ARTICLE IN PRESS

International Journal of Forecasting xxx (xxxx) xxx

Contents lists available at ScienceDirect

International Journal of Forecasting

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

IMF trade forecasts for crisis countries: Bias, inefficiency, and their origins ☆

Theo S. Eicher *, Reina Kawai

University of Washington, United States of America

ARTICLE INFO

JEL classification:
O19
O11
F47

Keywords:
Forecast evaluation
Economic crisis
Trade Forecasts
Exchange rate forecasts
IMF Crisis Forecasts

ABSTRACT

External sector surveillance and stabilization are core missions of the International Monetary Fund (IMF). Since 1992, the IMF approved over 600 crisis country loan programs, conditional on reforms and performance targets that are contingent on IMF crisis assessments and recovery forecasts. The literature evaluating IMF crisis forecasts has primarily focused on GDP, inflation, and fiscal budgets, but IMF programs often originate with the balance of payments crises. Our evaluation of IMF imports/exports/exchange rates in crisis countries reveals a surprising dichotomy: import forecasts are largely efficient and unbiased, while exports and exchange rate forecasts exhibit substantial biases and inefficiencies. We show forecast errors in the full sample are driven by deeply flawed IMF forecasts for LICs in crisis. Fixed exchange rate LICs (predominantly African franc zone countries) receive systematically inefficient import forecasts. Exchange rate forecasts for LICs with flexible exchange rates are so inefficient that they cannot outperform a naïve random walk, and over 30 percent of the forecasts cannot match the exchange rate's directional movement during the first year of the recovery. Examining the sources of biases and inefficiencies, we highlight effects of conditionality and geopolitics that were not fully accounted for in IMF forecasts, specifically those relating to arrears (domestic and foreign), fiscal finance (balance and credit limits), policy reforms (trade and government), (civil) wars, and elections.

1. Introduction

A core mission of International Monetary Fund (IMF) loan programs is "*to help countries restore macroeconomic stability by rebuilding their international reserves, stabilizing their currencies, and paying for imports - all necessary conditions for relaunching growth*" (IMF, 2020). IMF loans are accompanied by IMF conditionality is related to policy reforms and economic performance targets. These targets are jointly determined by IMF's crisis assessments and IMF recovery forecasts.[1] Reviews of IMF crisis loan performances suggest that a key indicator of the quality of program design is the absence of systematic bias in program forecasts (Baqir et al., 2006; Mody & Rebucci, 2006). Since most IMF loan programs initiate during the balance of payments crises, the viability of economic recoveries depends crucially on the quality of IMF external sector forecasts. Even if IMF programs do not originate with balance of payment crises, accurate external sector

☆ We thank the IMF for providing access to the Monitoring of Fund Arrangement (MONA) database and archived Executive Board Loan Program documents. We also benefited from insightful discussions with Chris Papageorgiou, Charis Christofides, Monica Gao Rollinson, David Moore, Olaf Unteroberdoerster, and David Kuenzel. Chengjun Zhang, Liyuan Zhang, Elena Zhu, Ziqing Wang, and Xuanming Da provided excellent research assistance.

* Corresponding author.
E-mail addresses: te@uw.edu (T.S. Eicher), rkawai7@uw.edu (R. Kawai).

[1] Musso and Phillips (2002) note the term "forecast" is sedulously avoided by IMF documents. We label the "future values" included in IMF loan documents "forecasts", since loan agreements represent formal contracts between the IMF and countries to implement policies and targets in exchange for funds.

https://doi.org/10.1016/j.ijforecast.2022.07.006
0169-2070/

forecasts are crucial to predict countries' financing needs to repay IMF loans. Hence, our study focuses on the IMF's external sector forecasts in times of crisis.

Only a single previous study of IMF trade forecasts for crisis countries includes import and export data. Most evaluations of IMF forecasts focused on the current account as a percent of GDP, conflating forecast errors that originate with imports, exports, and GDP.[2] Previous evaluations of IMF trade forecasts in crisis countries also employ surprisingly small and always different subsets of all available data, which may explain the surprising diversity of results. Forecast accuracy assessments range from exactly zero average forecast errors (Baqir et al., 2006) to forecasts that cannot outperform random walks that predict future growth with past growth – a harrowingly low bar in times of crisis (see Arora & Smyth, 1990; Artis, 1988; Musso & Phillips, 2002).[3] No previous forecast evaluation examines the accuracy of exchange rate forecasts, although exchange rate trajectories are arguably fundamental to all recovery trajectories in IMF programs, especially imports and exports. To address the lack of large sample forecast evaluations for imports, exports, and exchange rates, we analyze nearly 30 years of import, export, and exchange rate forecasts in a dataset of over 600 IMF loan programs. Our dataset is over three times larger and more than a decade longer than the most comprehensive study to date. The size of our dataset allows for an unprecedented analysis of drivers of forecast bias and inefficiency by subsamples (e.g., country-income status (LICs/non-LICs), crisis types (hyperinflation/non-hyperinflation/BOP) , exchange rate regimes (fixed/flexible), or openness (financial/trade openness).

The most recent paper to assess the accuracy of IMF trade forecasts in crisis countries also includes one of the largest datasets to date. It is also the only previous study that examines import and export forecasts (Eicher et al., 2019, EKPC from here on out). However, EKPC use estimated, not actual final data, which contaminates their assessments of forecast accuracy. EKPC are also transparent about their exclusion of a large number of observations for fear of errors in the IMF MONA loan database (IMF, 2021a). We audited the MONA database using original loan documents stored in the IMF archives to verify extreme values and correct database errors (see Appendix B), which allowed us to add 10 additional years and 476 additional country/program observations to the EKPC dataset. We also use actual, not estimated final data.

The corrected/audited data along with the additional data for additional years overturns the central findings of EKPC. First, EKPC suggested trade forecast errors were influenced by inefficient import growth forecasts. We find the opposite. Our expanded dataset shows IMF import growth forecasts for crisis countries are remarkably accurate, without bias or inefficiency. We document that this result is largely independent of country-income status (LICs/non-LICs), crisis types (hyperinflation/non-hyperinflation/BOP), or openness (financial/trade openness). The result does depend, however, on exchange rate regimes – IMF import forecasts for fixed exchange rate LICs feature substantial forecast inefficiency.

The second important finding of EKPC was that export growth forecasts were only inefficient but not biased. Our expanded dataset shows this EKPC result to be an artifact of the smaller dataset and with estimated rather than actual final data. For the corrected, full dataset, IMF export growth forecasts have, in fact, been systematically biased and inefficient across most subsamples: for non-LICs, (non)hyperinflation countries, BOP programs, fixed/floating exchange rates, and all degrees of trade and financial openness. Only LICs with floating exchange rate are unbiased and efficient. The stark difference in import vs. export forecast accuracies is noteworthy and surprising.[4] Being the first to examine exchange rate forecast accuracy, we can confirm that the divergence in forecast accuracies is not driven by inaccuracies in IMF exchange rate forecasts. Exchange rate forecasts for the full sample are unbiased and efficient, although LICs forecasts are spectacularly inaccurate. For LICs with floating exchange rates, IMF crisis forecasts cannot beat the naive random walk as they overestimate crisis depreciations by more than 100 percent on average.[5] Overall import, export, and exchange rate forecasts are shown to be systematically optimistic, on average, projecting larger than the realized growth rates. As we disentangle average forecast errors, we can also show that IMF forecasts underestimate growth rates for slow recovering countries and overestimate growth rates for high-growth recoveries. This is an important insight since it suggests that the most vulnerable, slow recovering countries receive systematically pessimistic, excessively cautious external sector forecasts.

To identify the sources of IMF forecast inefficiencies, we examine whether IMF forecasters properly integrate all information known at the time of forecasts. We group this information into three areas. First, we confirm the findings of the previous literature that loan size is indeed a factor that affects IMF forecast accuracy for crisis countries (see Beach et al., 1999; Dreher et al., 2008; Luna, 2014). Second, we document the effects of conditionality on forecast inefficiency (as previously suggested

[2] See Eicher et al. (2019) and Genberg and Martinez (2014) for surveys.

[3] For some financial data, especially if serially correlated, the inability to outperform the random walk is not an indictment. But for countries that careen into crisis, it is highly problematic when forecasts cannot beat the naïve random walk projection where past growth equals future growth.

[4] Boz et al. (2020) suggest a potential explanation for the differential bias and efficiency observed in IMF import and export forecasts: countries invoicing predominantly in foreign currency (dollars) may experience greater exchange rate pass-through to import prices and higher sensitivity of trade volumes to exchange rate fluctuations. Unfortunately, we cannot examine this hypothesis since the unique Boz et al. (2020) dataset covers only 40% of our full sample and 15% percent of our LICs sample.

[5] Meese and Rogoff (1983) and a voluminous subsequent literature point out that exchange rates are notoriously difficult to predict; out-of-sample forecasts seldom outperform random walks (Rossi, 2013, surveys the literature). Itskhoki (2020) highlights that real exchange rate forecasts are difficult in part because they are virtually uncorrelated with most other macroeconomic variables, nominal, or real, with the exception of the nominal exchange rate. In our findings, it is noteworthy that only LIC real exchange rate forecasts struggle to outperform the random walk, but forecasts for the full sample are unbiased, efficient, and outperform the random walk.

by Dreher et al. (2008) for GDP forecasts). Specifically, debt conditionality is shown to be central to explaining forecast inefficiencies for exports, while a better account of arrears conditionality (domestic and external) could improve imports and exchange rate forecasts. Fiscal conditionality is not integrated properly into import and export forecasts, while inaccurate estimates of the effects of trade reforms played an important role in the unusual bias observed in LICs exchange rate forecasts. Third, we investigate whether geopolitical events are adequately integrated into forecasts, since Aldenhoff (2007) and Dreher et al. (2008) suggest geopolitical events drive errors in IMF GDP forecasts. We find that effects of elections (executive/legislative) and conflicts (internal/external known at the time of forecasts) were not properly accounted for in IMF crisis forecasts.

It is important to understand our results in the context of the previous IMF forecast evaluation literature that covers crisis countries. Genberg and Martinez (2014) survey 75 studies that focus only on GDP growth, inflation, or fiscal balances when evaluating IMF forecast accuracy since 1983. IMF external sector forecasts are addressed in only 19 of these 75 studies, and only 7 of these include an analysis of import and export forecasts (see Table 5). Of these 7, only EKPC examine import and export data for crisis countries, while all other studies use IMF World Economic Outlook data (WEO, (IMF, 2021b)). WEO data does, however, not report forecasts for individual crisis countries, and before 2004, WEO data are instead aggregated into a single forecast for all developing countries per geographic region (Europe, Asia, Africa, Middle East, and Western Hemisphere).

The previous literature evaluating trade forecasts for IMF crisis countries is hampered by small and ever-different datasets that produce an unusual array of distinctly different results. Musso and Phillips (2002) are the first study to evaluate IMF forecasts for crisis countries. They examine current account (percent of GDP) forecasts for 69 IMF loans over 5 years (1993–1997) to find IMF forecasts hold "very little predictive power beyond that of the random walk alternative". Following the Asian Financial Crisis, the US House of Representatives (USGAO, 2003) commissioned a report on the accuracy of IMF forecasts from 1990–2001. The report covered 87 countries, 57 of which were crisis countries that received IMF loans. Using Theil's U statistic, the report finds that forecasts for non-crisis countries outperform forecasts for crisis countries, but neither sample can beat the naïve random walk.

More recent studies of current account forecasts (as a percent of GDP) use 175 crisis countries for 8 years from 1993 to 2001 (Atoyan et al., 2004) or 183 crisis countries for 17 years from 1993 to 2009 (Atoyan & Conway, 2011). Atoyan et al. (2004) find forecasts biased and inefficient, and Atoyan and Conway (2011) decompose the forecast error to identify an association of the forecast error with a single, aggregate conditionality indicator. We disaggregate conditionality into 18 different types of policy reforms and quantitative performance targets to highlight specific conditionalities associated with inefficient forecasts. No previous study correlated forecast inaccuracy with such detailed conditionality.

An unusual finding is reported by Baqir et al. (2006), who examined 94 IMF crisis loans over 13 years (1989–2002) to report a surprising zero average forecast error for current accounts (as a percent of GDP). Luna (2014) examines 103 program countries over 9 years (2002–2011) to find that current account forecasts (as a percent of GDP) consistently erred on the pessimistic side, but without statistically significant bias. The impressive range of forecast (in)accuracies throughout the previous literature thus ranges from zero information value to zero average errors, at times with and without bias and inefficiencies. Notice that none of these studies (other than EKPC) examine imports and exports, meaning that import, export, GDP, and inflation forecast errors were always conflated.

This diversity of results in the previous studies may not only be driven by the small and always different samples. Artis (1996) suggests "data for many of these [developing] countries are poor and tardy", while Atoyan and Conway (2011) surmise the lack of forecast accuracy may have its origins not only in (i) poor data quality but also in (ii) "country-specific differences" in the forecast models, (iii) "incomplete policy implementations", and (iv) "random errors". Random errors should balance out in bias calculations, however. We rule out data quality issues by painstakingly auditing the MONA database using the original loan documents in the IMF archives. We address the effects of incomplete policy implementation by examining whether canceled programs drive bias and inefficiency but find no evidence for that hypothesis. Timmermann (2007) and EKPC (2019) suggest IMF forecast accuracy is affected by outliers; we show, however, in Section 2.2 that large observed forecast variances mirror large variations in the actual final data. While extreme values may exhibit high leverage, we examined results without extreme values and found that they did not deviate qualitatively from those reported below.[6]

Section 2 presents the data and methodology. Section 3 introduces baseline and robustness results. Section 4 investigates if forecast accuracy varies by country-incomes, crisis-types, or exchange rate regimes. Section 5 explores sources of forecast inefficiencies, while Section 6 documents the evolution of IMF forecast bias and inefficiency over time. Section 7 investigates if the forecast horizon affects forecast accuracy, and Section 8 concludes.

2. Data and methodology

2.1. Data

IMF forecasts were obtained from the IMF's *Monitoring of Fund Arrangements* database (MONA, IMF, 2021a), which reports data from loan documents that are presented to the IMF's executive board at the time of program approval.[7] We examine only current-year forecasts

[6] Results without extreme values are available upon request.

[7] The IMF's WEO database provides forecasts in April and October of each year, but does not provide forecasts for countries in crisis, or forecasts set at the time program loans commence. WEO forecasts for individual developing countries are not available prior to 2004. We audited the MONA database; a list of errors/corrections is provided in Appendix B.

(in year t for year t) that reflect the most recent information used in the loan program design. Missing observations rule out a meaningful analysis of longer forecast horizons. Our focus is on the growth of imports and exports of goods and services in the US dollar and on real exchange rate depreciations vis-à-vis the dollar. As we prepared the paper, the MONA database reported data for 602 crisis countries programs in 123 countries over 29 years, from 1992 to 2020.[8]

Imports and exports in the US dollar, together with real exchange rate depreciations, avoid conflating forecast errors of these variables with local currency and inflation forecast errors. This is also why we do not examine the current account balance as a share of GDP, which conflates forecast errors for imports, exports, GDP, or inflation. Kunze (2020) suggests that exchange rate forecast accuracy differs by exchange rate regimes. Hence, we also examine the quality of IMF forecasts in fixed and flexible exchange rate regimes. To identify exchange rate regimes as "fixed" or "flexible", we follow Ilzetzki et al. (2019) and the exchange rate regime classifications of the IMF's *Annual Report on Exchange Arrangements and Exchange Restrictions* (IMF, 2021e). When announced exchange rate bands cannot exceed +/-2 percent, regimes are labeled "fixed".[9]

The actual final data were obtained from three IMF databases. The final inflation data were obtained from the IMF's *World Economic Outlook* database (WEO, (IMF, 2021b)), final imports and exports from the IMF's *International Financial Statistics* database (IFS, IMF, 2021c), and final exchange rate data from the IMF's *Balance of Payments and International Investment Position* database (BOP, IMF, 2021d). Between missing observations in WEO, BOP, IFS, and MONA, we managed to gather forecasts and final outcome data for 576 programs. Our dataset is thus over three times larger and at least 10 years longer than the largest previous study of IMF external sector forecasts for program countries. Appendix Table A.1. provides a detailed data description.

2.2. Methodology to evaluate forecast accuracy

One approach to measuring forecast accuracy is to produce forecast error statistics such as the mean absolute error (MAE) to compare different forecasts. Such statistics are useful only in the presence of two or more forecasts, but in our case, forecast evaluation involves only a single forecast, the IMF's. This is not by choice; instead, the IMF is the sole institution with access to country-level data in times of crisis. This highlights the crucial importance of IMF forecasts: they are required for the lender of last resort's loan documents, and the IMF is the only entity with access to country data in times of crisis.

One forecast statistic that can be employed in the presence of only a single forecast is Theil's U2, which is a relative accuracy measure. It compares IMF forecasts to a naive random walk forecast based on "no-change extrapolations", where future values are predicted as past values. U2 is the root sum-of-squared forecast errors divided by the naive forecasting error where the forecast in *t+1* equals the actual in period *t.* U2 $=0$ indicates perfect forecasts, while U2 > 1 indicates forecasts do not outperform the naïve random walk and have no informational value. In times of crisis, when programs are designed to reverse the economic trajectory of a country, the random walk is a low bar.

An assessment of forecast accuracy requires formal statistical tests. These tests are based on Mincer and Zarnowitz (1969), who extended the seminal work of Theil (1961). Theil's "*Prediction-Realization Diagram*" (Fig. 1) displays IMF forecasts for the current year, F_t, on the horizontal axis and official, actual final data for the current year, A_t, on the vertical axis. Mincer and Zarnowitz (1969) label the solid 45-degree line the "*Line of Perfect Forecasts*" as it represents coordinates where program forecasts equal actual final data.

Fig. 1 visualizes that the extraordinary variation in forecasts is mirrored by a similarly unusual variance in the outcome data; perhaps, this is not surprising for crisis countries that require IMF programs. Timmermann (2007) noted that outliers influence the evaluation of non-G7 countries' forecasts, and EKPC eliminated 30 percent of their sample worrying about data errors or unusual extreme values. Fig. 1 does highlight the existence of extreme values for import, export, and exchange rate forecasts. Such extreme values may or may not exert high leverage; we find they are not sufficiently influential to qualitatively change our results below. We see no reason to exclude them from the analysis.

Mincer and Zarnowitz (1969) suggest a formal test for unbiased and efficient forecasts, which was first employed by Kenen and Schwartz (1986) to evaluate IMF forecast accuracy. The technique has since been applied frequently in IMF forecast evaluations (albeit in much smaller samples), see Artis (1996), Musso and Phillips (2002), Timmermann (2007), Genberg and Martinez (2014), EKPC, and Eicher and Gao Rollinson (2022)

$$A_t = \alpha + \beta F_t + \varepsilon_t. \tag{1}$$

Forecasts are chosen as the "independent" variable in (1) only because they are available before the actual final data is published. Forecasts are efficient when the forecast error is random and uncorrelated with forecasts. Nordhaus (1987) notes the similarity between forecast efficiency and stock market efficiency – both imply that all relevant and available information was considered, and all errors are white noise. In this case, the slope parameter, β, is unity, and the intercept, α, is zero. Since estimates of α and β are generally correlated, individual T-statistics are

[8] A total of 26 programs provided only technical assistance through non-financing facilities (*Policy Support Instrument* (PSI), *Policy Coordination Instrument* (PCI) programs, and *Flexible Credit Line* (FCL) programs without financial assistance); we include these programs for completeness; results do not change if these programs are excluded.

[9] We chose Ilzetziki et al.'s (2019) "*coarse exchange rate regime classification*" for fixed exchange rates that included (i) no separate legal tender, (ii) preannounced pegs, (iii) currency boards, (iv) de factor pegs, and (v) preannounced horizontal bands $\leq$ 2%. If we use Ilzetziki's "fine classification" for fixed exchange rate regimes, we can include additional countries with some form of exchange rate rigidities (including bands up to 5%, and "moving bands" which allow for appreciation and depreciations). Results (available upon request) are qualitatively similar.

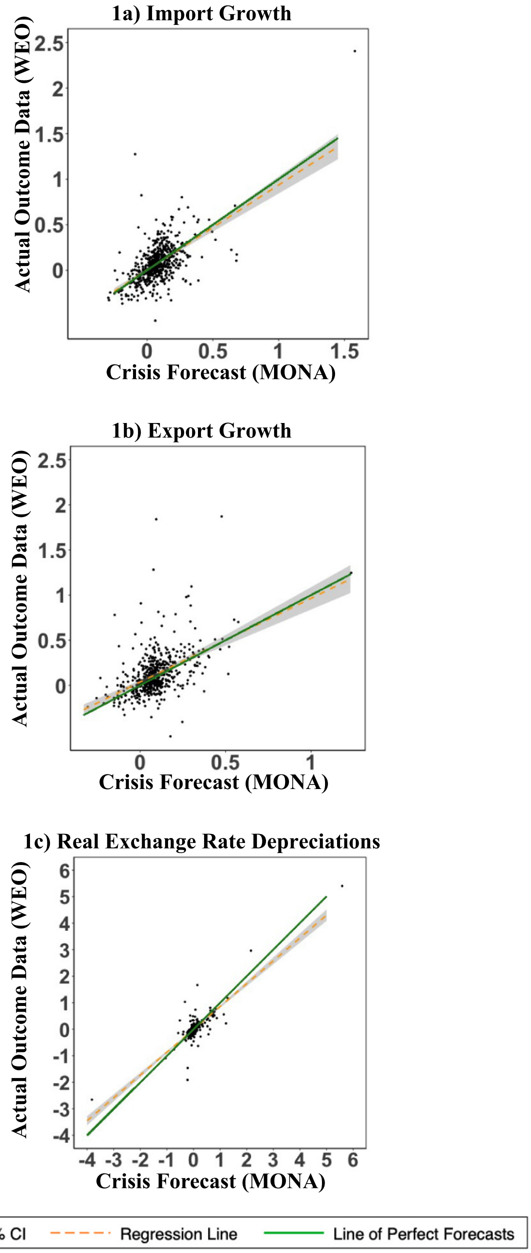

Fig. 1. Prediction-realization diagram, full sample.

inappropriate, and the joint null hypothesis, $\alpha = 0$ and $\beta = 1$, is tested.

If the Mincer and Zarnowitz null ($\alpha = 0\&\beta = 1$) is rejected, forecasts are inefficient, but they may not be biased in the sense that $E(A_t) \neq \mathrm{E}(F_t)$. Holden et al. (1987) demonstrate that $\alpha = 0$ is sufficient but not necessary for unbiased forecasts. Holden and Peel (1990) derived a necessary and sufficient condition for unbiased forecasts, which tests simply whether the regression line intersects the *Line of Perfect Forecasts* at $E(A_t) = E(F_t)$. When the Holden Peel test of $A_t - F_t = \gamma + \nu_t$ rejects the null of $\gamma = 0$, forecasts are said to be biased. Here, it is crucial to note that this metric of bias declares forecasts as "unbiased", when, for example, half are 40 percent higher, and half are 40 percent lower than the official final data. Bias is thus substantially less informative than efficiency.

3. Baseline results: the accuracy of IMF trade forecasts

Regressions (1a) - (3) in Table 1 present the results of Mincer and Zarnowitz regressions (Eq. (1)) associated with Fig. 1 (a-c). The regressions allow us to test for potential bias and inefficiency for IMF real exchange rate depreciations and import/export forecasts. To compare our results to the only previous analysis of IMF trade forecasts, regressions (1a) and (2a) restate EKPC results.

Table 1
Forecast bias and inefficiency for IMF program countries (1992–2020).

Dependent:actual final data	Import growth		Export growth		Real exchange rate depreciation
	EKPC	Full	EKPC	Full	Full
	1a	1b	2a	2b	3
Forecast (β)	0.63**	0.91	0.72***	0.93	0.86**
p-value ($\beta = 1$)	0.03	0.54	0.00	0.43	0.04
Constant (α)	0.02	0.00	0.03	0.04***	0.00
p-value ($\alpha = 0$)	0.50	0.80	0.18	0.00	0.62
Observations	110	578	110	576	565
Adj. R^2	0.47	0.35	0.55	0.24	0.73
MZ *p*-value ($\alpha = 0 \& \beta = 1$)	0.08*	0.77	0.00***	0.00***	0.11
HP *p*-value ($\gamma = 0$)	0.89	0.58	0.995	0.00***	0.40
Theil U2	0.65	0.76	0.55	0.79	0.54

EKPC Table (1) regressions include regional dummies, which we do not include to produce proper Mincer and Zarnowitz regressions to assess IMF forecast quality. Robust p-values, *** p<0.01, ** p<0.05, * p<0.1.

Their sample uses 110 of the 602 available programs in nominal, local currency (EKPC did not examine exchange rate forecasts), which conflates price and currency forecasts. EKPC also used estimated final data and omitted programs when data exceeded four standard deviations from the mean or when program durations did not exceed 18 months. This led EKPC to reject efficiency for both imports and exports. Results change as indicated in (1b) and (2b), once omitted and excluded programs along with the entire available time series from 1992–2020 are included in the dataset. Import forecasts are shown to be efficient and unbiased, and export forecasts are not only inefficient but also biased.

Theil coefficients in Table 1 are substantially smaller for EKPC than those observed in our sample. This is because EKPC's abundance of caution led them to exclude a sizable number of programs with data that represented large deviations from the mean. Many of these programs (46 for the EKPC time period, 2002–2015) were verified in our MONA audit and included in our dataset to produce substantially higher Theil coefficients. All Theil coefficients easily beat the random walk in the full sample.

Regression (3) indicates that the dichotomy in imports and exports forecast accuracy is not necessarily driven by inaccurate exchange rate forecasts. Neither the Holden Peel test nor the Mincer and Zarnowitz test for unbiased and efficient forecasts can be rejected in the full sample (EKPC did not examine exchange rate forecasts). However, the beta coefficient for exchange rate forecasts is significantly below unity, indicating that, on average, IMF forecasts systematically overestimate real exchange rate depreciations for program countries. When the regression line intersects the line of perfect forecast from above, as in the case of exchange rate forecasts, it implies that IMF forecasts are systematically different for programs with small vs. large depreciations. For countries with small forecasted/realized exchange rate changes, the IMF underestimates the resulting depreciation, while programs with large depreciations receive excessively large forecasts in times of crisis. Given the slope coefficient, we know the excessive depreciation forecasts dominate our sample. The low slope coefficient along with Fig. 1(c) also suggests that the forecast errors are substantial, and the confidence interval seldom includes the line of perfect forecast, so that errors simply equal out to highlight the weakness of the Holden Peel test that we discussed above.

The excessively large depreciation forecasts (on average) may explain the IMF's tendency to overestimate export growth (on average) in regression 2b, where the beta coefficient is also smaller than unity. On the other hand, imports are accurately forecasted, which thus presents a puzzle. Since the intercept for export growth in regression 2b is positive and significant, we know that exports for slow (fast) growth recoveries are under (over) estimated by IMF program forecasts.

Bias and inefficient export forecasts have important implications for IMF program countries. Forecasts that overestimated future exports may lead to underestimates of the required external capital inflows to jeopardize the recovery. Overly optimistic export forecasts also translate into optimistic program targets and performance criteria (e.g., trade balance, government revenues, unemployment rate, and exchange rate), which thus become more challenging to meet. When forecasts are overly optimistic, the country review will produce seemingly below-par country program performance and suggest a lack of conditionality implementation, when in fact, the differential may only be due to inaccurate export and exchange rate forecasts.[10]

[10] A number of authors (Aldenhoff, 2007; Beach et al., 1999; Luna, 2014) suggest IMF staff may face career incentives to forecast optimistic outcomes, while Musso and Phillips (2002) counter that the IMF Executive Board review process incentivizes staff to forecast initially pessimistic outcomes and to be able to characterize future results as "unexpectedly better" instead of "unexpectedly weak". Beaudry and Willems (2022) show that the quality of IMF forecasts may be contingent on the optimism/pessimism of individual IMF Mission Chiefs. Independent of career incentives or personal proclivities, Genberg and Martinez (2014) document that IMF desk economists with longer work experiences produce more accurate forecasts.

4. Forecast accuracy by country-income, crises-types, and exchange rate regimes

In a formal review of IMF program lending, Ghosh et al. (2006) suggest forecast accuracy may differ by programs-types or country-types since trade flows can recover at differential speeds depending on such country characteristics. Of course, the hope would be that forecasters appropriately incorporate country and crisis characteristics into their forecasts so that subsamples do not exhibit systematically different forecast performance. In this case, the external sector forecasts would not exhibit systematic differences by country or program-type. The unique size of our dataset allows, for the first time, an evaluation of country and program-specific subsample effects with a sufficient statistical power of inference. We thus examine the forecast accuracy by (i) crisis-types (inflation/BOP[11]), (ii) country-types (LICS/non-LICs)[12], (iii) exchange rate regimes (fixed/floating), and (iv) openness (financial/trade).

4.1. Import forecast accuracy by subsample

We examine import forecasts by subsample in Table 2a. Unbiased and efficient import forecasts are not universal across country-types. Specifically, fixed exchange rate regimes show a systematic pattern of inefficient import forecasts. This inefficiency is driven entirely by LICs. Once fixed exchange rate LICs are removed from the sample, IMF import forecasts are unambiguously unbiased and efficient. Our sample features 98 fixed exchange rate LICs that produce remarkably inefficient import forecasts. The vast majority of these loans (75) represent 14 African countries' programs that are members of the African Franc zones.[13] The low slope coefficient of 0.67 for fixed-exchange rate LICs implies forecasts overestimate import growth to the point where actual imports are on average 50 percent smaller (1/0.67–1) than predicted. Nevertheless, the forecast errors average out to indicate no systematic bias, highlighting that inefficiency is a more informative concept to assess forecast accuracy as we mentioned above.

The Theil U2 statistic confirms that import growth forecasts contain some informational value relative to naïve forecasts for all subsamples. We also include directional analysis introduced by Theil (1961) in the last row of Table 2. Directional analysis reports the percent of forecasts that do not match the direction of actual outcomes ($F_t > 0$ & $A_t < 0$, $F_t < 0$ & $A_t > 0$). Henriksson and Merton (1981) derived a test that pits this percentage against a coin flip to establish the null hypothesis that forecast directions and outcome directions are independent events.[14] Of course, this test sets a low bar, perhaps even lower than the threshold value of 1 in Theil's U2, but directional accuracy is nevertheless insightful for a number of reasons. First, instead of squaring or averaging positive and negative errors, the directional error analysis provides insight into the forecasting fundamentals: to what extent do forecasters assume a continuation of a trend or not. Second, when the IMF provides loans to countries in crisis in exchange for difficult conditionality, it is crucial to understand if IMF forecasts of future recoveries coincide at least with the observed directional changes for imports, exports, and exchange rates. Third, our particular interest in forecasts in times of crisis stems from the fact that turning points in crisis countries are to be expected after they accept IMF loans and conditions. Hence, one would expect IMF forecasters to do especially well against the random directional coin flips. Finally, we also examine the root mean squared error (RMSE), which turns out to correlate closely with Henriksson-Merton's directional inaccuracy, as higher directional errors correlated with higher RMSE (0.88 correlation).

For imports, between 20 percent and 29 percent of forecasts from 1992–2019 were not only off the mark in terms of their quantitative levels but in terms of the direction in which imports changed during the recovery. Here, it is of note that among the 4 subsamples (full, BOP-programs, floats, and fixed), LICs have a substantially higher rate of directional errors than other subsamples, reaching at times 29 percent for LICs with the balance of payments crises. Once LICs are purged from the sample, the directional error rate drops to about 20 percent for non-LICs.

4.2. Export forecast accuracy by subsample

Table 2b highlights that export and import forecast accuracies differ profoundly, especially at the subsample level. In contrast to import forecasts, all export forecasts other than LICs' are biased and inefficient, independent of exchange rate regimes, or crisis-types. Among LICs, fixed exchange rate export forecasts are again inefficient, and the low slope coefficient of 0.75 implies that IMF forecasts substantially exceed actual outcomes by 33 percent on average (1/0.75–1) as a one-unit increase in forecasts is associated with an increase in actual outcomes of 0.75 units only. The regression also suggests the mean actual export

[11] IMF BOP crisis loan programs include Extended Credit Facility (ECF), Extended Fund Facility (EFF), Exogenous Shock Facility (ESF), Flexible Credit Line (FCL), Stand-By Agreements (SBA), Standby Credit Facility (SCF), Precautionary Credit Line (PCL), and Precautionary Liquidity Line (PLL). The remaining programs in our dataset focus on structural reforms following crises: Structural Adjustment Facility (SAF), Policy Reform Instrument (PSI), Poverty Reduction and Growth Trust (PRGT), Policy Coordination Instrument (PCI), Enhanced Structural Adjustment Facility (ESAF), see IMF (2021a).

[12] Strictly speaking, when we refer to "the LICs sample" below, we are referring to 'non-hyperinflation-LICs.' Bias and inefficiency results do not change if we use the full LICs sample instead of non-hyperinflation-LICs, we believe, however, the latter sample is more informative. Results for 'all LICs' are available upon request.

[13] Benin, Burkina Faso, Guinea-Bissau, Côte D'Ivoire, Mali, Niger, Senegal, and Togo (West African CFA franc zone), and Cameroon, Central African Republic, Chad, Republic of the Congo, Equatorial Guinea, and Gabon (Central African CFA franc zone).

[14] Strictly speaking, the Merton-Henriksson test is based on Merton (1981) and Henriksson and Merton (1981), who examined whether market-timing forecasts of asset returns add informational value. Schnader and Stekler (1990) adapted Merton-Henriksson 2 × 2 contingency tables to assess directional accuracy and developed test statistics for the 2 × 2 case, see Diebold and Lopez (1996).

Table 2a
Import forecast bias and inefficiency by subsamples.

	Dependent variable: Actual final import growth data															
	Full sample				Non-hyperinflation				LICs				Non-LICs			
	Full	BOP	Float	Fixed	Full	BOP	Float	Fixed	Full	BOP	Float	Fixed	Full	BOP	Float	Fixed
Forecast (β)	0.91	1.08	0.96	0.81**	0.93	1.11	1.00	0.73**	0.95	1.25	1.05	0.67**	0.90	0.98	0.94	0.80
p-val. ($\beta = 1$)	0.54	0.61	0.85	0.05	0.64	0.53	0.99	0.01	0.83	0.32	0.87	0.01	0.31	0.86	0.61	0.33
Constant (α)	0.00	−0.01	−0.01	0.02*	0.00	−0.01	−0.01	0.03**	−0.01	−0.05*	−0.03	0.04**	0.01	0.00	0.00	0.02
p-val. ($\alpha = 0$)	0.80	0.52	0.72	0.07	0.97	0.30	0.39	0.01	0.77	0.09	0.33	0.02	0.39	0.93	0.99	0.16
Observations	578	308	371	207	512	263	324	188	269	63	171	98	243	200	153	90
Adj. R^2	0.35	0.53	0.35	0.34	0.38	0.57	0.43	0.26	0.37	0.71	0.43	0.22	0.37	0.41	0.42	0.26
MZ p-val. ($\alpha = 0 \& \beta = 1$)	0.77	0.81	0.65	0.07*	0.70	0.58	0.29	0.01***	0.52	0.23	0.24	0.03**	0.50	0.98	0.87	0.31
HP p-val. ($\gamma = 0$)	0.58	0.85	0.36	0.66	0.41	0.62	0.16	0.56	0.27	0.45	0.15	0.83	0.88	0.98	0.72	0.50
Theil U2	0.76	0.66	0.77	0.75	0.74	0.63	0.72	0.78	0.73	0.53	0.72	0.76	0.75	0.74	0.72	0.81
Directional inaccuracy	24%***	22%***	25%***	23%***	24%***	22%***	24%***	24%***	27%***	29%***	27%***	27%***	21%***	20%***	21%***	21%***
RMSE	0.18	0.16	0.19	0.15	0.16	0.15	0.17	0.15	0.19	0.19	0.20	0.17	0.14	0.13	0.13	0.14

Robust p-values, *** p<0.01, ** p<0.05, * p<0.1 Henriksson–Merton (H-M) $\chi 2$ test for directional inaccuracy null is no forecasting ability (outcomes are independent of the forecasts).

Table 2b
Export forecast bias and inefficiency by subsamples.

	Dependent variable: Actual final export growth data															
	Full sample				Non-hyperinflation				LICs				Non-LICs			
	Full	BOP	Float	Fixed	Full	BOP	Float	Fixed	Full	BOP	Float	Fixed	Full	BOP	Float	Fixed
Forecast (β)	0.93	1.03	1.00	0.81*	0.91	1.01	0.98	0.77**	0.95	1.23	1.04	0.75**	0.88	0.91	0.90	0.84
p-val. ($\beta = 1$)	0.43	0.83	0.97	0.05	0.38	0.93	0.90	0.02	0.71	0.59	0.84	0.07	0.22	0.39	0.42	0.31
Constant (α)	0.05***	0.02**	0.03**	0.05***	0.04***	0.02*	0.03*	0.06***	0.03	−0.01	0.01	0.05**	0.05***	0.03***	0.05**	0.06***
p-val. ($\alpha = 0$)	0.00	0.04	0.03	0.00	0.00	0.06	0.07	0.00	0.13	0.79	0.58	0.03	0.00	0.01	0.02	0.01
Observations	576	307	369	207	511	263	323	188	268	63	170	98	243	200	153	90
Adj. R^2	0.24	0.34	0.24	0.25	0.23	0.32	0.25	0.20	0.25	0.32	0.26	0.21	0.20	0.31	0.22	0.16
MZ p-val. ($\alpha = 0 \& \beta = 1$)	0.00***	0.03**	0.03**	0.00***	0.00***	0.04**	0.10	0.00***	0.21	0.83	0.63	0.07**	0.00***	0.03**	0.07*	0.02**
HP p-val. ($\gamma = 0$)	0.00***	0.02**	0.01**	0.01***	0.00***	0.03**	0.05*	0.01**	0.14	0.57	0.35	0.18	0.00***	0.01**	0.05*	0.01**
Theil U2	0.79	0.74	0.80	0.76	0.79	0.75	0.79	0.78	0.77	0.75	0.78	0.74	0.81	0.74	0.80	0.82
Directional inaccuracy	23%***	19%***	24%***	19%***	23%***	20%***	26%***	19%***	29%***	30%***	32%***	22%***	18%***	16%***	19%***	16%***
RMSE	0.20	0.18	0.22	0.18	0.20	0.18	0.21	0.18	0.22	0.25	0.23	0.18	0.19	0.15	0.19	0.18

Robust p-values, *** p<0.01, ** p<0.05, * p<0.1 Henriksson-Merton (H-M) $\chi 2$ test for directional inaccuracy null is no forecasting ability (outcomes are independent of the forecasts).

Table 2c
Exchange rate forecast bias and inefficiency by subsamples.

	Dependent variable: Actual final real exchange rate depreciation data															
	Full sample				Non-hyperinflation				LICs				Non-LICs			
	Full	BOP	Float	Fixed	Full	BOP	Float	Fixed	Full	BOP	Float	Fixed	Full	BOP	Float	Fixed
Forecast (β)	0.86**	0.81*	0.86*	0.80***	0.78**	0.95	0.75*	0.89	0.56**	0.88	0.49**	0.75	0.98	0.98	0.96	1.11
p-val. ($\beta = 1$)	0.04	0.06	0.09	0.00	0.05	0.60	0.06	0.37	0.03	0.78	0.05	0.10	0.76	0.67	0.49	0.43
Constant (α)	−0.00	−0.00	0.00	−0.00	0.00	0.02*	0.01	0.00	0.00	0.03	0.01	0.00	0.01	0.01	0.01	0.00
p-val. ($\alpha = 0$)	0.62	0.89	0.86	0.32	0.45	0.08*	0.36	0.72	0.88	0.36	0.74	0.82	0.28	0.14	0.20	0.87
Observations	565	302	367	197	500	257	321	179	262	59	170	92	238	198	151	87
Adj. R^2	0.73	0.61	0.75	0.84	0.25	0.38	0.21	0.47	0.09	0.14	0.05	0.34	0.59	0.59	0.57	0.68
MZ p-val. ($\alpha = 0 \& \beta = 1$)	0.11	0.17	0.21	0.00***	0.04*	0.03**	0.07*	0.45	0.03**	0.07*	0.08*	0.13	0.51	0.29	0.33	0.72
HP p-val. ($\gamma = 0$)	0.40	0.72	0.89	0.18	0.18	0.05*	0.29	0.27	0.35	0.20	0.58	0.14	0.26	0.13	0.19	0.77
Theil U2	0.54	0.65	0.52	0.46	0.88	0.79	0.90	0.68	0.98	0.93	1.003	0.79	0.64	0.64	0.66	0.53
Directional inaccuracy	22%***	18%***	23%**	21%***	24%***	20%***	24%**	23%***	31%***	29%	31%	32%***	16%***	17%***	17%***	15%***
RMSE	0.19	0.23	0.22	0.15	0.13	0.13	0.16	0.07	0.17	0.23	0.20	0.08	0.08	0.09	0.10	0.05

Robust p-values, *** p<0.01, ** p<0.05, * p<0.1 Henriksson-Merton (H-M) $\chi 2$ test for directional inaccuracy null is no forecasting ability (outcomes are independent of the forecasts).

growth is overpredicted by 0.05 percent when forecasts are zero. Together with the statistically significant positive constant, this implies that IMF forecasts under (over) estimate exports for low (high) growth recoveries. The full sample is biased and inefficient throughout, independent of crisis type or exchange rate regimes.

The Theil index indicates that export forecasts do hold informational value over naïve forecasts, although the directional analysis indicates a substantial share of forecasts feature the wrong direction. For the full sample, about 23 percent of export forecasts are not just quantitatively but also directionally inaccurate. This figure increases to over 30 percent for LICs (LICs-full, BOP-program-LICs, and float-LICs). Exports forecast RMSEs are even more closely related to export directional inaccuracy as for the case of import forecasts, as the correlation between directionally inaccuracy and RMSEs rises to 0.90 for IMF export forecasts.

4.2.1. Exchange rate forecast accuracy by subsample

Forecasts of real exchange rate depreciations in Table 2c exhibit a roughly similar pattern as imports, with some important exceptions. Similarly, once LICs are eliminated from the sample, exchange rate and import forecasts are unbiased and efficient for the remaining countries. For LICs, the inefficiency of exchange rate forecasts is driven by countries with floating exchange rates and countries with balance of payment crises. Of note is the staggeringly low slope estimate of 0.49 for LICs' floating exchange rate forecasts. It indicates that, on average, IMF forecasts predict 104 percent greater depreciation than realized (1/0.49–1). The average depreciation forecasts can be parsed further, as the low slope coefficient also indicates that low (high) depreciation events are systematically under (over) forecasted by the IMF.

These results imply that the IMF is much more optimistic than justified regarding the stability of exchange rates for low depreciation countries while expecting much

Table 2d
Contributions of nominal exchange rate and inflation to real exchange rate depreciation inaccuracies.

	Dependent variable: Actual final real exchange rate depreciation data															
Variables	Full sample				Non-hyperinflation				LICs				Non-LICs			
	Full	BOP	Float	Fixed	Full	BOP	Float	Fixed	Full	BOP	Float	Fixed	Full	BOP	Float	Fixed
Nom. E forecast (β_1)	0.82***	0.75***	0.84***	0.80***	0.79***	0.96***	0.75***	0.94***	0.56***	0.89**	0.49*	0.80***	0.99***	0.98***	0.96***	1.17***
p-val. (β_1=0)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.05	0.05	0.00	0.00	0.00	0.00	0.00
Inflation forecast (β_2)	0.74***	0.68***	0.75***	1.36***	0.70***	0.86***	0.68***	0.79***	0.52*	1.01**	0.46	0.68**	0.85***	0.86***	0.84***	0.96***
p-val. (β_2=0)	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.09	0.04	0.29	0.01	0.00	0.00	0.00	0.00
Constant (α)	−0.01	−0.01	−0.01	0.03	0.00	0.01	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	−0.01
p-val. ($\alpha = 0$)	0.11	0.34	0.24	0.23	0.89	0.45	0.91	0.71	0.95	0.20	0.96	0.99	0.73	0.97	0.97	0.42
Observations	559	297	364	195	495	253	319	176	258	56	168	90	237	197	151	86
Adj. R^2	0.70	0.55	0.76	0.44	0.25	0.38	0.21	0.48	0.09	0.13	0.05	0.34	0.59	0.59	0.57	0.69
p-value (β_1 = β_2)	0.16	0.33	0.05	0.19	0.56	0.56	0.71	0.35	0.87	0.77	0.90	0.62	0.42	0.49	0.61	0.32

Robust p-values, *** p<0.01, ** p<0.05, * p<0.1 Henriksson-Merton (H-M) $\chi 2$ test for directional inaccuracy null is no forecasting ability (outcomes are independent of the forecasts).

greater than realized depreciations for countries with large changes in exchange rates. These are concerning insights since IMF exchange rate forecasts provide crucial market guidance which has the potential to exacerbate crises if depreciations are much lower than forecast. At the same time, these results suggest the IMF imparts excessive confidence with regard to country performance for the most vulnerable crisis countries: LICs with larger than average depreciations. Such systematically optimistic real exchange rate depreciation forecasts have direct implications for the forecasted current account, reserve, and financing needs, rendering these program countries less likely to hit their actual quantitative performance targets.

LICs forecasts for floating exchange rate regimes are so inaccurate that the Theil coefficient indicates IMF forecasts hold zero information value as they cannot outperform the naïve, no-change random walk extrapolation. That sample also produces an unusually high RMSE. As noted in footnote 5, real exchange rate forecasts are notoriously difficult, and a voluminous literature outlines the associated forecasting challenges. Hence, perhaps, the most noteworthy insight derived here is that only LICs forecasts feature substantial large forecasting challenges, while forecasts for the full sample are indeed unbiased, efficient, and outperform the random walk.

The directional analysis is a lower bar than the Theil coefficient, since it examines only if the sign of the forecast is accurate. The Henriksson-Merton test for directional accuracy for the floating LICs subsample cannot reject the null that the direction of forecasts and outcomes are independent, with directional inaccuracies exceeding 30 percent for just about all LICs subsamples. Once the sample is purged of LICs, directional accuracy improves dramatically, and non-LICs exhibit only between 15 percent–17 percent of directional inaccuracies. For Non-LICs, we thus find the lowest level of forecast inaccuracy among either of the three forecasts that we examined (imports, exports, and exchange rates). Directional inaccuracies are less tightly correlated with the RMSE (correlation 0.38) than in the case of imports and exports, suggesting that directional errors are less of a driver of the forecast errors for real exchange rate depreciations.

4.2.2. Disaggregation of exchange rate forecast accuracy

A remaining question is whether the source of real exchange rate forecast errors derives from forecast errors for nominal exchange rates or inflation. In Table 2d, we show the contribution of inaccuracies in nominal exchange rate and inflation forecasts to real exchange rate forecasts. We find inaccuracies in the full sample are roughly equally explained by nominal exchange rates and inflation forecast errors. We cannot reject the null hypothesis that the two estimates ($\beta_1 = \beta_2$) are equal for all but one subsample (floating exchange rate regimes). The inability to reject ($\beta_1 = \beta_2$) is driven, however, entirely by hyperinflation countries. Once they are removed, the nominal exchange rate and inflation contribute equally. There is thus a clear pattern that observed statistically significant deviations from perfect forecasts in Table 2c are driven equally by nominal exchange rate and inflation forecast errors.

Our findings regarding import, export, and exchange rate accuracy reveal three important insights: First, import and exchange rate forecasts are unbiased and efficient once we eliminate LICs from the sample, while export forecasts are deeply flawed throughout all subsamples. Second, for import forecasts, fixed exchange rate LICs introduce noise, essentially produced by forecasts for African franc zones countries. Third, for exchange rate forecasts, flexible exchange rate LICs introduce substantial noise to the point that IMF forecasts cannot beat the naïve random walk, and the forecasts and outcomes are so directionally challenged that we cannot reject that the two are independent. All three insights combined suggest that greater efficiency of LICs program forecasts should be a key policy goal for future IMF loan programs.

4.3. Openness and forecast accuracy

The accuracy of external account forecasts may well be associated with trade and/or financial openness, if only because more open economies face greater exposure to contagion and idiosyncratic global shocks, whose advent and effects are difficult to forecast. In a panel of 29 OECD/BRIICS countries, Lewis and Pain (2014) document that OECD GDP growth forecast accuracy decreases with trade openness and with banking assets held by foreign banks. The authors surmise openness allows external imbalances and financial leverage to accumulate faster and to increase uncertainty and forecast inaccuracy. Chatterjee (2019) and Chatterjee and Nowak (2016) document that financial and trade openness impact forecasts since openness increases the sensitivity of real variables (including trade) to uncertainty shocks.

In this section, we examine whether forecast accuracy in IMF crisis countries differs systematically by GDP trade shares and financial openness (Chinn-Ito index, 2006). We use the global sample mean for each measure

Table 2e
Import forecast bias and inefficiency by subsamples.

	Dependent Variable: Actual Final Import Growth Data																			
	Full Sample					Non-Hyper Inflation					LICs					Non-LICs				
	Full	Fin. Open		Trade Share		Full	Fin. Open		Trade Share		Full	Fin. Open		Trade Share		Full	Fin. Open		Trade Share	
	Full	Large	Small	Large	Small	Full	Large	Small	Large	Small	Full	Large	Small	Large	Small	Full	Large	Small	Large	Small
Forecast (β)	0.91	0.90	0.91	0.84	0.93	0.92	0.91	0.94	0.81	0.95	0.95	0.82	0.96	0.77	0.98	0.90	1.00	0.85	0.86	0.91
p-val. ($\beta = 1$)	0.54	0.35	0.67	0.28	0.71	0.64	0.40	0.77	0.26	0.81	0.83	0.42	0.90	0.43	0.94	0.31	1.00	0.38	0.43	0.47
Constant (α)	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	−0.01	−0.01	−0.01	0.00	−0.01	0.01	0.00	0.01	0.01	0.00
p-val. ($\alpha = 0$)	0.80	0.65	0.87	0.53	0.97	0.97	0.79	0.93	0.48	0.86	0.77	0.60	0.86	0.93	0.73	0.39	0.77	0.35	0.34	0.69
Observations	578	181	361	162	416	512	173	316	141	371	269	59	204	48	221	243	114	112	93	150
Adj. R2	0.35	0.38	0.34	0.28	0.37	0.38	0.40	0.37	0.27	0.41	0.37	0.29	0.39	0.20	0.41	0.37	0.48	0.29	0.33	0.37
MZ p-val. ($\alpha = 0\&\beta = 1$)	0.77	0.42	0.86	0.56	0.85	0.70	0.55	0.90	0.53	0.75	0.52	0.24	0.79	0.63	0.62	0.50	0.96	0.57	0.61	0.74
HP p-val. ($\gamma = 0$)	0.58	0.26	0.64	0.85	0.59	0.41	0.37	0.65	0.72	0.45	0.27	0.12	0.51	0.50	0.37	0.88	0.79	0.78	0.80	1.00
Theil U2	0.76	0.75	0.77	0.79	0.76	0.74	0.74	0.74	0.79	0.72	0.73	0.81	0.72	0.85	0.71	0.75	0.69	0.79	0.72	0.76
Directional Inaccuracy	24%***	21%***	26%***	27%***	23%***	24%***	21%***	27%***	25%***	24%***	27%***	29%***	27%***	33%***	26%***	21%***	18%***	26%***	20%***	21%***
RMSE	0.18	0.13	0.19	0.18	0.18	0.16	0.13	0.18	0.16	0.18	0.19	0.15	0.19	0.21	0.18	0.14	0.11	0.15	0.12	0.14
RMSPE	28.68	44.87	10.49	15.13	32.46	27.98	45.68	11.21	16.21	31.31	36.76	76.89	8.5	11.22	40.22	12.39	10.37	14.94	18.27	6.46

Robust p-values, *** p<0.01, ** p<0.05, * p<0.1 Henriksson-Merton (H-M) $\chi 2$ test for Directional inaccuracy null is no forecasting ability (outcomes are independent of the forecasts).

Table 2f
Export forecast bias and inefficiency by subsamples.

	Dependent Variable: Actual Final Export Growth Data																			
	Full Sample					Non-Hyper Inflation					LICs					Non-LICs				
	Full	Fin. Open		Trade Share		Full	Fin. Open		Trade Share		Full	Fin. Open		Trade Share		Full	Fin. Open		Trade Share	
	Full	Large	Small	Large	Small	Full	Large	Small	Large	Small	Full	Large	Small	Large	Small	Full	Large	Small	Large	Small
Forecast (β)	0.93	0.82	0.91	1.02	0.85	0.91	0.83	0.91	1.00	0.84	0.95	0.94***	0.95***	1.02***	0.89***	0.88	0.79***	0.87***	1.01***	0.79***
p-val. ($\beta = 1$)	0.43	0.14	0.44	0.74	0.33	0.38	0.16	0.52	0.96	0.34	0.71	0.80	0.78	0.80	0.66	0.22	0.14	0.38	0.92	0.10
Constant (α)	0.05***	0.02*	0.05***	0.04**	0.04***	0.04***	0.02	0.05***	0.03**	0.04***	0.03	−0.02	0.03*	0.02	0.03	0.05***	0.03**	0.07***	0.04**	0.05***
p-val. ($\alpha = 0$)	0.00	0.10	0.00	0.02	0.00	0.00	0.14	0.00	0.01	0.01	0.13	0.50	0.08	0.53	0.20	0.00	0.02	0.00	0.02	0.01
Observations	576	181	360	162	414	511	173	315	141	370	268	59	203	48	220	243	114	112	93	150
Adj. R2	0.24	0.28	0.21	0.48	0.16	0.23	0.29	0.20	0.50	0.15	0.25	0.25	0.25	0.55	0.16	0.20	0.34	0.11	0.42	0.13
MZ p-val. ($\alpha = 0\&\beta = 1$)	0.00***	0.18	0.00***	0.01**	0.00***	0.00***	0.24	0.00***	0.03**	0.02**	0.21	0.61	0.11	0.73	0.25	0.00***	0.05*	0.01**	0.01***	0.01**
HP p-val. ($\gamma = 0$)	0.00***	0.78	0.00***	0.00***	0.01**	0.00***	0.85	0.00***	0.01***	0.03**	0.14	0.34	0.06	0.44	0.20	0.00***	0.18	0.01***	0.00***	0.05*
Theil U2	0.79	0.77	0.81	0.63	0.84	0.79	0.76	0.8	0.61	0.84	0.77	0.8	0.76	0.58	0.82	0.81	0.72	0.87	0.65	0.87
Directional Inaccuracy	23%***	23%***	24%***	17%***	25%***	23%***	24%***	18%***	18%***	25%***	29%***	37%***	27%***	19%***	31%***	18%***	16%***	21%***	18%***	17%***
RMSE	0.2	0.14	0.23	0.16	0.22	0.2	0.14	0.23	0.15	0.22	0.22	0.17	0.23	0.19	0.22	0.19	0.11	0.24	0.13	0.21
RMSPE	116.2	6.82	146.8	5.81	137	15.96	6.84	19.69	6.1	18.38	21.8	10.73	24.37	9.16	23.68	3.39	3.38	3.64	3.63	3.23

Robust p-values, *** p<0.01, ** p<0.05, * p<0.1 Henriksson-Merton (H-M) $\chi 2$ test for Directional inaccuracy null is no forecasting ability (outcomes are independent of the forecasts).

Table 2g
Exchange rate forecast bias and inefficiency by subsamples.

	Dependent Variable: Actual Final Real Exchange Rate Depreciation Data																			
	Full Sample					Non-Hyper Inflation					LICs					Non-LICs				
	Full	Fin. Open		Trade Share		Full	Fin. Open		Trade Share		Full	Fin. Open		Trade Share		Full	Fin. Open		Trade Share	
	Full	Large	Small	Large	Small	Full	Large	Small	Large	Small	Full	Large	Small	Large	Small	Full	Large	Small	Large	Small
Forecast (β)	0.86**	0.38***	0.89	0.92	0.69***	0.78**	1.01	0.61***	0.85	0.74***	0.56**	1.10	0.43***	0.62	0.49***	0.98	0.95	0.92	1.01	0.96
p-val. ($\beta = 1$)	0.04	0.00	0.39	0.25	0.00	0.05	0.97	0.00	0.57	0.00	0.03	0.90	0.00	0.43	0.00	0.76	0.70	0.17	0.87	0.62
Constant (α)	0.00	0.01	−0.01	0.03	−0.01*	0.00	0.01	0.00	0.04**	−0.01	0.00	0.03	0.00	0.09**	−0.02**	0.01	0.00	0.01	0.01	0.01
p-val. ($\alpha = 0$)	0.62	0.51	0.29	0.11	0.07	0.45	0.42	0.65	0.05	0.26	0.88	0.43	0.67	0.05	0.03	0.28	0.75	0.30	0.41	0.45
Observations	565	178	350	158	407	500	170	308	135	365	262	58	198	48	214	238	112	110	87	151
Adj. R2	0.73	0.08	0.56	0.82	0.48	0.25	0.20	0.19	0.22	0.29	0.09	0.10	0.08	0.06	0.11	0.59	0.49	0.48	0.61	0.57
MZ p-val. ($\alpha = 0\&\beta = 1$)	0.11	0.00***	0.25	0.17	0.00***	0.04*	0.35	0.01**	0.05*	0.00***	0.03**	0.46	0.01***	0.04**	0.00***	0.51	0.74	0.23	0.70	0.63
HP p-val. ($\gamma = 0$)	0.40	0.34	0.13	0.12	0.00***	0.18	0.28	0.32	0.02**	0.60	0.35	0.35	0.54	0.03**	0.29	0.26	0.60	0.31	0.42	0.43
Theil U2	0.54	1.07	0.67	0.44	0.79	0.88	0.89	0.94	0.9	0.85	0.98	0.95	1.03	0.99	0.97	0.64	0.71	0.71	0.62	0.65
Directional Inaccuracy	22%***	24%***	23%***	17%***	24%***	24%***	24%***	25%***	21%*	26%***	31%***	0.36	31%***	21%*	34%***	16%***	17%***	15%**	17%**	15%***
RMSE	0.19	0.2	0.19	0.26	0.16	0.13	0.15	0.13	0.19	0.11	0.17	0.23	0.15	0.29	0.12	0.08	0.07	0.09	0.09	0.08
RMSPE	381	24.66	483.8	2.02	448.9	14.94	25.23	3.3	2.17	17.43	13.99	28.87	3.85	1.39	15.46	15.91	23.12	1.96	2.5	19.89

Robust p-values, *** p<0.01, ** p<0.05, * p<0.1 Henriksson-Merton (H-M) $\chi 2$ test for Directional inaccuracy null is no forecasting ability (outcomes are independent of the forecasts).

to parse our sample into "more open" and "less open" trade/financial subsamples. Tables 2e–2g report results suggesting forecast accuracy is remarkably stable across subsamples. Generally, both more and less open subsamples for both trade and capital flows show little difference in terms of bias and inefficiency when compared to the full sample for imports, exports, and the exchange rate depreciations. This result holds even when we compare subsamples by income levels. For example, if non-LICs forecasts cannot reject the null that imports are unbiased and efficient, the same is true for non-LICs with more/less open trade/financial openness.

While the tables convey a strong overall sense that IMF forecast accuracy does not differ systematically for more/less trade/financial openness, there are two exceptions. For real exchange rate depreciations, the full sample is never biased for any country-income subsample, but countries with small trade shares are inefficient in all but the non-LICs sample. Once hyperinflation countries are removed, financial openness is improving the accuracy of real exchange rate depreciation forecasts. For exports, all full samples for all income levels are biased and inefficient except for LICs, but countries with high degrees of financial openness buck the trend as they are unbiased in all subsamples, and efficient in all but the non-LICs country-income subsample. Financial openness is again clearly improving the accuracy of IMF export forecasts.

5. Sources of forecast inefficiency

After establishing the extent to which specific subsamples suffer from systematic bias and inconsistency, we turn our attention to identifying possible sources of IMF

Table 3
Forecast bias and inefficiency: Canceled programs excluded.

Dependent: actual final data	Import growth	Export growth	Real exchange rate depreciation
	No canceled	No canceled	No canceled
	1c	2c	3c
Forecast (β)	0.71***	0.848*	0.875*
p-value ($\beta = 1$)	0.00	0.08	0.07
Constant (α)	0.02*	0.05***	−0.00
p-value ($\alpha = 0$)	0.09	0.00	0.71
Observations	511	509	500
Adj. R^2	0.20	0.18	0.74
MZ *p*-value ($\alpha = 0 \& \beta = 1$)	0.00***	0.00***	0.17
HP *p*-value ($\gamma = 0$)	0.34	0.00***	0.53
Theil U2	0.85	0.81	0.53

Robust p-values, *** p<0.01, ** p<0.05, * p<0.1.

forecast inaccuracies. We examine if all tangible information available at the time of the forecast was adequately integrated into the forecast. The unprecedented size of our dataset allows us to introduce a substantial number of covariates that represent possible sources of forecast errors for the samples with highly inaccurate forecasts.

5.1. Program cancellations

Mussa and Savastano (1999) and Stiglitz (2011) note that MONA forecasts may not reflect actual forecasts but negotiated compromises between the IMF and country authorities. If this were the source of forecast errors, it would be difficult to explain why authorities systematically argue for excessively optimistic export growth or excessively large depreciations; both set up countries for debt, reserve, and overall program shortfalls.

Atoyan and Conway (2011), Luna (2014), and IMF (2019) point out that IMF forecasts are conditional on the assumption that conditionality is implemented; hence, implementation failures may well explain overly optimistic IMF forecast bias. If this were the case, it is difficult to explain the asymmetric bias and inefficiency in IMF forecasts, where imports are unbiased and efficient, but exports and exchange rate forecasts show systematic forecast errors. Nevertheless, since this is an important line of reasoning, we explore the effects of implementation failures by dropping programs from our dataset that experienced program cancellations.

Programs are canceled when IMF loan performance reviews do not produce sufficient evidence of conditionality implementation. The IMF monitors conditionality implementation in regular intervals as a part of its program surveillance (monthly, quarterly, or biannually depending on the program), and waivers may be obtained if conditionality or policy implementation is lagging. Lack of implementation may initially lead to "conditionality waivers", but eventually, too many waivers may trigger program cancellation. In our dataset, cancellations occurred in 69 programs. In Table 3, we reproduce Table 1 without canceled programs to document that forecast inaccuracy was not driven by canceled programs. Indeed, results without canceled programs suffer, and forecast accuracy deteriorates. This is because most of the canceled programs (65 percent) are non-LICs, which we now know to receive on average better forecasts than LICs. The absence of these non-LICs thus weighs down the sample with a larger share of inefficiently forecasted LICs, so that overall forecast accuracy without these non-LICs deteriorates. Limited conditionality implementations and cancellations thus cannot explain forecast inaccuracies.

5.2. Conditionality, loan size, and geopolitics as drivers of inefficiency

In addition to program cancellation, the literature suggests three areas where information available at the time of the forecasts may not have been integrated properly into IMF forecasts. We examine all three areas below, which include IMF conditionality, loan size, and geopolitical events. A voluminous literature examines the effect of conditionality of forecast errors (see, for example, the survey contained in Stubbs et al. (2020)), but this literature is hampered by selection bias since conditionality is included only selectively. For example, IMF (2019) associates monetary conditionality with inflation forecast errors, while Ismail et al. (2020) examine whether the number of conditions affects GDP growth forecast errors. Carriere-Swallow and Marzluf (2021) examine whether credit and fiscal conditionality affects GDP growth forecasts. Only Eicher and Gao Rollinson (2022) examine the effect of all conditions on GDP growth and inflation. They examine whether any of the entire palette of conditions affect forecast accuracy GDP growth and inflation. We follow their approach and, instead of selecting particular conditions, we test the entire range of conditionality as categorized into thematic topics by the IMF in the MONA database.[15]

Second, we investigate whether loan size affects IMF forecast accuracy since an important strand of the literature consistently links IMF forecast optimism to the size of the loan a program received (see, e.g., Beach et al., 1999; Dreher et al., 2008; Luna, 2014). Finally, we ex-

[15] The IMF MONA database provides a list of conditions for each program, grouped by IMF-specified programmatic objectives (see Table A.1).

Table 4
Sources of forecast inefficiencies.

Dependent variable:		Import growth (fixed exchange rate LICs)	Export growth (full sample)	Real exchange rate depreciation (float LICs)
Forecast	**IMF forecast** (β)	0.67**	0.90	0.58*
	p-value ($\beta = 1$)	0.02	0.31	0.07
	Loan size: Loan/Quota	−7.81***	−0.06	6.61
		0.00	0.66	0.24
Conditionality: Quantitative Performance Criteria	**Arrears:** External Ceilings	0.02	0.02	−0.06*
		0.74	0.27	0.07
	Arrears: Domestic Ceilings	0.09*	−0.03	−0.02
		0.05	0.15	0.49
	Fiscal: Gov't/Pub. Credit	−0.00	−0.05**	0.04
		0.92	0.04	0.45
	Fiscal: Gov't Balance Limits	−0.07*	−0.01	0.02
		0.09	0.65	0.5
	Debt: Short-Term	0.02	0.04**	−0.03
		0.73	0.05	0.25
Conditionality reforms	**Reforms: General Gov't**	0.20***	0.00	0.08
		0.01	0.91	0.35
	Reforms: Trade	−0.03	0.03	0.08**
		0.51	0.24	0.02
Geopolitics	**Elections:** Executive/Legisl.	−0.01	−0.03*	0.08*
		0.81	0.05	0.09
	Wars: International	Na	−0.06**	0.00
		Na	0.03	0.94
	Wars: Intranational	0.07*	−0.03	0.03
		0.09	0.25	0.54
	Constant (α)	−0.14	0.025	0.03
		0.13	0.03	0.19
	Observations	98	576	170
	Adj. R^2	0.31	0.24	0.18
	SJS F-test (α=δ=0 & $\beta = 1$)	5.34***	1.46*	1.33
	p-value	0.00	0.07	0.16

Examining drivers of biased and inefficiency for import, export, and exchange rates (In Table 2).
Note: Robust p-values, *** p<0.01, ** p<0.05, * p<0.1
Variables included in the regressions include (insignificant variables are not reported in the table to economize on space): General Gov't Reform, Central Bank Reforms, Civil Service Wage/Employment Reforms, Pension Reforms, Gov't Enterprise Pricing Reforms, Financial Sector Reforms, Current Capital Account Openness/Reforms, Tariff/Quota Reductions/Reforms, Labor Market Wage/Employment Reforms, Statistics Reforms, Legal/Market Reforms as well as quantitative reform criteria, such as Domestic Credit Ceiling, Gov't/Public Sector Credit Ceilings, BOP Reserve Tests, Debt Ceilings (short, medium, and long term), Arrears Ceilings (domestic and external), Fiscal Deficit Ceilings.

plore whether the economic effects of geopolitical events, known to IMF forecasters at the time of forecasts, were adequately integrated into IMF forecasts. In their reviews of IMF program design and conditionality, IMF (2019) note that forecast errors are impacted by political transitions, conflicts, and natural disasters (see also Przeworski & Vreeland, 2000, Park, 2006, Kentikelenis et al., 2016; Mody & Rebucci, 2006). Hence, we consider variables related to elections (executive and legislative) up to one year before the start of a program. For conflicts, we include indicators for (civil) wars that were initiated up to one year prior to the program starts. Finally, we consider natural disasters that occurred up to one year prior to the program starts (see Table A.1 for election, conflict, and disaster data). Crucial is that all geopolitical events were known at the time of forecasts, so IMF forecasters were well aware of their potential effects on economic performance.

To investigate drivers of inefficiencies, we utilize the methodology of Sinclair et al. (2010), who extended Mincer and Zarnowitz regressions (1) to include additional covariates, X_t, that are thought to represent information available to forecasters at the time of the forecast:

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

If any entries in the vector δ are non-zero, the information contained in the associated covariates can then, in part, explain forecasters' bias and inefficiencies. Significant covariates thus represent areas the IMF may consider with special caution in future IMF program forecasts. Sinclair et al. (2010) propose the joint null hypothesis of $\beta = 1$ & $\alpha = \delta = 0$ as a formal test of whether the information contained in the additional covariates was properly included in the forecast. If the null is rejected, Sinclair et al. (2010) note that the information contained in X was not fully integrated into the forecast, which then identifies possible sources of inefficiency. We apply the methodology to those three subsamples that were identified as drivers of forecast errors for imports, exports, and exchange rate forecasts (Table 4).

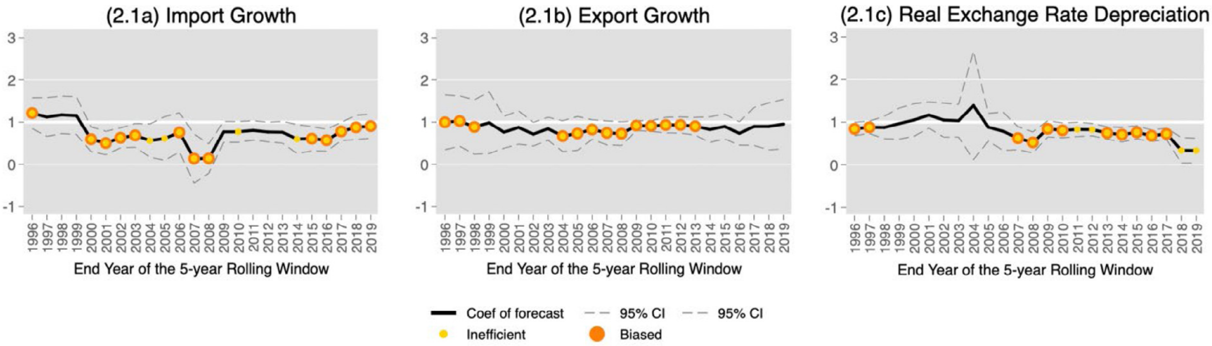

Fig. 2. Evolution of forecast bias and efficiency over time 5-year rolling windows, full sample.

In Table 4, we are considering (i) fixed exchange rate LICs subsample for import forecasts, (ii) the full sample for export forecasts, and (iii) the floating exchange rate LICs subsample for exchange rate depreciation forecasts. The Sinclair et al. (2010) test indicates that the additional, significant regressors cannot be rejected as information known at the time of the forecast that was not properly integrated for import and export growth. For real exchange rate depreciations forecast, the Sinclair et al. (2010) test rejects the hypothesis that the additional regressors could have improved forecast accuracy; this finding is not surprising given the problems encountered by exchange rate forecasts in general (see, e.g., Meese & Rogoff, 1983).

Given the sizable number of covariates, Table 4 reports only significant regressors in either subsample to economize on space. Significant coefficients are found for 11 covariates, whose effects were thus not properly integrated into IMF forecasts. Five of these regressors relate to quantitative conditionality: arrears ceilings (domestic & foreign), government finance (credit and balance limits), and short-term debt, while two relate to structural reform conditionality (trade and government). Geopolitical effects are also significant, specifically elections, wars, and civil wars.

Subsamples do not share statistically significant regressors, which is not surprising since the regressions cover not only different country-types but also different dependent variables (export/import/exchange rates) and subsample characteristics (fixed/flex exchange rates). Nevertheless, several common themes present themselves as LICs subsamples share significant regressors in the arrears category (domestic arrears for imports and external arrears for exchange rates), and trade forecasts share significant regressors in the fiscal category (fiscal balance limits for imports and government credit limits for exports). Truly broad-based effects are established only for geopolitical events where either elections or (civil) wars are significant in all three subsamples. Finally, we also confirm the findings in the previous literature that loan size is indeed a factor that affects IMF forecast accuracy, but only for import forecasts in LICs with fixed exchange rates.

6. Did forecast bias and efficiency change over time?

Forecasts for IMF program countries are available dating back to 1992, which raises the question of whether the accuracy of IMF external sector forecasts has changed over time. Instead of reporting forecast accuracy for individual years, we report results for 5-year rolling windows to avoid large fluctuations in the number of observations across samples and retain sufficiently large samples for each time interval. Fig. 2(1–3) provide visual summaries of the evolution of forecast accuracy over time, based on Mincer and Zarnowitz regressions reported in Appendix Table A.2. The solid black line in Fig. 2(1) represents the values of β estimates in Table A.2, while the dotted lines represent 95 percent confidence intervals. Yellow dots indicate years with inefficient forecasts; bias years receive an orange dot.

Only export growth forecasts show a marked improvement in both bias and efficiency since 2013. Import growth forecasts and real exchange rate deprecations continue to exhibit substantial inaccuracies even in recent years. Real exchange rate forecasts were indeed much more accurate prior to the arrival of 2007 in the 5-year rolling window. It is fascinating to see that, even in windows when real exchange rate forecasts are accurate, import and export forecasts can exhibit substantial bias and/or inefficiency. The slope coefficients for real exchange rate depreciations spike in the 2000–2004 window to exceed unity while substantially declining thereafter.

7. Do forecast horizons affect forecast accuracy?

Forecast accuracy is generally thought to decrease as forecast horizons increase (Armstrong, 2001; U.S. Government Accountability Office (USGAO), 2003). One might suspect this to be particularly relevant for IMF program forecasts, as forecasters' information sets grow substantially in size and accuracy toward the end of the year as ever more data vintages are released. Timmermann (2007) previously found evidence that IMF forecast errors increase with time horizons (using WEO data, which does not include crisis countries). In this section, we examine whether forecast horizons drive bias and inefficiency for IMF trade forecasts. Fig. 3(a-c) provides a visual summary of forecast accuracy by the month in which programs

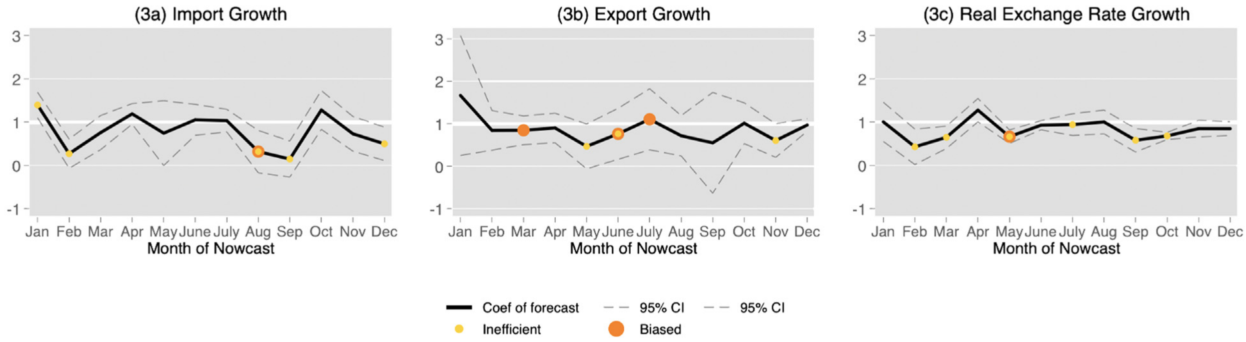

Fig. 3. Forecast horizons and forecast accuracy full sample.

were approved. The figure is a visual representation of the Mincer and Zarnowitz regressions in Appendix Table A.3.

Examining whether forecasts produced earlier in the program year exhibit a greater propensity toward bias and inefficiency than those formed later in the year does not produce the expected result. There is no clear pattern of improved forecast accuracy for programs that are approved later in the year. Import and export forecasts show inefficiencies as late as December and November, respectively. Exchange rate forecasts show inefficiencies as late as October. Biases and inefficiencies are distributed roughly evenly across the year without a clear pattern of either bias or efficiency improvements as forecast horizons change.

8. Conclusion

We analyze the accuracy of IMF's external sector forecasts for crisis countries. The previous literature on trade forecast evaluations was hampered by small-sample studies that cover ever-different samples and time periods and often conflated import, export, and exchange rate forecast errors. This may have been the reason for the vastly different results in the literature, ranging from exactly zero average forecast errors to zero informational values of IMF forecasts. We audited the IMF MONA database which contains external sector forecasts of crisis countries to overcome the small-sample issues; our sample is thus three times larger than any previous study, and it contains the first evaluation of real exchange rate depreciation forecasts. The size of our dataset also allows us to drill down to different subsample levels (country-income levels, program-types, exchange rate regimes, financial openness, and trade shares) to understand the drivers of forecast inaccuracies.

Our results indicate import forecasts are broadly unbiased and efficient while export forecasts are broadly biased and inefficient. We can confirm that this surprising asymmetry in import and export forecast accuracy is not driven by exchange rate forecast inaccuracies. Once the sample is purged of LICs exchange rate forecasts are unbiased and efficient in contrast to what one might expect given the Meese and Rogoff (1983) puzzle. Nevertheless, for the full sample and for floating exchange rate countries, unusually large forecast errors hamper exchange rate forecasts and at times cannot outperform the random walk. The exchange rate is a crucial variable and central to equilibrium conditions in both goods and asset markets of the macroeconomic model predicting crisis recovery. Hence, we suggest the explicit acknowledgment of uncertainties involved in IMF exchange rate forecasts and how these uncertainties translate to debt and reserve adequacy projections. This recommendation contrasts with the recent IMFs policy of not disclosing the exchange rate forecasts that underlie the entire crisis recovery program.

We also examine whether IMF conditionality is a driver of bias and inefficiency and find that IMF conditionality relates to trade and government reforms, ceilings on domestic/foreign arrears, and limits on fiscal finance (government credit/fiscal balance) that are not properly integrated into IMF forecasts. The policy implication is that IMF forecasts must be sensitive to the effects of IMF conditionality, especially as they relate to fiscal finance, arrears, and reforms. We also find that geopolitical events that are known at the time of the forecast (elections and armed conflict) were not properly accounted for in IMF forecasts to cause systematic forecast inaccuracies. Future forecasts should thus be more sensitive to the economic impact of non-economic events in crisis countries' recoveries.

Contrary to conventional wisdom, we find the accuracy of IMF current-year forecasts does not improve as the time horizon shrinks, but our longest forecast horizon is only one year. While forecast accuracy did improve over the past 28 years for exports, the accuracy of imports and exchange rate forecasts remains a challenge even in the most recent forecasts. Clearly, forecast accuracy is affected by exogenous shocks, such as the COVID crisis or the 2008 global financial crisis. However, if the forecast errors we observed were due to idiosyncratic shocks only, we would expect them to cancel instead of producing the systematic bias and inefficiency.

Our findings of systematically optimistic, biased, and inefficient export and exchange rate forecasts, especially

Table 5
Survey of IMF trade forecast accuracy studies sample sizes, data sources, variable coverage.

Author/year	Scope of study	Years	Data source	Trade variables
Arora and Smyth (1990)	Developing countries aggregated into 1 forecast per continent	1981–1988	WEO	Import and export, current account % GDP
Artis (1988)	7 industrial countries. Developing countries aggregated into 1 forecast per continent	1971–1986	WEO	Import and export, current account % GDP
Artis (1996)	7 industrial countries. aggregated into 1 forecast per continent	1972–1994	WEO	Import and export, current account % GDP
Artis and Zhang (1990)	G7 countries	1980–1987	WEO	Current account % GDP
Atoyan and Conway (2011)	183 IMF loan programs	1993–2009	MONA	Current account % GDP
Atoyan et al. (2004)	175 IMF loan programs	1993–2001	MONA	Current account % GDP
Baqir et al. (2006)	94 IMF loan programs	1989–2002	MONA	Current account % GDP
Beach et al. (1999)	14 countries	1971–1998	WEO	Current account % GDP
EKPC (2019)	110 IMF loan programs	2002–2015	MONA	Import and export
Fratianni and Pattison (1991)	G7 countries	1980–1987	WEO	Current account % GDP
Frenkel et al. (2013)	G7 countries	1989–2010	WEO	Current account % GDP
IMF (2012)	148 IMF loan programs	2002–2011	MONA	Current account % GDP
Kenen and Schwartz (1986)	7 industrialized countries. Non-oil LDCs are aggregated into 1 forecast per continent. Imports (exports) feature 6 (8) years of forecasts	1971–1985	WEO	Import and export
Luna (2014)	103 IMF loan programs	2002–2011	MONA	Current account % GDP
Musso and Phillips (2002)	69 IMF loan programs	1993–1997	MONA	Current account % GDP
Timmermann (2007)	29 industrialized countries, 149 developing countries aggregated into 1 forecast per continent	1990–2003	WEO	Import and export, current account % GDP
US Government Accounting Office (2003)	87 emerging markets, of which 57 received loan programs, G7 countries	1990–2001	WEO	Current account % GDP
Verbeek (1999)	23 countries	1991–1997	World Bank (Unified Survey Projections)	Imports, exports, current account % GDP

for low-income countries, point a direction toward further research as to the origins of the forecast errors. One approach is to not publish exchange rate forecasts anymore (as evidenced in recent IMF loan documents); perhaps, the better approach would be an upfront acknowledgment of the uncertainty surrounding exchange rate forecasts along with the inclusion of alternative exchange rate forecast scenarios for the associated macroeconomic implications.

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.

Appendix A

See Tables A.1–A.3.

Table A.1
Variables/sources/definitions.

Variable	Data source	Description/database codenames
Current-year forecasts	IMF MONA database, (IMF, 2021a). For the construction of our dataset and MONA error corrections, see Appendix B	"t-1" to "t" period growth rates for **Real exchange rate (per $)** • MONA: ENDA - PCPIC, • If ENDA is not available, MONA: ENDE • If PCPIC is not available, MONA: PCPIE **Imports of goods and services in USD** Pre 2002 MONA data • MONA: BMT+ BMS_O • If BMS_O are not available, MONA: BMG Post-2002 MONA data: • MONA: BMGS • If BMGS is not available, MONA: BMG +BMS • IF BMGS, BMG, BMS are not available, MONA: NM/ENDA **Export of goods and services in USD** Pre 2002 MONA data: • MONA: BXT+ BXS_O • IF BXS_O is not available, MONA: BXG Post-2002 MONA dataset: • MONA: BXGS, • If BXGS is not available, MONA: BXG + BXS • If BXGS, BXG, BXS are not available, MONA: NX/ENDA
Actual final data	IMF IFS database, (IMF, 2021c), IMF BOP database, (IMF, 2021d), IMF WEO database, (IMF, 2021b)	"t-1" to "t" period growth rates for **Real exchange rate (per $)** • IFS: ENDA_XDC_USD_RATE – WEO: PCPIE • If ENDA_XDC_USD_RATE is not available IFS: ENDE_XDC_USD_RATE • If PCPIE is not available, WEO: PCPI **Imports of goods and services in USD** • IFS: BMGS_BP6_USD, • If BMGS_BP6_USD is not available, BOP: BM.GSR.GNFS.CD **Export of goods and services in USD** • IFS: BXGS_BP6_USD • If BXGS_BP6_USD is not available, BOP: BX.GSR.GNFS.CD
Elections data	Beck et al. (2001) pre1998; IFES (2020) post1998	Election dummy covers head of state, government, and legislative election. Program received a "1" if election occurred up to 1 year prior to program start date. Details at http://www.electionguide.org
Conflicts data (UCDP/PRIO)	Harbom et al. (2009)	Conflict dummy covers intra-state & inter-state conflicts. Program received a "1" if country experienced a conflict up to one year prior to program start date.
Disasters data	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.
Conditionality: Quantitative Performance Criteria	IMF MONA database, (IMF, 2021a).	Dummy variables defined by MONA's Glossary (IMF, 2021a) for quantitative performance targets: Domestic Credit Ceiling, Gov't/Public Sector Credit Ceilings, BOP Reserve Tests, Debt Ceilings (short, medium, and long term), Arrears Ceilings (domestic and external), Fiscal Deficit Ceilings.
Conditionality: Structural Performance Criteria	IMF MONA database, (IMF, 2021a).	Dummy variables defined based on the MONA Glossary (IMF, 2021a) for structural performance criteria: General Gov't Reform, Central Bank Reforms, Civil Service Wage/Employment Reforms, Pension Reforms, Gov't Enterprise Pricing Reforms, Financial Sector Reforms, Current Capital Account Openness/Reforms, Tariff/Quota Reductions/Reforms, Labor Market Wage/Employment Reforms, Statistics Reforms, Legal/Market Reforms
Trade Share	World Bank (2022)	Trade (% of GDP), World Bank national accounts data, and OECD National Accounts data files
Financial Openness	Chinn and Ito (2006)	Index measuring a country's degree of capital account openness, based on restrictions on cross-border financial transactions reported in the IMF's *Annual Report on Exchange Arrangements and Exchange Restrictions (AREAER).*

Table A.2
Regressions for Figure 2.

	1992–1996	1993–1997	1994–1998	1995–1999	1996–2000	1997–2001	1998–2002	1999–2003	2000–2004	2001–2005	2002–2006	2003–2007
IMPORTS												
MONA forecast, β	1.21	1.12	1.17	1.16	0.6***	0.51***	0.63***	0.69**	0.56**	0.61	0.76	0.14***
p-value ($\beta = 1$)	0.25	0.60	0.44	0.50	0.01	0.00	0.00	0.03	0.03	0.14	0.30	0.00
Constant, (α)	0.03	0.01	−0.01	−0.02	−0.01	−0.02	−0.02*	−0.02	0.03	0.06**	0.08**	0.17***
p-value ($\alpha = 0$)	0.29	0.75	0.52	0.32	0.28	0.14	0.07	0.24	0.25	0.03	0.01	0.00
Observations	144	173	185	183	177	165	148	132	110	92	83	72
Adj. R^2	0.48	0.40	0.46	0.46	0.13	0.12	0.21	0.24	0.11	0.15	0.15	−0.01
MZ p-value (α=0, $\beta = 1$)	0.03**	0.53	0.74	0.61	0.00***	0.00***	0.00***	0.00***	0.07*	0.09*	0.00***	0.00***
HP p-value (γ=0)	0.02**	0.29	0.91	0.64	0.00***	0.00***	0.00***	0.00***	0.33	0.20	0.00***	0.00***
EXPORTS												
MONA forecast, β	1.00	1.03	0.89	0.99	0.76	0.87	0.72**	0.84	0.67*	0.73	0.83	0.75*
p-value ($\beta = 1$)	0.99	0.93	0.73	0.97	0.22	0.53	0.05	0.26	0.08	0.19	0.14	0.08
Constant, (α)	0.07**	0.06**	0.05*	0.03	0.01	0.00	0.01	0.02	0.06**	0.08***	0.08***	0.12***
p-value ($\alpha = 0$)	0.02	0.04	0.08	0.35	0.48	0.91	0.57	0.28	0.02	0.01	0.00	0.00
Observations	143	172	184	182	177	165	147	131	109	91	82	72
Adj. R^2	0.20	0.19	0.14	0.12	0.09	0.11	0.08	0.10	0.11	0.15	0.42	0.44
MZ p-value (α=0, $\beta = 1$)	0.00***	0.00***	0.03**	0.31	0.46	0.81	0.14	0.35	0.06*	0.01***	0.00***	0.00***
HP p-value (γ=0)	0.00***	0.00***	0.00***	0.16	0.67	0.65	0.55	0.67	0.07*	0.00***	0.00***	0.00***
REAL EXCHANGE RATE DEPRECIATION												
MONA forecast, β	0.84**	0.88*	0.88	0.97	1.05	1.17	1.05	1.03	1.39	0.88	0.79	0.62***
p-value ($\beta = 1$)	0.05	0.10	0.35	0.85	0.80	0.29	0.81	0.86	0.54	0.47	0.35	0.01
Constant, (α)	−0.03	−0.03	−0.02	0.00	0.02	0.01	0.02	0.01	0.00	−0.01	−0.02	−0.06***
p-value ($\alpha = 0$)	0.12	0.15	0.24	0.91	0.48	0.47	0.27	0.67	0.83	0.56	0.11	0.00
Observations	145	175	186	185	177	166	146	129	107	90	79	71
Adj. R^2	0.83	0.83	0.56	0.49	0.44	0.57	0.31	0.39	0.22	0.27	0.18	0.38
MZ p-value (α=0, $\beta = 1$)	0.02**	0.05**	0.23	0.98	0.71	0.34	0.32	0.60	0.79	0.67	0.27	0.00***
HP p-value (γ=0)	0.05*	0.06*	0.11	0.89	0.43	0.26	0.18	0.59	0.74	0.79	0.25	0.00***

	2004–2008	2005–2009	2006–2010	2007–2011	2008–2012	2009–2013	2010–2014	2011–2015	2012–2016	2013–2017	2014–2018	2015–2019
IMPORTS												
MONA forecast, β	0.14***	0.77*	0.77*	0.80*	0.77**	0.76*	0.60**	0.61***	0.57***	0.78**	0.88	0.91
p-value ($\beta = 1$)	0.00	0.07	0.07	0.10	0.05	0.07	0.02	0.01	0.00	0.05	0.39	0.52
Constant, (α)	0.19***	0.03	0.03*	0.02	0.01	0.00	0.04**	−0.01	−0.03**	−0.03**	−0.03**	−0.04**
p-value ($\alpha = 0$)	0.00	0.12	0.08	0.29	0.58	0.98	0.05	0.54	0.02	0.04	0.03	0.01
Observations	71	75	89	85	92	85	78	59	66	62	62	62
Adj. R^2	0.00	0.41	0.38	0.43	0.43	0.43	0.24	0.35	0.31	0.38	0.37	0.38
MZ p-value (α=0, $\beta = 1$)	0.00***	0.13	0.08*	0.19	0.14	0.19	0.05**	0.01***	0.00***	0.00***	0.01***	0.01***
HP p-value (γ=0)	0.01***	0.62	0.58	0.89	0.45	0.40	0.82	0.04**	0.00***	0.01***	0.01***	0.00***
EXPORTS												
MONA forecast, β	0.73*	0.92	0.91	0.93	0.93	0.90	0.83	0.90	0.73*	0.90	0.90	0.95
p-value ($\beta = 1$)	0.06	0.17	0.16	0.46	0.50	0.36	0.25	0.49	0.06	0.65	0.72	0.88
Constant, (α)	0.12***	0.05***	0.07***	0.06***	0.04**	0.04**	0.05*	0.01	0.00	0.01	0.01	−0.01
p-value ($\alpha = 0$)	0.00	0.00	0.00	0.00	0.02	0.03	0.06	0.74	0.85	0.77	0.76	0.69
Observations	71	75	89	85	92	85	78	59	66	62	62	62
Adj. R^2	0.42	0.70	0.54	0.47	0.44	0.42	0.18	0.26	0.15	0.20	0.17	0.33
MZ p-value (α=0, $\beta = 1$)	0.00***	0.01***	0.00***	0.00***	0.05*	0.08*	0.17	0.78	0.16	0.87	0.91	0.88
HP p-value (γ=0)	0.00***	0.00***	0.00***	0.00***	0.03**	0.09*	0.17	0.99	0.49	0.97	0.96	0.62
REAL EXCHANGE RATE DEPRECIATION												
MONA forecast, β	0.53***	0.84	0.81**	0.83**	0.83**	0.75***	0.71***	0.76***	0.68***	0.72***	0.34***	0.33***
p-value ($\beta = 1$)	0.00	0.10	0.03	0.05	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Constant, (α)	−0.08***	−0.03***	−0.02**	−0.01	0.00	0.01**	0.01	0.02***	0.04***	0.03***	0.04***	0.03**
p-value ($\alpha = 0$)	0.00	0.00	0.02	0.12	0.73	0.04	0.32	0.01	0.00	0.00	0.01	0.01
Observations	71	76	90	86	92	84	75	55	61	55	54	53
Adj. R^2	0.30	0.58	0.56	0.62	0.64	0.71	0.63	0.70	0.54	0.46	0.30	0.30
MZ p-value (α=0, $\beta = 1$)	0.00***	0.01***	0.02**	0.09*	0.05*	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***
HP p-value (γ=0)	0.00***	0.01***	0.06*	0.36	0.85	0.02**	0.02**	0.00***	0.00***	0.01***	0.75	0.88

Robust p-values. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.3
Regressions for Figure 3.

	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec
						IMPORTS						
MONA Forecast, β	1.39***	0.27***	0.76	1.19	0.74	1.05	1.03	0.32***	0.15***	1.28	0.73	0.5**
p-value ($\beta = 1$)	0.01	0.00	0.22	0.12	0.49	0.77	0.80	0.01	0.00	0.21	0.18	0.01
Constant, (α)	−0.04*	0.07*	−0.02	−0.01	0.05	0.02	−0.02	−0.01	0.09**	−0.04	0.02	0.03
p-value ($\alpha = 0$)	0.08	0.05	0.32	0.56	0.11	0.25	0.13	0.91	0.05	0.41	0.46	0.43
Observations	69	27	60	55	48	67	61	32	37	22	32	68
Adj. R^2	0.70	0.04	0.27	0.52	0.17	0.37	0.51	−0.01	−0.02	0.55	0.34	0.12
MZ *p*-value (α=0, $\beta = 1$)	0.03**	0.00***	0.27	0.27	0.28	0.28	0.31	0.01***	0.00***	0.43	0.40	0.00***
HP *p*-value (γ=0)	0.97	0.90	0.15	0.96	0.39	0.12	0.21	0.09*	0.53	0.93	0.63	0.54
						EXPORTS						
MONA Forecast, β	1.67	0.84	0.85	0.90	0.47**	0.76	1.10	0.72	0.55	1.01	0.60**	0.97
p-value ($\beta = 1$)	0.35	0.49	0.36	0.57	0.05	0.43	0.77	0.23	0.45	0.96	0.05	0.69
Constant, (α)	−0.04	0.05	0.05**	0.01	0.07**	0.12**	0.04	0.01	0.09	0.01	0.03*	0.01
p-value ($\alpha = 0$)	0.47	0.17	0.04	0.72	0.04	0.02	0.11	0.79	0.14	0.87	0.07	0.46
Observations	69	27	60	55	48	65	61	32	37	22	32	68
Adj. R^2	0.31	0.21	0.29	0.36	0.07	0.02	0.31	0.25	0.02	0.41	0.27	0.57
MZ *p*-value (α=0, $\beta = 1$)	0.62	0.39	0.1*	0.85	0.07*	0.05**	0.14	0.43	0.28	0.96	0.08*	0.73
HP *p*-value (γ=0)	0.46	0.27	0.07*	0.99	0.47	0.02**	0.09*	0.43	0.11	0.79	0.70	0.59
						REAL EXCHANGE RATE DEPRECIATION						
MONA Forecast, β	1.00	0.43***	0.65***	1.27**	0.66***	0.93	0.94	1.00	0.58***	0.68***	0.85	0.85*
p-value ($\beta = 1$)	0.99	0.01	0.01	0.05	0.00	0.17	0.64	0.99	0.00	0.00	0.12	0.06
Constant, (α)	0.04	0.01	0.04	−0.04	−0.06***	−0.02	−0.02	0.02	−0.03*	0.03	−0.01	0.00
p-value ($\alpha = 0$)	0.21	0.62	0.27	0.34	0.01	0.28	0.39	0.39	0.07	0.52	0.44	0.81
Observations	68	27	56	53	48	63	59	29	35	23	33	71
Adj. R^2	0.27	0.17	0.49	0.73	0.55	0.93	0.58	0.49	0.49	0.86	0.61	0.64
MZ *p*-value (α=0, $\beta = 1$)	0.19	0.00***	0.00***	0.11	0.00***	0.29	0.08*	0.66	0.01***	0.00***	0.25	0.17
HP *p*-value (γ=0)	0.17	0.12	0.95	0.41	0.01***	0.20	0.24	0.41	0.13	0.22	0.63	0.95

Robust p-values. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix B. Auditing the MONA database

The MONA database presents challenges as it contains a wide range of errors. Unlike the WEO database, MONA does not include release dates, hence it is unclear if/when revisions/updates to the database take place. To prevent data errors from deriving our results, we audited MONA and corrected the following 11 kinds of errors that fall into three major categories:

Data Entry Errors
- B.1. Data Entered with Wrong Signs
- B.2. Temporal Errors: Correct Data Entered for the Wrong Program Year
- B.3. Zeros Identify Missing Values
- B.4. Typos and Spelling Mistakes
- B.5. Wrong Line Items Entered

Inconsistencies
- B.6. Currency Unit Magnitude Inconsistencies
- B.7. Indicator Variable Inconsistencies
- B.8. Rates vs Level Inconsistencies
- B.9. Base Year Inconsistencies

Corrected Data from IMF Archives (Executive Board Documents)
- B.10. Missing Data Corrected
- B.11. Outliers Verified and Corrected

B.1. Data Entered with Incorrect Signs

We corrected 5070 observations (2536 for *t-1* and 2534 for *t*) where had been entered with incorrect signs. Most errors affect trade data (Mona defines imports and exports to be entered as positive values).

B.2: Temporal Errors

MONA reports data from *t-3* to *t+4*, where "*t*" is the program year. For example, if the program year is 1997, then MONA reports data from 1994 to 2001. Sometimes, data entry confused the program year and generated temporal errors associating the correct data with the wrong year. Seven programs suffered this error (see Table B.1).

B.3: Zeros Identify Missing Values

MONA does not possess a consistent indicator for missing values. At times missing values are presented as "NA", ""., "0", or "NULL". Exact zero levels of imports, exports, or exchange rates are suspicious, so if we could not find the data in the archives, we had to assign 106 observations missing data status instead of accepting the value of exact zero for several program years (see Table B.2).

B.4: Typos and Spelling Mistakes

We adjusted 31 observations when (i) a series is misspelled, (ii) one decimal is incorrect, (iii) one additional integer is added in the wrong place, (iv) one integer is missing, (v) the wrong country is being identified as the program country, (vi) the wrong year is identified as the program year, or (vii) when the variable contained typos. In total there are 31 of these typos and spelling errors that were corrected based on the original IMF *Executive Board Special* (EBS) loan documents (see Table B.3).

B.5: Wrong Line Item Entered

At times, data entry inadvertently fell into the wrong line and entered the wrong line item. For example, instead

Table B.1
Temporal errors.

Program	Country name	Prog. year	Mnemonic	Review type	Correction
7	Estonia	1993	All	Last	Using EBS (moved data one year forward)
15	El Salvador	1993	All	Last	Using EBS (moved data one year forward)
256	Indonesia	1998	PCPIC	All	Using EBS (moved data one year forward)
275	Indonesia	1999	ENDA, PCPIC	All	Using EBS (moved data one year forward)
552	Dominican Republic	2005	All	All	Using EBS (moved data one year forward)
571	Madagascar	2006	All	Last	Using EBS (moved data one year forward)
579	Gabon	2007	All	R1-Last	Using EBS (moved data one year forward)

Table B.2
Zeros identify missing values.

Count	Program	Country name	Prog. year	Mnemonic	Review type
1	11	Honduras	1992	BMS_O	R0
2	16	Kyrgyz Republic	1993	BXS_O	R0
3	16	Kyrgyz Republic	1993	BMS_O	R0
4	25	Poland	1993	ENDA	R0
5	108	Kazakhstan	1994	BXS_O	R0
6	108	Kazakhstan	1994	ENDA	R0
7	114	Mozambique	1994	ENDA	R0
8	153	Kazakhstan	1995	ENDA	R0
9	153	Kazakhstan	1995	BXS_O	R0
10	159	Azerbaijan	1996	BMS_O	R0
11	163	Moldova	1996	BMS_O	R0
12	250	Mauritania	1997	PCPIC	R0
13	264	Central African Republic	1998	PCPIC	R0
14	281	Argentina	1998	PCPIC	R0
15	337	Indonesia	2000	PCPIC	R0
16	507	Albania	2002	PCPIE	R6
17	560	Benin	2005	PCPIE	R0
18	564	Iraq	2005	BXS	R0
19	572	Haiti	2006	ENDA	R0
20	590	Liberia	2008	PCPIE	R7
21	610	Sao Tome and Principe	2009	ENDA	R0
22	632	Malawi	2010	ENDA	R0
23	642	Tanzania	2010	ENDA	R0
24	649	Armenia	2010	ENDA	R0
25	651	Haiti	2010	ENDA	R0
26	656	Senegal	2010	PCPIE	R6-R7
27	658	North Macedonia	2011	ENDA	R0
28	678	Burundi	2012	ENDA	R0
29	679	Guinea	2012	ENDA	R0
30	681	Niger	2012	PCPIE	R2-R7
31	683	Georgia	2012	ENDA	R0
32	685	The Gambia	2012	ENDA	R0
33	685	The Gambia	2012	PCPIE	R0
34	686	Central African Republic	2012	ENDA	R0
35	687	Tanzania	2012	ENDA	R0
36	688	Sao Tome and Principe	2012	ENDA	R0
37	689	Malawi	2012	ENDA	R0
38	691	Morocco	2012	ENDA	R0
39	692	Bosnia and Herzegovina	2012	ENDA	R0
40	693	Liberia	2012	NX	R0
41	697	Jamaica	2013	ENDA	R0
42	698	Cyprus	2013	ENDA	R0
43	699	Tunisia	2013	ENDA	R0
44	701	Mozambique	2013	ENDA	R0

(*continued on next page*)

of entering the inflation data, data entry entered GDP data from one line below inflation in the report (see Table B.4).

B.6: Inconsistent Currency Units Entered

We corrected eight instances when data magnitudes within a program were internally inconsistent. One entry might be in millions the other in thousands (see Table B.5).

B.7: Rates vs. Levels Inconsistencies

MONA occasionally reports the price index instead of the inflation rate (specified by the IMF MONA descriptor document). We corrected 22 inconsistencies within a program (see Table B.6).

B.8: Unit Inconsistency: Base Years

MONA occasionally reports data with different base years within a program. One entry might have *t-4* as

Table B.2 (*continued*).

Count	Program	Country name	Prog. year	Mnemonic	Review type
45	702	Uganda	2013	ENDA	R0
46	703	Pakistan	2013	ENDA	R0
47	705	Sierra Leone	2013	ENDA	R0
48	709	Albania	2014	PCPIE	R2-R3
49	710	Armenia	2014	ENDA	R0
50	712	Seychelles	2014	ENDA	R0
51	714	Tanzania	2014	ENDA	R0
52	714	Tanzania	2014	PCPIE	R0
53	717	Chad	2014	ENDA	R0
54	722	Kenya	2015	ENDA	R0
55	723	Serbia, Republic of	2015	ENDA	R0
56	724	Ukraine	2015	NX	R0
57	724	Ukraine	2015	NM	R0
58	724	Ukraine	2015	ENDA	R0
59	726	Kyrgyz Republic	2015	ENDA	R0
60	729	Senegal	2015	ENDA	R0
61	729	Senegal	2015	PCPIE	R3
62	730	Guinea-Bissau	2015	ENDA	R0
63	731	Sao Tome and Principe	2015	ENDA	R0
64	731	Sao Tome and Principe	2015	PCPIE	R0
65	735	Tunisia	2016	ENDA	R0
66	738	Sri Lanka	2016	ENDA	R0
67	739	Rwanda	2016	ENDA	R0
68	739	Rwanda	2016	PCPIE	R0
69	741	Iraq	2016	PCPIE	R0
70	742	Madagascar	2016	ENDA	R0
71	747	Bosnia and Herzegovina	2016	ENDA	R0
72	749	Moldova	2016	ENDA	R0
73	750	Cote d'Ivoire	2016	BXGS	R0
74	750	Cote d'Ivoire	2016	ENDA	R0
75	750	Cote d'Ivoire	2016	BMGS	R0
76	751	Niger	2017	ENDA	R0
77	753	Poland	2017	ENDA	R0
78	754	Benin	2017	ENDA	R0
79	755	Georgia	2017	ENDA	R0
80	756	Togo	2017	ENDA	R0
81	757	Mongolia	2017	ENDA	R0
82	757	Mongolia	2017	NX	R0
83	757	Mongolia	2017	NM	R0
84	758	Sierra Leone	2017	ENDA	R0
85	760	Cameroon	2017	ENDA	R0
86	761	Chad	2017	ENDA	R0
87	764	Mauritania	2018	ENDA	R0
88	765	Guinea	2017	ENDA	R0
89	766	Seychelles	2017	ENDA	R0
90	767	Burkina Faso	2018	ENDA	R0
91	768	Malawi	2018	ENDA	R0
92	769	Colombia	2018	ENDA	R0
93	771	Serbia, Republic of	2018	ENDA	R0
94	772	Barbados	2018	ENDA	R0
95	778	Armenia	2019	ENDA	R0
96	780	Rwanda	2019	ENDA	R0
97	781	Pakistan	2019	ENDA	R0
98	782	Honduras	2019	ENDA	R0
99	783	Cabo Verde	2019	ENDA	R0
100	784	Congo, Rep.	2019	ENDA	R0
101	785	Mali	2019	ENDA	R0
102	786	Sao Tome and Principe	2019	ENDA	R0
103	787	Mexico	2019	ENDA	R0
104	789	Liberia	2019	ENDA	R0
105	789	Liberia	2019	NX	R0
106	790	Central African Republic	2019	ENDA	R0

the base year, another entry for the same variable might have *t-2* as the base year. We corrected 23 such errors by converting the data from levels to growth rates (see Table B.7).

Table B.3
Typos and spelling mistakes.

Count	Prog.	Country name	Prog. year	Mnemonic	Review yype	Correction
1	15	El Salvador	1993	programyear	R2	programyear corrected using EBS
2	18	Latvia	1993	programyear	R1-Last	programyear corrected using EBS
3	75	Turkey	1994	PCPIC	R0	data corrected using EBS
4	117	Albania	1994	countryname	All	countryname corrected using EBS
5	117	Albania	1994	countryncode	All	countryname corrected using EBS
6	132	Sierra Leone	1995	programyear	R0	programyear corrected using EBS
7	136	Haiti	1995	PCPIC	R0	data corrected using EBS
8	143	Pakistan	1996	programyear	R0-R1	programyear corrected using EBS
9	160	Russian Federation	1995	PCPIC	R0	data corrected using EBS
10	205	Vietnam	1996	boarddocno	R1	board document corrected using EBS
11	207	Ethiopia	1997	reviewtype	All	reviewtype corrected using EBS
12	212	Kyrgyz Rep.	1997	ENDA	Last	data corrected using EBS
13	230	Burkina Faso	1996	programyear	R0	data corrected using EBS
14	250	Mauritania	1997	PCPIC	R0	data corrected using EBS
15	274	Ukraine	1998	PCPIC	R0	data corrected using EBS
16	274	Ukraine	1998	programyear	R5-R6	programyear corrected using EBS
17	402	Moldova	2000	BXS_O	R0	data corrected using EBS
18	510	Argentina	2003	PCPIE	R0	data corrected using EBS
19	521	Ghana	2003	PCPIE	R0	data corrected using EBS
20	527	Nicaragua	2002	programyear	R10	programyear corrected using EBS
21	535	Uruguay	2002	PCPIE	R0	data corrected using EBS
22	539	Dominican Republic	2003	PCPIE	R0	data corrected using EBS
23	545	Peru	2004	PCPIE	R0	data corrected using EBS
24	560	Benin	2005	PCPIE	R0	data corrected using EBS
25	560	Benin	2005	boarddocno	R0	board document corrected using EBS
26	628	Kyrgyz Rep.	2008	reviewtype	All	reviewtype corrected using EBS
27	681	Niger	2012	programyear	R8	programyear corrected using EBS
28	724	Ukraine	2015	reviewtype	All	reviewtype corrected using EBS
29	734	Kenya	2016	reviewtype	All	reviewtype corrected using EBS
30	764	Mauritania	2017	programyear	R0-R4	programyear corrected using EBS
31	All	All	All	initialenddate	All	spelling error corrected using EBS

Table B.4
Wrong line item entered.

Count	Program	Country name	Prog. year	Mnemonic	Review type	Correction
1	560	Benin	2005	PCPIE	R0	data corrected using EBS

Table B.5
Inconsistent currency magnitudes.

Count	Program	Country name	Prog. year	Mnemonic	Review type	Correction
1	70	Poland	1994	ENDA	R0	divided by 1000
2	75	Turkey	1994	ENDA	Last	divided by 1000
3	84	Algeria	1995	ENDA	R0	divided by 1000
4	164	Russian Federation	1996	ENDA	R0	divided by 1000
5	199	Croatia	1997	ENDA	Last	divided by 1000
6	398	Bulgaria	2002	ENDA	R0	divided by 1000
7	517	Croatia	2003	PCPIE	R0-R1	changed to index
8	580	Mozambique	2007	PCPIE	R0	changed to index

Table B.6
Rate vs Level inconsistencies.

Count	Program	Country name	Prog. year	Mnemonic	Correction
1	510	Argentina	2003	PCPIE	corrected to rates
2	521	Ghana	2003	PCPIE	corrected to rates
3	527	Nicaragua	2002	PCPIE	corrected to rates
4	535	Uruguay	2002	PCPIE	corrected to rates
5	539	Dominican Republic	2003	PCPIE	corrected to rates
6	545	Peru	2004	PCPIE	corrected to rates
7	556	Turkey	2005	PCPIE	corrected to rates
8	560	Benin	2005	PCPIE	corrected to rates
9	562	North Macedonia, Rep	2005	PCPIE	corrected to rates
10	564	Iraq	2005	PCPIE	corrected to rates

(*continued on next page*)

Table B.6 (*continued*).

Count	Program	Country name	Prog. year	Mnemonic	Correction
11	566	Grenada	2006	PCPIE	corrected to rates
12	572	Haiti	2006	PCPIE	corrected to rates
13	580	Mozambique	2007	PCPIE	corrected to rates
14	588	Iraq	2007	PCPIE	corrected to rates
15	591	Honduras	2008	PCPIE	corrected to rates
16	685	The Gambia	2012	PCPIE	corrected to rates
17	709	Albania	2014	PCPIE	corrected to rates
18	714	Tanzania	2014	PCPIE	corrected to rates
19	718	Yemen	2014	PCPIE	corrected to rates
20	731	Sao Tome and Principe	2015	PCPIE	corrected to rates
21	739	Rwanda	2016	PCPIE	corrected to rates
22	741	Iraq	2016	PCPIE	corrected to rates

Table B.7
Base year errors.

Count	Program	Country name	Prog. year	Mnemonic	Review type	Correction
1	16	Kyrgyz Republic	1993	All	All	unresolved, dropped
2	108	Kazakhstan	1994	All	All	unresolved, dropped
3	532	Sierra Leone	2001	PCPIE	Last	corrected to rates
4	533	Tanzania	2000	PCPIE	Last	corrected to rates
5	538	Burundi	2004	PCPIE	Last	corrected to rates
6	547	Zambia	2004	PCPIE	Last	corrected to rates
7	549	Bulgaria	2004	PCPIE	Last	corrected to rates
8	554	Kyrgyz Republic	2005	PCPIE	Last	corrected to rates
9	561	Sao Tome and Principe	2005	PCPIE	Last	corrected to rates
10	565	Albania	2006	PCPIE	Last	corrected to rates
11	567	Moldova	2006	PCPIE	Last	corrected to rates
12	568	Paraguay	2006	PCPIE	Last	corrected to rates
13	596	Burundi	2008	PCPIE	Last	corrected to rates
14	617	Romania	2009	PCPIE	Last	corrected to rates
15	619	Ghana	2009	PCPIE	Last	corrected to rates
16	620	Sri Lanka	2009	PCPIE	Last	corrected to rates
17	623	Angola	2010	PCPIE	Last	corrected to rates
18	625	Congo, Democratic Rep.	2010	PCPIE	Last	corrected to rates
19	635	El Salvador	2010	PCPIE	Last	corrected to rates
20	678	Burundi	2012	PCPIE	Last	corrected to rates
21	697	Jamaica	2013	PCPIE	Last	corrected to rates
22	704	Romania	2013	PCPIE	Last	corrected to rates
23	712	Seychelles	2014	PCPIE	Last	corrected to rates

Table B.8
Missing MONA data filled using IMF archives.

Count	Program	Country name	Prog. year	Mnemonic	Correction
1	95	Ukraine	1995	ENDA	data from EBS
2	153	Kazakhstan	1995	ENDA	data from EBS
3	250	Mauritania	1997	RENDA	data from EBS
4	261	Mozambique	1997	ENDA	data from EBS
5	273	Ukraine	1997	ENDA	data from EBS
6	377	Pakistan	2001	ENDA	data from EBS
7	419	Congo, Dem. Rep.	2002	ENDA	data from EBS
8	421	Rwanda	2001	ENDA	data from EBS
9	502	Tajikistan	2002	NM	data from EBS
10	502	Tajikistan	2002	NX	data from EBS
11	506	Bosnia and Herzegovina	2002	NM	data from EBS
12	506	Bosnia and Herzegovina	2002	NX	data from EBS
13	537	Serbia	2002	NX	data from EBS
14	537	Serbia	2002	NM	data from EBS
15	572	Haiti	2006	ENDA	data from EBS

(*continued on next page*)

Table B.8 (*continued*).

Count	Program	Country name	Prog. year	Mnemonic	Correction
16	610	Sao Tome and Principe	2009	ENDA	data from EBS
17	624	Maldives	2009	ENDA	data from EBS
18	648	Rwanda	2010	NM	data from EBS
19	651	Haiti	2010	ENDA	data from EBS
20	678	Burundi	2012	ENDA	data from EBS
21	685	The Gambia	2012	RENDA	data from EBS
22	691	Morocco	2012	ENDA	data from EBS
23	698	Cyprus	2013	ENDA	data from EBS
24	699	Tunisia	2013	ENDA	data from EBS
25	702	Uganda	2013	ENDA	data from EBS
26	703	Pakistan	2013	ENDA	data from EBS
27	705	Sierra Leone	2013	ENDA	data from EBS
28	711	Ukraine	2014	NX	data from EBS
29	711	Ukraine	2014	NM	data from EBS
30	712	Seychelles	2014	ENDA	data from EBS
31	713	Grenada	2014	NX	data from EBS
32	714	Tanzania	2014	ENDA	data from EBS
33	726	Kyrgyz Republic	2015	ENDA	data from EBS
34	731	Sao Tome and Principe	2015	ENDA	data from EBS
35	735	Tunisia	2016	ENDA	data from EBS
36	739	Rwanda	2016	ENDA	data from EBS
37	742	Madagascar	2016	ENDA	data from EBS
38	749	Moldova	2016	ENDA	data from EBS
39	750	Cote d'Ivoire	2016	RENDA	data from EBS
40	756	Togo	2017	RENDA	data from EBS
41	761	Chad	2017	RENDA	data from EBS
42	764	Mauritania	2017	ENDA	data from EBS
43	765	Guinea	2017	ENDA	data from EBS
44	766	Seychelles	2017	ENDA	data from EBS
45	768	Malawi	2018	ENDA	data from EBS
46	769	Colombia	2018	ENDA	data from EBS
47	770	Argentina	2018	ENDA	data from EBS
48	772	Barbados	2018	ENDA	data from EBS
49	778	Armenia	2019	ENDA	data from EBS
50	781	Pakistan	2019	ENDA	data from EBS
51	782	Honduras	2019	ENDA	data from EBS
52	786	Sao Tome and Principe	2019	ENDA	data from EBS
53	787	Mexico	2019	ENDA	data from EBS
54	789	Liberia	2019	RENDA	data from EBS

Table B.9
Outliers corrected/verified.

Count	Program	Country	Prog. year	Mnemonic	Correction
1	1	Albania	1993	NM	data from EBS
2	14	Jamaica	1992	RENDA	data from EBS
3	14	Jamaica	1992	ENDA	data from EBS
4	16	Kyrgyz Republic	1993	PCPIC	verified in EBS
5	17	Lao PDR	1993	NX	data from EBS
6	19	Lithuania	1993	RENDA	data from EBS
7	19	Lithuania	1993	ENDA	data from EBS
8	29	Kenya	1993	NX	data from EBS
9	29	Kenya	1993	NM	data from EBS
10	75	Turkey	1994	PCPIC	typos and spelling mistakes. Fixed with EBS data
11	82	Moldova	1994	RENDA	data from EBS
12	82	Moldova	1994	ENDA	data from EBS
13	93	Cambodia	1994	NX	data from EBS
14	108	Kazakhstan	1994	PCPIC	verified in EBS, but dropped due to undocumented base year issues
15	114	Mozambique	1994	RENDA	data from EBS
16	114	Mozambique	1994	ENDA	data from EBS
17	118	Republic Of Congo	1994	PCPIC	WEO data is likely a typo. MONA last review data used as actual
18	129	Papua New Guinea	1995	RENDA	data from EBS
19	129	Papua New Guinea	1995	ENDA	data from EBS
20	132	Sierra Leone	1995	NX	data from EBS

(*continued on next page*)

Table B.9 (*continued*).

Count	Program	Country	Prog. year	Mnemonic	Correction
21	134	Belarus	1995	RENDA	data from EBS
22	134	Belarus	1995	ENDA	data from EBS
23	134	Belarus	1995	NX	data from EBS
24	134	Belarus	1995	NM	data from EBS
25	136	Haiti	1995	PCPIC	typos and spelling mistakes. Fixed with EBS data
26	160	Russian Federation	1995	PCPIC	typos and spelling mistakes. Fixed with EBS data
27	170	Cambodia	1995	NX	data from EBS
28	202	Bulgaria	1997	RENDA	data from EBS
29	202	Bulgaria	1997	ENDA	data from EBS
30	218	Haiti	1997	NX	data from EBS
31	228	Sierra Leone	1997	NM	data from EBS
32	228	Sierra Leone	1997	NX	data from EBS
33	242	Bosnia and Herzegovina	1998	NM	data from EBS
34	256	Indonesia	1998	PCPIC	temporal issue. Fixed with EBS data
35	275	Indonesia	1999	PCPIC	temporal issue. Fixed with EBS data
36	307	Mozambique	1999	NM	data from EBS
37	316	Albania	1999	NM	data from EBS
38	316	Albania	1999	NX	data from EBS
39	347	Kenya	2000	NM	data from EBS
40	508	Argentina	2003	PCPIE	no data available in EBS
41	510	Argentina	2003	PCPIE	typos and spelling mistakes. Fixed with EBS data
42	521	Ghana	2003	PCPIE	typos and spelling mistakes, correct data entered from loan document
43	532	Sierra Leone	2001	NM	data from EBS
44	535	Uruguay	2002	RENDA	data from EBS
45	535	Uruguay	2002	ENDA	data from EBS
46	535	Uruguay	2002	PCPIE	typos and spelling mistakes. Fixed with EBS data
47	538	Burundi	2004	NX	data from EBS
48	539	Dominican Republic	2003	PCPIE	typos and spelling mistakes, correct data entered from loan document
49	547	Zambia	2004	NX	data from EBS
50	552	Dominican Republic	2005	PCPIE	temporal issue. Fixed with EBS data
51	574	Mauritania	2006	NM	data from EBS
52	579	Gabon	2007	NM	data from EBS
53	603	Iceland	2008	NX	data from EBS
54	625	Congo, Dem. Rep.	2010	NX	data from EBS
55	645	Burkina Faso	2010	NX	data from EBS
56	733	Mozambique	2016	ENDA	data from EBS
57	765	Guinea	2017	NX	data from EBS
58	791	Ethiopia	2019	NM	data from EBS

B.9: Missing Data

Missing data encountered in the MONA database was filled in using the IMF archives' *Executive Board Special* (EBS) loan documents when available. We filled in 54 observations listed in Table B.8.

B.10: Outliers Verification/Correction

We audited observations that fell three or more standard deviations from the mean. Since the distribution changes with each outlier correction, we conducted two rounds of outlier checks. Outliers were checked using the original IMF *Executive Board Special* (EBS) loan documents. We corrected 58 observations listed in Table B.9.

References

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

Armstrong, J. S. (Ed.), (2001). *The principles of forecasting*. Norwell, Mass: Kluwer Academic Publisher.

Arora, H. K., & Smyth, D. J. (1990). Forecasting the developing world. An accuracy analysis of the IMF's forecasts. *International Journal of Forecasting*, *6*(3), 393–400.

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

Artis, M. J. (1996). *How accurate are the WEO's short-term forecasts? An examination of the world economic outlook*: *IMF working paper No. 96/89*, Washington, DC: International Monetary Fund.

Artis, M. J., & Zhang, W. (1990). BVAR forecasts for the G-7. *International Journal of Forecasting*, *6*, 349–362.

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

Atoyan, R., Conway, P., Selowsky, M., & Tsikata, T. (2004). *Macroeconomic adjustments in IMF-supported programs: projections and reality*: *IEO background paper BP/04/2*, Washington D.C.: International Monetary Fund.

Baqir, R., Ramcharan, R., & Sahay, R. (2006). Imf programs and growth: Is optimism defensible?. In M. A. Mody, & Rebucci M.A. (Eds.), *IMF-supported programs: recent staff research*. Washington DC: International Monetary Fund.

Beach, W., Schavey, A., & Isidro, I. (1999). *How reliable are IMF economic forecasts?*: *Heritage foundation center for data analysis report 99-05*, Washington D.C: Heritage Foundation.

Beaudry, P., & Willems, T. (2022). On macroeconomic consequences of over-optimism. *American Economic Journal: Macroeconomics*, *14*(1), 38–59.
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.
Boz, E., Casas, C., Georgiadis, G., Gopinath, G., Le Mezo, H., Mehl, A., & Nguyen, T. (2020). Patterns in invoicing currency in global trade.
Carriere-Swallow, Yan, & Marzluf, José (2021). Macrofinancial Causes of Optimism in Growth Forecasts. IMF Working Paper No. 2021/275.
Chatterjee, P. (2019). Asymmetric impact of uncertainty in recessions: are emerging countries more vulnerable? *Studies in Nonlinear Dynamics & Econometrics*, *23*, 21–27.
Chatterjee, P., & Nowak, S. (2016). Forecast Errors and Uncertainty Shocks IMF Working Paper 16/228.
Chinn, M. D., & Ito, H. (2006). What matters for financial development? Capital controls, institutions, and interactions. *Journal of Development Economics*, *81*(1), 163–192.
Diebold, F. X., & Lopez, J. A. (1996). Forecast evaluation and combination. In G. S. Maddala, & C. R. Rao (Eds.), *Handbook of statistics*. Amsterdam: North-Holland.
Dreher, A., S, Marchesi., & Vreeland, J. (2008). The political economy of IMF forecasts. *Public Choice*, *137*(1–2), 145–171.
Eicher, T. S., & Gao Rollinson, Y. (2022). The accuracy of IMF crises nowcasts. *International Journal of Forecasting*.
Eicher, T. S., Kuenzel, D. J., Papageorgiou, C., & Christofides, C. (2019). Forecasts in times of crises. *International Journal of Forecasting*, *35*, 1143–1159.
EM-DAT (2020). The OFDA/CRED international disaster database. URL http://www.emdat.be/. (Accessed 30 October 20).
Fratianni, M., & Pattison, J. (1991). International institutions and the market for information. In Vaubel R., & Willett T.D. (Eds.), *The political economy of international organizations: A public choice approach*. Boulder: Westview Press.
Frenkel, M., Ruelke, J. C., & Zimmermann, L. (2013). Do private sector forecasters chase after IMF or OECD forecasts? *Journal of Macroeconomics*, *37*(April), 217–229.
Genberg, H., & Martinez, A. (2014). *On the accuracy and efficiency of imf forecasts: a survey and some extensions*: *Independent evaluation Office BP/14/04*, Washington DC: International Monetary Fund.
Ghosh, A., Lane, T., Thomas, A., & Zalduendo, J. (2006). Program objectives and program success. In A. Mody, & A. Rebucci (Eds.), *IMF-supported programs*. Washington, D.C: International Monetary Fund.
Harbom, L., Strand, H., & Nygård, H. M. (2009). UCDP/PRIO armed conflict dataset codebook. Version 20.1. URL http://www.ucdp.uu.se/database. (Accessed 30 October 20).
Henriksson, R. D., & Merton, R. C. (1981). On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills. *Journal of Business*, *54*(4), 513–533.
Holden, K., & Peel, D. A. (1990). On testing for unbiasedness and efficiency of forecasts. *The Manchester School*, *58*(2), 120–127.
Holden, K., Peel, D., & Sandhu, B. (1987). The accuracy of OECD forecasts. *Empirical Economics*, *12*, 175–186.
IFES (2020). Electionguide by the international foundation for electoral systems (IFES). URL https://www.electionguide.org/elections. (Accessed 15 October 20).
Ilzetzki, E., Reinhart, C. M., & Rogoff, K. S. (2019). Exchange rate arrangements entering the 21st century: which anchor will hold? *Quarterly Journal of Economics*, *134*(2), 599–646.
IMF (2012). *2011 review of conditionality*. Washington DC: International Monetary Fund.
IMF (2019). *2018 review of program design and conditionality*. Washington DC: International Monetary Fund.
IMF (2020). Our work: the IMF's fundamental mission. https://www.imf.org/external/about/ourwork.htm#:~:text=The%20IMF's%20fundamental%20mission%20is,giving%20practical%20help%20to%20members. (Accessed 1 June 2021).
IMF (2021a). *Monitoring of fund arrangements, various issues*. Washington DC: International Monetary Fund.
IMF (2021b). *World economic outlook, various issues*. Washington DC: International Monetary Fund.
IMF (2021c). *International financial statistics, various issues*. Washington DC: International Monetary Fund.
IMF (2021d). *Balance of payments statistics yearbook (balance of payments and international investment position), various issues*. Washington DC: International Monetary Fund.
IMF (2021e). *Annual report on exchange arrangements and exchange restrictions, various issues*. Washington DC: International Monetary Fund.
Ismail, K., Perrelli, R., & Yang, J. (2020). Optimism bias in growth forecasts—The role of planned policy adjustments. IMF working paper WP/20/229.
Itskhoki, O. (2020). The story of the real exchange rate. NBER 28225.
Kenen, P. B., & Schwartz, S. B. (1986). *An assessment of macroeconomic forecasts in the IMF's WEO*: *Working papers in international economics*, Princeton University: International Finance Section.
Kentikelenis, A. E., Stubbs, T. H., & King, L. P. (2016). IMF conditionality and development policy space, 1985–2014. *Review of International Political Economy*, *23*(4), 543–582.
Kunze, F. (2020). Predicting exchange rates in Asia: New insights on the accuracy of survey forecasts. *Journal of Forecasting*, *39*(2), 313–333.
Lewis, C., & Pain, N. (2014). Lessons from OECD forecasts during and after the financial crisis. *OECD Journal: Economic Studies*, *2014*(1), 9–39.
Luna, F. (2014). *IMF forecasts in the context of program countries*: *Independent evaluation office BP/14/05*, Washington DC: International Monetary Fund.
Meese, R., & Rogoff, K. S. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, *14*(1–2), 3–24.
Merton, Robert C. (1981). On market timing and investment performance. I. An equilibrium theory of value for market forecasts. *J. Bus.*, 363–406.
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). New York: National Bureau of Economic Research.
Mody, M. A., & Rebucci, M. A. (2006). Overview. In M. A. Mody, & Rebucci M.A. (Eds.), *IMF-supported programs: recent staff research*. Washington DC: International Monetary Fund.
Mussa, M., & Savastano, M. (1999). The IMF approach to economic stabilization. *NBER Macroeconomics Annual*, *14*, 79–122.
Musso, A., & Phillips, S. (2002). Comparing projections and outcomes of IMF-supported programs. *IMF Staff Papers*, *49*(1), 22–48.
Nordhaus, W. (1987). Forecasting efficiency: concepts and applications. *The Review of Economics and Statistics*, *69*, 667–674.
Przeworski, A., & Vreeland, J. R. (2000). The effect of IMF programs on economic growth. *Journal of Development Economics*, *62*(2), 385–421.
Schnader, M. H., & Stekler, H. O. (1990). Evaluating predictions of change. *The Journal of Business, Part 1*, *63*(1), 99–107.
Sinclair, T. M., Joutz, F., & Stekler, H. O. (2010). Can the fed predict the state of the economy? *Economics Letters*, *108*(1), 28–32.
Stiglitz, J. (2011). *A short note on surveillance and how reforms in surveillance can help the IMF to promote global financial stability*. Washington DC: International Monetary Fund.
Stubbs, T., Reinsberg, B., Kentikelenis, A., & King, L. (2020). How to evaluate the effects of IMF conditionality. *The Review of International Organizations*, *15*(1), 29–73.
Theil, H. (1961). *Economic forecasts and policy*. Amsterdam: North-Holland.
Timmermann, A. (2007). An evaluation of the world economic outlook forecasts. *IMF Staff Papers*, *54*(1), 1–33.
U. S. Government Accountability Office (USGAO) (2003). *International financial crises: challenges remain in IMF's ability to anticipate, prevent, and resolve financial crises*: *GAO-03-734*, Washington: Government Accountability Office.
Verbeek, J. (1999). *The World Bank's unified survey projections, how accurate are they? An ex-post evaluation of US91-US97*: *World bank policy research working paper 2071*, Washington D.C: World Bank.
World Bank (2022). World bank national accounts data, and OECD national accounts data files. https://data.worldbank.org/indicator/NE.TRD.GNFS.ZS.