



Are IMF rescue packages effective? A synthetic control analysis of macroeconomic crises[☆]

Kevin Kuruc

Department of Economics, University of Oklahoma, 332 Cate Center One, Norman, Oklahoma, 73019, USA



ARTICLE INFO

Article history:

Received 16 February 2021

Revised 31 January 2022

Accepted 1 February 2022

Available online 3 February 2022

JEL classification:

E3

E6

F4

O5

Keywords:

IMF

Financial crises

Synthetic controls

Business cycle policy

ABSTRACT

Whether, and to what degree, IMF lending succeeds in stabilizing economies remains an open question. Here, a synthetic control analysis of macroeconomic crises with IMF intervention is performed—leveraging the existence of similar crises without intervention—that finds positive recovery effects. In the first five years following a crisis, output differences are, on average, nearly two percent of GDP per year. Consistent with a liquidity channel, effects are hump-shaped and fade in the medium run. An analysis of historical IMF forecasts provides evidence against selection as a spurious driver of this result, suggesting that these positive estimates are indeed causal.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

The International Monetary Fund was designed and primarily functions as a source of financing for countries in short-term distress. In times of financial and other macroeconomic shocks, it is through the IMF that the international community attempts to provide the aid necessary to restore stability. The volatility of output in low- and middle-income countries leads the IMF to perform this function regularly: between 1970 and 2013, on average, 20 new programs began each year with total credit access equal to 3% of recipient GDP.

Despite the importance of this role, and the frequency at which the IMF plays it, a consensus does not exist on the organization's effectiveness. A simple model of a liquidity injection would predict that these interventions are weakly useful; in the worst case, subsidized lending substitutes for more expensive capital on private markets. In practice, skeptics have pointed to misguided policy imposed by the IMF (Stiglitz, 2002) and the negative signaling effect of using a lender of last resort (Reinhart and Trebesch, 2016) as countervailing forces against these liquidity benefits. It is frequently argued that these negative effects even dominate, making these engagements harmful on net. Settling this debate empirically has proven challenging due to the well-known econometric problems of this setting: IMF programs are not randomly allocated.

[☆] I would like to thank Oli Coibion and Dean Spears for invaluable mentorship throughout this project. Additionally, I thank David Beheshti, Saroj Bhattarai, Firat Demir, Dan Hicks, Cooper Howes, Niklas Kroner, Amartya Lahiri, Melissa LoPalo, Shinji Takagi, Tom Vogl, Tim Willems and the seminar participants at the University of Texas at Austin, the University of Oklahoma, and the Centre For Advanced Financial Research and Learning at the Reserve Bank of India, for helpful comments, feedback, and suggestions. All remaining errors are my own.

E-mail address: kkuruc@ou.edu

This study brings new facts, new data, and a new estimator to this problem and finds that IMF programs during crises have large, positive, recovery effects. I begin by documenting a stylized fact that has been thus far overlooked: on average, growth rates follow a sharp “V” around stabilization loans, bottoming out at the time of loan receipt. Past work has smoothed over these rich high-frequency dynamics in attempts to detect sustained growth effects (hence, focusing on multi-year averages) for the set of all IMF programs, including longer-run support for structural adjustments. Stabilization is a unique objective with potentially distinct effects over the time-horizons of interest during acute crises. When these programs are isolated, the challenges of the empirical setting resemble the well-known “Ashenfelter Dip” (Ashenfelter, 1978) of labor economics. It is not that IMF loans come in times of poor performance—a negative level effect—but that they are timed exactly during a reversal in falling growth rates—a positive dynamic effect.

To ensure that the empirical analysis allows for these recovery effects, attention is restricted to identifiable macroeconomic crises. By directly asking whether crises with IMF involvement have recovery dynamics that differ from crises without IMF involvement, any crisis-specific effects that could account for the dip and reversal in growth rates are present in both the treated and untreated observations. Information on the dates and locations of macroeconomic crises are drawn from Laeven and Valencia (2018) where standardized criteria are used to identify the onset of banking, currency, and sovereign debt crises around the world.

The IMF effect is then estimated using the synthetic control method (SCM) on this crisis sample: each crisis with IMF intervention has a synthetic control drawn from convex combinations of similar crises without IMF intervention. This estimator, designed in Abadie and Gardeazabal (2003) and Abadie et al. (2010), is a data-driven approach to selecting the appropriate control units for a given treated observation. Reweighting the untreated observations is particularly useful in this setting: the crises that receive IMF programs are more severe, on average, than the universe of untreated crises. An additional advantage of the SCM for this application is that it can be easily modified to account for non-linearities—a defining feature of the crisis dynamics studied here. I formalize a recommendation in Abadie et al. (2015) to drop untreated observations that appear qualitatively different than the treated observation as a relaxation of standard linearity assumptions. In the main specification, this is implemented by restricting synthetic controls for any given observation to come only from crises that are “local” (in the covariate space) to that observation.

The main empirical exercises indicate that IMF programs are effective at promoting recovery: in the first five years following a crisis, treated observations outperform their synthetic controls. The peak response—three years following the onset of a crisis—suggests that GDP in treated crises is about 2.8 percent larger than it otherwise would have been. A back of the envelope estimate of total output gains during the recovery divided by total credit access implies an associated “IMF multiplier” of roughly 2.6. These results are not driven by outliers, nor are they sensitive to the exact specification or assumptions underlying the SCM implementation. Six years following the onset of a representative crisis, the level effect is near-zero and noisily estimated across robustness exercises. Despite the absence of parametric restrictions on the dynamics of the IMF effect, the estimated impulse response function takes on a standard hump shape. These results are consistent with model-implied effects of a liquidity injection as well as the large literature studying the effects of more conventional business cycle policy.

The primary concern in assigning a causal interpretation to these results is that the IMF may have, and act on, information that has been omitted in this analysis. Indeed, it is a precondition of IMF financing that recipients have the “capacity to repay” (IMF, 2021c). It may then merely be lending funds where it forecasts strong recoveries, leading to a spurious positive relationship. The organization’s historical forecasts can be leveraged to study the plausibility of this threat. I demonstrate that, conditional on variables included in the SCM, IMF forecasts have little predictive power as to which recoveries will be unusually strong. If the IMF cannot predict differential recoveries, it is unlikely that its programs are correlated with unobservable factors that would produce them.

Following the evidence in support of a positive average effect of these programs, I conclude the paper with an analysis of effect size heterogeneity. Uncovering heterogeneity, or lack thereof, can inform future program design and point towards potential mechanisms for the main effects. For example, there appears to be (i) no correlation between the number/strictness of the policy conditions attached to the loan and (estimated) program success and (ii) a negative correlation between governance indicators and program success. Evidence for these takeaways is weaker than for the positive unconditional effect, but the contrast with well-known hypotheses in the development literature (e.g., Burnside and Dollar, 2000; Stiglitz, 2002), and their indirect support for a liquidity channel, makes them worth highlighting.

This paper contributes to a body of work analyzing the causes and consequences of IMF programs. Two particularly similar papers are Newiak and Willems (2017) and Essers and Ide (2019). These authors both (i) use the synthetic control method and (ii) select a sub-sample of programs to avoid confounding crisis effects. However, both papers take an inverse approach to the one taken here—they analyze episodes of explicit *non-crisis* IMF intervention. Newiak and Willems (2017) does so by studying a non-lending program, finding positive growth effects; Essers and Ide (2019) studies a “precautionary” program and finds more limited effects. This paper draws lessons from their respective methodologies but instead directs attention exclusively to crises, thereby studying a more widespread, and arguably consequential, set of interventions.

In general, more work exists on IMF effectiveness than can be properly accounted for here. Notable studies focusing on the growth effects of IMF lending include: Gündüz (2016), which uses a propensity score matching technique and finds that IMF loans are growth-promoting in low-income countries; Bas and Stone (2014), which estimates a two-stage structural model designed to account for adverse selection and likewise finds pro-growth estimates; Barro and Lee (2005), which uses political variables to instrument for IMF intervention and finds that the IMF harms growth; and Vreeland (2003), which finds

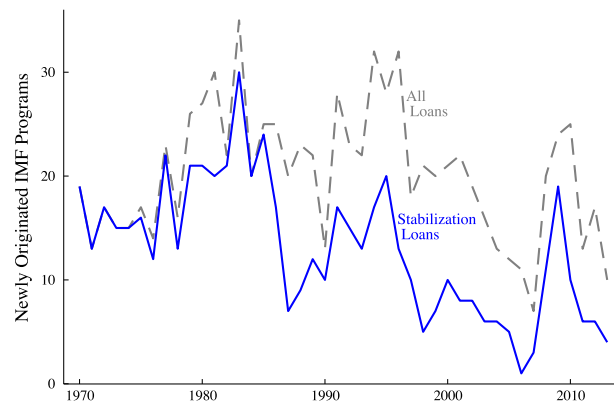


Fig. 1. The IMF is actively involved in global economy. *Notes:* Number of new IMF programs over time IMF (2019). Dashed line measures all programs; solid line measures the subset of loans classified as having a stabilization objective (see Section 2.1 or Appendix A for classification details).

that the IMF harms growth using a Heckman selection correction. Hutchison (2003) (in the spirit of the current analysis) includes a currency crisis dummy variable in an otherwise standard panel approach but continues to find negative effects of IMF involvement. Bordo and Schwartz (2000) uses a structural approach to ask if IMF lending during the Latin American and Asian crises was effective; they find it was not. Steinwand and Stone (2008) provides a more complete review of the literature that came prior to its publication and concludes that little consensus exists; work since then has failed to resolve this. Alongside this research primarily focused on the output effects of IMF lending, there is a related literature studying whether IMF programs prevent crises in the first place (Dreher and Walter, 2010; Jorra, 2012; Papi et al., 2015). Similarly mixed results have been documented.

Insofar as IMF loans are net-injections of liquidity financed through, albeit subsidized, public-sector debt, the implications of this paper are also related to the broader literature on business cycle policy. These programs are distinct in many ways from conventional domestic policy—they primarily cover balance of payments shortfalls and have external effects, including catalyzing additional sources of international support or investment (Collins et al., 2021)—but the large estimates here may provide indirect evidence on this broader topic. In particular, the crisis subsample this estimate is derived on speaks in part to the open question of whether the effects of debt-financed liquidity injections are larger during economic contractions (Auerbach and Gorodnichenko, 2012; Jordà and Taylor, 2016; Ramey and Zubairy, 2018).

The remainder of the paper is structured as follows. Section 2 presents the data and documents the stylized fact motivating the focus on crises. Section 3 details the estimator used in the analysis, including placebo exercises that guide the selection of a main specification. Section 4 presents main results, robustness, and studies the threat of IMF selection. Section 5 then explores which country and program characteristics predict program success. Section 6 discusses the implications of the analysis and concludes.

2. Empirical setting: IMF loans and crises

This section defines and presents characteristics of IMF programs and the crises used throughout the paper. A new stylized fact arises from a higher-frequency, unconditional, analysis of the setting than is typically performed: IMF stabilization programs are preceded by falling rates of economic growth and followed by rapid increases. Using the dates of crises rather than an IMF program as the event of interest, I show that this pattern could plausibly arise from the IMF becoming involved at the onset of acute macroeconomic crises—a similarly fast recovery in growth rates follows these events.

2.1. IMF programs

The IMF is actively involved in the global economy. Documented in Fig. 1, between 1970–2013 approximately 20 new IMF country-programs began per year. These come in a variety of instruments: Stand-by Arrangements, Extended Credit Facility, Rapid Financing Instrument, etc, that differ in their purpose. For example, the Extended Credit Facility is described as being for “Protracted BoP [Balance of Payments] need/medium-term assistance,” in contrast to the Rapid Financing Instrument which is designed for “Actual and urgent BoP needs” (IMF, 2021b). This paper is focused on the IMF’s effectiveness during economic turbulence, so the legal definitions of these instruments have been used to categorize a subset as “stabilization

Table 1
Summary statistics for stabilization programs & recipient country-years.

	Mean	Median	St. Dev	N
Loan Size (% GDP)	2.1	1.3	2.4	533
Conditions (count)	33.0	30.0	20.7	411
GDP Growth Rate (%)	2.3	3.2	5.6	524
Inflation (%)	44.8	11.7	154.1	466
External Debt (% GDP)	64.3	50.4	64.1	416
Current Account Balance (% GDP)	−4.5	−3.8	7.0	422

Notes: Mean, median, standard deviation and sample size for select characteristics of IMF stabilization loans and recipient countries. *N* indicates how many country-years with an IMF stabilization have non-missing values for the characteristic. GDP growth refers to real, local currency unit, total, GDP. Condition data is only available from 1980-on (Kentikelenis et al., 2016). Country characteristics from the World Development Indicators (World Bank, 2019) are occasionally missing because of imperfect national accounts systems in low- and middle-income countries.

loans” for the purposes of describing the empirical setting.¹ As is also documented in Fig. 1, the set of these stabilization loans makes up a substantial fraction of all lending activity.

Table 1 documents the characteristics of these loans as well as the economic situations they are provided towards. First, note the substantial number of programs in this sample (533). With this level of activity it cannot be the case that a representative stabilization loan is directed towards a high-profile rescue like Argentina in the early 2000s or the large loans initiated during the European Debt Crisis of the 2010s. Most programs are instead designed for less abrupt events in lower income countries.² A second feature of this data to note before presenting average country-program characteristics is the tremendous variation underlying each indicator. The sample of IMF stabilization programs is very heterogeneous.

In general, these stabilization programs come with substantial credit access at 2 percent of GDP, on average, though in two-thirds of cases the full amount is not disbursed. This is either because programs go “off-track” if recipients do not meet the conditions of a program (described below) or authorities voluntarily do not fully draw down available financing. A small subset of precautionary programs has no disbursements at all, despite large open lines of credit. In completed programs, disbursement of funds is typically complete within 2 years, though some later programs have extended this range to 3 years. These loans must be, and are in practice, repaid³; it is a precondition of IMF financing that recipients have the capacity to repay any withdrawn funds (IMF, 2021c).

Programs typically come with policy conditions, both quantitative and structural, related to economic management. Quantitative conditions are targets for indicators within the authority’s control, such as monetary aggregates or fiscal balances, but with flexibility over how they are met; structural conditions are non-quantitative directives, such as privatizing a specific sector (IMF, 2021a). Within these subsets, both have historically included conditions that are strictly necessary for funding continuation (“hard” conditions), but in recent years these are primarily of the quantitative type.⁴ Kentikelenis et al. (2016) provides a comprehensive dataset detailing the number, scope, and strictness of the conditions attached to individual programs starting in or after 1980. The average number of these conditions at the start of each loan is rather large at 33, though these conditions evolve throughout a program as some are completed, modified, or added. For the purposes of this paper, this information is used in Section 5 where I analyze whether the burden of conditions attached to a loan predicts estimated program success.

As for the country situations that receive stabilization loans, average real GDP growth rates (2.3%) are higher than might be expected, and even this is pulled down by extreme events as evidenced by the higher median. This corroborates the previous suggestion that high-profile collapses make up a minority of IMF interventions. Current account balances average −4% of GDP in the year a stabilization loan is enacted. Such a deficit is predictable given the packages were originally designed to alleviate negative imbalances in an effort to retain exchange rate stability. Average inflation is skewed by a moderate number of very high inflation episodes, but even the median is rather high (12%). The most interesting feature of these country situations, however, is the dynamic behavior of output surrounding the start of a program.

¹ Any loan from the following program is included as a stabilization loan: Standby Arrangements, Standby Credit Facility, Rapid Financing Instrument, Rapid Credit Facility, Precautionary Lending Line, Flexible Credit Line, Exogenous Shocks Facility; see Appendix A.

² In the main analysis a plurality of treated crises are in Africa, for example.

³ Only 3 countries are currently in arrears with the IMF (Somalia, Sudan, and Zimbabwe).

⁴ Specifically, quantitative conditions can come as Quantitative Performance Criteria (hard conditions) and Indicative Targets (soft conditions); structural conditions can come as Structural Performance Criteria (hard conditions), Prior Actions (hard conditions) or Structural Benchmarks (soft conditions) (Kentikelenis et al., 2017). In recent years Structural Performance Criteria have been eliminated, so that most hard conditions are quantitative.

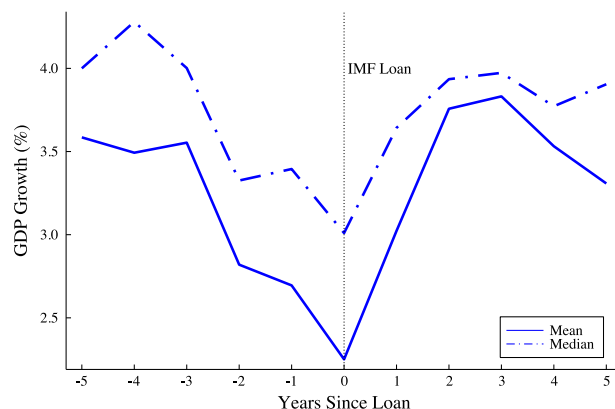


Fig. 2. Growth dynamics around IMF programs. *Notes:* Unconditional mean and median real (total) output growth rates surrounding the 467 stabilization loans in the sample with complete growth rate data. A “V” shape characterizes the process: IMF lending is either successful or systematically timed at the trough of macro-crises.

2.2. Average recoveries: An ashenfelter dip

Country growth rates increase following receipt of an IMF stabilization program. This is a critical first step towards understanding the empirical regularities of the treatment variable and the challenges posed by this setting. Fig. 2 depicts this pattern in an unconditional event study. Growth rates fall in the years preceding a program and recover rapidly following its inception, a pattern that is seemingly strong evidence in favor of IMF effectiveness.

Complicating this takeaway is the fact that national governments and the IMF may agree to begin programs at country-specific troughs. In this case, estimation issues would be analogous to the well-known Ashenfelter Dip of labor economics (Ashenfelter, 1978). In the setting of job retraining programs, Ashenfelter (1978) observes that wage earnings fall just prior to these programs and would have rebounded even in their absence; earnings were unusually low in the treatment period relative to these individuals' latent potential. The general point is that traditional methods—which correct for level, not dynamic, differences—are ineffective in settings with selection near an individually-specific nadir. In the case of interest here, GDP growth rates at or just before treatment are not a good counterfactual for post-treatment performance by exactly this logic. Identification requires accounting for the fact that these loans may originate at moments when growth rates were set to rebound regardless of IMF involvement.

2.3. Treated and untreated macroeconomic crises

The question arising from Fig. 2 is whether the types of events the IMF typically involves itself with are the same types of events that are naturally followed by strong reversals in growth slides. To address this issue, data is taken from Laeven and Valencia (2018) which systematically provides start dates for three types of crises, defined in the following way: *Banking Crises* (N=127) are years with significant bank runs, losses or liquidations, and banking policy intervention; *Currency Crises* (N=191) are years when the domestic currency depreciates 30% or more relative to the U.S. dollar (only the first year if this happens in consecutive years); *Sovereign Debt Crises* (N=51) are years with sovereign default or debt rescheduling; *Twin/Triple Crises* (N=30) are years with some combination of the above crises, defined as a mutually exclusive category here. N refers to the total number of crises in the sample period used for the main analysis, 1970–2013, with the temporal constraint being the six years of recovery studied. Note that these criteria are independent of growth rates; output dynamics around these events will not arise merely by construction.

With these crises defined, it is possible to study whether the pattern observed in Fig. 2 could be caused by the fact that IMF loans are designed to solve short-term, crisis-like, problems. Panel (a) of Fig. 3 is similar to the unconditional event study that produced Fig. 2; however the event is now the start date of one of these three crises, regardless of IMF intervention. A similar ‘V’ pattern in growth rates emerges. Much or all of the recovery following IMF loans in Fig. 2 may be driven by this effect, illustrating the importance of conditioning on experiencing one of these unique events.

Panel (b) takes a preliminary step towards such an analysis by splitting the sample of crises into those that I call *treated* throughout the paper and those that are *untreated*. Treated observations are country-years experiencing a crisis that receive any IMF program in the year of, or year following, this crisis. In contrast to the prior subsection, both stabilization and non-stabilization loans are included. Once the focus is restricted to these identifiable crises, it is no longer necessary to discard non-stabilization loans out of concern that they were initially designed for non-crisis contexts. If an Extended Credit Facility is the instrument offered to a country in crisis, in practice, that is still an IMF stabilization loan, regardless of the instrument's legally specified motive. Additionally, I retain precautionary programs where authorities do not ultimately request any disbursement as treated observations. As with something like the European Central Bank's famous “whatever it

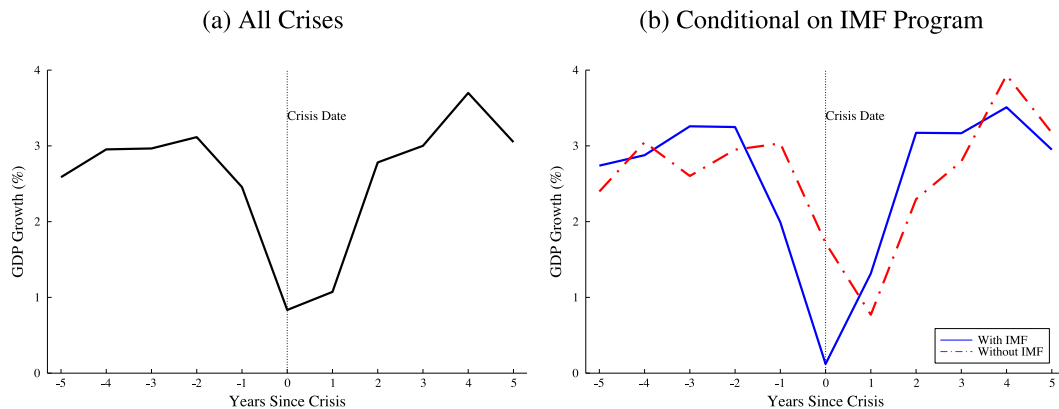


Fig. 3. Growth dynamics around crises. Notes: (a) Unconditional mean real (total) output growth rates surrounding the various crises as classified by [Laeven and Valencia \(2018\)](#) ($N = 269$). (b) Crises split into those with an IMF program ($N = 149$) in year of or year following crisis date, and those without an IMF program ($N = 120$). See [Section 2.3](#) for formal treated/untreated definition.

takes” promise or deposit insurance designed to prevent bank runs, crisis management relies on expectations management, so even a credible promise of IMF financing is likely to have positive effects if actual financing does.

If two crises occur in subsequent years (i.e., a debt crisis happens in the year following a currency crisis) the second of the two crises is dropped from the analysis—the second is considered to be an extension of a particularly severe initial crisis. If instead a country experiences two crises with at least one year between them, these are considered separate crisis observations with overlapping data in their outcomes. Dynamic panel and time-series analyses often use observations with overlapping data; I follow this convention here. In total, there are 149 treated crises with full growth rate data.

Untreated observations have no IMF program arranged in the year of or year following their crisis. The same rule applies for dropping the second of directly subsequent crises. Additionally, I drop from the analysis any remaining untreated observation that had an IMF program begin in the year prior to, or two years following, this crisis. These observations are partially treated in a way that makes them difficult to conceptualize as belonging to either of these discrete groups. There are 120 untreated observations.

Two facts arise from the conditional dynamics in [Fig. 3b](#). First, recipients of an IMF program experience more severe output collapses prior to their crisis. This is not surprising; the IMF exists in part to stabilize the most extreme economic events. Second, the crises with an IMF program have a rebound in growth rates a year earlier than their counterparts. This may be directly caused by the first fact—larger collapses by definition allow for larger recoveries—but could indicate that IMF intervention quickens the pace of recovery. To preview the main results, when pre-crisis differences are eliminated using synthetic controls this post-crisis timing difference remains, driving the positive estimated effects.

3. Empirical strategy: Constructing synthetic controls

The SCM was first used in [Abadie and Gardeazabal \(2003\)](#) and was formalized thereafter in [Abadie et al. \(2010\)](#). This section draws from these studies as well as [Dube and Zipperer \(2015\)](#), [Doudchenko and Imbens \(2016\)](#), and [Powell \(2021\)](#), which generalize the original estimator. The main departure from traditional regression methods is that each treated unit’s counterfactual comes from a synthetic control—a weighted average of untreated observations—rather than being inferred from an estimated model. This gives the method intuitive appeal and adds a layer of transparency: each observation’s counterfactual can be easily and directly observed ([Abadie, 2021](#)).

The primary advantage of the SCM is that it provides a data-driven approach to reweight the untreated group in a way that increases its similarities with the treated observations. [Heckman et al. \(1998\)](#) argues—in an empirical setting with a similar pre-program dip—that one of the most acute problems for non-experimental econometric techniques is that the empirical distributions of the treated and untreated samples often do not have a common support. Even where their supports do overlap, their mass may be concentrated in different regions. Stated simply by the authors, “Comparing incomparable people contributes substantially to selection bias as conventionally measured” ([Heckman et al., 1998](#)). [Billmeier and Nannicini \(2009\)](#), studying trade liberalizations, shows this “parametric extrapolation” problem to be likewise serious in a standard cross-country setting.

[Abadie et al. \(2010\)](#) originally mitigates this problem by restricting synthetic controls to come from convex combinations of untreated units. By forcing the SCM to use events that surround the treated observation, cases of extrapolation are prevented. Here, I additionally follow the prescription in [Abadie et al. \(2015\)](#) for settings where a subset of untreated units is observationally distinct enough from the treated unit that they are unlikely to usefully serve in constructing a counterfactual, even when quantitatively up- or down-weighted. The authors suggest tightening the convex combination restriction such that synthetic controls can only be drawn from local observations (i.e., those with sufficiently similar pre-treatment

characteristics and experiences). Restricting to local observations relaxes global linearity assumptions, a well-studied advantage of matching estimators that has been explicitly shown to improve performance in the face of pre-program dips (Dehejia and Wahba, 2002).

To formalize, suppose that the data generating process can be written as an unbiased forecasting equation as in (1):

$$y_{i,h} = F^h(X_{i,0}, \mathbf{y}_{i,0}^l) + \theta_h IMF_i + u_{i,h}. \quad (1)$$

Here $y_{i,h}$ represents real (total) GDP growth rates h years following crisis i ; h is normalized to 0 at the date of the crisis. $F^h(\cdot)$ is h specific but maps only inputs known at time 0 to future outcomes—a direct, rather than iterated, forecast. IMF_i is an indicator for IMF involvement. It has no time subscript because each i is a crisis, such as Kenya-1992, not a country; being treated or untreated is therefore a fixed characteristic. The vector X is some set of crisis characteristics that predict recovery dynamics and may be correlated with IMF program receipt, for example, the type of crisis or the level of government debt at its onset. $\mathbf{y}_{i,0}^l$ is a vector of lagged values of the outcome variable for all years prior to the crisis. The treatment effect, θ_h , varies with the horizon and is separable. The error, $u_{i,h}$, is mean-zero and uncorrelated with IMF assignment (i.e., all information that is jointly correlated with outcomes and IMF lending are represented in X, \mathbf{y}^l).

Aside from the standard conditional independence assumption on $u_{i,h}$, two further assumptions are necessary for constructing an unbiased counterfactual by synthetic controls.

Assumption 1. For all treated observations i , there exists a local linear approximation of $F^h(X_{i,0}, \mathbf{y}_{i,0}^l)$ in a neighborhood around $(X_{i,0}, \mathbf{y}_{i,0}^l)$, denoted $\hat{F}_i^h(\cdot)$.

Assumption 2. In this neighborhood of i , there exists a set of J_i untreated observations and a vector of weights $\lambda_i^j \in [0, 1]$ such that:

$$\sum_{j \in J_i} \lambda_i^j X_{j,0} = X_{i,0} \quad \sum_{j \in J_i} \lambda_i^j \mathbf{y}_{j,0}^l = \mathbf{y}_{i,0}^l \quad \sum_{j \in J_i} \lambda_i^j = 1$$

Assumption 1 states that there is a first-order linear approximation of the data generating process at each point. In practice, the analysis strays from an infinitesimally small neighborhood, so there is an implicit assumption that there exists a “good” linear approximation as the space of interest expands. Assumption 2 requires that within this neighborhood there exists a convex combination of untreated crises that can match $X_{i,0}$ and $\mathbf{y}_{i,0}^l$. Notice that efforts to satisfy these two assumptions push against one another; the smaller the local neighborhood, the more reasonable Assumption 1 becomes, but the harder it is to satisfy Assumption 2.

Denote the counterfactual outcome as $y_{i,h}^c$. If the above assumptions are met, then the typical synthetic control result obtains. See Appendix B for details.

$$\begin{aligned} \sum_j \lambda_i^j y_{j,h} &= y_{i,h}^c + e_{i,j,h} \Rightarrow \\ y_{i,h} - \sum_j \lambda_i^j y_{j,h} &= \theta_h - e_{i,j,h} \end{aligned}$$

The difference between a treated outcome and its synthetic control's outcome is an unbiased estimator for θ_h . The error, $e_{i,j,h}$, depends on disturbances to the treated observation and all J_i observations plus an error arising from the linear approximation. This non-parametric forecasting approach shares advantages of the widely used local projection method (Jordà, 2005) and other direct forecasting approaches. For each horizon, h , the effect estimate is the average difference between the treated and their synthetic controls at that horizon. No structure is imposed on the dynamic shape of the effect, nor are errors compounded as they are in an iterated forecast.

In practice, each synthetic control is found by solving the following minimization problem.

$$\begin{aligned} \Lambda^i &= \text{argmin}_{\Lambda^i \in [0,1]^J} (l'Z_j - Z_i)W(l'Z_j - Z_i)' \\ &\text{subject to } \sum_{j \in J} l_j = 1 \end{aligned} \quad (2)$$

Here Z_i is a row vector of the pre-crisis variables that are targeted for matching. Z_j is a matrix where each row contains these same variables for one of the J_i local untreated observations. The assumption of local, rather than global, linearity on the data generating process (Assumption 1) requires that only local crises are used to generate synthetic controls. W is a weighting matrix that rescales the errors of each targeted variable by that variable's inverse variance. This ensures that the influence of squared errors in the minimization problem is invariant to the scale of the respective distributions. Λ^i is the column vector of weights assigned to the untreated observations.

3.1. Main specification

The baseline implementation is intentionally simple, though I note at the outset that the main results are robust to changes in any of the three important choices detailed here. First, the crises in the synthetic control must be of the same

crisis type as the treated observation: banking (currency) (debt) crises will only have other banking (currency) (debt) crises underlying their synthetic control. Second, untreated observations must fall within a ± 9 percentage point growth rate band in *each* of the pre-periods; crises with a growth rate of 2% one year prior to their crisis can only be matched with crises that had growth rates between -7% and 11% one year prior to their crisis. These two restrictions are implemented to increase the similarity of the crises being compared to the treated observation, in accordance with the local linearity of Assumption 1. Third, the SCM tries to match six pre-crisis real GDP growth rates (five years preceding and contemporaneous) using the restricted subset of local untreated observations.

This specification choice is in part driven by prior literature—for example, the finding that different crisis types have different recoveries (Reinhart and Rogoff, 2009) and the popularity of lagged dependent variable SCM specifications (Powell, 2021)—but it is also independently supported by placebo exercises on the universe of untreated crises. Placebo exercises are common in SCM analyses as a method for quantifying the range of outcomes expected under the null hypothesis. This measure is useful for inference purposes,⁵ but can also be informative of how to best generate synthetic controls (Dube and Zipperer, 2015). The untreated observations can serve as training data for generating an SCM specification because the coefficient of interest is known: there is no IMF program in these crises, so the IMF effect must be zero. If, then, for a given j untreated observation, the other $-j$ untreated observations are used to generate a synthetic control, the expected θ_h is zero. Iterating this procedure over all untreated observations provides both a mean error and a root mean squared error (RMSE) for each synthetic control specification. Under ideal conditions, specifications can be empirically ranked according to how precisely the trajectories of untreated crises are forecasted when other untreated crises are used to generate synthetic controls. The pseudo-algorithm and corresponding numerical results are available in Appendix C.

In this application, there are various dimensions by which specifications can be ranked (the RMSE for different forecasting horizons, for example), so this procedure is not perfectly discriminatory as in an ideal case. The results of this exercise therefore only informally guide the choice of specification rather than providing a precise mapping between desirability and numerical RMSEs. Despite this imprecision, these exercises unequivocally favor imposing some restrictions on which crises are considered local to a treated observation, and hence available for matching. There is necessarily a trade-off between generating good pre-treatment fits and restricting the similarity of local crises—the more observations it can draw from, the better the SCM can find convex combinations that reproduce pre-crisis characteristics. How to balance these considerations is an empirical question.

Restricting the controls to have sufficiently similar pre-period growth rates (± 7 p.p, 9 p.p, 11 p.p, etc.) significantly reduces forecasting errors on the untreated training data relative to an SCM with wider bounds (± 13 , 15 p.p.). Growth boundaries prevent cases where, say, mild crises have synthetic controls comprised of unusually high growth episodes averaged with episodes of extreme collapses. It is not surprising that convex combinations of these qualitatively different crises would fail to predict the outcome of the hypothetical mild crisis. Conversely, it is perhaps surprising that additional restrictions beyond ± 9 – 11 p.p. do not seem to improve model fit. Likewise, restricting synthetics to come from same-crisis-type observations provides surprisingly little improvement in fit. That restriction is retained in the main analysis as it does not empirically reduce performance and better reflects well-known findings about heterogeneous recovery dynamics (Reinhart and Rogoff, 2009).

A notable implication of restricting which untreated crises can contribute to a synthetic control is that some treated crises will have an empty set of local crises. In these cases, the observation gets no synthetic control, and hence is dropped from the analysis. Of the 149 treated crises available for the exercise, 31 are dropped for this reason in the main specification. The omitted crises are mostly comprised of twin/triplet crises and debt crises as these start with the smallest number of available untreated observations, and hence have a higher probability that none fall within the growth bounds.⁶ This feature may be viewed as an advantage. If there are no untreated crises within a liberally defined neighborhood around some observation, it is unlikely that much can be learned about IMF effectiveness by studying its outcome. The approach here is then a data driven way to eliminate crises lacking a reasonable counterfactual in the data. As the sample size of this setting is already on the smaller side, I hedge towards specifications that disqualify the fewest observations.

Once the sample has been restricted to these local crises, a simple specification that targets only lagged growth rates works well according to the placebo exercises in Appendix C.⁷ In the context of Eq. 1, this implies the $X_{i,0}$ vector is left empty. This special case is a common application of the SCM⁸; it is intuitively appealing and lagged outcomes are thought to capture much of the predictive capability with the fewest data constraints (Powell, 2021). This general conjecture appears to hold in this setting and has the further benefit of dropping the fewest observations due to missing non-GDP data. Ultimately, the choice of a baseline model for estimating post-crisis growth rates is inconsequential: the results are robust to a battery of alternative SCM specifications based on relaxing/altering the choices detailed here.

Before proceeding to the results, it is worth discussing the choice of estimating the system in growth rates rather than levels, as is common in other cross-country SCM applications (e.g., Billmeier and Nannicini, 2013; Newiak and Willems, 2017). In a broad sense, this can be reduced to the fact that this is a business cycle analysis rather than a traditional devel-

⁵ The asymptotic variance of this estimator has not been analytically characterized.

⁶ Appendix Table A1 contains all eligible treated crises with an indicator for whether they are used or dropped in the main analysis.

⁷ Current account deficit, inflation, and external debt to GDP ratio (all in the year of the crisis) were considered for matching. This is discussed in Section 4.2 where these variables are included in robustness exercises.

⁸ So common, in fact, that Doudchenko and Imbens (2016) have a specific name for this version of the SCM: “constrained regression.”

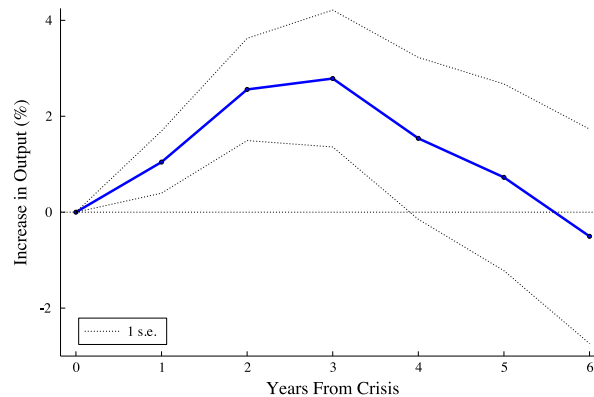


Fig. 4. Output gains associated with IMF involvement. *Notes:* Implied real (total) GDP level effects from main specification computed as the average treated observation's cumulative growth minus their respective synthetic control's cumulative growth at that horizon. Standard errors are approximated using the sample standard deviation from the distribution of these effect size estimates (see Appendix C for details).

opment analysis, despite the focus on low- and middle-income countries. More concretely, it helps to notice that estimating a counterfactual for the level of GDP in time $t + h$ or a counterfactual path of growth rates between t and $t + h$ both identify the cumulative output effects over that period. The question then is only whether the path of GDP leading into a crisis is more informative, when used as the matching variable, than the path of GDP growth rates for predicting post-crisis cumulative growth.

Asymptotically, it would be preferable to match the level in all periods. Exact level matches additionally imply exact growth rate matches, and so dominate matching growth rates directly. In this finite sample where matches are imperfect, however, there is a trade-off between imperfectly matching levels—and generating potentially large differences in growth rates depending on the sequence of level errors—or directly matching trajectories leading into the crisis. The choice here to match trajectories at the expense of level differences is in line with the logic of a traditional difference-in-differences analysis requiring parallel trends. A distinct further benefit is that growth rates are more naturally comparable across time; because there are so few crises each year, synthetic controls are necessarily drawn from crises occurring across the entire temporal sample. Matching on levels raises the question of whether GDP per capita of, say, \$5,000 (constant USD) meant the same thing in 1970 as it meant in 2013 regarding whether these economies are structurally similar and should still be expected to serve as a good counterfactual. Whatever the underlying reason, in Appendix C I again use the placebos to verify these concerns by demonstrating that the variance of forecast errors with this specification is approximately three times larger than when growth rates are used. For completeness, Appendix D shows that the results in levels are consistent with, though noisier than, the main results.

4. Results

This section presents the results, performs robustness checks, and studies whether the main results could be driven by unobservable selection. I find that crises with an IMF program have significantly faster recoveries than their synthetic counterparts, though similar medium-run outcomes. This pattern is robust to a battery of specification changes, not driven by outliers, and appears unlikely to be attributable to selection on the part of the IMF.

4.1. Main results

Fig. 4 presents the results from the main specification where treated and untreated observations are defined in the way described in Section 2.3. The impulse response function is the average difference in cumulative growth (i.e., implied level differences) between IMF program recipients and their synthetic controls for the 118 crises (41 Banking, 58 Currency, 19 Debt) where the data allowed a synthetic control to be constructed.⁹ The synthetics come from a pool of 120 (46 Banking, 62 Currency, 10 Debt, 2 Twin Crises) untreated crises. Differential recoveries result in treated economies that are larger for up to 5 years. The integral of this function, approximately 8.7 p.p., is the total output difference as a percent of crisis year GDP. For loans that average only 3.3% of crisis year GDP on this sub-sample, an 8.7% increase in output is a large return on investment. Along with being economically significant, the probability that the entire path of coefficients is jointly zero is low ($p < .01$) under the estimated covariance matrix (see Appendix C for details on hypothesis testing).

⁹ As noted in Section 3.1, of the 149 treated crises, 31 are dropped from the main analysis because there are no local crises for them to draw a synthetic control from.

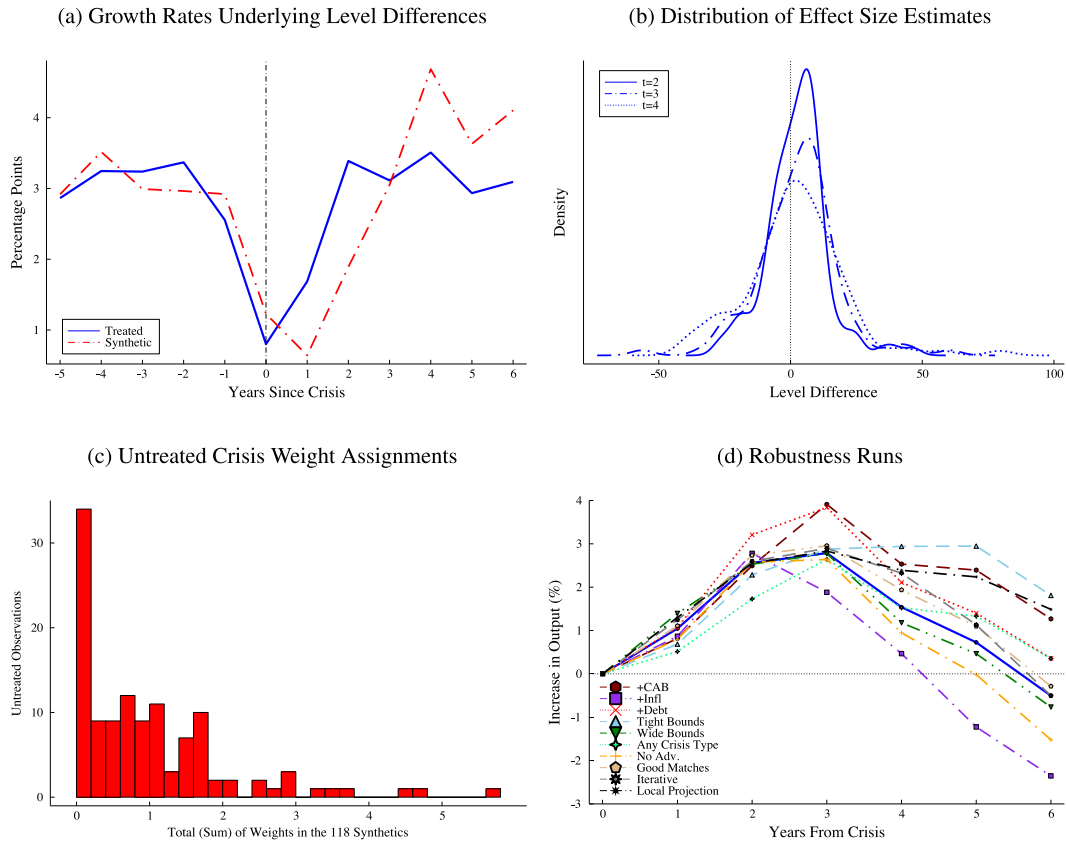


Fig. 5. Analysis of main results. *Notes:* Components underlying main effect in Fig. 4. (a) Real (total) GDP growth rate differences underlying level differences. Solid blue line: average from the 118 crises with a non-empty set of “local” crises to match with; Dotted red line: average of their synthetic controls; pre-crisis rates matched by construction. (b) Density plot of the 118 differences in outcomes between the treated and their synthetic controls at various horizons (i.e., x-axis measures cumulative growth since the start of the crisis of interest; a value of 20 on the $t = 3$ distribution corresponds to 20% cumulative growth over the first three post-crisis years). Densities can be conceptualized as horizontally overlaid around each horizon’s point estimate in Fig. 4. (c) Histogram of total weight for each untreated observation in the main analysis, summing to 118 as each synthetic control has total weight 1.0 by construction. The three crises with weights >4 are: Venezuela 1982, Iceland 1989, and Belize 2012; none have particularly bad crises that disproportionately drive the main results. (d) Robustness tests: “+ CAB” adds contemporaneous current account deficit (as % of GDP) as a variable the SCM targets; “+ Infl” adds contemporaneous inflation; “+ Debt” adds contemporaneous external debt to GDP ratio; “Tight Bounds” shrinks the growth bounds defining local crises from 9 to 7; “Wide Bounds” increases these bounds to 11; “Any Crisis” drops the same-crisis-type restriction; “No Adv.” removes advanced economies from the potential controls; “Good Matches” computes the average after dropping 10% treated-synthetic combinations with the worst pre-crisis matches; “Iterative” recomputes synthetic controls for each horizon to take advantage of all crises with h years of post-crisis data for the h -period estimate; “Local Projection” runs horizon-specific linear regressions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

These level differences come from an underlying comparison of growth rates between the recipients and synthetic controls. Fig. 5a plots these growth dynamics. The hump-shaped level response is the result of recoveries that begin much stronger in the treated crises but are followed by a period of catch-up growth in the untreated crises. The unconditional result in Fig. 3—that crises with IMF intervention have recoveries of a similar magnitude, but that begin earlier—holds under this conditional analysis. Fig. 5a also demonstrates that the SCM has generated good matches, on average. The pre-crisis experiences of the untreated group now closely resemble those of the treated group. The SCM, by construction, oversamples from the untreated crises in a way that generates these similar trajectories.

Before demonstrating the robustness of this result to the estimator, it should be noted that it is not driven by a small number of outliers, a problem known to commonly plague cross-country analyses (Easterly, 2005). First, Fig. 5b demonstrates that the average main effects are not the result of a few particularly high-performing treated observations. This density plot is made up of the 118 underlying effect estimates (treated minus synthetic control) that, when averaged, produce the various point estimates in Fig. 4. For example, the average of the $t = 3$ density function is the peak point estimate on the impulse response function. Note that this is not the distribution where hypothesis testing would occur; it can simultaneously be likely that the mean of these distributions is above zero—which would correspond to a low p-value on a non-zero effect estimate—even while a non-trivial fraction of outcomes are below zero due to the random disturbance,

$u_{i,h}$. What is highlighted here is that the entire distribution of estimated effects appears shifted right from zero: the mean, median, and mode are all positive at these various horizons.

Fig. 5b does not rule out that a single untreated observation with a particularly bad outcome could be included in many synthetic controls and thus disproportionately driving the positive effects. Fig. 5c plots the distribution of total synthetic weights for the untreated observations across their respective contributions to all synthetic controls. For example, a value of 2.0 for an untreated observation could be generated if it was used with weights {0.5, 0.5, 1.0} in three synthetic controls, and 0.0 in all others; the integral of the histogram in Fig. 5c equals 118.0 as the 118 synthetic controls have weights that sum to 1.0 by construction. It is apparent from this figure that a wide variety of untreated observations contribute to the analysis. At the same time, a few observations receive substantial weight—three contribute total weight greater than 4.0 across synthetics. Manual verification of these crises (Belize 2012, Iceland 1989, and Venezuela 1982) confirms that none have unusually bad outcomes that could themselves account for the main results. For further detail, Appendix Table A2 contains the full set of untreated crises and their respective weights. In general, the unequal weights displayed in this histogram—including the one-third of untreated crises that are completely unused—is expected when the untreated group is quite different than the treated group. The SCM has identified and oversampled from the untreated crises that look most like they would have gotten an IMF program, leaving those without similarity to the treated group with little or no weight.

4.2. Robustness to alternative matching

A series of modifications to the SCM confirms that the main results are not driven by the exact specification choice. Fig. 5d summarizes the collection of robustness exercises by plotting the corresponding collection of impulse response functions from these alternative specifications. The recovery effects are stable.

First, the characteristics that the SCM targets for matching is expanded to include inflation ('+Infl'), external debt to GDP ratios ('+Debt'), and current account deficits ('+CAB') in the year of the crisis. Each is independently included as an additional target variable (weighted by the variable's inverse variance) in separate analyses; in Appendix D a single analysis incorporating all three is presented as well. These variables are likely used by the IMF in its lending decisions and, despite much evidence to the contrary in the placebo exercises, may be correlated with recoveries conditional on pre-crisis growth paths.

Then, returning to the main "growth only specification, the structure of the SCM is altered. First, growth boundaries are shrunk and expanded by 2 p.p. in either direction (from ± 9 to 7, 'Tight Bounds'; to 11, 'Wide Bounds') as a way of ensuring the exact local definition does not account for the results.¹⁰ I then let the SCM use any crisis type to generate synthetics, so that, for example, banking crises can have currency and debt crises in their synthetic control ('Any Crisis'). This exercise substantially alters the variation the SCM utilizes and reduces the observations dropped for lack of any local match from 31 to just 8 because it introduces more potential matches for each treated observation. Similarly, I drop advanced economies (as defined by the IMF; 'No Adv.') from the pool of potential matches. These economies may be structurally different and are overwhelmingly untreated. Then, to make use of the most recent available data, I iteratively estimate a new SCM for each horizon and plot the corresponding point estimates ('Iterative'). In contrast to the main analysis where data must be available for all six post-crisis years for inclusion (and therefore only includes crises occurring in 2013 and earlier), this estimation computes new synthetic controls for each horizon such that crises occurring as late as 2018 are used for the horizon one estimate, crises occurring as late as 2017 are used to estimate horizon two, and so on. This increases the number of observations at early horizons to 124 from 118 and likewise adds additional untreated observations. Next, the main specification is re-run, but the average outcome is computed after dropping the 10% of treated observations with the worst pre-period synthetic control matches (as measured by the squared errors of pre-period growth rates; 'Good Matches'). And finally, to ensure that there is nothing peculiar about the SCM in general, I replicate the specification as closely as possible in a local projection framework ('Local Projection'). The local projection regressions—which independently estimate the effect at each horizon using a linear model—condition on the same pre-crisis growth rates and crisis-type indicators that the baseline SCM matches on. By imposing global linearity assumptions, this estimation removes the local restrictions used in the other specifications.

In each case the general shape of the response function is unchanged. Nearly all point estimates fall within the one standard error band of the main results. Despite the uncertainty underlying any given run, a clear pattern arises: crises with IMF intervention have substantially stronger recoveries than observationally similar crises without.

4.3. IMF Forecasts support causal interpretation

The identification argument to this point has dealt exclusively with the observable challenges of the setting (i.e., the pre-program dip). Implicitly, the reasoning has been that if a synthetic control looks enough like a treated observation in measurable ways, it serves as a good estimate for how that program would have behaved without an IMF program. However, if program assignment is correlated with a variable omitted from this analysis, and with recoveries, the estimates would not have a causal interpretation. The most plausible worry of this sort is that the IMF—because of its requirement regarding

¹⁰ Appendix D does this for bounds other than 7 and 11.

Table 2
IMF forecasts lack marginal predictive power.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
a. Forecasts published in spring									
	Y_1			Y_2			Y_3		
γ^f	0.20 (0.26)	-0.17 (0.26)	-0.18 (0.26)	0.38** (0.19)	0.11 (0.18)	0.09 (0.19)	0.48*** (0.17)	0.27 (0.17)	0.26 (0.17)
Crisis Controls		✓	✓		✓	✓		✓	✓
Program Indicator			✓			✓			✓
N	156	156	156	156	156	156	156	156	156
b. Forecasts published in fall									
	Y_1			Y_2			Y_3		
γ^f	0.50** (0.21)	0.23 (0.21)	0.23 (0.22)	0.51*** (0.16)	0.30* (0.16)	0.29* (0.17)	0.51*** (0.15)	0.35** (0.16)	0.34** (0.16)
Crisis Controls		✓	✓		✓	✓		✓	✓
Program Indicator			✓			✓			✓
N	157	157	157	157	157	157	157	157	157

Notes: Results from regressions of IMF forecasts in year of crisis on actual cumulative recoveries at multiple horizons. Y_t columns represent cumulative growth rates t periods following crisis. γ^f is the estimated coefficient on the IMF forecast in that particular regression. For each horizon, regressions with and without controls are run. "Crisis Controls" are simply contemporaneous growth rates and crisis type indicators, "Program Indicator" is an indicator for whether or not the crisis is treated. The sample size, 156/157, is smaller than the total sample of crises—historical forecasts are only made publicly available starting in 1990. Two iterations of these forecasts are produced throughout the year—in the spring (panel a), and in the fall (panel b). *, **, *** represent significance at the 0.10, 0.05, and 0.01 levels, respectively.

recipients' capacity to repay—is lending into crises that it (rightly) anticipates will have a strong recovery for some reason not included in the SCM or robustness exercises. This would produce an erroneous positive estimate.

This concern is now directly addressed using historical forecasts published by the IMF at the time of these crises (IMF, 2018). I show that there is little to no independent information in these forecasts. This is true even after accounting for the possibility that forecasts are differentially produced for countries with and without IMF programs. If IMF forecasts have little marginal predictive value, there is limited scope for the organization to perform the selection that threatens a causal interpretation of the estimates. What this exercise cannot rule out is the existence of a variable that (i) drives recoveries, but (ii) the IMF does not know drives recoveries, and yet (iii) is nonetheless correlated with lending decisions—an unlikely, but not impossible, scenario.

To formalize, suppose there is some variable ϕ that drives recoveries, $Y_{i,h}$ (measured as cumulative growth h years following crisis i), through a monotonic function $G(\phi)$, but is absent from the current model of the data generating process, $H(\mathbf{y}_{i,0}^f, \mathbf{X}_{i,0})$. Here $H(\cdot)$ can be seen as an incomplete version of the true data generating process, $F^h(\cdot)$ in Eq. 1; it is a model of recoveries using data included in the SCM, but omitting the hypothesized ϕ variable.

$$Y_{i,h} = H(\mathbf{y}_{i,0}^f, \mathbf{X}_{i,0}) + G(\phi_{i,0}) + \omega_{i,h} \quad (3)$$

If the IMF observes and is aware of the positive effects of ϕ , then the following regression will return a positive coefficient on its forecasts, $\gamma_{i,h}^f$.

$$Y_{i,h} = H(\mathbf{y}_{i,0}^f, \mathbf{X}_{i,0}) + \gamma^f Y_{i,h}^f + \xi_{i,h} \quad (4)$$

IMF forecasts are assumed to have no direct effect on growth, but $\xi_{i,h}$ includes $G(\phi_{i,0})$ and is therefore positively correlated with $Y_{i,h}^f$. This produces a $\hat{\gamma}^f > 0$.

In practice, it is possible that IMF forecasts are produced in a systematically different way for country-years where programs begin, for example by building in some positive effect of the intervention. The regression is therefore separately implemented both with and without an indicator for whether the observation receives an IMF program. This instead asks whether the IMF can predict, within the treated and untreated groups, which recoveries will be unusually strong. The logic presented above holds when this indicator is included as long as the ϕ variable is not perfectly correlated with program receipt. $H(\cdot)$ is simplified to be linear; simplifying $H(\cdot)$ (or incorrectly specifying in other ways) will, if anything, overstate the marginal information in the forecasts as it leaves more variation for forecasts to explain.

Table 2 presents the estimation results. Forecasts are available starting in 1990 (spring and fall iterations are published), so this subset of later crises makes up the sample of 156. The columns are separated by which horizon of cumulative growth is forecasted (where Y_h is h -year cumulative growth following the crisis). Forecasts made in the spring better represent the information available throughout the year to the IMF, so I focus the discussion first on panel (a). Columns (1), (4), and (7) include no controls as verification that the forecasts are at least unconditionally correlated with future performance. Beyond the first post-crisis year there is a reasonably strong and significant relationship between forecasted and actual outcomes when no covariates are included. While the organization does not meet a full information rational expectations (FIRE)

benchmark, $\hat{\gamma} = 1$ (Mincer and Zarnowitz, 1969), these forecasts are certainly more than noise. Once basic controls of crisis year growth rates and crisis type indicators are added, coefficients shrink substantially in magnitude and all horizons lose statistical significance at even the 10% level. It is this marginal predictive power that would be necessary to threaten a causal interpretation. For context, in the presence of covariates a FIRE forecast would continue to return $\hat{\gamma} = 1$ (see Appendix E), so these regressions indicate an absence of both economic and statistical significance. Using forecasts published in October (panel b) or reproducing the analysis with the Penn World Tables (Appendix Table A5) changes little.

In sum, these coefficients are far from what would be expected under a FIRE benchmark with an omitted variable. This suggests that the IMF has little ability to forecast differential recoveries once simple controls are included and makes selection on such differential recoveries unlikely. While this result is not surprising—forecasting the recoveries of crises is notoriously challenging (Eicher et al., 2018)—it is critical for the interpretation of the results in this paper.

5. Heterogeneity and mechanisms

This section takes steps towards contextualizing the results and understanding potential mechanisms by analyzing if, and which, country-program characteristics predict estimated program success. The data constraints of the setting coupled with the lack of a natural extension of the SCM to other outcomes leaves this short of a full investigation into the underlying drivers. Nonetheless, the correlations presented below serve as preliminary evidence for future research on IMF program design.

The formal heterogeneity analysis uses a simple regression approach. The dependent variable is the individual estimate of output effects across all horizons, $\sum_h \hat{\beta}_{i,h}$, for observation i as a percent of crisis-year GDP. Equivalently, it is a discrete approximation of the integral of each respective i 's estimated impulse response function. This cumulative effect size is studied along six dimensions of interest to academics and policymakers: crisis type, world region, monetary size/percent completed/conditionality of the program, and state capability. These various dimensions enter as the independent variable Θ_i in distinct regression analyses.

Est. Cumulative Effect for i

$$\left(\sum_h \hat{\beta}_{i,h} \right)_i = \mathbf{A} \Theta_i + v_i \quad (5)$$

Formally, Eq. 5 is estimated separately for each of the six country-program characteristics. The resulting \mathbf{A} estimates are displayed in Fig. 6 with quantitative regression results presented in Appendix D. Unfortunately, these are inherently low-powered tests. The individual outcomes here are noisy estimates on a high-variance series, which makes the small number of observations available (118) a constraint. For this reason, only conditional means are presented with the blanket caveat that none of the results are statistically significant at conventional levels, though in one case of interest I note where an estimate is more than one standard error from zero.

Crisis and regional differences (panels a, b) uncover interesting features of the crises that drive the positive average effects. Although the modal crisis the IMF engages in is a currency crisis, the average effect on this subsample is estimated to be just above zero. Instead, banking crises (and less so debt crises) drive the positive effects. Regionally, the effects are primarily driven by the 84 total programs in African and Latin American countries, with East Asia and Pacific countries performing extraordinarily poorly following receipt of an IMF loan. This latter finding is consistent with case-study critiques of the East Asian crisis of the late '90s (Stiglitz, 2002).

Loan size and conditions (panels c, e) have no notable relationship with outcomes. There is one outlier (not plotted) with a loan-to-GDP ratio of over 30% that fully accounts for the positive relationship between loan size and estimated effects. A line of best fit omitting that single point (the gray dashed line in panel c) demonstrates that the positive relationship is not a robust feature of the estimates. That loan size is not predictive of success says little about the liquidity channel; it is consistent with a scenario where the IMF scales programs approximately to their need.

The non-negative estimate for conditions is more interesting. This measure is a weighted average of conditions—constructed to account for relative strictness (Kentikelenis et al., 2016)—and so the positive point estimate implies that more burdensome agreements predict weakly better outcomes. While quantitatively small and far from statistical significance, that the estimate is non-negative leaves this analysis unsupportive of the primary critique of the IMF, that the conditions imposed with loans are harmful for recipients. This non-relationship is not an artifact of pooling quantitative and structural conditions which may have distinct effects; in Appendix Figure A2, I show that neither of these condition-types independently have a meaningful relationship with program outcomes.

Percent of agreed funds that are disbursed throughout the program, a monetary measure of program completion (panel d), displays a weakly positive relationship with outcomes. However, there are nine purely precautionary programs with zero disbursements. The zero values are not indicative of failure to complete a program; rather, in these cases, authorities chose not to draw on available funds. When these observations are omitted in the regression analysis, the positive correlation becomes much stronger. This provides evidence for the claim that, conditional on participating in a program where authorities request some disbursement, the more of the program that is completed, the better. Coupled with the observation that conditionality has little relation with program success, this further supports the idea that it is the liquidity itself, and not program design or strictness, that promotes recoveries.

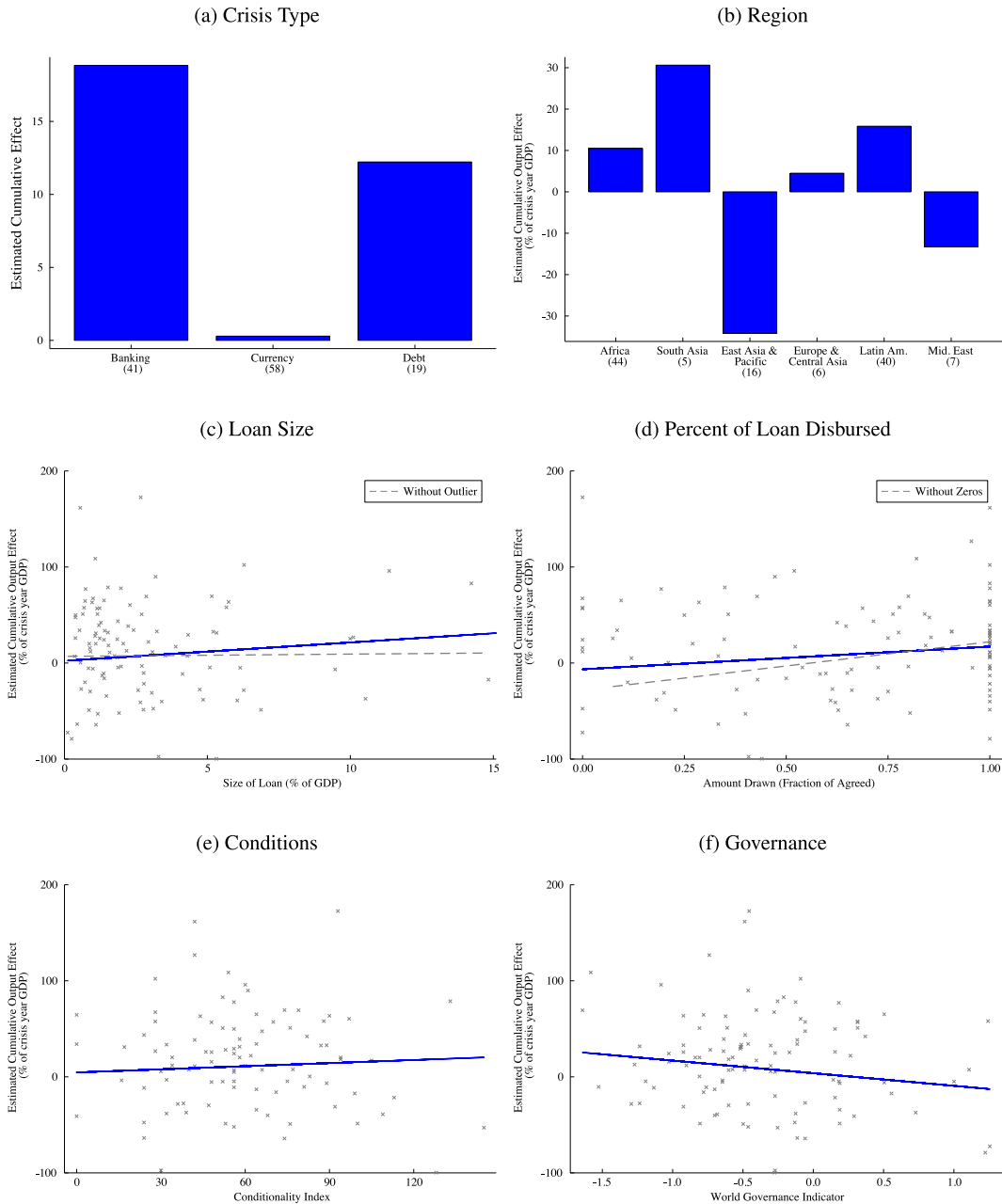


Fig. 6. Effect estimates along various dimensions. *Notes:* Average cumulative growth effects plotted along different dimensions. Numbers in parenthesis in panels (a,b) denote treated observations in the respective subcategory. Loan size (c) is as a percent of contemporaneous GDP. Percent of loan disbursed (d) measures the degree of program completion, monetarily. Conditions (e) are a weighted (2:1 “hard” to “soft”) sum of total conditions (Kentikelenis et al., 2016). Governance indicators (f) are taken from the World Governance Indicators (Kaufmann et al., 2011). Limits on the axes are imposed to make regression slopes (or lack thereof) visible; this omits outliers from the scatter. Appendix A has more details on the data used; Appendix D contains the underlying formal regression results.

Finally, and perhaps most interestingly, governance indicators (panel f) appear negatively correlated with outcomes: the relationship visible in the scatter is estimated to be 1.2 standard errors from zero.¹¹ This negative relationship runs counter to well-known findings in the development literature that international transfers lead to increases in economic activity only in well-governed countries (Burnside and Dollar, 2000). As stressed throughout the paper, the focus on IMF loans during crises creates a unique setting which may account for the contrasting findings. Weaker states tend to run fiscal policy that is

¹¹ The corresponding p-value is 0.23 (see Appendix D).

procyclical (Frankel et al., 2013); if financing during recessions prevents contractionary policy in these countries, the effects of lending could plausibly be much larger. This correlation also serves to mitigate remaining selection concerns. If the main results were in fact due to country-specific qualities that attracted both IMF loans and promoted recovery, it should not be the countries scoring lowest on governance indicators driving them.

Further research on potential mechanisms is necessary before concrete lessons can be drawn for IMF program design. Nonetheless, the fact that policy conditions have a weak positive correlation with performance, and that programs in countries with the weakest governance were most successful, are interesting in their own right and provide additional evidence that a liquidity channel is responsible for the main results.

6. Conclusion

Analyzing the rich high-frequency dynamics surrounding IMF loans into macroeconomic crises reveals that these programs have substantial benefits for recipient countries: cumulative output gains are estimated at more than 8 percent during the first 6 years of recovery. In some sense, this is a low bar for the organization to clear. It would be surprising if large liquidity injections into crises did not achieve the goal of, at least partially, stabilizing economies. Yet, convincing evidence for any such effect has been absent in the academic and policy literature, resulting in considerable doubt regarding the usefulness of the organization. The analysis here fills this gap and points towards a more optimistic reality: if the estimates in this paper are any guide, there are large gains to the IMF actively engaging with the global economy in pursuit of its stabilization objective.

Declaration of Competing Interest

None.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at [10.1016/j.jmoneco.2022.02.002](https://doi.org/10.1016/j.jmoneco.2022.02.002)

References

- Abadie, A., 2021. Using synthetic controls: feasibility, data requirements, and methodological aspects. *J Econ Lit* 59 (2), 391–425.
- Abadie, A., Diamond, A., Hainmueller, J., 2010. Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. *J Am Stat Assoc* 105 (490), 493–505.
- Abadie, A., Diamond, A., Hainmueller, J., 2015. Comparative politics and the synthetic control method. *Am J Pol Sci* 59 (2), 495–510.
- Abadie, A., Gardeazabal, J., 2003. The economic costs of conflict: a case study of the basque country. *American Economic Review* 93 (1), 113–132.
- Ashenfelter, O., 1978. Estimating the effect of training programs on earnings. *Rev Econ Stat* 60 (1), 47–57.
- Auerbach, A.J., Gorodnichenko, Y., 2012. Measuring the output responses to fiscal policy. *American Economic Journal: Economic Policy* 4 (2), 1–27.
- Barro, R.J., Lee, J.-W., 2005. IMF programs: who is chosen and what are the effects? *J Monet Econ* 52 (7), 1245–1269.
- Bas, M.A., Stone, R.W., 2014. Adverse selection and growth under IMF programs. *The Review of International Organizations* 9 (1), 1–28.
- Billmeier, A., Nannicini, T., 2009. Trade openness and growth: pursuing empirical glasnost. *IMF Staff Papers* 56 (3), 447–475.
- Billmeier, A., Nannicini, T., 2013. Assessing economic liberalization episodes: asynthetic control approach. *Review of Economics and Statistics* 95 (3), 983–1001.
- Bordo, M.D., Schwartz, A.J., 2000. Measuring real economic effects of bailouts: historical perspectives on how countries in financial distress have fared with and without bailouts. National Bureau of Economic Research. Working Paper No. 7701.
- Burnside, C., Dollar, D., 2000. Aid, policies, and growth. *American Economic Review* 90 (4), 847–868.
- Collins, C., Kuruc, K., Takagi, S., 2021. Assessing the role of the IMF in fragile states. In: Chami, R., Espinoza, R., Montiel, P.J. (Eds.), *Macroeconomic Policy in Fragile States*. Oxford University Press, pp. 521–547.
- Dehejia, R.H., Wahba, S., 2002. Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics* 84 (1), 151–161.
- Doudchenko, N., Imbens, G.W., 2016. Balancing, regression, difference-in-differences and synthetic control methods: a synthesis. National Bureau of Economic Research. Working Paper No. 7701.
- Dreher, A., Walter, S., 2010. Does the IMF help or hurt? the effect of IMF programs on the likelihood and outcome of currency crises. *World Dev* 38 (1), 1–18.
- Dube, A., Zipperer, B., 2015. Pooling multiple case studies using synthetic controls: an application to minimum wage policies. IZA. Discussion Paper No. 8944.
- Easterly, W., 2005. National policies and economic growth: a reappraisal. In: Aghion, P., Durlauf, S.N. (Eds.), *Handbook of Economic Growth Volume 1*. Elsevier, pp. 1015–1059.
- Eicher, T., Kuenzel, D., Papageorgiou, C., Christofides, C., 2018. Forecasts in times of crisis. *IMF Working Papers* 18/48.
- Essers, D., Ide, S., 2019. The IMF and precautionary lending: an empirical evaluation of the selectivity and effectiveness of the flexible credit line. *J Int Money Finance* 92, 25–61.
- Frankel, J.A., Vegh, C.A., Vuletin, G., 2013. On graduation from fiscal procyclicality. *J Dev Econ* 100 (1), 32–47.
- Gündüz, Y.B., 2016. The economic impact of short-term IMF engagement in low-income countries. *World Dev* 87, 30–49.
- Heckman, J., Ichimura, H., Smith, J., Todd, P., 1998. Characterizing selection bias using experimental data. *Econometrica* 66 (5), 1017–1098.
- Hutchison, M., 2003. A cure worse than the disease? currency crises and the output costs of IMF-supported stabilization programs. In: Dooley, M., Frankel, J. (Eds.), *Managing Currency Crises in Emerging Markets*. University of Chicago Press, pp. 321–360.
- IMF, 2018. Historical WEO forecasts database. <https://www.imf.org/external/pubs/ft/weo/data/WEOhistorical.xlsx> (Accessed: February, 2018).
- IMF, 2019. Member's financial data: Lending arrangements. <https://www.imf.org/external/np/fin/tad/extarr1.aspx> (Accessed: September, 2019).
- IMF, 2021a. Factsheet: IMF conditionality. <https://www.imf.org/en/About/Factsheets/Sheets/2016/08/02/21/28/IMF-Conditionality>. Accessed 8/4/2021.
- IMF, 2021b. Factsheet: IMF lending. www.imf.org/en/About/Factsheets/IMF-Lending. Accessed 8/3/2021.
- IMF, 2021c. Factsheet: IMF stand-by arrangements. www.imf.org/en/About/Factsheets/Sheets/2016/08/01/20/33/Stand-By-Arrangement. Accessed 8/3/2021.
- Jordà, Ò., 2005. Estimation and inference of impulse responses by local projections. *American Economic Review* 95 (1), 161–182.
- Jordà, Ò., Taylor, A.M., 2016. The time for austerity: estimating the average treatment effect of fiscal policy. *The Economic Journal* 126 (590), 219–255.

- Jorra, M., 2012. The effect of IMF lending on the probability of sovereign debt crises. *J Int Money Finance* 31 (4), 709–725.
- Kaufmann, D., Kraay, A., Mastruzzi, M., 2011. The worldwide governance indicators: methodology and analytical issues. *Hague Journal on the Rule of Law* 3 (2), 220–246.
- 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.
- Kentikelenis, A.E., Stubbs, T.H., King, L.P., 2017. IMF Conditionality 1980–2014: codebook & uses of data. University of Cambridge. Available at: <http://www.imfmonitor.org/conditionality.html>
- Laeven, L., Valencia, F., 2018. Systemic banking crises revisited. IMF Working Paper No. 18/206.
- Mincer, J.A., Zarnowitz, V., 1969. The evaluation of economic forecasts. In: Mincer, J.A. (Ed.), *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*. Nation Bureau of Economic Research, pp. 3–46.
- Newiak, M., Willems, T., 2017. Evaluating the impact of non-financial IMF programs using the synthetic control method. IMF Working Papers No. 17/109.
- Papi, L., Presbitero, A.F., Zazzaro, A., 2015. IMF Lending and banking crises. *IMF Economic Review* 63 (3), 644–691.
- Powell, D., 2021. Synthetic control estimation beyond case studies: does the minimum wage reduce employment? *Journal of Business & Economic Statistics*.
- Ramey, V.A., Zubairy, S., 2018. Government spending multipliers in good times and in bad: evidence from us historical data. *Journal of Political Economy* 126 (2), 850–901.
- Reinhart, C.M., Rogoff, K.S., 2009. *This Time is Different: Eight Centuries of Financial Folly*. Princeton University Press.
- Reinhart, C.M., Trebesch, C., 2016. The international monetary fund: 70 years of reinvention. *Journal of Economic Perspectives* 30 (1), 3–28.
- Steinwand, M.C., Stone, R.W., 2008. The international monetary fund: a review of the recent evidence. *The Review of International Organizations* 3 (2), 123–149.
- Stiglitz, J.E., 2002. *Globalization and its Discontents*. W.W. Norton & Company.
- Vreeland, J.R., 2003. *The IMF and Economic Development*. Cambridge University Press.
- World Bank, 2019. World development indicators. <https://data.worldbank.org/> (Accessed: September, 2019).