Are IMF Rescue Packages Effective? A Synthetic Control Analysis of Financial Crises*

Kevin Kuruc†

University of Texas at Austin

November 5, 2018
Click Here For Most Recent Version

Abstract

This paper estimates the output effects of IMF loans during acute macroeconomic crises. Using the universe of financial crises from 1975-2010, I study whether recovery dynamics differ across crises that do and do not receive IMF intervention. I condition on the *type* of financial crisis, employ a new estimator to find the most relevant controls units—the synthetic control method—and use forward looking variables to address the selection issues associated with IMF lending. In contrast to much of the existing literature, I find that IMF lending has large short-run effects. Countries that receive an IMF loan have GDP that is, on average, 1-2 percent larger in the 2-3 years following the onset of a crisis than what is predicted by their synthetic controls. Consistent with either a liquidity effect or policy advice specific to managing a crisis, the difference fades in the medium run. Likewise, I find the recovery effects are largest in countries with weak institutions: places where policy advice and an "international lender of last resort" may be most useful.

*I would like to thank Oli Coibion and Dean Spears for invaluable mentorship throughout this project. Additionally I thank David Beheshti, Saroj Bhattarai, Cooper Howes, Niklas Kroner, Melissa LoPalo, Shinji Takagi, Tom Vogl and the seminar participants at the University of Texas for helpful comments, feedback and suggestions.

†Ph.D. Candidate, University of Texas at Austin. Email: kevinkuruc@utexas.edu

1 Introduction

Lending to countries experiencing financial and other macroeconomic crises is a unique role served by the International Monetary Fund (IMF) in the global economy. These rescue packages—such as were employed in the Latin American Debt Crisis, the Asian Financial Crisis and more recently the global financial crisis—can be politically contentious. They include large sums of pooled international money and come with a strong push for the structural adjustments the IMF sees fit. However, as depicted in Figure 1, there are many more of these programs[1] than just the high profile events; in the 1980's, for example, all but 2 years saw more than 15 new short-term loans. The importance and frequency of these events has resulted in broad economic and political interest in the question of whether IMF intervention into crises does in fact help stabilize macroeconomic conditions. This remains an open question.

Theoretically, a simple economic model of a liquidity transfer would predict these loans must be weakly useful; in the worst case it substitutes for more expensive capital on private markets. In practice, skeptics have pointed to the strict countercyclical policy requirements of the IMF (Stiglitz, 2002) and the negative signaling effect of using a "lender of last resort" (Reinhart and Trebesch, 2016) as counterveiling forces against the liquidity benefits. These counterveiling effects, if large enough, could go as far as to make these loans harmful for the recipient. Settling this empirically has proven challenging given the well-known selection issues: countries do not randomly ask for loans, nor does the IMF randomly approve them.

This paper revists this question with an empirical strategy that directly overcomes selection using a combination of new data and a new estimator, and finds IMF loans into crises have large positive output effects. The approach, generally, will be to estimate the differential recoveries of well-defined macro-crises with and without IMF financing. For the sample of "well-defined" crises I take advantage of recent work by Valencia and Laeven (2012) who systematically define and date country-years experiencing the onset of a financial crisis. Within this sample, I generate counterfactual recoveries for the treated group with a new matching estimator—the synthetic control method (SCM). For each observation, the SCM chooses a convex combination of untreated crises to be that crisis' "synthetic control." These controls are chosen by searching for a weighted average of untreated crises that reproduce targeted pre-crisis characteristics of the treated unit. For example, the main specification will find synthetic controls by attempting to replicate the path of GDP growth rates leading into the crisis. Treatment effects are then simply the difference in outcomes

[1] A *program* is what the IMF calls lending packages since they come with policy reforms as well as liquidity.

between the treated crisis and its synthetic control in the post-periods of interest.

Using this method I find IMF involvement in a financial crisis is associated with a significantly faster recovery than would be otherwise anticipated. Figure 2 plots an impulse response function that summarizes the main results of the paper. In the first 3 years following a financial crisis treated observations substantially outperform their synthetic controls. These differences are both economically signficant and robust across specifications. The point estimate two years following the onset of the crisis—that IMF lending is associated with a 2 percent increase in GDP—remains large across a wide range of robustness checks. I further find government spending rises faster in treated countries than their synthetic controls. This verifies what might be viewed as an expected "first-stage" mechanism. In the medium-run (horizons of 4-5 years following the crisis) differences are smaller, estimated with substantially less precision, and turn out not to be robust. This pattern is qualitatively consistent with the IMF helping countries through the crisis phase, but not systematically changing long-run potential GDP.

The strategy employed here overcomes the most challenging identification problems of this setting. First, the IMF lends into country-years that may be expected to recover with or without IMF lending. I formally document this challenge with a new stylized fact: IMF loans are preceded by a pre-program, or Ashenfelter (1978), dip. Output growth rates are falling in the years leading into an IMF loan and rapidly recover to (or slightly higher than) the rates they were 5-6 years prior to the loan. The pattern is a striking "V" with the IMF entering at the trough. In light of this, it is necessary to use observations experiencing a similar macro-crisis as controls to account for the recovery dynamics that come along with these events. I take care of this by using a "within crisis" strategy: only comparing the recoveries of financial crises with other country-years experiencing a financial crisis.

Second, even within the sample of financial crises, the treated observations have different pre-crisis trends. On average, financial crises receiving IMF loans experience a more severe crash; the SCM is designed to take account for this. By constructing weighted averages of the untreated crises that replicate each pre-crisis growth path, the SCM is over sampling from the crises that "look" more like the treated observations. The SCM can—and will—further mitigate extrapolation by restricting the synthetic controls to only draw these convex combinations from qualitatively similar crises. Here, qualitatively similar crises are defined as those that fall in a neighborhood of the pre-crisis growth values of the treated crisis of interest.

The final challenges are standard concerns regarding selection on unobservables. This setting has a two-sided selection process, either of which could be confouding: countries choose to apply for IMF loans and then the IMF chooses which of these to accept. While the SCM is not designed explicitly to deal with this, I will provide evidence that in properly controlling for observable characteristics the SCM has indirectly left little space for selection to be confounding.

To study whether selection on the IMF's part is likely to be problematic I will take advantage of publicly available historical forecasts produced by the organization. A standard OLS regression with actual recoveries as the dependent variable and the IMF's forecasted recovery (at the time of the crisis) as an independent variable can shed light on whether the organization is able to predict unusually good (or bad) recoveries. If there is additional information contained in their forecasts that the SCM is not accounting for, the estimated coefficient on these forecasts should remain positive when variables used in the SCM are included as covariates. There is no evidence this is the case.

With regards to country selection, characteristics of the observations that drive the positive results can be informative as to whether selection remains problematic on this end. As an example of what sort of issues I am concerned about, suppose it were the case that only governments planning to pursue counter cyclical fiscal policy are the ones attempting to generate outside financing. The SCM would misattribute well-managed crises to IMF lending in this case. Here, and in other plausible stories that drive an upward bias in estimates, it would be *positive* selection on the country side. However, measured by the World Bank's Country Policy Institutional Assessment (CPIA), I find the estimated effects sizes[2] are *negatively* correlated with measures of institutional quality and economic policy. That is, the countries with the weakest institutions are the observations driving the positive results. Regardless of the exact country specific selection mechanism that can be conjured up, most seem unlikely to square with this fact.

In fact, I find the negative correlation between effect sizes and outcomes to be an interesting dimension of heterogeneity that may provide evidence on the mechanisms at work. Despite being in stark contrast to findings in the literature studying foreign aid more broadly and its effects on growth (Burnside and Dollar, 2000) the negative interaction is not necessarily surprising here. Countries with weak institutions and/or below average policy are likely most in need of both advice on managing such crises and an international lender of last resort. Consistent with this line of reasoning, my own past work has shown for countries with extremely low levels of state capacity—"Fragile States"—IMF lending and the fiscal oversight it brings can have catalytic effects on outside inflows of development financing (Kuruc, 2018).

Finally, I find that effect sizes are larger for countries with fixed exchange rates than for coun-

[2]Effect sizes are measured by how much a treated observation outperforms its synthetic control.

tries with more flexible regimes. This is in line with theoretical results that spending multipliers—especially from *external* financing—are large under fixed exchange regimes (Farhi and Werning, 2016). Output is estimated to be 4-5 percent larger (cumulatively) in response to IMF loans that are between 1-2 percent of GDP, and so the implied "IMF multiplier"[3] is in fact large here.

The paper continues as follows. The next subsection puts this work in the context of the existing literature. Section 2 introduces the empirical setting and formalizes the challenges to overcome. Section 3 describes the synthetic control method, both generally and how it is specifically used in this paper. Section 4 presents the results under the main specification and shows these findings are robust. Section 5 examines the response of other aggregates as well as heterogeneity in the effect sizes to try and understand the mechanisms underlying the main results.

1.1 Related Literature

This paper offers a resolution to the challenges that have limited past work estimating the effects of IMF lending. Up until this point, the foremost concern has been to find methods to overcome endogenous selection. The general difficulty of obtaining strong, excludable aggregate instruments (Deaton, 2010) has induced most papers to employ a method that requires specifying a parametric first-stage equation that predicts the probability of obtaining an IMF loan (Bordo and Schwartz, 2000; Hutchison, 2003; Vreeland, 2003; Bas and Stone, 2014; Gündüz, 2016). Methods that control for this probability of selection, such as propensity score matching or heckman corrections, rely on having a well-estimated and properly specified first-stage equation which has presented a formidable challenge to these authors (Gündüz, 2016). The way forward has been to augment these first-stages to include variation in political variables that can help explain IMF financing, such as "share of trade with US." Barro and Lee (2005), for example, identify a few political variables with this property and use these directly as instruments. Many papers using other first-stage estimators have built on this approach. Although a majority of these papers estimate negative effects of IMF loans, it is hard to know what to make of these estimates in the case that their first-stages are misspecified (or in the case of instruments that exclusion restrictions are not satisfied).[4] This is not just a theoretical concern, results in this literature are sensitive to exact choice of first-stage variables and functional form (Gündüz, 2016).

[3]What I call an *IMF multiplier* is not a spending multiplier persey, just a back of the envelope calculation suggesting of how much output is created per IMF dollar lent.

[4]While failure of exclusion restrictions is in theory not a unique problem in this setting, Deaton (2010) suggests it is in practice especially problematic in the case of cross-country aggregate instruments.

This paper approaches the problem instead by directly attempting to make comparisons among similar experiences with and without IMF financing.[5] The literature spawned by Ashenfelter (1978) suggests that in many settings using "like" control units that properly account for observables can go a long way towards alleviating selection concerns (Heckman et al., 1998). One additional benefit of the approach in this paper is that even if the exogeneity requirements needed for causality fail, the conditional differences are easily interpretable and interesting for moving the debate forward.[6] These benefits come at the cost of only estimating IMF output effects local to financial crises. Past work has been more ambitious in attempting to estimate the global effects of IMF lending. The more limited scope of this paper, however, seems to be a reasonable starting point for a literature that has failed to converge on a consensus.

Methodologically, this paper also draws on and contributes directly to the literature developing and applying the synthetic control estimator. While first used in Abadie and Gardeazabal (2003), it was formally developed by Abadie, Diamond and Hainmueller (2010) and extended by Dube and Zipperer (2015). Applications include Peri and Yasenov (2015), Cavallo et al. (2013), Acemoglu et al. (2016), Abadie, Diamond and Hainmueller (2015) and many others. Despite its emphasis on estimating counterfactual dynamics the SCM has yet to be widely taken up in macroeconomics, though one notable exception is Billmeier and Nannicini (2013) who study trade liberalizations.

Finally, the implications of this paper relate to work on spending multipliers. An especially similar line of work to this one is Kraay (2012, 2014) who estimates government spending multipliers using World Bank lending as an instrument and finds relatively small multipliers ($\approx .5$). While this paper estimates an "IMF multiplier" which is not analogous, the results here are different enough to warrant mention. The combination of results here and in Kraay (2012) turns out not to be puzzling in light of the results in Auerbach and Gorodnichenko (2012). These authors show, using US data, multipliers may be substantially larger in times of recession. Even Ramey and Zubairy (2018), who dispute the conclusions of Auerbach and Gorodnichenko (2012), show that when monetary policy is constrained, as it is given the exchange rate regimes of many countries in my sample, spending multipliers can be well-over 1.[7] In short, the large results in this paper seem consistent with and contribute to the literature on spending multipliers more generally.

[5] In contrast to fitting a global regression model that uses model based counterfactuals.

[6] Clemens et al. (2012) argue for simple estimators in the foreign aid literature precisely for this reason.

[7] They study the zero-lower bound rather than fixed exchange rates, but there are many theorertical similarities in these settings.

2 Empirical Setting: IMF Loans and Financial Crises

This section defines and presents characteristics of IMF programs and financial crises. I describe the characteristics of these programs and document the stylized fact that IMF loans are preceeded by falling rates of economic growth and experience rapid increases following their introduction. Then, using the dates of financial crises rather than IMF program as the "event," I show this pattern could plausibly arise from a setting in which the IMF becomes involved at the onset of an acute macroeconomic crisis. A similarly fast recovery in growth rates follows financial crises. This leads me to proceed by using a "within crisis" strategy for the main analysis: comparing whether financial crises with an IMF loan have better recovery dynamics than otherwise similar financial crises without a loan.

2.1 IMF Programs

The IMF is extremely involved in the global economy both in supplying credit and guiding economic policy. Figure 1 plots the number of newly originated IMF programs per-year that I've classified as "short-term." A "program" is an agreement between the IMF and a member country that involves extending credit (in rare cases only a line of credit is opened that is not ultimately drawn from) and comes with some policy conditions the IMF imposes on the country. These come from a variety of instruments at the IMF's disposal: Stand by Arrangements (SBAs), Extended Credit Facility (ECF), Rapid Financing Instrument (RFI), etc, that differ slightly in their purpose. For example, the Extended Credit Facility's purpose 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."[8] As this paper is focused on the short-term effects of IMF loans I have categorized only a subset of loans as "short-term" for the purposes of presenting summary statistics. This classification is based on the IMF's description where, for example, ECFs would not be short-term but RFIs would be classified as such.[9] This split is far from perfect, but it is only used to produce summary statistics to roughly understand the empirical regularities. In the main analysis I will measure the differences in outcomes between crises that receive *any* IMF program; a longer-term loan is still clearly a liquidity treatment.

While the IMF began issuing programs before 1970, the empirical analysis will be restricted to programs beginning in 1975 and beyond. The mid-70's were a turning point in IMF operations

[8] Source: https://www.imf.org/en/About/Factsheets/IMF-Lending

[9] Described in the data appendix.

as membership increased to include many low and middle income countries, and its operations began to look much more similar to present day programs (see Reinhart and Trebesch (2016) for a more complete history of this evolution; also note the pick-up in programs at this time in Figure 1). Unsurprisingly, the level of IMF activity is relatively counter-cyclical; for example, the early 2000's showed a large dip in lending which quickly reversed during the global financial crisis.

Even among the subset of IMF programs classified as "short-term" there is a wide range of country situations and loan sizes. Table 1 presents moments for the distributions of various short-term indicators in country-years recieving a short-term program. GDP growth is calcuated from the Penn World Table, the financial crisis indicator comes from Valencia and Laeven (2012) and will be described in more detail in the following sub-section, the size of IMF programs comes from the MONA database at the IMF and the other indicators are pulled from the *World Economic Outlook* 2017 edition. The primary takeaway from Table 1 is that while situations are by no means good, they vary significantly. *Most* have slow growth, but some do not; *most* have low to moderate inflation, but some are hyper-inflationary; *most* are running large current account deficits, but certainly not all. Nearly 20% are facing at least one of the financial crises to be studied in this paper, which of 476 loans turns out to make up significant share of all financial crises in the data. Finally, the average loan is large at 2.4% of GDP; the American Reinvestment and Recovery Act during the recent global financial crisis was around 4% of the US economy. Taken together, it becomes difficult to label a situation and IMF response "typical," and goes a long way in demonstrating why models relying on first-stage selection methods have a poor fit: these situations fall all over the distribution of economic indicators.

The policy requirements on these loans, too, are highly idiosyncratic and country specific. For example, a 2010 Jamaican program came with the condition to sell Air Jamaica and this level of specificity is not unusual. As has been written about in great detail prior, the IMF policy conditions typically are related to increasing privitization, liberalizing trade, reducing fiscal burdens, and restraining the monetary authorities issuance of cash (Stiglitz, 2002). These conditions will ultimately not play a role in the statistical analysis. For one, they are difficult to categorize in a clean way. Not only are they highly country specific, but there are many attached to each loan. A further complication is that waivers are occasionally issued for countries that fail to implement certain condition so it is not obvious how binding they are. As a result, much doubt has arisen over whether these conditions do anything in practice (Easterly, 2005; Gokmen et al., 2018). That being said, while policy conditions will not explicitly enter the estimation of effects, the average effect of IMF intervention and how long it persists can provide evidence as to whether these conditions are important.

2.2 Average Recoveries: An Ashenfelter Dip

Average growth paths around IMF short-term loans indicate country growth rates recover rapidly following IMF intervention. This is a critical first step towards understanding the empirical regularities of the "treatment" variable and the challenges posed by the setting. Figure 3 depicts this pattern in an unconditional event study, methodologically following Bruno and Easterly (1998), Gourinchas and Obstfeld (2012) and Kuruc (2018). The exercise is simple: the unconditional mean (or median) growth rate at each horizon from the start of an IMF program is plotted ($t = -2$, for example, is the average growth rate 2 years prior to recieving a program).[10] Mean and median growth rates are falling going into a program and recover rapidly following it. While the mean growth rate begins to slip again 4 years following a program, the median suggest this is driven by outliers. On the surface, this summary plot is quite positive in terms of IMF effectiveness, but it becomes more complicated given the economic conditions that are commonly associated with the start of IMF programs.

Countries and the IMF may agree to begin programs at country-specific low points, thus making estimation issues here very similar to that of Ashenfelter (1978). In the setting of job retraining programs, Ashenfelter (1978) makes the point that panel estimation becomes substantially more challenging with selection at individual troughs. There has been much subsequent work on this problem in the non-experimental evaluation literature of microeconometrics. Traditional panel data methods correct for level differences, not dynamic differences, making them ineffective in this setting; a comparison against growth rates before treatment is not a good counterfactual for what growth rates would have been afterwards. The panel and selection correction methods of past work primarily rely on level differences ("growth was slow") and not dynamic differences ("growth rates were falling") despite this being an important and robust feature of the data that is almost certainly important for accurately generating a counterfactual. This paper will account for both level and trend differences at the time of the crisis.

In this specific circumstance, controlling for the dynamic path on its own is likely to be inadequate. Country growth slides are not typically associated with such rapid reversals universally

[10] Studying financial crises in general Gourinchas and Obstfeld (2012) instead use HP-filtered output and include country fixed effects in their formalization, with a slightly different objective. I do not detrend or remove fixed effects following Bruno and Easterly (1998) and Kuruc (2018) who plot completely unconditional moments surrounding their event of interest in order to formalize interpretable stylized facts.

(Pritchett, 2000). These appear to be growth slides culminating in some event that marks the bottom of the trough and start to recovery. I will verify this concern by documenting a similar pattern surrounds financial crises and ultimately end up conditioning on these events.

2.3 Financial Crises

I define financial crises as the years compiled in Valencia and Laeven (2012) who systematically provide start dates for three types of crises: banking, currency and sovereign debt crises. These are defined in the following way.

- **Banking Crisis (N=134)** Years with significant bank runs, losses or liquidations and banking policy intervention.[11]
- **Currency Crisis (N=199):** 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 Crisis (N=64):** Years with sovereign default or debt rescheduling.

There are many crises of each type in the sample period of interest (1975-2010). These categories are not mutually exclusive, so the number of country-years experiencing a financial crisis is less than the sum of the three crisis types; there are 372 unique crises as I have defined them. This data is consistent with the observations of Reinhart and Rogoff (2009)—financial crises were not rare events outside of developed nations in recent history.

It is now possible to gather evidence on whether the pattern observed in Figure 3 is likely to be driven by IMF loans going disproportionately into situations that would have recovered regardless. Figure 4 runs a similar unconditional event study to Figure 3 but uses the onset of a financial crisis (pooling all types) as the event of interest. Financial crises share the sharp-"V" pattern observed surrounding IMF loans, in fact it is even more extreme. This fast recovery in *rates* is not inconsistent with the conventional wisdom that the *level* effects of financial crises are long-lived (Reinhart and Rogoff, 2009). Given the low growth rates leading into the crash, economies will remain below trend as long as growth rates are not substantially higher in the post-period.

The evidence here suggests that conditioning on the experience of an acute crisis will be necessary to construct a plausible counterfactual that shares the recovery properties of the events of interest. To this end I will study the recoveries of financial crises treated by an IMF program

[11] This definition is admittedly more qualitative than the other two, but I defer to the definitions chosen by the original authors throughout the paper.

and use untreated financial crises to approximate the recovery dynamics that would have arisen otherwise.

2.4 Average Outcomes for Treated Vs. Untreated Financial Crises

I define a *treated* crisis as a financial crisis that receives a new IMF program in the year of, or year following, the onset of a financial crisis. An *untreated* crisis is a financial crisis that does not receive an IMF loan in the year of, or year following the onset of their crisis. I also eliminate crises that are at the tail end of IMF programs started prior to their crisis. There are not many of these, but they are partially treated in a way that the estimator is not well-suited to handle. For "twin crises," I discard—as a unique observation—any crisis that comes in the year following a separately dated crisis; if a sovereign debt crisis happens in the year following a currency crisis, I only treat the currency crisis as an event of interest. The debt crisis in the following year is considered a negative outcome of this unfolding event.

With these definitions, I can plot the average path of treated versus untreated crises as this will ultimately be the comparison of interest. Figure 5 plots the average path of these crises and reveals two important features in this sample. First, recoveries begin a year earlier for crises treated by the IMF. While this is not a sophisticated comparison, this pattern will end up being robust to more careful comparisons and drive the main results of the paper (Figure 2). Second, the crises the IMF lends to have different pre-period trends than the untreated crises; the IMF is lending into crises that have more extreme crashes. The synthetic control method, described in the next section, will attempt to appropriately condition on the pre-growth path in order to correct for this difference and produce a more convincing comparison.

3 Empirical Strategy: Constructing Synthetic Controls

In this section the synthetic control method is introduced. As it is a newly developed estimator I take some time to discuss details and properties of the method. I determine the specification for the main analysis by following the method of Dube and Zipperer (2015) who take advantage of the untreated events as "placebos." The placebos are used to find a matching procedure that constructs synthetic controls which can best predict post-crisis recoveries in a group where it is known the coefficient of interest is zero, a "training sample" of sorts. This exercise indicates it is important to restrict the synthetic control to only draw from untreated crises that have similar values for the pre-crisis target variables of interest (as opposed to constructing averages from *any* untreated crisis). Conditional on trimming the potential control units, a simple procedure of just targeting pre-period growth rates is effective.

3.1 Synthetic Control Method

This section draws heavily on both Abadie, Diamond and Hainmueller (2010) and Dube and Zipperer (2015) to make explicit the details of the SCM. The main idea is that a convex combination of untreated, but similar, crises can serve as an estimate of the counterfactual outcome for the treated unit of interest. There are various advantages, especially in this setting, of this estimator relative to standard regression—including vector autoregression (VAR)—methods. First, globally linear assumptions can be easily relaxed. While it is true that a fully saturated regression model can replicate this feature, it involves estimating many parameters with little data. Second, by forecasting a counterfactual at each horizon it shares the benefits of the local projection method developed in Jordà (2005); the mean difference between treated and their synthetics can be directly compared at each horizon to produce a non-parametric dynamic estimate of the outcome of interest.

The primary advantage, however, comes in having a data-driven approach to reweight the control group in a way that increases its similarities with the treated observations. Heckman et al. (1998) argue—in an empirical setting with a similar pre-program dip coupled with selection into treatment—that perhaps the most serious problem of non-experimental econometric techniques is using units with near-zero probabilities of being treated to construct counterfactuals for the treated. Billmeier and Nannicini (2009), studying trade liberalizations, show this problem is likely as severe in many cross-country studies. Abadie, Diamond and Hainmueller (2010) originally restrict the synthetic control to convex combinations (weights between (0,1)) of untreated units as a way to prevent serious extrapolation. I am slightly more restrictive; the cost of further restricting matches to be drawn only from "local" crises (those that have sufficiently similar values for the pre-crisis variables the SCM is targeted to match) allows me to relax a global linearity assumption used in papers employing the SCM.[12] This trade-off is well-known and originally motivated the use of matching estimators generally (Rubin, 1977; Dehejia and Wahba, 2002). I extend this logic to the case of synthetic controls.

To formalize the discussion, suppose the data generating process can be written as a mean-zero

[12]The difference with past papers is described in detail in Appendix 2.

forecasting equation as in (1):

$$y_{i,t} = F^t(X_{i,0}, y_{i,0}, y_{i,-1} \dots y_{i,-\infty}) + \theta_t IMF_i + u_{i,t} \tag{1}$$

Here t is normalized at 0 to be the date of the crisis with $t > 0$ being the time period following. Notice, the forecasting data generating process is t specific but only relies on inputs known at time 0; these are ex-ante forecasts at the time of the crisis. IMF_i has no time-subscript because each i is a crisis, such as Kenya 1992, *not a country*. Each crisis is either treated or untreated as a fixed characteristic and the differential evolution between these types is analyzed. The treatment effect, θ_t, varies with the horizon and I impose $\theta_k = 0 \; \forall k \leq 0$; there is no IMF effect before the IMF has intervened. The vector X is some set of characteristics about the financial crisis that may affect the recovery dynamics, such as the level of government debt at the time of crisis, and $y_{i,-k}$ is the outcome variable k periods prior to the onset of the crisis. The treatment effect is linear and separable. This linearity can be easily relaxed but I maintain it for ease of notation here.

The two assumptions necessary for constructing a mean-zero counterfactual by synthetic controls in this setting are the following.

Assumption 1. *For all treated observations i, there exists a local linear approximation of $F^t(X_{i,0}, y_{i,0}, y_{i,-1}, \dots y_{i,-\infty})$ in a neighborohod around $(X_{i,0}, y_{i,0}, y_{i,-1}, \dots y_{i,-\infty})$ denoted $\hat{F}^t_i()$*

Assumption 2. *In this neighborhood of i, there exists J_i potential controls and a vector of weights λ_i^j such that*

$$\sum_{j \in J_i} \lambda_i^j X_{j,0} = X_{i,0} \qquad\qquad \sum_{j \in J_i} \lambda_i^j y_{j,k} = y_{i,k} \;\; \forall \, k < 0$$

Assumption 1 just says there is a first-order linear approximation of the data generating process at each point. If it is continuous and differentiable this will be satisfied by Taylor's thereom. In practice, as I will stray from an infinitesimally small neighborhood there is an implicit assumption that I have a "good" linear approximation as the space of interest expands. Assumption 2 says that within some neighborhood there exists a convex combination of untreated crises that can match the $X_{i,0}$ and all $y_{i,-k}$.[13]

[13] Notice these two assumptions will push against one another in practice; the smaller I define a neighborhood the more reasonable Assumption 1 becomes, but the harder it is to satisfy Assumption 2.

Denote the counterfactual under $IMF_i = 0$ as $y^c_{i,t}$. If these are met then the following result obtains.[14]

$$\sum_j \lambda_i^j y_{j,t} = y^c_{i,t} + e_{i,J,t} \Rightarrow$$
$$y_{i,t} - \sum_j \lambda_i^j y_{j,t} = \theta_t + e_{i,J,t}$$

The convex combinations of outcomes is equal to the counterfactual for i, $y^c_{i,t}$, plus some mean-zero error, $e_{i,t,J}$. The error, $e_{i,J,t}$ depends on disturbances to the treated unit and all J observations plus an error arising from the linear approximation. This implies that subtracting this convex combination from the actual outcomes is a mean-zero estimator for the effect of interest, θ_t. This non-parametric forecasting approach shares advantages of the widely used local projection method (Jordà, 2005). For each horizon, t, the average effect estimate is simply the average of the differences between the treated and their synthetic controls at that horizon. No structure is imposed on the dynamic shape of the effects.

Now, let Z_i be defined as a row vector of the variables to be targeted for the treated observation. $Z_{\mathbb{J}}$ is a matrix where each row contains the same variables for one of the J potential controls and Λ^i is a column vector of weights across these untreated observations. The weights, Λ^i, (and corresponding synthetic controls) will be generated by solving the following minimization problem.

$$\Lambda^i = \underset{l \in [0,1]^J}{\operatorname{argmin}} (l' Z_{\mathbb{J}} - Z_i)(l' Z_{\mathbb{J}} - Z_i)'$$
$$\text{subject to} \sum_{j \in J} l_j = 1$$

Prior to solving this problem the data is restricted to the neighborhood of interest which is why the local consideration previously disccused is absent when this problem is defined.[15]

3.2 Utilizing Placebo Data

The estimator presented leaves two practical choices to make: what contemporaneous conditions and lags ($X_{i,0}, y_{i,-k}$) to include as targets for matching and how to choose the neighborhood of

[14] See Appendix 2 for details.

[15] This sum of squared errors can be generalized to have non-uniform weights across variables if generating good matches for certain targeted variables is seen to be more important than others.

untreated units considered by the minimization problem. This subsection describes how to utilize the *untreated* units to determine an appropriate matching procedure.

All untreated units, by definition, are such that $\theta_t IMF_i = 0$. If a synthetic control is created for some untreated crisis *using the other untreated crises as the potential controls* it should return an estimated of difference of zero, in expectation. There is no “IMF effect” in countries not treated by the IMF, and this should be reflected in the synthetic controls created from other untreated units.

More importantly, for a given specification, an empirical distribution of forecast errors can be created by repeating this process for all untreated crises. To be explicit, the pseudo-algorithm is as follows.

(1) Choose a set of pre-crisis target variables to match between treated and synthetic and a rule for defining a neighborhood of “qualitatively similar” crises.

(2) For each untreated crisis, j:

   a. Remove crisis j from the pool of potential control observations, as if it is treated, leaving crises $-j$ as potential controls to create matches for j.

   b. Use the specification chosen in step (1) to generate a synthetic control from the set of $-j$ by minimizing the SSE over the qualitatively similar crises.

   c. Track and store the outcome differences in post-period between j and its synthetic control.

   d. Place j back in the set of potential control observations.

(3) Analyze the distribution of forecast errors: confirm mean-zero, examine empirical variance, $\sigma^2_{placebo}$.

Generating a mean-zero forecast here is trivial, but different specifications will lead to very different emprical variances, $\sigma^2_{placebo}$. As an example, suppose specification R generates synthetic controls by randomly drawing among all potential control countries while specification M targets pre-crisis growth rates as variables to match in its construction of synthetics. In this placebo exercise, the same data utimately make up both the treated *and* control group once this has been repeated over all untreated crises. Therefore, the synthetic recoveries even under R should be the same as the treated recoveries, on average. If M is in reality a better specification for generating synthetic controls it will not show up as a smaller absolute value for the mean difference. But if specification M makes better individual predictions it will be reflected in a forecast error distribution that is tighter around zero. The second moment of these distributions will be used to distinguish between potential specfications.

While this exercise appears to be informative only about precision, it can also provide evidence concerning potential omitted variables. To see this, suppose there is some specification O that omits variable x_{omit} that may be correlated with IMF lending. An omitted variable bias only arises if the omitted variable is correlated with the forecast errors generated by specifcation O. If x_{omit} is included as a target variable and it does not lead to better forecasts in the placebo exercise then it would appear *not* to have marginal predictive power.[16] While this argument provides good reason to choose the procedure that minimizes $\sigma^2_{placebo}$, I will present robustness checks to show the main results remain similar when explicitly targeting variables that do not appear to add value in the placebeo exercise.

Aside from guiding the specification of the estimator, this empirical distribution of placebo errors is used to construct standard errors for the main analysis. Using these errors is necessary because the asymptotic variance of the SCM has not been characterized (Dube and Zipperer, 2015). This is one of the main drawbacks of the method. However, these placebo runs give us an estimate for σ^2_e, the variance of the errors due to synthetic controls being unable to perfectly forecast recoveries even in the absence of a treatment effect. Intuitively, performing inference in this way leverages the fact there is a subsample for which the null-hypothesis is true. Knowing the mean—assumed to be zero—and the variance of errors for a group in which the null is true allows for constructing the likelihood of all possible outcomes under a normal distribution.[17]

3.3 Main Specification

After experimenting with the placebo exercises the main specification settled on is a relatively simple one. Growth rates for periods $t \in \{-5, -4, \ldots 0\}$ are the targeted variables for matching. I experiment with adding three potential contemporaneous variables to target: government debt to GDP ratios, inflation rates and current account deficits at the time of crisis. After controlling for growth rates, additionally targeting these produces forecasts that are no better than omitting them according to the placebo runs. These will be included in robustness checks to show the results are not driven by omitting any of these. Likewise, overfitting on growth rates does not appear to be an issue; matching on subsets of the available pre-period growth rates fails to produce forecasts as good as using all available pre-years.

[16]In practice it can actually lead to lower quality forecasts if matching on x_{omit} reduces O's ability to match the variables it was previously matching that are important for predicting outcomes.

[17]The central limit theorem gives us normality here since the distribution of interest is one of sample means. This is not an assumption about the distribution of errors for individual observations which does in fact have non-normal properties.

In terms of restricting the neighborhood of potential matches I rely both on (i) using only the same crisis types (ie. Banking/Currency/Debt) and (ii) crises within a ± 7 percentage point "growth boundary" in the pre-period. Figure 10 provides a stylized example to make explicit these restrictions. The goal in this illustration is to create a synthetic control for a treated banking crisis (the thick olive colored line in this figure) from the various potential control countries (the dashed lines B1-B3,C1,D1). The restrictions imply I would only use a subset of them to try and replicate the pre-crisis trends. C1 and D1 represent a currency and debt crisis, respectively. As the treated crisis is a banking crisis, and I require banking synthetics only draw from banking controls, I will discard these observations. Further, crises must fall within the "Growth Boundaries" labelled on this figure (7 percentage points around each growth rate). Crisis B3 is a banking crisis, but falls outside of this window in period $t = -1$ and so will be discarded as well.

The use of growth boundaries eliminates convex combinations for mild crises to come from some weighted average of severe collapses and episodes where growth remains high throughout the pre-period. Allowing this to happen makes for bad forecasts in the placebo exercises that are substantially improved by adding relatively weak restrictions like the ones here.[18] Likewise, if the specification ignores the fact that currency crises are different from debt crises which are different from banking crises there is a significant reduction in the quality of post-period predictions in the placebo exercises. This further illustrates why it is critical to condition on the general fact that country-years experiencing a financial crisis have recoveries that differ from what would be expected from only conditioning on pre-period growth.

4 Results

This section presents the results of the main analysis, discusses the characteristics of the synthetic controls used to generate those results, performs robustness checks and provides evidence the main effect is not driven by unobservable selection. I find financial crises that receive an IMF program have significantly faster recoveries than their synthetic counterparts that do not. The effect is large: point estimates in the main specifcation suggest 2 years following a crisis the treated observations have GDPs that are nearly 2 percentage points larger (in levels) than if the recovery had instead followed the growth rates of the synthetic control. This short-run effect is robust to a battery of specification changes and comes from synthetic controls that draw from a wide range of untreated crises. In the medium-run I find little evidence the effect persists: 5 years out standard errors are

[18] I consider 7 percentage points, the band chosen for the main analysis, to be a wide window.

extremely wide and alternative specifications do not provide a consistently positive estimate at this horizon. The dynamic path is consistent with the IMF stabilizing economies but not necessarily changing long-run potential GDP. I conclude this section by studying whether this estimate could be the result of factors unobservable to the econometrician but correlated with IMF lending. Using publicly available historical forecasts from the time of these crises I provide evidence the IMF is not able to predict which crises will recover faster once relatively few variables are conditioned on. If they are unable to predict differential recoveries it is unlikely their lending is correlated with the unobservable factors that produce them.

4.1 Main Results

Figure 2 presents the results from the main specification detailed in the subsection 3.3.[19] Recoveries in treated crises result in economies that are on average larger for 5 years following the start of a financial crisis. Horizon 2 is point-wise significant at the 5% level and the null hypothesis that the entire path is zero can be rejected at the 1% level. As the null is "crises treated by the IMF have recoveries that are no different than untreated crises" it is this joint-significance that is relevant.[20] These level differences come from an underlying comparison of growth rates between the treated and synthetic control directly. Figure 7 plots the differences in growth dynamics. The treated crises have much higher growth rates in the first 2 years of recovery, but in years 3-5 the synthetic control has higher growth in a period of catch-up; this accounts for the decreasing differences in the later periods of Figure 2.

The SCM compares the growth rates between *each* treated and its synthetic control which makes it is easy to verify that outliers are not driving this mean difference. This is a common concern in regressions using cross-country growth rates as these data are known to have thick tails with many extreme values.[21] Figure 8 plots the density of level differences after 2 years (chosen because it is the largest point estimate in Figure 2). The mean effect size, about 1.9 percent, comes from an underlying distribution with a large fraction of its mass above zero, not from a few large outliers. Most crises that are treated with an IMF loan beat their synthetic counterpart.[22] The

[19] 5 years of pre-period growth rates are the targeted variables. Potential control crises come from a pool within the 7 percentage point growth bands and that are of the same crisis-type (ie, banking, currency, or debt).

[20] Since these estimates do not come from a single regression I cannot run a traditional F-test. Instead I compute joint significance using a Hotelling T^2 test which extends a univariate t-test to testing the probability that a multivariate distribution has the zero vector as its mean.

[21] See Burnside and Dollar (2000), for example. These authors carefully consider observations they find to lie 4-5 standard deviations away from the mass of the distribution in their data.

[22] Note that this is *not* a distribution of the average effect, which would depict that the estimated effect is not

large variance of this distribution should be noted as an important feature of the data. This is not inherently a problem. As in any statistical analysis, some treated observations do better than their counterfactual, some worse, and the distribution is analyzed to estimate the parameters of the data generating process. However, the large variance here creates power problems for any sub-sample analyses that cut the data into smaller bins.

4.2 Analyzing the Quality of the Synthetic Controls

The results rely on creating "similar" crises to serve as synthetic controls. Here, I investigate the degree to which the SCM has been successful at this. Most critically, according to the placebo exercises, the average growth path heading into crises should be similar to generate good growth forecasts in the post-period. As can be seen in the pre-period of Figure 7, the synthetic controls are fairly successful at replicating the pre-crisis growth path. Recall, however, that this is by construction. I minimize the sum of squared errors over this pre-crisis path to create the synthetic controls. While this figure only shows the average growth rate for each group, Appendix 3 shows the full distribution of matches achieved by the minimization problem. Some crises and their synthetic controls have markedly different pre-period growth rates. In some cases only one or two untreated crises survive the trimming process (being within the growth bands and of the same crisis type). With so few potential controls avaible locally it is difficult to replicate the pre-period growth rates. Since these misses happen in both directions the average pre-period growth rates—reported in Figure 7—continue to be comparable and so it is not obvious this would bias the results one way or another. Nonetheless, as a robustness exercise in the next subsection I will test whether these badly matched crises drive the pattern seen in Figure 2. They do not.

A second important point to make about the synthetic controls is that they draw from a large fraction of the entire pool of untreated countries. It is a feature, not a bug, of the SCM that it will oversample from crises that look more like treated observations. It would be concerning, however, if a few crises had extreme representation. Under the specification chosen nearly 50% of untreated crises contribute a total weight of at least 0.5 to the synthetic group (calculated by summing across the weights of all synthetics). As a point of reference, under random assignment each control would account for around 0.65 total synthetic controls.[23] Only 30% of untreated observations are

significant. This is the distribution of outcome differences used to conclude the *mean* of the data generating process for differences is (highly) unlikely to be zero.

[23]There are 157 untreated crises for 101 synthetic controls that need to be generated so each would get a weight less than 1 even in this case.

not used at all and the maximum contribution of any untreated unit is one with a total weight of 3.1. Appendix 3 discusses in detail this assignment and displays the full set of weights. The wide range of untreated crises that contribute to the synthetic group provides confirmation the positive estimate is not an artifact of utilizing only a small fraction of the total variation in controls.

Finally, these crises can be compared to their synthetic controls for untargeted characteristics of interest. Table 2 displays such comparisons. Along contemporaneous characteristics in the year of the financial crisis (external debt, current account deficit, terms of trade), no consistent story arises about one group doing observably "better" along these measures. While debt levels are lower in the treated observations and terms of trade are stronger (defined here as export price over import price), the current account deficit appears worse. The IMF officially is tasked with helping countries manage balance of payments problems, but it appears (conditional on growth rates) that this doesn't necessarily come along with other issues. In terms of structural characteristics, the economies have a similar level of government involvement (measured as the share of GDP in the Penn World Tables accounted for by government consumption). However, the countries receiving financing are poorer per person, and smaller. On average, the crises getting financing had GDP per capita (in PPP terms) that was 42% of the cross-sectional average for their respective year relative to control crises that had GDP levels 72% of the average. While missing on some of these metrics is not ideal, recall that the placebo exercises indicated that improving fit here (by explicitly including one or more of these variables as target variables along with growth rates) does not necesssarily result in better post-period predictions on that sample. I will provide evidence in the following subsection that targeting the contemporaneous variables that may be of interest to the IMF does not change the results.

4.3 Robustness to Alternative Matching

This subsection performs a battery of specification modifications and shows that the main result is robust. Figure 9 summarizes the results of these exercises by reporting the original impulse response along with impulse responses from alternative specifications.

I begin by varying the data the SCM uses for targeting. Before adding new variables to the matching process, I first verify the results hold under growth rates from other sources. This alternative run uses the World Bank's World Development Indicators as opposed to the Penn World Tables. Johnson et al. (2013) argue checks like this are necessary when using growth data given the large suspected errors in these poorly measured aggregates.[24] I then experiment with having the SCM additionally target inflation, external debt to GDP ratios and current account deficits in the year of the crisis. Each is employed one at a time as an additional target variable along with the original specification of just lagged growth. These variables are transparent and likely used by the IMF to determine lending decisions and, despite evidence to the contrary in the placebo exercises, may be correlated with recovery conditional on growth paths.

I then return to the main, "growth only," specification and alter the exact structure of the SCM. First, I vary the growth boundaries from being ± 7 to ± 6 and 8, respectively. Finally, I make sure the worst matches are not driving the results by discarding the 10% of the sample with the biggest errors in pre-period match quality.

The main estimate for the first 3 years of recovery is remarkably stable. Interestingly, the stability of this coefficient is not driven by the SCM generating the same synthetic controls with small changes in the specification. If that were the case the later periods would be stable as well. These different specifications are reweighting the control sample enough to drastically change the later horizon averages, but this new identifying variation continues to tell a similar story in the early phase of recovery.[25]

4.4 Identification Check Using Historical Forecasts

Up to this point I have shown that the SCM is an effective method to address selection on observables, I will now present evidence that selection on unobservables is unlikely to contaminate the results. Recall this has two sides: countries demand loans and the IMF chooses whether or not to supply them. I begin by studying supply side selection using historical forecasts produced by the IMF. The main idea is as follows. If, conditional on the inputs to the synthetic control specification, IMF forecasts have additional predictive power as to how the recovery will progress then it is likely there is an important omitted variable (or variables) that could be driving the estimated effects. I will present evidence this is not the case.

I do this by estimating a simple regression model as in Equation (2):

$$Y_{i,t} = \gamma_0 + \gamma^f Y_{i,t}^f + \gamma_1 X_i^{SCM} + \xi_{i,1} \tag{2}$$

[24]They show just using different vintages of the Penn World Tables can reverse signs on significant coefficients in growth regressions.

[25]These late horizon differences are well within the one standard error bands of Figure 2 so I do not want to read into the point estimate differences.

The regression in (2) fits a linear model to predict cumulative growth rates at some horizons, t, following a crisis. Here I will show evidence for growth rates one year following the crisis, $Y_{i,1}$, two years cumulative growth $Y_{i,2}$ and three years, $Y_{i,3}$. For exposition, consider the case where Equation (2) is used to predict $Y_{i,1}$. The regression includes, as independent variables, the IMF's one-year ahead forecast at the date of the crisis, $Y_{i,1}^{f}$, and some subset of the variables used as matches in the main specification, X_i^{SCM}. If it is the case that both (a) the IMF makes mean-zero, rational, forecasts and (b) there is some information available to the IMF not included in X_i^{SCM} that is informative about future growth, then $\gamma^f \approx 1$. In an extreme case where the IMF has *no* additional predictive power once the variables in X^{SCM} are accounted for, then $\gamma^f \approx 0$.

One complication with using these forecasts is that it is not clear whether the IMF is forecasting their own effect for treated crises. In the year of the crisis these forecasters may know, whether it has been offically announced or not, the likelihood there will be an IMF program into a specific crisis. It is likely the staff at the IMF believe, or at least have incentives to project that they believe, crises the organization intervenes in will recover quickly. I want to avoid comparing forecasted recoveries of crises with IMF programs to crises without and then concluding the differences found were in fact forecastable if it only arises for this reason. To avoid this I estimate these effects separately for regressions on both the treated and untreated samples.

Table 3 shows that the estimates for γ^f are close to zero in the 6 cases considered. These regressions are relatively conservative to leave some variation that may be plausibly forecastable. With only 65 treated observations (the forecasts are only available starting in 1990) a regression with too many covariates would make obtaining any precision difficult. Two lagged years of growth rates (the year of the crisis and 5 years prior to capture growth rates at the end points) rather than the entire pre-period path will be included in X^{SCM} along with dummy variables for crisis types. Columns 1, 3 and 5 run these regressions for the different forecasting horizons on the treated sample. IMF forecasts are in fact *negatively* correlated with residualized recoveries one-year following the onset of the crisis (coefficient of -0.48). Over two and three year horizons small positive estimates are obtained (≈ 0.25), but even one-standard error confidence bands would continue to overlap zero. Columns 2, 4 and 6 run this same exercise on the untreated sample. Here, the coefficients on IMF forecasts are estimated to be nearly 0 (0.07), moderately positive (0.44), and severely negative (-0.66) over the respective horizons. While this exercise is not nearly well-powered enough to provide conclusive evidence the IMF has no additional predictive power (for that stronger conclusion the results would need to be tightly estimated zeros, not just confidence intervals that include zero), it does much to address these concerns.

There is not an analogous test for whether unobservable country differences drive the estimated positive effect, but given the results here it must be the case that these differences are unobservable to the IMF as well. I additionally show in the following section that countries with the weakest institutions and economic policy measures appear to account for much of the positive main result. This is precisely the opposite of what would be expected if unaccounted for country differences could explain the estimated differences. While the SCM was designed explicitly to address observable differences between crises it appears to have done enough to purge the environment of selection issues entirely. I would argue for these reasons the results in Figure (2) should be interpretted as causal evidence.

5 Transmission & Heterogeneity

In this section I will attempt to examine, both directly and indirectly, the transmission mechanisms for the positive output effects that have been estimated in Figure (2). Using the SCM specification designed to create counterfactual growth rates is not guaranteed to produce good counterfactual evolutions of other variables.[26] For this reason directly disentangling the effects by comparing the treated and synthetic controls along other dimensions can only reveal so much. One key difference that does arise is that government spending increases more in the treated countries than would be predicted by their synthetic controls, an intuitive "first-stage" that might be expected. I will then leverage heterogeneity in the treatment effects with the objective to verify both the credibility of the results and indirectly gather evidence on the transmission mechanism.

5.1 Government Spending Response

The most obvious place to look for evidence on the transmission of this effect is the evolution of government spending, G. This is analogous to the first-stage in Kraay (2012, 2014). His work runs a similar analysis using World Bank lending decisions to construct an instrument for government spending in order to estimate the associated multiplier. Importantly for the case here, Table 2 verified that government spending (as a share of GDP) was nearly identical in the treated and

[26] Regressions that estimate an outcome for other variables re-estimate all parameters of the model and so implicitly use different observations as counterfactuals. Here this issue becomes more acute. I would need to re-generate all synthetic controls and it is no longer clear that this is the relevant comparison. This new group of synthetics may not even predict the same post-period differences in the main outcome, so it is not really "explaining" why it would happen.

synthetics making the evolution of G for the synthetic group a reasonable counterfactual.

Figure 11 tracks the percent increase in government spending over the four years following the onset of the crisis. The increase for the treated group is substantially larger than for the synthetics in these years. Three years out G is nearly 3 percent larger than in the synthetic controls. The IMF grants these loans to country governments and leaves the disbursement to them. The fact these loans expand the government's budget in these years makes it unsurprising an increase in G is observed.[27]

While not surprising, this is an important fact in light of typical business cycle policy in low-income countries that is over-whelmingly procyclical (Frankel, Vegh and Vuletin, 2013). One leading hypothesis for why fiscal policy would be conducted in a way that intensifies business cycles is imperfect credit availability in developing countries (Caballero and Krishnamurthy, 2004; Riascos and Vegh, 2003). If this is the case, it is not surprising that access to IMF financing—a direct relaxation of credit constraints—would help induce government spending increases. Further, I show in the following subsection the effects are estimated to be largest in the environments Frankel, Vegh and Vuletin (2013) argue are the ones in which procyclical fiscal policy is most likely.

5.2 Heterogeneity

In this subsection I study which countries drive the main results in order to understand what potential channels the effects could be coming through. The specific exercise will be to see which, if any, country characteristics predict how much a treated unit beats its specific synthetic control by (recall the full distribution of differences was presented in Figure 8, this is the variation I use). I find and discuss two major dimensions of heterogeneity: along institutional quality and between exchange rate regimes.

Using the World Bank's Country Policy and Institutional Assessment (CPIA), which ranks 16 policy, corruption and institutional measures, I find that effect sizes are inversely correlated with state capacity. Figure 12 (a) displays the scatter plot of effects along this dimension with a simple linear regression fit through it. I follow the literature studying "Fragile States" (countres with the weakest state capacity) and take the average among the 16 underlying indicators as an

[27]Evidence on other transmission mechanisms is unfortunatley far noisier. I find some suggestive evidence extreme collapses in exchange rates are less likely with an IMF loan, but essentially no differences show up in consumption, investment or net exports. This does not imply differences do not exist, but the SCM as it has been specified to estimate output effects just fails to generate any other clear results.

overall measure of state capacity (IMF, 2018). This effect is in stark contrast to highly influential work in the literature studying foreign development assistance and medium to long run economic growth. Burnside and Dollar (2000) famously found that foreign aid is only effective in promoting growth in countries with good policy. Their result spawned a large subsequent literature verifying and further disentangling where aid can be useful. While the effects here are specifically for IMF financing—and so do not directly contradict this result—it is nonetheless interesting to see the opposite correlation for IMF aid. This effect manifests itself geographically in unsurprising ways. Table A1 shows the average effect size by region and finds Africa and Small Island Economies to have the largest estimated effects from IMF lending.[28][29] These are regions that typically suffer from problems of weak state capacity (IMF, 2018).

The inverse correlation of effect sizes with state capacity has two implications. First, it makes a story about selection on unobservable country characteristics much less threatening. If the concern was that only countries organized to fight the crisis and conduct counter cyclical fiscal policy, for example, were the ones even applying for financing the correlation of effects with state capacity would almost certainly be positive (and at the very least non-negative).

Second, this correlation is consistent with some combination of the two most obvious channels of IMF financing being active. For one, if generating outside sources of non-IMF liquidity during a crisis is especially difficult for low-capacity countries then these are the places the IMF has an opportunity to make a substantial difference.[30] My own prior work further verifies the possibility of this channel (Kuruc, 2018). In that paper, and replicated in Figure A2, I show that for Fragile States[31] the start of an IMF program is associated with large increases in outside development financing.[32] This is consistent with informal evidence from authorities in low capacity states who claim it is substantially easier to generate outside financing when the IMF plays an active role in fiscal oversight (IMF, 2018). Another possibility is that the IMF in fact has provided useful policy advice in places with poor measures of policy to begin with. There are arguments that both in

[28]These numbers are going to be estimated with large error bands that are complicated to compute (recall the placebo inference discussion in Section 3). For this reason I just present the averages and read into them with caution.

[29]Interestingly, the SCM employed estimates that countries in Asia, on average, did far worse following IMF lending than they would have without it. Critics of the IMF, and eventually the IMF itself, point to Asia (and specifically the Asian financial crisis) as an episode where the IMF made major policy mistakes (Stiglitz, 2002; IMF, 2001). The results here constitute addtional empirical evidence for this narrative.

[30]The results in (Kraay, 2012) that World Bank lending makes up a substantial fraction of total government financing in very low-income countries makes this story seems plausible.

[31]Fragile States are binarily defined as country-years with a CPIA score below a certain threshold

[32]In that paper no distinction between financial crises and "normal times" is made, and no identification strategy is employed. It just an unconditional event study, but it has a stark enough break with past trends that it likely provides some evidence on causality.

Asia (Stiglitz, 2002), and more recently Europe (Blanchard and Leigh, 2013), that the IMF did not succeed in promoting good policy. But this does not necessarily need to be universal, especially if policy would have been obviously bad in absence of IMF involvement. I am not able to separately identify the respective roles of these channels. However, the fact both channels are likely operative in the countries with the lowest state capacity and that the estimated effects are largest in these places lends some credibility to a causal interpretation aside from being generally interesting for development reasons.

An additional source of heterogeneity comes from the exchange rate regime of the treated countries. Farhi and Werning (2016) show that under fixed exchange rates transfers from outside sources (as opposed to internally financed government spending) can have large multipliers. Figure 12 (b) shows the average effect size by a measure of exchange rate flexibility from Ilzetzki, Reinhart and Rogoff (2018) (higher values implying more flexibility). There is little pattern other than the large jump for fixed exchange rates (a value of 1 on the x-axis in this figure). This result is consistent with theory and provides evidence in support of it.

The results in this section help corroborate the main results and further advance a few alternative lines of inquiry with three key pieces of evidence. First, government spending increases more in countries that are treated than would be predicted by their counterparts; an obvious first-stage that would be expected by most mechanisms posited. Second, effects are largest in countries with low state-capacity; I argue these are places most likely to benefit from the services an IMF program offers. Third, and finally, effects are largest in fixed exchange regimes; this is consistent with theoretical work on the size of multipliers for outside sources of financing.

6 Conclusion

As recent history has made clear, financial crises are not merely events of the past (Reinhart and Rogoff, 2009). Understanding whether existing international structures designed to combat these crises, and seek macroeconmic stabilization generally, is critically important for the design of future policy. Despite being the central pillar of international coordination towards these goals, there is little convincing evidence that actions taken by the International Monetary Fund have been effective.

In this paper I bring new empirical evidence regarding the output effects of IMF involvement in macro-crises. Looking within the sample of financial crises directly and using a new estimator, I find IMF lending to be associated with significantly faster recoveries than their estimated counterfactual. Making use of direct and indirect tests I show this does not appear to be driven by underlying selection biases and so provides causal evidence as to the effectiveness of these loans. Further corroborating the plausibility of a causal channel I show these effects are strongest in settings that there is good reason to anticipate large effects: countries with low-state capacity and countries with fixed exchange regimes appear to benefit the most.

The importance of these findings are clear, especially in the face of mixed (and even primarily negative) results that have so far dominated the literature. Liquidity during times of crisis appears to be useful and, in light of this evidence, it is important the IMF continue serving this unique role in global markets.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program." *Journal of the American Statistical Association*, 105(490): 493–505.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2015. "Comparative politics and the synthetic control method." *American Journal of Political Science*, 59(2): 495–510.

Abadie, Alberto, and Javier Gardeazabal. 2003. "The economic costs of conflict: A case study of the Basque Country." *American Economic Review*, 93(1): 113–132.

Acemoglu, Daron, Simon Johnson, Amir Kermani, James Kwak, and Todd Mitton. 2016. "The value of connections in turbulent times: Evidence from the United States." *Journal of Financial Economics*, 121(2): 368–391.

Ashenfelter, Orley. 1978. "Estimating the effect of training programs on earnings." *The Review of Economics and Statistics*, 47–57.

Auerbach, Alan J, and Yuriy Gorodnichenko. 2012. "Measuring the output responses to fiscal policy." *American Economic Journal: Economic Policy*, 4(2): 1–27.

Barro, Robert J, and Jong-Wha Lee. 2005. "IMF programs: Who is chosen and what are the effects?" *Journal of Monetary Economics*, 52(7): 1245–1269.

Bas, Muhammet A, and Randall W Stone. 2014. "Adverse selection and growth under IMF programs." *The Review of International Organizations*, 9(1): 1–28.

Billmeier, Andreas, and Tommaso Nannicini. 2009. "Trade openness and growth: Pursuing empirical glasnost." *IMF Staff Papers*, 56(3): 447–475.

Billmeier, Andreas, and Tommaso Nannicini. 2013. "Assessing economic liberalization episodes: A synthetic control approach." *Review of Economics and Statistics*, 95(3): 983–1001.

Blanchard, Olivier J, and Daniel Leigh. 2013. "Growth forecast errors and fiscal multipliers." *American Economic Review*, 103(3): 117–20.

Bordo, Michael D, and Anna J Schwartz. 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.

Bruno, Michael, and William Easterly. 1998. "Inflation crises and long-run growth." *Journal of Monetary Economics*, 41(1): 3–26.

Burnside, Craig, and David Dollar. 2000. "Aid, policies, and growth." *American Economic Review*, 90(4): 847–868.

Caballero, Ricardo J, and Arvind Krishnamurthy. 2004. "Fiscal policy and financial depth." National Bureau of Economic Research.

Cavallo, Eduardo, Sebastian Galiani, Ilan Noy, and Juan Pantano. 2013. "Catastrophic natural disasters and economic growth." *Review of Economics and Statistics*, 95(5): 1549–1561.

Clemens, Michael A, Steven Radelet, Rikhil R Bhavnani, and Samuel Bazzi. 2012. "Counting chickens when they hatch: Timing and the effects of aid on growth." *The Economic Journal*, 122(561): 590–617.

Deaton, Angus. 2010. "Instruments, randomization, and learning about development." *Journal of Economic Literature*, 48(2): 424–55.

Dehejia, Rajeev H, and Sadek Wahba. 2002. "Propensity score-matching methods for nonexperimental causal studies." *Review of Economics and statistics*, 84(1): 151–161.

Dube, Arindrajit, and Ben Zipperer. 2015. "Pooling multiple case studies using synthetic controls: An application to minimum wage policies."

Easterly, William. 2005. "What did structural adjustment adjust?: The association of policies and growth with repeated IMF and World Bank adjustment loans." *Journal of Development Economics*, 76(1): 1–22.

Farhi, Emmanuel, and Iván Werning. 2016. "Fiscal Multipliers: Liquidity Traps and Currency Unions." In *Handbook of Macroeconomics*. Vol. 2, 2417–2492. Elsevier.

Frankel, Jeffrey A, Carlos A Vegh, and Guillermo Vuletin. 2013. "On graduation from fiscal procyclicality." *Journal of Development Economics*, 100(1): 32–47.

Gokmen, Gunes, Tommaso Nannicini, Massimiliano Gaetano Onorato, and Chris Papageorgiou. 2018. "Policies in Hard Times: Assessing the Impact of Financial Crises on Structural Reforms."

Gourinchas, Pierre-Olivier, and Maurice Obstfeld. 2012. "Stories of the twentieth century for the twenty-first." *American Economic Journal: Macroeconomics*, 4(1): 226–65.

Gündüz, Yasemİn Bal. 2016. "The economic impact of short-term IMF engagement in low-income countries." *World Development*, 87: 30–49.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica*, 66(5): 1017–1098.

Hutchison, Michael. 2003. "A cure worse than the disease? Currency crises and the output costs of IMF-supported stabilization programs." In *Managing currency crises in emerging markets*. 321–360. University of Chicago Press.

Ilzetzki, Ethan, Carmen M Reinhart, and Kenneth S Rogoff. 2018. "Exchange arrangements entering the 21st century: which anchor will hold?" *Quarterly Journal of Economics*.

IMF. 2001. "Asian Financial crises: Origins, implications and solutions." IMF.

IMF. 2018. "The IMF and Fragile States." Independent Evaluation Office of the IMF.

Johnson, Simon, William Larson, Chris Papageorgiou, and Arvind Subramanian. 2013. "Is newer better? Penn World Table revisions and their impact on growth estimates." *Journal of Monetary Economics*, 60(2): 255–274.

Jordà, Òscar. 2005. "Estimation and inference of impulse responses by local projections." *American Economic Review*, 95(1): 161–182.

Kraay, Aart. 2012. "How large is the government spending multiplier? Evidence from World Bank lending." *The Quarterly Journal of Economics*, 127(2): 829–887.

Kraay, Aart. 2014. "Government spending multipliers in developing countries: evidence from lending by official creditors." *American Economic Journal: Macroeconomics*, 6(4): 170–208.

Kuruc, Kevin. 2018. "The IMF and Fragile States: Assessing Macroeconomic Outcomes." *Background Papers: Independent Evaluation Office of the IMF*.

Peri, Giovanni, and Vasil Yasenov. 2015. "The labor market effects of a refugee wave: Applying the synthetic control method to the Mariel boatlift." National Bureau of Economic Research.

Pritchett, Lant. 2000. "Understanding patterns of economic growth: searching for hills among plateaus, mountains, and plains." *The World Bank Economic Review*, 14(2): 221–250.

Ramey, Valerie A, and Sarah Zubairy. 2018. "Government spending multipliers in good times and in bad: evidence from US historical data." *Journal of Political Economy*, 126(2): 850–901.

Reinhart, Carmen M, and Christoph Trebesch. 2016. "The International Monetary Fund: 70 Years of Reinvention." *Journal of Economic Perspectives*, 30(1): 3–28.

Reinhart, Carmen M, and Kenneth S Rogoff. 2009. *This time is different: Eight centuries of financial folly.* princeton university press.

Riascos, Alvaro, and Carlos A Vegh. 2003. "Procyclical government spending in developing countries: The role of capital market imperfections." International Monetary Fund.

Rubin, Donald B. 1977. "Assignment to treatment group on the basis of a covariate." *Journal of Educational Statistics*, 2(1): 1–26.

Stiglitz, Joseph E. 2002. *Globalization and its Discontents.* Vol. 500, New York Norton.

Valencia, Fabian, and Luc Laeven. 2012. "Systemic Banking Crises Database: An Update." *IMF Working Papers*, 12(163).

Vreeland, James Raymond. 2003. *The IMF and Economic Development.* Cambridge University Press.

Table 1: Summary Statistics for Short-Term Loans

	Mean	Median	St. Dev	10%	90%	N
Growth Rates (%)	1.4	2.5	6.2	-5.6	7.4	461
Inflation (%)	45	10	231	1	61	392
External Debt (% GDP)	58	48	49	18	101	452
Terms of Trade	112	103	60	75	143	425
Current Account Balance (% GDP)	-5	-4	7	-13	5	470
Financial Crisis (Dummy)	.17	.	.	.	.	476
Size of Loan (% GDP)	2.4	1.4	2.8	0.4	5.6	476

Notes: Summary statistics at the time of initiation of IMF short-term lending programs. While conditions are not great, on average, there is a wide distribution for each of the indicators presented. *Source*: MONA Database & World Economic Outlook, IMF; World Development Indicators, World Bank; Valencia and Laeven (2012)

Table 2: Summary Statistics for Treated vs. Synthetics

Variable	Treated	Synthetics
External Debt (% GDP)	63	80
Government Spending (% GDP)	21	20
Current Account Deficit (% GDP)	-5	-3
Terms of Trade	115	103
GDP/Capita (% of Global Average)	.42	.75
Population (Millions)	25	39

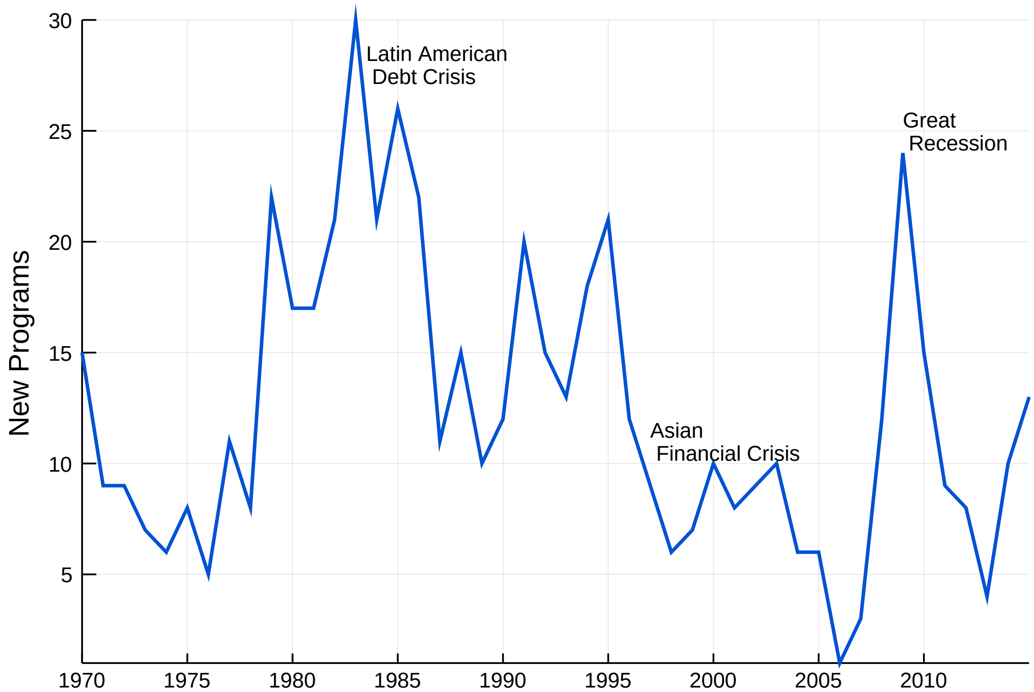
Notes: Average values for the 101 treated observations and the corresponding average for the synthetic controls. Care must be taken with missing data in the countries that make up synthetic controls. Here, for each synthetic control, I temporarily set the weight of observations with missing data to 0 and scale up the weights on observations with the relevant information such that they continue to sum to 1. This allows for a more informative average than discarding all synthetic controls such that a single underlying input country has missing data.
Source: MONA Database & World Economic Outlook, IMF; Penn World Tables; World Development Indicators, World Bank; Valencia and Laeven (2012)

Table 3: IMF Forecasts On Actual Outcomes

	(1) Y_{t+1}	(2) Y_{t+1}	(3) Y_{t+2}	(4) Y_{t+2}	(5) Y_{t+3}	(6) Y_{t+3}
Y^f	**-0.48**	**-0.08**	**0.23**	**0.44**	**0.27**	**-0.66**
	(0.40)	(0.26)	(0.25)	(0.83)	(0.40)	(1.90)
Sample	Treated	Control	Treated	Control	Treated	Control
N	64	89	64	89	89	64
R^2	.17	.56	.12	.07	.11	.16

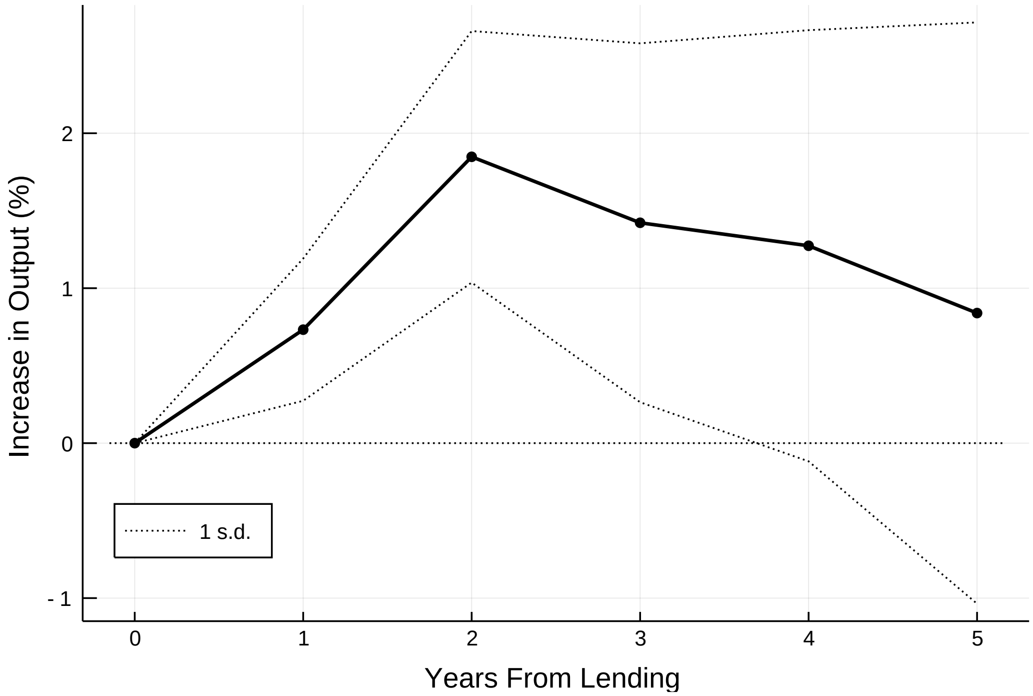
Notes: Coefficient estimates for IMF forecasts, Y^f, at the time of crisis on cumulative output growth either 1 year Y_{t+1}, 2 Y_{t+2} or 3 $Y_{i,3}$, after the date of the financial crisis from a regression with SCM inputs as covariates. This specification includes as covariates a crisis type dummy (ie, Banking, Currency and Debt) as well as growth rates in $t = \{0, -5\}$ to account for the lowest growth rate and some measure of "steady-state" growth. Results are qualitatively similar for different combinations of growth rates used and are available upon request.
Source: MONA Database & World Economic Outlook (+ Historical Forecast Dataseries), IMF; Penn World Tables; Valencia and Laeven (2012)

Figure 1: Time-Series of IMF Programs



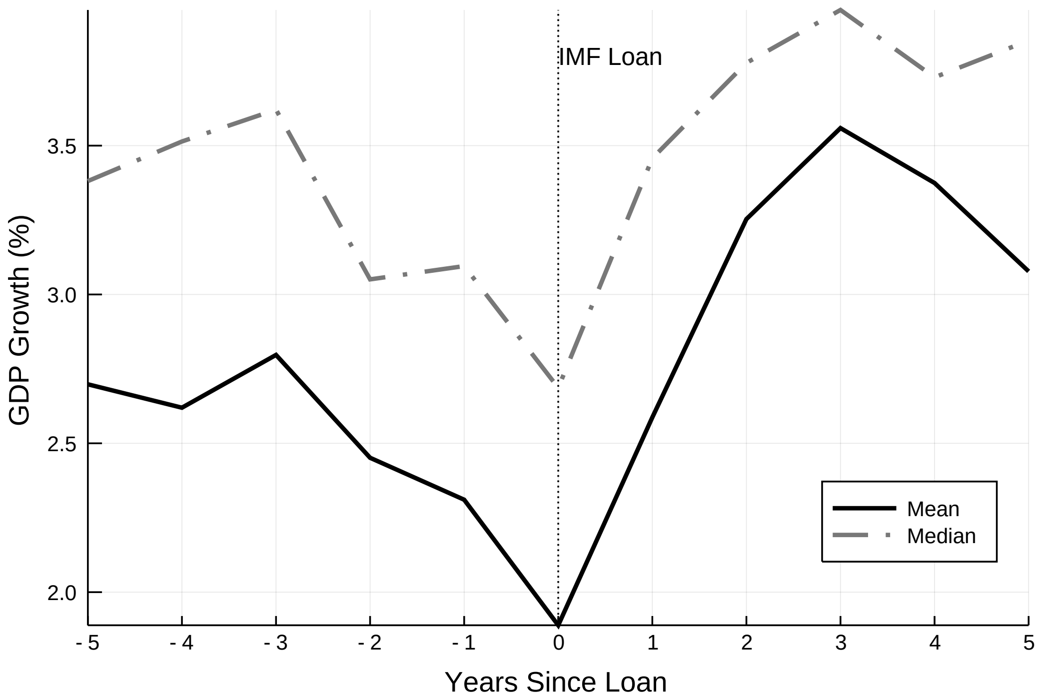
Notes: All *newly initiated* IMF "short-term" programs. The loans included as short-term are: *Stand By Arrangement, Stand By Credit Facility, Rapid Financing Instrument, Rapid Credit Facility, Precautionary Liquidity Line, Flexible Credit Line* and the *Exogenous Shock Facility*.
Source: MONA Database, IMF.

Figure 2: Baseline Results: Impulse Response of GDP to IMF Loan During Crisis



Notes: Measures the implied level difference in GDP at each horizon under growth rates followed by treated units versus synthetic controls. IMF lending is associated with a faster recovery leading to an initial difference in ouptput that fades at longer-horizons as the synthetic observation ultimately recovers. Standard errors are calculated using the variance from empirical distribution of errors in placebo runs (see Section 4), joint significance computed as a Hoetelling T^2 test.

Figure 3: Summary of Growth Paths Surrounding Short-Term Programs

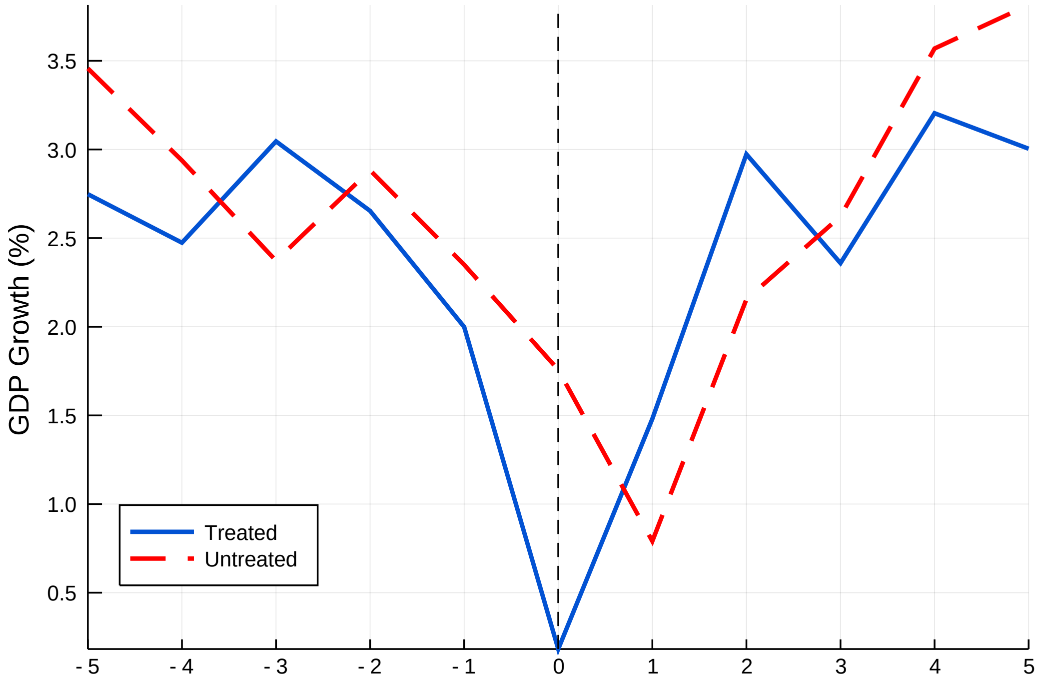


Notes: Unconditional mean and median output growth rates surrounding short-term (using the same classification as Figure 1) programs. A sharp "V" shape characterizes the process suggesting either successful IMF intervention or lending that is timed at the trough of macro-crises.
Source: Author's Calculations using World Development Indicators, World Bank and MONA Database, IMF.

Figure 4: Growth Path Surrounding Financial Crises

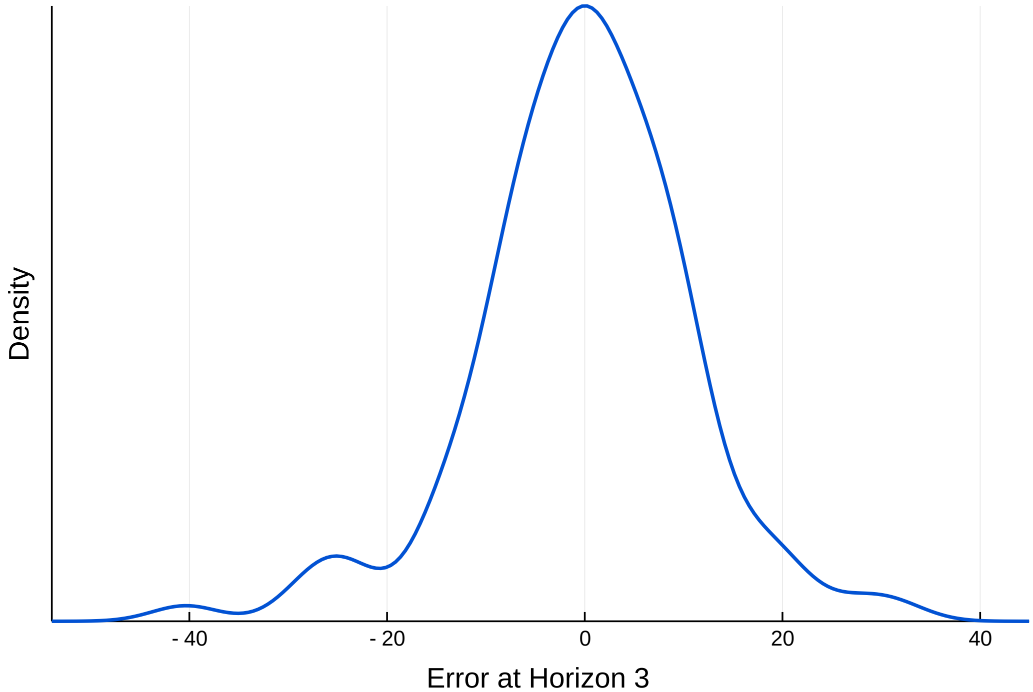
Notes: The average path of IMF short-term programs compared on the same axes as the average path during a financial crisis. Given the large number of financial crises in the data (200+ between 1975 and 2011) it is plausible this large "V" drives the shape of the IMF response.
Source: Author's Calculations from Figures 3-A1.

Figure 5: Growth Paths for Crises With and Without IMF Lending



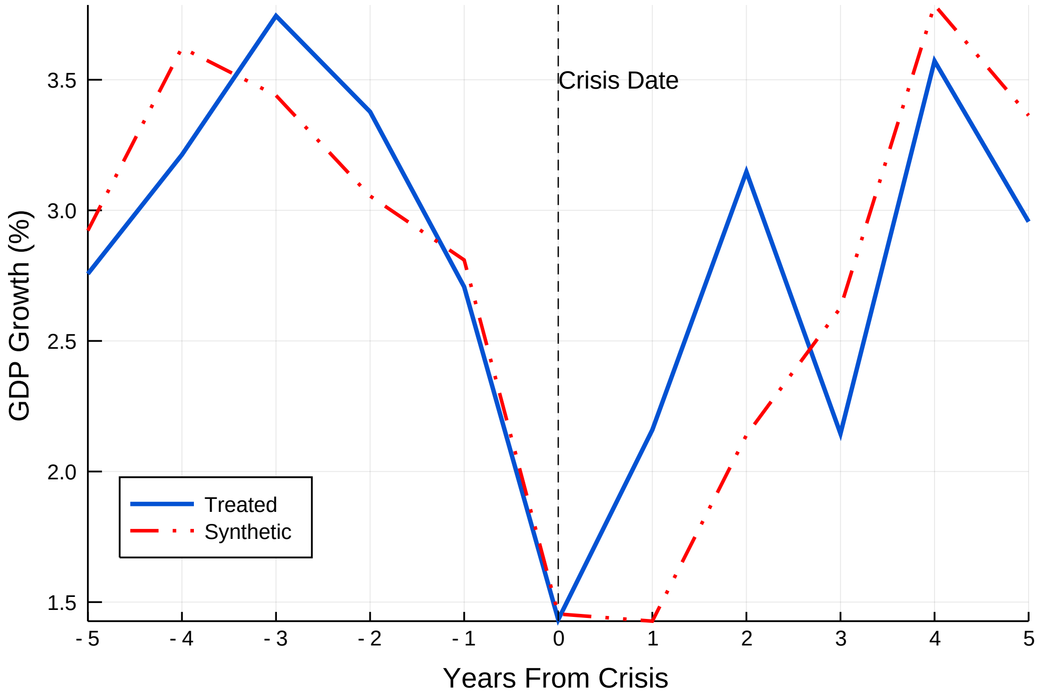
Notes: Average path of growth rates for financial crises with IMF loans (solid blue line) and without (dotted red line). Growth rates recover faster for crises with IMF financing, but these crises also have a substantially worse crashes leaving more space for recovery.
Source: Author's Calculations.

Figure 6: Placebo Runs: Mean-Zero Estimates + Large Variance



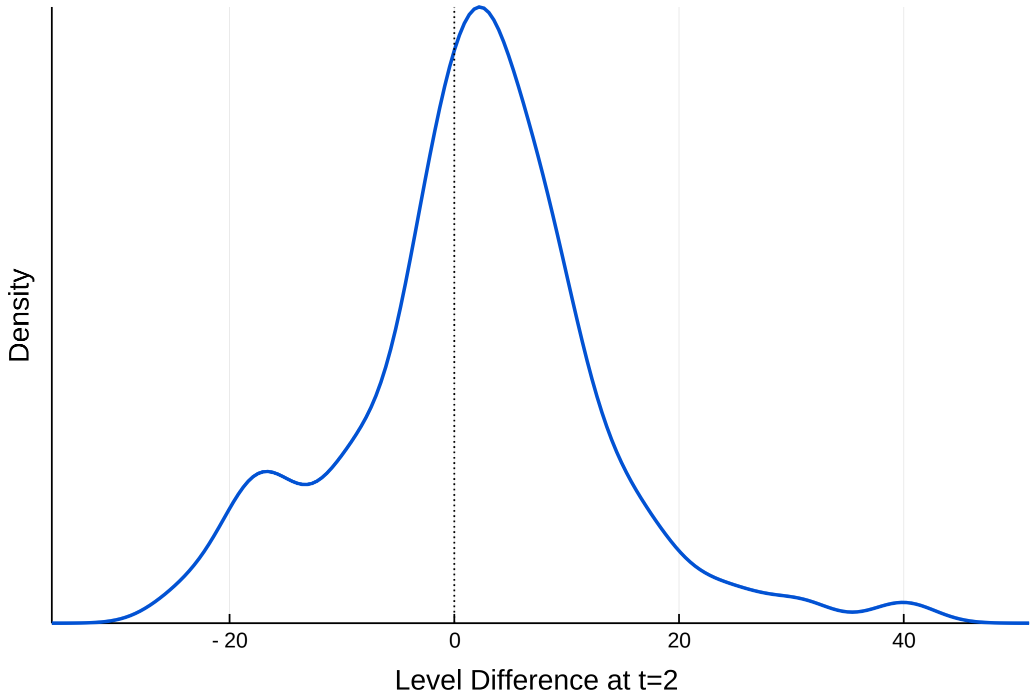
Notes: Density of forecast errors for "placebo" runs. Computed by using the control units as if they were treated, one at a time, and creating synthetic controls from the other untreated crises. This distribution should be mean-zero, but better synthetic control specifications will reduce the variance of this distribution (indicating, on average, better post-period forecasts). Additionally, this can be thought of as an estimate for the distribution of errors under the null-hypothesis and is used for computing the confidence intervals in Figure 2.
Source: Author's Calculations.

Figure 7: Growth Rates Estimates via Synthetic Control



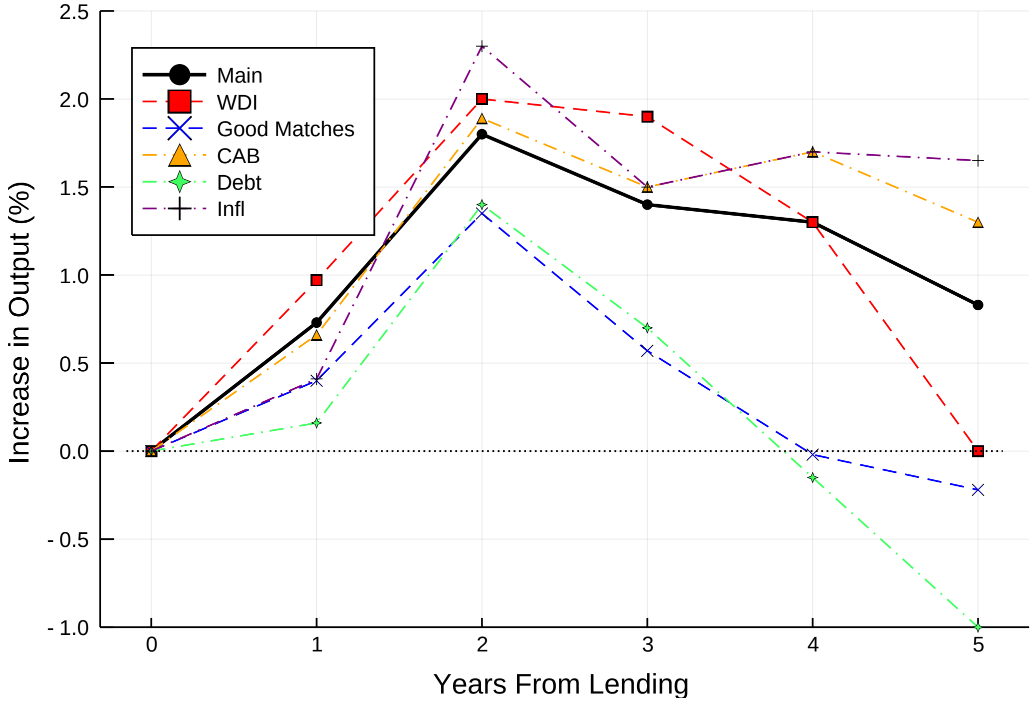
Notes: Main results for growth rates. The solid line is now the average from the 101 crises with a non-empty set of eligible donors; the dotted red is the average of the synthetic controls constructed for each of these 101 crises. The pre-period is matched by construction. The crises treated by the IMF, however, grow faster than their synthetic counterparts for 2 years in the recovery phase.
Source: Author's Calculations from World Development Indicators, World Bank; Valencia and Laeven (2012) and MONA Database, IMF.

Figure 8: Distribution of Effect Sizes



Notes: Full distribution of underlying variation that drives main (average) results. Each point is computed as the implied GDP level difference between the treated observation and its synthetic control 2 years following the onset of its crisis. Notice, hypothesis testing is done on the *mean* of this distrubtion; it is not the case that because 30-40% of the observations fall below zero that the mean is not significantly different from zero.
Source: Author's Calculations from World Development Indicators, World Bank; Valencia and Laeven (2012) and MONA Database, IMF.

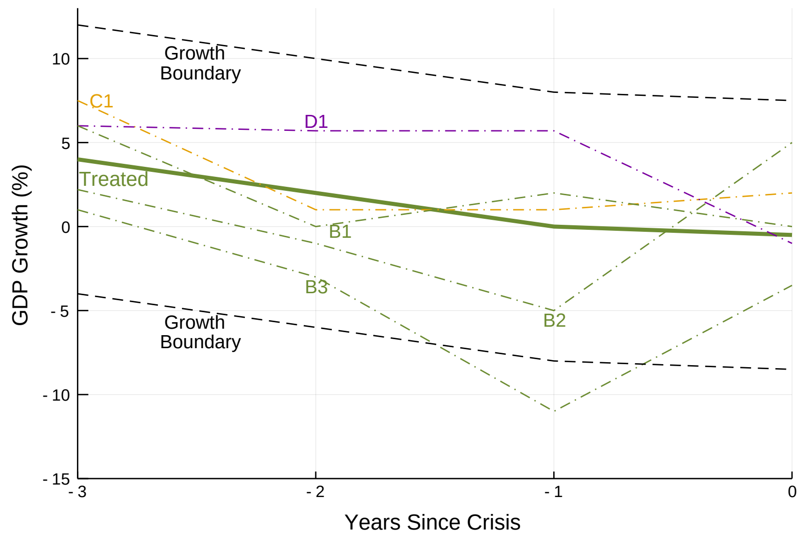
Figure 9: Main Effects Are Robust to Alternative Matching Specifications



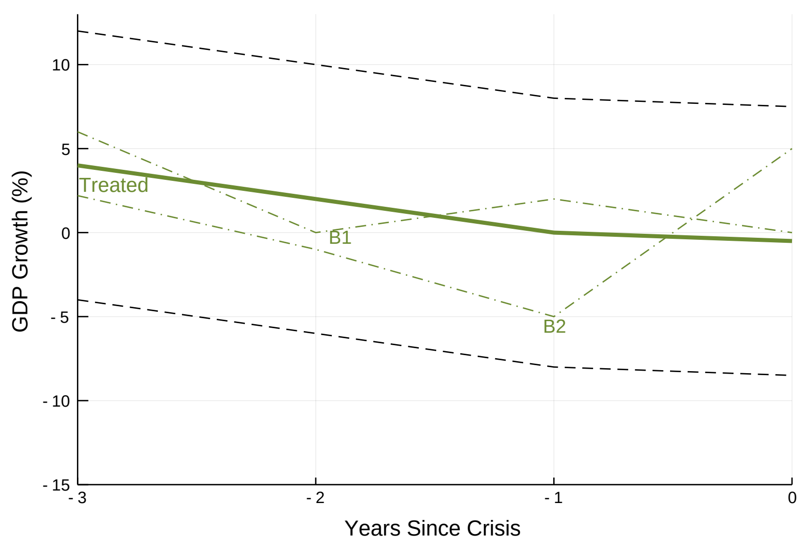
Notes: Robustness IRFs: each IRF comes from changing the main specification as detailed in section 4.3.
Source: Author's Calculations from World Development Indicators, World Bank; Valencia and Laeven (2012) and MONA Database, IMF.

Figure 10: Stylized Example of Local Restrictions

(a) All Donors

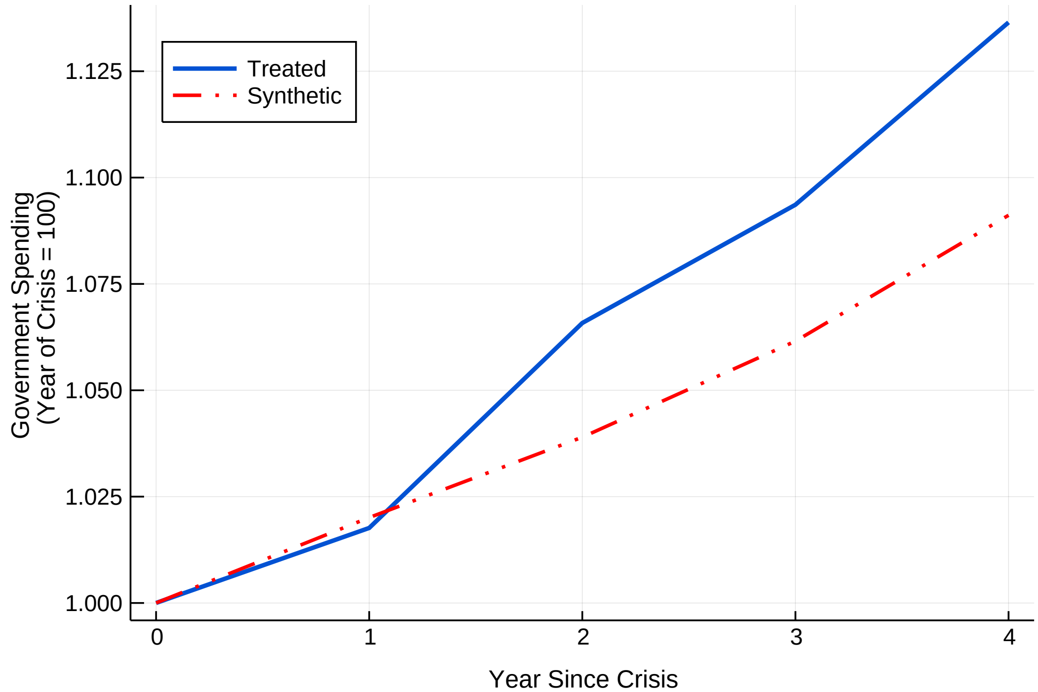


(b) Eligible Donors



Notes: A stylized example of the actual trimming process that seeks eligible donors for constructing a convex combination to match the treated observation. As an example, here the thick mint line is a (made-up) treated banking crisis and D1-D5 are the full sample of donors (mint being denoting banking donors; gold are currency donors; purple are debt donors). The growth boundary lies ± 8 percentage points from the treated countries of interest. D1 and D2 are eliminated because they are a different crisis type. D5 is eliminated because it falls outside of the growth boundaries so is no longered considered a "local" crisis. Panel (b) shows the eligible donors that fit both criteria, D3 and D4, that would then be used for constructing a convex combination to match the characteristics of the treated observation. This is repeated *for each* treated observation.
Source: Author's Calculations (Data manufactured for example).

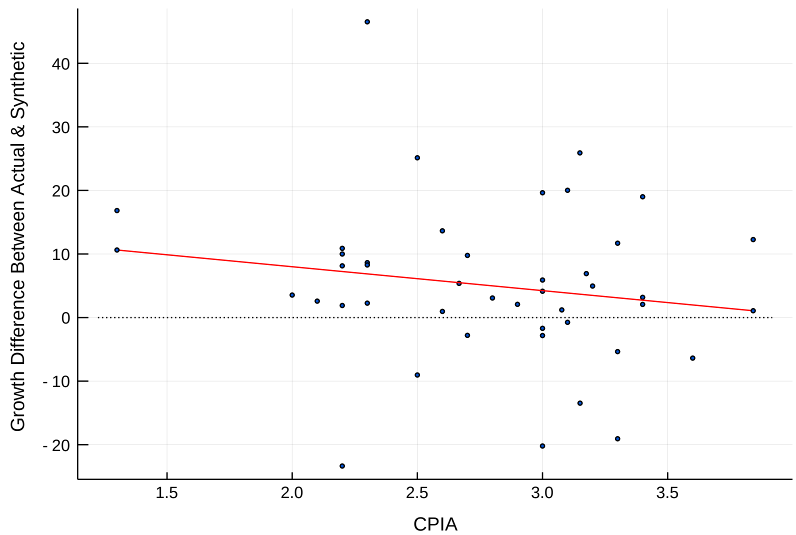
Figure 11: Government Spending Increases More in Treated Countries



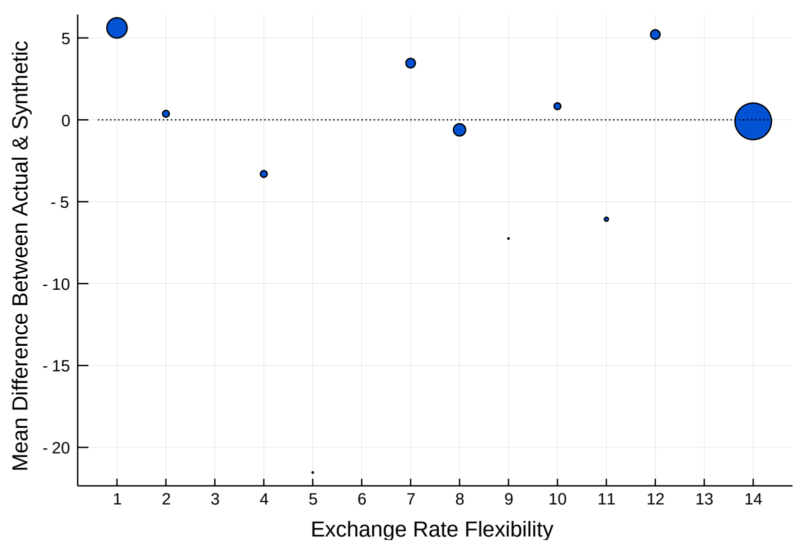
Notes: Evolution of government spending for treated versus synthetic observations. The y-axis is relative to a value of 100 in the year of the crisis; 102, for example, indicates a cumulative 2 percent increase over the years since $t = 0$.
Source: Author's Calculations from World Development Indicators, World Bank; Valencia and Laeven (2012) and MONA Database, IMF.

Figure 12: Effect Sizes Larger in Predictable Settings

(a) Effect Sizes Inversely Correlated With State Capacity



(b) Effect Sizes Largest in Fixed Exchange Rate Regimes



Notes: Panel (a) plots the relationship between World Bank Country Policy Institutional Assessment (CPIA) Scores and estimated effect sizes. Effect sizes being the difference between a treated observation and its synthetic control. Following past work, I take state capacity to be the average of 16 underlying CPIA measures regarding both economic policy, corruption and institutional strength generally. Countres with weaker state capacity are estimated to benefit the most from these loans. Panel (b) performs a similar exercise using exchange rate regimes as classifed by Ilzetzki, Reinhart and Rogoff (2018) where higher values correspond to more flexible exchange rates. Scatter points are weighted by the number of underlying observations in that exchange rate bin. The plot indicates fixed exchange regimes (x-values of 1) have the largest effects with the estimates having little clear pattern throughout the rest of the distribution.
Source: Author's Calculations from World Development Indicators, World Bank; CPIA, World Bank; Valencia and Laeven (2012) and MONA Database, IMF.

A Data Appendix

This appendix makes precise the definition and sources of the data pulled for this analysis.

A.1 IMF Loans

IMF loans come from the Monitoring of Fund Arrangements (MONA) database which tracks all IMF programs and their details.

A.2 Definition and Sources for Outcomes and Covariates

- Growth Rates: Growth rates are constructed using logged differences of the levels from the Penn World Table's (v 9.0) real GDP (in local currency) from national accounts. Robustness is checked using analogous estimates from the World Bank's Word Development Indicators.
- From the World Economic Outlook Spring 2017 (variable in paper: variable name in WEO):
  - Current Account Balance: BCA_GDP_BP6
  - Inflation: Percent Change in PCPIE
  - External Debt: D_GDP

B Details of Synthetic Control Method

Suppose at some horizon, t, following a financial crisis (at time $t = 0$) Equation 3 determines $y_{i,t}$.

$$y_{i,t}(IMF_i) = F^t(X_{i,0}, y_{i,0}, y_{i,-1}, \dots y_{i,-\infty}) + \theta_t IMF_i + u_{i,t} \tag{3}$$

Here $F^t()$ is a function only of outcomes in the year of the crisis and prior, so it can be thought of as a mean-zero forecasting equation (in the absence of IMF lending) from the time of the crisis on. It can in theory incorporate any characteristics known at time 0, $X_{i,0}$, as well as an arbitrary number of lags for the outcome variable with a fully non-linear structure. IMF_i is a dummy variable for whether the IMF began a program in response to a crisis.[33] As in any policy analysis the goal is to estimate $y_{i,t}(0)$ for crises treated by IMF lending in order to identify θ_t.

[33]This will be empirically identified as a financial crisis that received an IMF program in that same year or following year.

The only assumption necessary on (3) for constructing a good counterfactual using the SCM is that F^t can be well approximated *locally* by a linear function, $\hat{F}_i^t()$.[34] To simplify notation let (X, Y) represent the vectors of $X_{i,0}$ and all lags of y_i that F^t is a function of.

$$F^t(X, Y) \approx \hat{F}_i^t(X, Y) = \mathbf{A}_i X + \mathbf{B}_i Y \quad \text{if } (X, Y) \in \mathbb{L}_i \tag{4}$$

$\mathbb{L}_i$ is defined as the set of all points in a ball of radius δ surrounding the (X_i, Y_i) vectors. Notice this function is i-dependent: local linear approximations will be different depending on what they are local to. This is not problematic. Now suppose there exists a set of crises untreated by the IMF, the donors $\mathbb{D}$, and a subset of these donors $\mathbb{P} \in \mathbb{D}$ that are close (technically defined by 5).

$$p \in \mathbb{P} \iff (X_p, Y_p) \in \mathbb{L}_i \quad \cap \quad p \in \mathbb{D} \tag{5}$$

I call $\mathbb{P}$ the set of *eligible donors* for crisis i. Suppose further that among the eligible donors there exists a weighting vector $\lambda^i = (\lambda_1^i, \ldots, \lambda_p^i, \ldots, \lambda_P^i)$ such that conditions (6)-(8) hold.

$$Y_i = \sum_{p \in \mathbb{P}} \lambda_p^i Y_p \tag{6}$$

$$X_i = \sum_{p \in \mathbb{P}} \lambda_p^i X_p \tag{7}$$

$$\sum_{p \in \mathbb{P}} \lambda_p^i = 1 \tag{8}$$

Conditions in (6)-(8) require having a convex combination of eligible donors that matches i on the variables that determine F^t. It can then be shown that this convex combination of eligible donors *also* approximates the outcomes of the treated i *had it not received treatement* through the

[34] Since all functions have a 1st order taylor series that approximates them linearly this is not a restrictive assumption, it must be continuously differentiable. "Well approximated" is the only restriction, then.

following logic. (Denote ν_i as the error coming from using the local linear approximation.)

$$\begin{aligned}
\sum_{p\in\mathbb{P}} \lambda_p^i y_{p,t} =& \sum_{p\in P} \lambda_p^i F^t(X_{p,0},\ldots,y_{p,0}) + \sum_{p\in\mathbb{P}} \lambda_p^i u_{p,t} \\
=& \sum_{p\in\mathbb{P}} \lambda_p^i (\hat{F}_i^t(X_p,Y_p) + \nu_p) + \sum_{p\in\mathbb{P}} \lambda_p^i u_{p,t} \\
=& \hat{F}_i^t\Big(\sum_{p\in\mathbb{P}} \lambda_p^i X_p, \sum_{p\in\mathbb{P}} \lambda_p^i Y_p\Big) + \sum_{p\in\mathbb{P}} \lambda_p^i (u_{p,t} + \nu_p) \\
=& \hat{F}_i^t(X_i,Y_i) + \sum_{p\in\mathbb{P}} \lambda_p^i (u_{p,t} + \nu_p) \\
=& y_{i,t}(0) - \nu_i - u_{i,t} + \sum_{p\in\mathbb{P}} \lambda_p^i (u_{p,t} + \nu_p) \Rightarrow \\
y_{i,t}(0) =& \sum_{p\in\mathbb{P}} \lambda_p^i y_{p,t} + \underbrace{(u_{i,h} - \sum_{p\in\mathbb{P}} \lambda_p^i u_{p,t})}_{\text{0 in Expecation}} + \underbrace{(\nu_i - \sum_{p\in\mathbb{P}} \lambda_p^i \nu_p)}_{\approx 0 \text{ if locally linear}} \qquad (9)
\end{aligned}$$

Notice the advantages of this result relative to traditional regressions.

1. $\mathbf{A}_i, \mathbf{B}_i$ can vary for each crisis depending on its pre-conditions *and* never needs to be estimated
2. The underlying structure only requires *local* linearity rather than the much more restrictive *global* linearity assumption; SCM removes the need for extreme parametric extrapolation
3. For each treated country the counterfactual is directly observable as the convex combination of actual untreated observations making it highly transparent

C Quality of Synthetic Controls

Along with Figure 7 there are two additional figures I present here as a way to understand the quality of the synthetic controls. First, Figure A3 shows the distribution of weights on each untreated crisis, measured as the sum of it's contribution to each synthetic. For example, suppose there was an application where only 2 synthetic controls were going to be created for 2 treated countries. If untreated country D contributed a weight of 0.07 and 0.23 to these synthetics, respecitively, it would have a *total weight* of 0.3. On average, if there are more untreated units than treated each untreated will get a total weight less than 1. In the main example here, there are 101 synthetics with 157 untreated units, so this will be the case. It is the case here that many—nearly $\frac{1}{3}$—are not used at all; a few crises have a total weight of approximately 3. But recall the point of the synthetic control is to over sample from crises that "look" more like IMF crises, so this is by design. It would be troubling if, for instance, 2 or 3 countries made up nearly all of the variation in the synthetic controls, but this is clearly not the case.

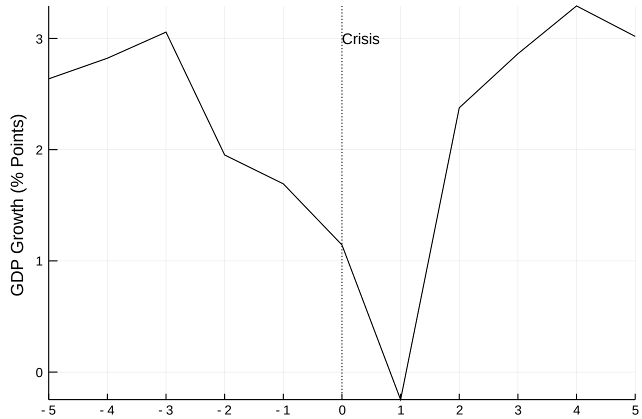
Second, Figure A4 measure the misses for the *target* growth rate one year from the crisis—an arbitrary choice that looks similar regardless of target variable. This is a variable the synthetic control is attempting to match, so big misses here indicate that there may not be a good "synenthetic control" available anywhere in the untreated sample. This is not a problem unique to the synthetic control method, in regressions it is commonly the case we extrapolate a counterfactual from observations far—in a generalized distance sense—from the treated unit. However, an advantage of the synthetic control is that it is easy to see when such extrapolation is taking place and test whether it is important in generating the main results. The robustness check used in Figure 9 is to discard the 10% worst matches—defined as distance from target growth rates. It can be seen there that these do not drive the results.

Table A1: Heterogeneity in Effect Sizes By World Region

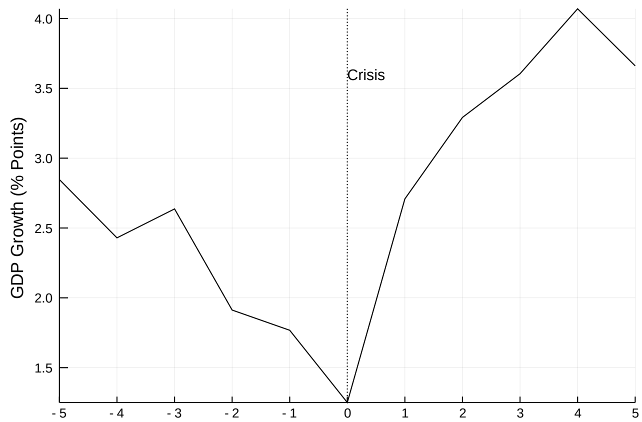
Region	Average Effect Size
Africa	3.5
Asia	-4.5
Latin America	2.1
Small Islands	4.2

Figure A1: Growth Paths Around Financial Crises

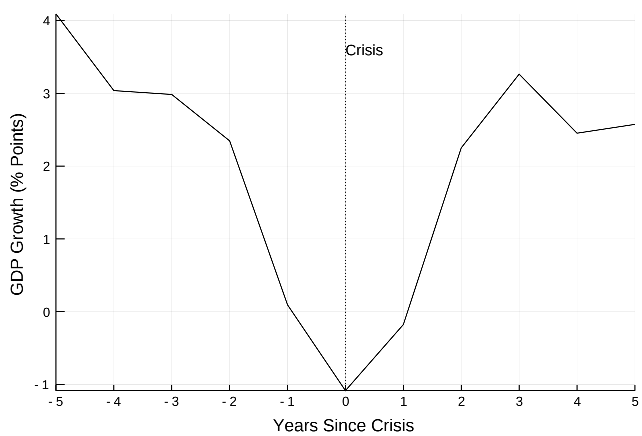
(a) Banking



(b) Currency

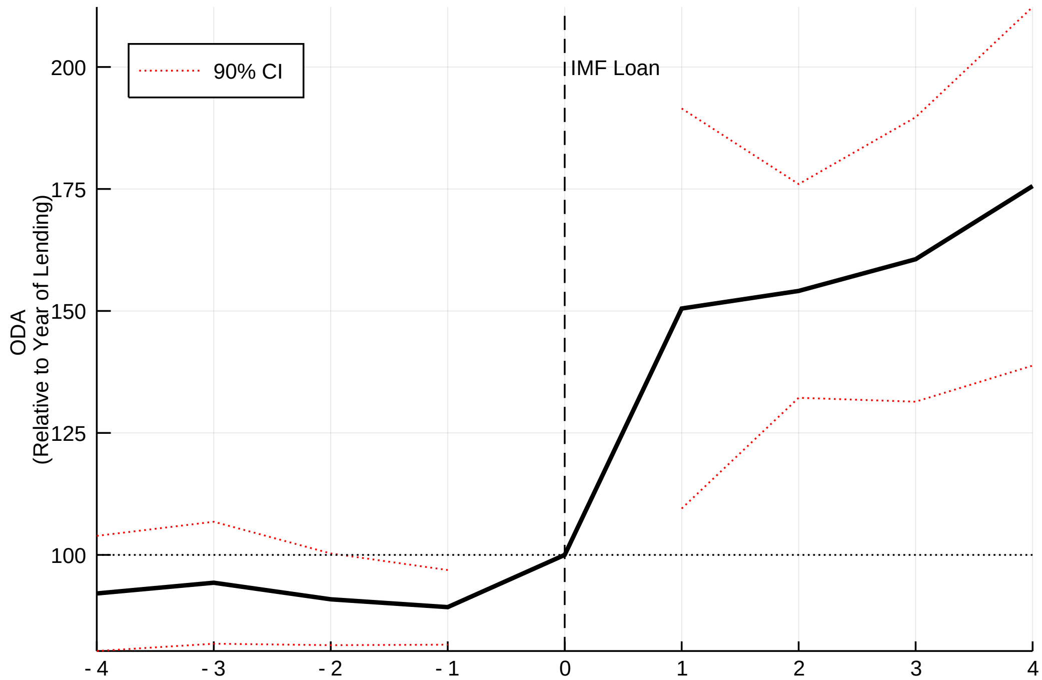


(c) Debt



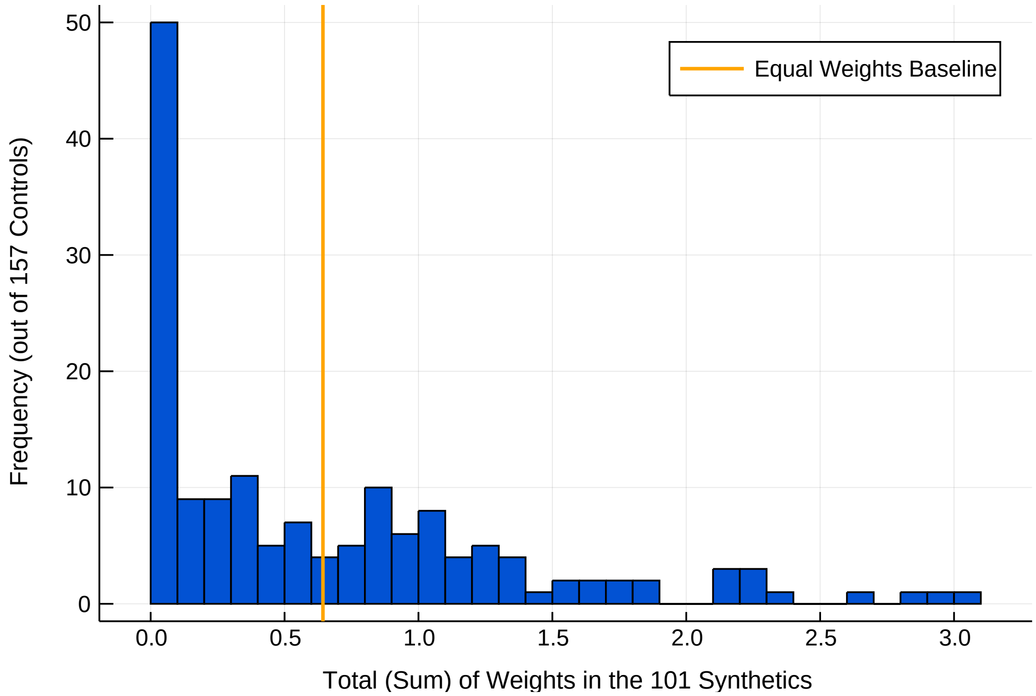
Notes: Unconditional mean and median output growth rates surrounding various financial crisis. An even more extreme "V" characterizes these paths suggesting Figure 3 may be due to this rather than a causal IMF effect.
Source: Author's Calculations using World Development Indicators; crisis dates come from Valencia and Laeven (2012).

Figure A2: IMF Lending Associated with Large Increases in Foreign Aid (Kuruc,2018)



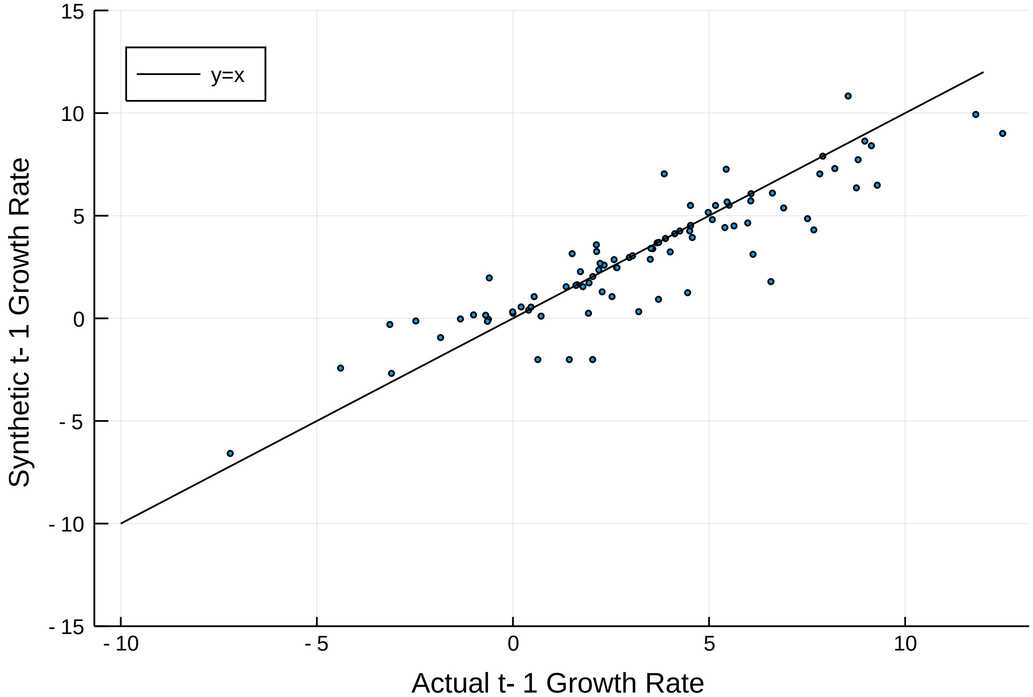
Notes: Response of Official Developmental Assistance in Fragile States following an IMF program. ODA taken from the OECD development financing database; “Fragile States” defined as country-years with CPIA scores below 3.2. All loans (in crisis and out of crisis) constitute the event.
Source: Kuruc (2018).

Figure A3: Histogram of Synthetic Control Weights



Notes: Distirubiton of the sum of weights from contributions to all synthetic controls in main run. *Equal weights baseline* just indicates where the entire mass of the distribution would lie under random assignment; 157 controls equally distributed to 101 treated observations implies each would receive a total weight less than 1.

Figure A4: Synthetic vs. Actual for Targeted Growth Rates



Notes: Comparison of targeted variable (growth in $t-1$) for actual observation versus the synthetic control. 45-Degree line represents what would be achieved under the SCM perfectly targeted this variable; the actual observations stray a bit from this.
Source: Main Synthetic Control Run.