up to one year prior to program start, based on Harbom et al. (2009). The disaster dummy covers natural disasters. Programs received a "1"

4.1. Sources of nowcast inefficiency in the full sample

Table 4 regressions (4.1.a)–(4.3.c) display GDP growth and inflation results for the three key subsamples. We start by discussing the full sample results. For real GDP growth (regression 4.1.a), the null hypothesis that nowcasts could not have been improved by the consideration of additional variables (SJS F-test) is rejected at the 10% level. Since the covariates that we added to the Mincer and Zarnowitz regression were known to IMF nowcasters at the time of crisis, their coefficients would be zero, if their effects had been fully integrated into the nowcast. Any statistically significant deviation from zero indicates that the effect of a variable was not fully integrated into the nowcast. For example, for real GDP growth (regression 4.1.a), the full model indicates that program countries may expect IMF nowcasts during times of global crises to be more imprecise than usual, which is entirely understandable since the depth of these crises is, by definition, unknown even as the crisis has commenced. Perhaps, this expectation could be paired with a wider range of conditionality outcomes, given the uncertainty associated with the estimates.

The crisis dummy indicates only programs that were approved with nowcasts that were produced after the 2008 crisis had commenced. So forecasters knew an unusual event had occurred, but the data indicate unusual events such as global crises produce more imprecise forecasts, which is not unique to the IMF but to all forecasts at that time. It is informative, however, that the coefficient for the 2008 crisis is negative, indicating that, on average, IMF nowcasts overestimated real GDP growth for countries that started their loan during the 2008 crisis. It proves the difficulties that IMF nowcasters face: the 2008 crisis and the 1997 Asian crisis (for LICs' real GDP growth) introduced substantial nowcast errors since the depths and lengths of the crises and hence their effects on program countries were difficult to predict.

Other covariates tell a similar story, but for different reasons. For regions, the nowcast of nominal GDP growth in the full sample is systematically high for the Americas, although the dummy in effect captures Latin American programs since the USA and Canada had no programs. Here, we find that over the course of 28 years, IMF systematically over-predicts nominal GDP growth in Latin America, and the finding may be a useful indicator to reexamine the IMFs models for Latin American countries to understand the sources of the inefficient optimism in this region of the world.

Similar is the case for conditionality. IMF nowcasters are acutely aware of which type of conditionality they impose and should integrate the effects of such conditions into the nowcasts. For real GDP growth, as indicated in regression (4.1.a), conditions relating to the floor on international reserves, ceiling on government credit, and improved government statistics lead to systematic inefficiencies of IMF nowcasts. The regressions indicate the benefit of efforts to revisit the IMF's Financial Programming model to examine why projections related to reserves, and government credit may cause systematic inefficiencies in nowcasts.

The effects of structural policy reforms on real GDP growth are equally important but more difficult to predict. Depending on the subsample, improved statistics, financial sector reform, price liberalizations for government enterprises, and wage reforms all show statistically significant effects (see regression 4.1.b-c), suggesting that the effect of these reforms are systematically not fully captured in the nowcast. For example, financial liberalization has a positive significant coefficient for non-hyperinflation LICs' real GDP growth (0.013), indicating that these reforms exert a larger, positive effect on GDP growth than IMF economists systematically anticipate. The nature of these inefficiencies indicates that nowcast errors are not "mistakes", but simply indicators where further work integrating recent advances in economic theory (e.g., financial liberalization) may aid future nowcast accuracy.

Similar to policy reforms, the effects of non-economic factors such as elections and conflicts are certainly hard the gauge for IMF nowcasters, although the dummies are coded such that IMF economists knew about the election result and the existing conflicts at the time of the crisis. In both cases, we can show that IMF nowcasts improperly account for elections (for non-hyperinflation LICs) and conflicts (for hyperinflation countries) in the sense that IMF nowcasts are still too optimistic. In other words, elections and conflicts have a systematically larger negative impact on real GDP growth than IMF nowcasters predict.

In sum, the extended Mincer and Zarnowitz regressions for GDP growth indicate that systematic nowcast errors have their roots in region-specific models (Latin America), in the IMF's Financial Programming model (reserve and credit ceilings), and in the assessments of economically exogenous events that exert an important influence on the predicted outcome, such as the 2008 crisis, wars, or elections. Nowcasts can be improved through troubleshooting of the country/region-specific models, of general IMF Financial Programming identities related to credit and reserves, and though a general awareness that past program nowcasts were too optimistic about the economic effects of elections and wars.

For nominal GDP growth in the full sample, the SJS F-test is rejected at the 1% level (regression 4.2.a), indicating that IMF nowcasts could have been improved if the effects of regions (America) and the effects of the international crises (the 2008 financial crisis) had been properly considered. In addition, there is strong evidence that several conditionality dimensions were also not properly integrated into IMF nominal GDP nowcasts in the full sample. These dimensions include QPC relating to (a) reserves, (b) external arrears, (c) fiscal deficit, (d) external debt (both short-term and medium-/long-term), and SPC relating to current/capital account restriction. In addition, nominal GDP growth nowcasts could have been improved through better consideration of the implications of program types that address the balance of payments stabilization problems (SBA programs) and poverty reduction and growth (ESAF, SAF, and PRGF programs).

indicator if disasters occurred up to one year prior to program start, based on EM-DAT (2020).

Table 4
Sources of nowcast inefficiency by subsamples.

		Real GDP Growth			Nominal GDP Growth			Inflation		
		(4.1.a)	(4.1.b)	(4.1.c)	(4.2.a)	(4.2.b)	(4.2.c)	(4.3.a)	(4.3.b)	(4.3.c)
		Full Sample	Non-Hyper LICs	Hyper Inflation	Full Sample	Non-Hyper LICs	Hyper Inflation	Full Sample	Non-Hyper LICs	Hyper Inflation
Nowcast	Constant, α	−0.007	−0.035	**0.084****	0.034	−0.099	−0.008	0.037	−0.155	0.176
	p-value ($\alpha = 0$)	0.533	0.491	0.050	0.336	0.161	0.982	0.553	0.124	0.609
	IMF Nowcast, β	**0.764***	**0.558*****	0.946	**0.649*****	0.947	**0.630*****	**0.837***	1.122	**0.802****
	p-value ($\beta = 1$)	0.015	0.003	0.673	0.000	0.699	0.001	0.082	0.688	0.011
Intl Crises	**2008 Crisis**	**−0.013***	0.007	−0.022	**−0.046*****	−0.053	0.323	**−0.035***	**−0.068***	1.370
		(0.007)	(0.015)	(0.046)	(0.014)	(0.033)	(0.292)	(0.021)	(0.041)	(0.876)
	1997 Crisis	−0.012	**−0.037****	0.011	−0.028	0.037	−0.194	−0.033	0.013	−0.289
		(0.015)	(0.018)	(0.022)	(0.037)	(0.065)	(0.250)	(0.053)	(0.076)	(0.199)
Regions	Africa	0.006	0.041	0.014	−0.038	0.043	−0.169	−0.022	0.081	−0.331
		(0.006)	(0.041)	(0.022)	(0.025)	(0.039)	(0.195)	(0.031)	(0.062)	(0.200)
	Americas	0.005	0.032	0.014	**−0.056****	0.063	0.299	−0.045	0.084	0.140
		(0.006)	(0.041)	(0.042)	(0.022)	(0.046)	(0.190)	(0.035)	(0.058)	(0.235)
Quantitative Conditionality (QPCs)	Ceiling_External_Debt(MT&LT)	0.007	0.020	0.063	**0.040*****	0.037	−0.054	0.006	0.011	−0.453
		(0.005)	(0.020)	(0.046)	(0.014)	(0.039)	(0.263)	(0.013)	(0.037)	(0.682)
	Floor_Int'l_Reserves	**0.010***	0.014	0.005	**0.037***	0.027	0.136	0.023	**0.043***	0.133
		(0.005)	(0.011)	(0.024)	(0.021)	(0.021)	(0.164)	(0.016)	(0.024)	(0.146)
	Ceiling_External_Arrears	−0.001	−0.003	0.006	**0.033****	0.013	0.029	0.010	0.019	−0.239
		(0.004)	(0.006)	(0.022)	(0.014)	(0.014)	(0.132)	(0.021)	(0.026)	(0.165)
	Ceiling_Gov't_Credit	**−0.010****	−0.005	**−0.160*****	−0.003	0.009	0.026	**0.028****	0.011	0.724
		(0.005)	(0.007)	(0.042)	(0.012)	(0.026)	(0.435)	(0.013)	(0.017)	(0.720)
	Ceiling_Gov't_Deficit	−0.004	−0.008	**0.041***	**−0.058*****	**−0.021****	−0.302	−0.021	−0.014	−0.026
		(0.004)	(0.005)	(0.022)	(0.018)	(0.011)	(0.203)	(0.021)	(0.024)	(0.156)
	Ceiling_External_Debt(ST)	−0.001	−0.002	**−0.085****	**−0.020***	−0.009	0.109	0.006	0.014	−0.066
		(0.004)	(0.005)	(0.040)	(0.012)	(0.013)	(0.237)	(0.010)	(0.014)	(0.242)
	Ceiling_CB_Net_Dom_Assets	0.002	−0.009	**0.089*****	0.010	−0.013	−0.012	**0.041****	−0.006	**0.524***
		(0.006)	(0.007)	(0.032)	(0.020)	(0.016)	(0.413)	(0.017)	(0.019)	(0.292)
Policy/Structual Reform Conditionality (SPCs)	Civil_Service_Wage/Empl.	0.003	0.010	0.009	−0.002	−0.002	**0.719***	−0.008	−0.009	0.989
		(0.004)	(0.007)	(0.080)	(0.009)	(0.013)	(0.380)	(0.009)	(0.014)	(0.721)
	Improve_Econ_Statistics	**0.008***	0.001	−0.144	−0.009	−0.002	0.001	**−0.028*****	−0.006	0.159
		(0.005)	(0.006)	(0.090)	(0.010)	(0.011)	(0.731)	(0.010)	(0.012)	(1.016)
	Open_Current&Capital_Account	−0.002	0.006	0.002	**−0.043****	**−0.036****	**−0.493***	−0.012	−0.019	−0.230
		(0.005)	(0.006)	(0.021)	(0.021)	(0.015)	(0.261)	(0.018)	(0.023)	(0.158)
	Financial_Sector_Reform	0.002	**0.013***	0.031	0.003	0.018	−0.240	−0.001	0.004	−0.244
		(0.005)	(0.007)	(0.025)	(0.015)	(0.015)	(0.209)	(0.018)	(0.022)	(0.256)
	Reduce_Trade_Tariff/Quota	0.001	0.006	−0.025	0.023	−0.005	0.402	**−0.029***	**−0.051***	0.039
		(0.004)	(0.005)	(0.024)	(0.024)	(0.014)	(0.257)	(0.017)	(0.026)	(0.126)
	Gov't_Enterprise_Pricing	−0.002	0.005	**−0.062***	0.017	0.001	0.190	**0.033***	−0.004	0.130
		(0.004)	(0.007)	(0.032)	(0.011)	(0.013)	(0.214)	(0.018)	(0.013)	(0.264)
	CB_Stats_Regs_Indep.	−0.002	0.001	**−0.126*****	−0.010	−0.013	−0.074	−0.007	−0.022	−0.244
		(0.004)	(0.008)	(0.038)	(0.011)	(0.016)	(0.250)	(0.013)	(0.021)	(0.335)
	Labor_Mkt_Wage/Empl.	0.009	**−0.031***	–	−0.025	−0.003	–	−0.025	−0.009	–
		(0.007)	(0.018)		(0.016)	(0.030)		(0.019)	(0.049)	

(*continued on next page*)

Table 4 (*continued*).

		Real GDP Growth			Nominal GDP Growth			Inflation		
		(4.1.a)	(4.1.b)	(4.1.c)	(4.2.a)	(4.2.b)	(4.2.c)	(4.3.a)	(4.3.b)	(4.3.c)
		Full Sample	Non-Hyper LICs	Hyper Inflation	Full Sample	Non-Hyper LICs	Hyper Inflation	Full Sample	Non-Hyper LICs	Hyper Inflation
Program Objective	BOP_Stablization[a]	−0.005	−0.012	−0.015	**0.042****	**0.087***	**0.398****	**0.053*****	0.179	0.167
		(0.005)	(0.013)	(0.025)	(0.018)	(0.044)	(0.163)	(0.020)	(0.118)	(0.132)
	BOP_Shocks_Precautionary[b]	0.009	0.009	–	0.012	0.020	–	−0.006	0.048	–
		(0.006)	(0.018)		(0.019)	(0.032)		(0.021)	(0.052)	
	Struct_Adj_Poverty_Growth[c]	0.004	0.002	–	**−0.023***	−0.007	–	0.009	**0.048***	–
		(0.005)	(0.011)		(0.014)	(0.018)		(0.017)	(0.025)	
	Non_Financial_Reforms[d]	0.001	−0.004	−0.038	0.008	0.006	**0.432****	0.011	**0.041*****	0.211
		(0.005)	(0.007)	(0.033)	(0.016)	(0.018)	(0.200)	(0.012)	(0.015)	(0.177)
	Elections	−0.001	**−0.010***	0.013	0.012	−0.019	0.149	−0.007	−0.020	0.012
		(0.004)	(0.005)	(0.015)	(0.018)	(0.013)	(0.174)	(0.012)	(0.018)	(0.116)
	Conflicts	−0.005	−0.007	**−0.038****	0.000	−0.019	0.005	−0.003	−0.007	−0.033
		(0.005)	(0.006)	(0.016)	(0.012)	(0.012)	(0.120)	(0.013)	(0.019)	(0.149)
	Observations	597	276	71	596	276	71	595	275	69
	Adjusted R-squared	0.406	0.199	0.521	0.842	0.643	0.823	0.819	0.231	0.859
	SJS F-test ($\alpha = \delta = 0 \& \beta = 1$)	1.368*	2.044***	13.64***	3.263***	1.433**	5.618***	1.573**	1.369*	13.84***
	p-value ($\alpha = \delta = 0 \& \beta = 1$)	0.081	0.001	0.000	0.000	0.063	0.000	0.021	0.091	0.000

Notes: Robust standard errors in parentheses unless otherwise indicated; $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Insignificant regressors included in the regressions are not reported, these variables include Regions (Asia), Qualitative Conditionality (Ceiling_Domestic_Arrears, Ceiling_New_Arrears/Default); Structural Conditionality (Gen_Gov't_Reform, Legal/Market_Reforms, Pension_Reform), NaturalDisasters, LoanAmount (approved or actually-drawn amounts). Dummies excluded to avoid singularity: Regions (Europe) and Program Objectives ("Long-Term BOP & Structural Reforms" consisting of ECF – Extended Credit Facility, EFF – Extended Fund Facility).

[a]Stand-By Agreements (SBA).

[b]Includes Exogenous Shock Facility (ESF), Standby Credit Facility (SCF), Flexible Credit Line (FCL), Precautionary Credit Line (PCL), Precautionary Liquidity Line (PLL).

[c]Includes Structural Adjustment Facility (SAF), Enhanced Structural Adjustment Facility (ESAF), Poverty Reduction and Growth Trust (PRGT).

[d]Policy Reform Instrument (PSI), Policy Coordination Instrument (PCI).

For inflation, the SJS F-test in the full sample is rejected at the 5% level (regression 4.3.a), indicating that the efficiency of IMF nowcasts could have been improved through the consideration of additional covariates such as the effect of the 2008 global financial crisis and conditionality. Specifically, the effects of quantitative targets on government and central bank credit were not properly integrated into IMF nowcasts. In addition, inflation nowcasts could have been improved through consideration of the effects of structural reforms in economic statistics, trade openness, and state enterprises. As in the case of nominal GDP growth, the effects of balance of payments stabilization programs were also not properly accounted for by IMF inflation nowcasts.

In summary, for the full sample, the effects of the 2008 global financial crisis are the only common factors that were not properly accounted for by IMF nowcasts for GDP growth and inflation, and a result previously reported by Genberg and Martinez (2014) for the global sample using WEO data. In addition to the common factors, real/nominal GDP growth and inflation all have their distinct factors that were known to nowcasters and whose proper consideration could have improved the nowcast. Since we learned in the previous section that the inefficiency of the full sample is decisively driven by inefficiencies in the non-hyperinflation LIC sample (and to a lesser extent by hyperinflation events in the early part of the sample), we examine the effects of additional covariates for these two subsamples in Table 4.

4.2. Sources of nowcast inefficiency in the non-hyperinflation LIC sample

Regressions (4.1.b)–(4.3.b) in Table 4 indicate that the efficiency of GDP growth and inflation nowcasts for non-hyperinflation LICs could have been improved substantially. Real GDP growth nowcast inefficiency (regression 4.1.b) was driven by overestimates of growth during the 1997 Asian crisis, as indicated by the negative coefficient. In addition, labor and financial market reforms, as well as the non-economic effect of elections, were not properly accounted for by the nowcasts. For nominal GDP growth, nowcasts efficiency could have been improved if quantitative limits on fiscal deficits, reforms of current/capital accounts, and program types (BOP stabilization/SBA program) had been properly integrated into the nowcasts. Inflation nowcasts could have been improved through better integration of the effects of reserves requirement, trade reforms, and non-financial reforms along with effects of the poverty and growth programs and the 2008 crisis.

4.3. Sources of nowcast inefficiency in the hyperinflation sample

For hyperinflation countries, we find the strongest evidence that inefficient real/nominal GDP growth nowcasts could be been improved along with inflation nowcasts. For real GDP growth, we find three highly statistically significant factors: quantitative limits on government credit and deficit, as well as limits on external debt and central bank credit. In terms of structural reforms, we find evidence (regression 4.1.c) that state enterprise reform and central bank statistical/regulatory reforms could have improved the nowcasts. Even the effect of conflicts is now indicated as a factor that could have improved the nowcasts. For nominal GDP growth nowcasts in hyperinflation countries, the most important factors that could have improved the nowcast were non-financial reforms (PSI and PCI) and BOP stabilization (SBA) program types. In addition, reforms of public employment and current/capital account openness are statistically significant. For inflation nowcasts, the SJS test is also rejected at the 1% level and the ceiling on central bank credit could have improved efficiency.

5. Did nowcast bias and efficiency change over time?

Our MONA nowcast data covers over a quarter-century of IMF programs. It is natural to ask whether the accuracy of nowcasts changed over time. It may well be that the advent of better modeling and improved data collection produced successively better nowcasts. Instead of reporting nowcasts accuracy for each individual year, we report rolling 5-year period results. This allows us to keep the number of observations per period roughly similar and of sufficient size.

Fig. 4.1 provides visuals of nowcast accuracy over time for the full sample, based on the Mincer and Zarnowitz regressions that are reported in Appendix Table B.1. The black dots in Fig. 4.1 represent the values of the β estimates in Table B.1, and the gray dashed lines represent 95% confidence intervals. The orange cross markers denote inefficient nowcasts, while the red empty circles denote biased nowcasts. We observe four distinct periods. First, all nowcasts struggled with bias and/or inefficiency until about 2001. Second, all nowcasts saw a reprieve with unbiased and efficient nowcasts until 2005. Third, another period of bias and/or inefficiency occurred until 2009 (2012 for real GDP growth). Fourth, after 2012, GDP growth nowcasts become unbiased and efficient (with one exception in 2013), but inflation nowcasts continue to struggle with efficiency as recently as 2018. It is fascinating to see that inflation still struggles with inefficiency in recent years, while nominal GDP growth nowcasts have become efficient.

In addition to our assessment of nowcast accuracy, we observe that the slope coefficients for nowcasts, β, are almost always smaller than unity until 2014. This implies a long-enduring pattern of overly optimistic nowcasts, on average with excess optimism for high-growth countries and overly pessimistic nowcasts for low-growth countries. The pattern reverses after 2015 when the slope coefficients start to exceed unity for both inflation and real GDP growth. This suggests that since about 2015 IMF nowcasts become excessively pessimistic (optimistic) for high (low) growth outcomes. We also note that the width of the error bands suggests standard errors are roughly similar throughout. The exception is real GDP growth, which experienced a widening of the confidence interval during the 2008 global financial crisis (producing inefficient nowcasts). Nominal GDP growth exhibits

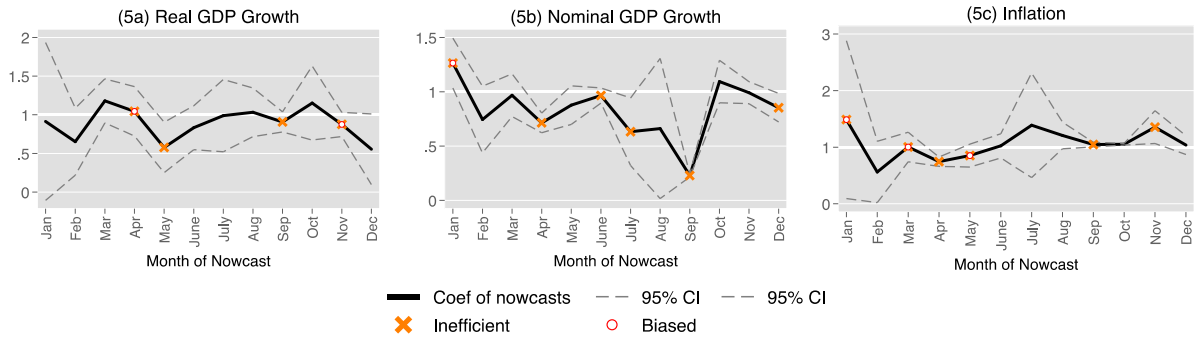

Fig. 5. Nowcast horizons and nowcast accuracy (full sample).
Source: Figures based on regressions in Appendix B2.

extraordinarily large errors in the early 1990s and also during the financial crisis. Inflation bucks the trend with tight standard errors until 2007 and stable error bands for the remaining years.

Above we noted the importance of two subsamples in our study of bias and inefficiency, especially non-hyperinflation LICs introduced nowcasts errors that translated into nowcast inaccuracy for the full sample. Given the results above, we are also interested in this subsamples' pattern of nowcast accuracy over time. Even though the hyperinflation sample contains 76 observations, it is too concentrated in the early years of our 28-year sample period, rendering too few observations to produce meaningful 5-year rolling time periods throughout.

For the non-hyperinflation LIC sample (Fig. 4.2 and Appendix Table B.2), we find a roughly similar pattern for the full sample, in that earlier nowcasts are more likely to be inefficient, and later nowcasts (since 2014) have become unbiased and efficient for GDP growth. Even inflation nowcasts are unbiased and efficient in the non-hyperinflation LIC sample in recent years. This is good news, especially given the nearly unbroken string of biased and/or inefficient nominal GDP growth nowcasts in this sample from 2001–2013.

6. Do nowcast horizons affect nowcast accuracy?

In general, forecast accuracy is expected to decrease as forecast horizons increase (Armstrong, 2001). One may suspect this insight to be particularly relevant for IMF nowcasts, as information sets are larger at the end of the year due to the accumulation of scheduled releases of additional data vintages. Hence, one might well expect nowcast bias and efficiency to improve for programs designed and approved later in the year. In this section, we examine whether bias and inefficiency are driven by nowcast horizons. Fig. 5 provides the visual summary of results for the full sample, and the regression outputs are reported in Appendix Table B.3. We examine whether nowcasts produced earlier in the program year exhibit a greater propensity towards bias and inefficiency than those formed later in the year.

The results are surprising, as there is no clear pattern of improved nowcast accuracy as the time horizon shortens. Both real GDP growth and inflation nowcasts exhibit greater variances early in the year, but these do not translate into greater bias and/or inefficiency. Indeed, real GDP growth and inflation poignantly produce inefficient and even biased nowcasts late in the year. Surprisingly, inflation nowcasts are the most stable around the slope parameter of unity throughout the year, while nominal GDP growth produces the largest deviations from unity. This could be due to the fact that inflation information is much more readily available (on a monthly basis) than GDP growth (at best quarterly). The divergence in GDP growth and inflation accuracy as the time horizon shortens also implies that GDP growth nowcast errors are not driven by inflation nowcast errors. This finding is supported by the fact that there is no similarity in the pattern of real GDP growth nowcast inaccuracies and either inflation or nominal GDP growth nowcasts.

Biased and inefficient real GDP nowcasts are found mid-year in April and May and, somewhat surprisingly, at the end of the year in November when nowcasts turn excessively optimistic. Nominal GDP growth has the expected bias and inefficiency in January, but the next two months are both unbiased and efficient with four additional inefficiencies throughout the year. For inflation, bias is again early in the year, but bias and inefficiencies are concentrated mid-year. Overall we see no pattern of either bias and efficiency improvements as the nowcast horizon shrinks. IMF-IEO (2014) previously found evidence that IMF forecast errors increase with time horizons in WEO data, but their study horizons far exceeded the time horizon covered in this paper and did not cover crisis nowcasts.

7. Conclusion

IMF nowcasts established at the time of crisis are the basis for IMF program conditions for countries that request assistance from the lender of last resort. Instead of examining the IMF program efficacy, we investigate the accuracy of these nowcasts that predicate IMF program design in a dataset that is six times larger than the largest previous study on the subject. We find that (real and nominal) GDP growth and inflation nowcasts are inefficient in the full sample, a result driven by substantial bias and inefficiency in LICs' nowcasts. We show that these inaccuracies are not a function of the nowcast horizon, and document that GDP growth nowcasts have improved

in recent years. In contrast, inflation nowcasts continue to struggle with accuracy until recently. On the upside, once the full sample is purged off hyperinflation countries and LICs, the remaining forecasts for all variables are unbiased and efficient.

Instead of documenting the uniform optimism in IMF forecasts that had largely been accepted as a stylized fact in the previous literature, we dissect the structure of nowcast bias and inefficiency and highlight that only the most vulnerable, low-growth recovering LICs are subject to excessively pessimistic nowcasts. Nowcasts for fast-growing countries are excessively optimistic, overestimating the speed of their recovery. Once purged of LICs, the remaining sample exhibits no statistically significant optimistic and pessimistic bias. Our findings have important implications for LIC crisis countries. IMF conditionality based on overly optimistic nowcasts may affect the likelihood that the country can achieve the conditions and affect future loan disbursements and program evaluation. In addition, the nowcast bias may produce quantitative performance targets that are impossible to reach. As the COVID-19 pandemic has raised the demand for IMF programs, improved nowcasts are thus more important than ever.

The dichotomy between countries that are optimistically or pessimistically assessed raises the question regarding the drivers of the inefficiency in IMF nowcasts. We investigate the sources of nowcast inefficiencies by country subsamples to highlight the factors that were in the IMF forecasters' information sets but were improperly integrated into nowcasts. Each type of nowcast (GDP/inflation) and each subsample of countries (Full/LICs/hyperinflation) produces a different set of conditions and program types that were improperly integrated into IMF nowcasts. Our work has been made possible through the merger of several IMF databases as well as a comprehensive audit of the data, which was found to include an inordinate amount of errors in the IMF MONA database. This is noteworthy since the databases are the basis of a substantial number of research papers.[20]

We leave for future research the pesky question as to how researchers best forecasts recoveries. There exists an important literature that notes that models do not provide detailed recovery dynamics and researchers rely instead on their forecasts on previous patterned recoveries. Loungani and An (2020) note the general tendency to assume forecasts are V-shaped. When the V-shaped pattern materializes forecasters are slow to update believes, perhaps, for lack of guidance from the models? In future research projects, we plan to investigate the shapes of the recoveries associated with program countries' crises to understand how much of the forecast bias and inefficiency is driven by patterned recovery forecasts rather than predicted modeled recovery dynamics.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

[20] A quick search produces over 2000 papers that have been published based on IMF MONA database. Results of our audit and corrected errors are documented in Appendix A.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2021.12.007.

References

Aldenhoff, F. O. (2007). Are economic forecasts of the international monetary fund politically biased? A public choice analysis. *Review of International Organizations*, *2*(3), 239–260.

Armstrong, J. S. (2001). In J. Scott Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic Publishers.

Artis, M. J. (1988). How accurate is the world economic outlook? A post mortem on short-term forecasting at the international monetary fund. In *Staff studies for the world economic outlook*. Washington, DC: International Monetary Fund.

Artis, M. J. (1997). How accurate are the WEO's short-term forecasts? An examination of the world economic outlook. In *Staff studies for the world economic outlook*. Washington, DC: International Monetary Fund.

Atoyan, R., & Conway, P. (2011). Projecting macroeconomic outcomes: Evidence from the IMF. *Review of International Organizations*, *6*(3), 415–441.

Baqir, R., Ramcharan, R., & Sahay, R. P. (2003). Imf program design and growth: What is the link. In *IMF working papers*. Washington, DC: International Monetary Fund.

Baqir, R., Ramcharan, R., & Sahay, R. (2005). IMF programs and growth: Is optimism defensible? *IMF Staff Papers*, *52*(2), 260–286.

Barrionuevo, J. M. (1993). How accurate are the world economic outlook projections? In *Staff studies for the world economic outlook*. Washington, DC: International Monetary Fund.

Barro, R. J. (1991). Economic growth in a cross section of countries. *Quarterly Journal of Economics*, *106*(2), 407–443.

Barth, R., Hemphill, W., Aganina, I., George, S., Greene, J., & McNeilly, C. (2000). *Financial programming and policy: the case of Turkey*. Washington DC: IMF Institute, International Monetary Fund.

Batchelor, R. (2001). How useful are the forecasts of intergovernmental agencies? The IMF and OECD versus the consensus. *Applied Economics*, *33*(2), 225–235.

Beach, W., & Schavey, A. (1999). *How reliable are IMF economic forecasts? heritage foundation report 99-05, August 27, 1999*.

Beaudry, P., & Willems, T. (2022). On the macroeconomic consequences of over-optimism. *American Economic Journal: Macroeconomics*, *14*(1), 38–59. http://dx.doi.org/10.1257/mac.20190332.

Beck, T., Clarke, G., Groff, A., Keefer, P., & Walsh, P. (2001). New tools in comparative political economy: The database of political institutions. *The World Bank Economic Review*, *15*(1), 165–176.

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, *71*(356), 791–799.

Clemens, M. P., & Hendry, D. F. (2011). *The oxford handbook of economic forecasting*. Oxford: Oxford University Press.

Dornbusch, R., & Fischer, S. (1986). Stopping hyperinflations past and present. *Weltwirtschaftliches Archiv*, *122*(1), 1–47.

Dreher, A. (2006). Imf and economic growth: The effects of programs. *Loans, and Compliance with Conditionality, World Development*, *34*(5), 769–788.

Easterly, W. (2006). An identity crisis? Examining IMF financial programming. *World Development*, *34*(5), 964–980.

Eicher, T. S., Kuenzel, D. J., Papageorgiou, C., & Christofides, C. (2019). Forecasts in times of crises. *International Journal of Forecasting*, *35*(3), 1143–1159.

Elliott, G., & Timmermann, A. (Eds.), (2013). *Handbook of economic forecasting: Vol. 2*, Elsevier, Part A.

EM-DAT (2020). The OFDA/CRED international disaster database. Date accessed: 10/30/20 URL http://www.emdat.be/.

Feldstein, M. (1998). Refocusing the IMF. *Foreign Affairs*, *77*(2), 20.

Fernandez, C., Ley, E., & Steel, M. F. J. (2001). Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics*, *16*(5), 563–576.

Fischer, S. (1999). On the need for an international lender of last resort. *Journal of Economic Perspectives*, *13*(4), 85–104.
Genberg, H., & Martinez, A. (2014). On the accuracy and efficiency of IMF forecasts: A survey and some extensions. *IMF-IEO background paper BP/14/04*, Washington, DC: Independent Evaluation Office of the IMF.
Granger, C. W. J., & Newbold, P. (2014). *Forecasting economic time series*. Academic Press.
Gultekin, I., Lahiri, K., & Loungani, P. (2006). How quickly do forecasters incorporate news? Evidence from cross-country surveys. *Journal of Applied Econometrics*, *21*, 703–725.
Harbom, L., Strand, H., & Nygård, H. M. (2009). UCDP/PRIO armed conflict dataset codebook. Version 20.1. Date Accessed: 10/30/20. URL http://www.ucdp.uu.se/database.
Holden, K., & Peel, D. A. (1990). On testing for unbiasedness and efficiency of forecasts. *The Manchester School*, *58*(2) 120–127.
Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. Melbourne, Australia: OTexts, URL https://otexts.com/fpp3.
IFES (2020). *ElectionGuide by the international foundation for electoral systems (IFES)*. Date accessed: 10/15/20 URL https://www.electionguide.org/elections.
IMF (2019). 2018 Review of program design and conditionality. *IMF policy paper, press release no. 19/174*, Washington, DC: International Monetary Fund.
IMF (2020a). IMF crisis lending fact sheet. Date accessed: 10/30/20. URL https://www.imf.org/en/About/Factsheets/IMF-Lending.
IMF (2020b). IMF MONA database glossary. Date accessed: 10/30/20. URL https://www.imf.org/external/np/pdr/mona/glossary.aspx.
IMF (2020c). *International financial statistics (IFS) database*. Date accessed: 10/30/20. URL https://data.imf.org/ifs.
IMF (2020d). *Monitoring of fund arrangements (MONA) database*. Date accessed: 10/30/20. URL https://www.imf.org/external/np/pdr/mona/index.aspx.
IMF (2020e). What we do. Date accessed: 11/21/20. URL https://www.imf.org/en/About.
IMF (2020f). *World economic outlook (WEO) database*. Version: 2020. URL https://www.imf.org/en/Publications/WEO/weo-database/2020/October.
IMF (2020g). IMF Covid-19 financial assistance and debt service relief. Date accessed: 10/30/2020. URL: https://www.imf.org/en/Topics/imf-and-covid19/COVID-Lending-Tracker.
IMF-IEO (2014). *IMF forecasts: process, quality, and country perspectives*. Washington, DC: Independent Evaluation Office of the IMF.
Kaminsky, G. L., Reinhart, C. M., & Vegh, C. A. (2003). The unholy trinity of financial contagion. *Journal of Economic Perspectives*, *17*(4), 51–74.
Kenen, P. B., & Schwartz, S. B. (1986). An assessment of macroeconomic forecasts in the IMF's WEO. In *Working Papers in International Economics*. Princeton University: International Finance Section.
Loungani, P. (2001). How accurate are private sector forecasts? Cross-country evidence from consensus forecasts of output growth. *International Journal of Forecasting*, *17*(3), 419–432.
Loungani, P., & An, Zidong (2020). *There will be growth in the spring: how well do economists forecast recoveries?*. IMF Independent Evaluation Office.
Luna, F. (2014). IMF forecasts in the context of program countries. *IMF-IEO background paper BP/14/05*, Washington: Independent Evaluation Office of the IMF.
Masanjala, W. H., & Papageorgiou, C. (2008). Rough and lonely road to prosperity: A reexamination of the sources of growth in Africa using Bayesian model averaging. *Journal of Applied Econometrics*, *23*(5), 671–682.
Mincer, J., & Zarnowitz, V. (1969). The evaluation of economic forecasts. In J. Mincer (Ed.), *Economic forecasts and expectations: analysis of forecasting behavior and performance* (pp. 3–46). NBER. New York: National Bureau of Economic Research.
Nordhaus, W. D. (1987). Forecasting efficiency: Concepts and applications. *The Review of Economics and Statistics*, *69*(4), 667–674.
Park, Y. C. (2006). Chapter 9: A re-evaluation of the IMF reform program: Wrong diagnosis and wrong prescription. In Y. C. Park (Ed.), *Economic liberalization and integration in east asia: A post-crisis paradigm*. Oxford University Press.
Pons, J. (2000). The accuracy of IMF and OECD forecasts for G7 countries. *Journal of Forecasting*, *19*(1), 53–63.
Romer, C. D., & Romer, D. H. (2000). Federal reserve information and the behavior of interest rates. *American Economic Review*, *90*(3), 429–457.
Rossi, B., & Sekhposyan, T. (2011). Understanding models' forecasting performance. *Journal of Econometrics*, *164*(1), 158–172.
Sinclair, T. M., Joutz, F., & Stekler, H. O. (2010). Can the fed predict the state of the economy? *Economics Letters*, *108*(1), 28–32.
Sinclair, T. M., Stekler, H. O., & Carnow, W. (2012). A new approach for evaluating economic forecasts. *Economics Bulletin*, *32*(3), 2332–2342.
Stiglitz, J. E. (2002). *Vol. 500*, *Globalization and its discontents*. New York Norton.
Theil, H. (1961). *Economic forecasts and policy*. North-Holland Pub. Co., 2nd Revised Edition.
Timmermann, A. (2007). An evaluation of the world economic outlook forecasts. *IMF Staff Papers*, *54*(1), 1–33.
Wallis, K. F. (1989). Macroeconomic forecasting: A survey. *The Economic Journal*, *99*(394), 28–61.