The Dissertation Committee for Kevin J. Kuruc
certifies that this is the approved version of the following dissertation:

Essays on Macroeconomic Development Policies

Committee:

Olivier Coibion, Co-Supervisor

Dean E. Spears, Co-Supervisor

Saroj Bhattarai

Shinji Takagi

Essays on Macroeconomic Development Policies

by

Kevin J. Kuruc

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2019

Acknowledgments

I would like to enthusiastically thank Olivier Coibion, Dean Spears and Saroj Bhattarai for their invaluable support throughout graduate school. I also benefited greatly from discussions with Shinji Takagi, Melissa LoPalo, David Beheshti, Cooper Howes and innumerable seminar participants at the University of Texas at Austin, the International Montery Fund, Texas A&M University, the Reserve Bank of India, and the University of Oklahoma.

Essays on Macroeconomic Development Policies

Kevin J. Kuruc, Ph.D.
The University of Texas at Austin, 2019

Supervisors: Olivier Coibion
Dean E. Spears

This dissertation examines the policies of developed countries and international financial institutions on developing countries.

The first chapter 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 different 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.

The second chapter (joint with Melissa LoPalo, Dean Spears and Mark Budolfson) asks how costly climate change will be for India. We first draw on microeconometric estimates of

the impacts of heat waves on important social indicators. This analysis demonstrates that India is uniquely climate vulnerable given the high levels of humidity in south Asia. Then, using a modified regional Integrated Assessment Model (RICE, Nordhaus (2010)), we perform a welfare exercise in which we quantify total future damages in terms of consumption equivalent near-term losses: how much would consumption need to be reduced for the next 20 years to be equivalently bad (in a welfare sense) as projected climate damages? We find damages are as costly for welfare as a near-term humanitarian crisis (30% GDP per capita reduction over 20 years), but that the relationship is convex: if India can spur even minimal international coordination we estimate there would be large social returns.

The third chapter presents a quantitative analysis of the macroeconomic characteristics and performance of fragile states, especially in the context of their engagement with the International Monetary Fund. It finds, among other things: (i) fragility may be a more fluid state than previously documented; (ii) while in fragile states GDP growth is more volatile, it is only slightly slower, on average, than growth in nonfragile states; (iii) fragile states' GDP appeared to grow about 1 percentage point faster following approval of an IMF lending arrangement; and (v) foreign aid flows to fragile states increased by about 60 percent in the years following approval of IMF program engagement, with or without IMF financing (no such increase was observed for non-fragile states), illustrating the IMF's catalytic role. While this analysis provides a positive overall assessment of the IMF's role in fragile states, care must be exercised in interpreting the results, especially concerning causality.

Table of Contents

List of Tables

List of Figures

Chapter 1

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

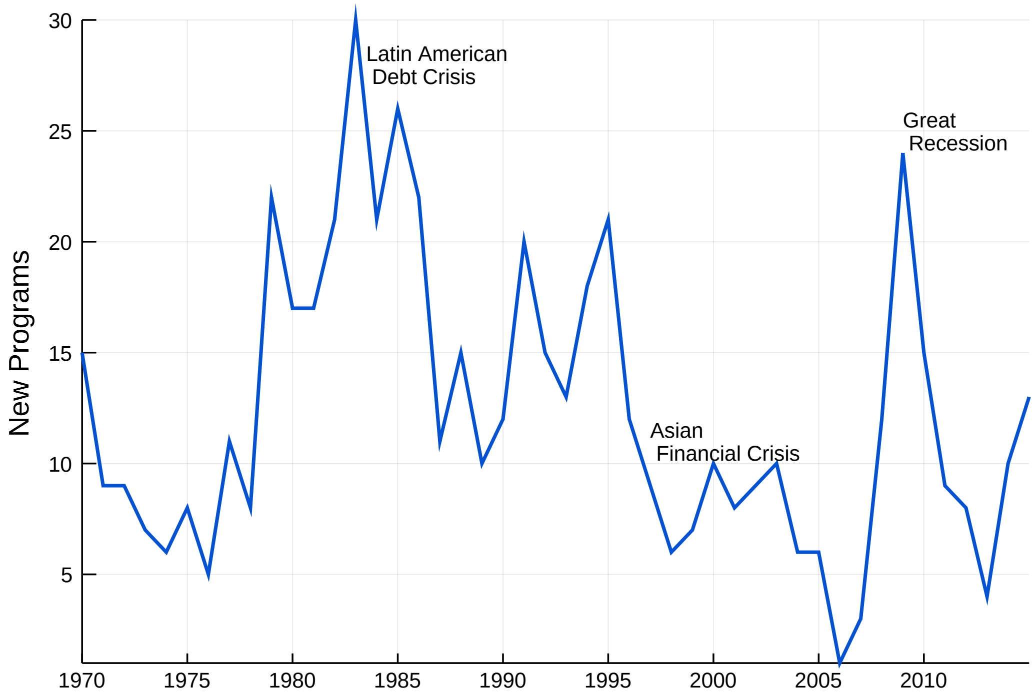
1.1 Introduction

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 the ones 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 structural reforms. However, as depicted in Figure 1.1, 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 fiscal policy advocated by the IMF (Joseph E Stiglitz, 2002) and the negative signaling effect of using a "lender of last resort"

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

Figure 1.1: Time-Series of IMF Programs



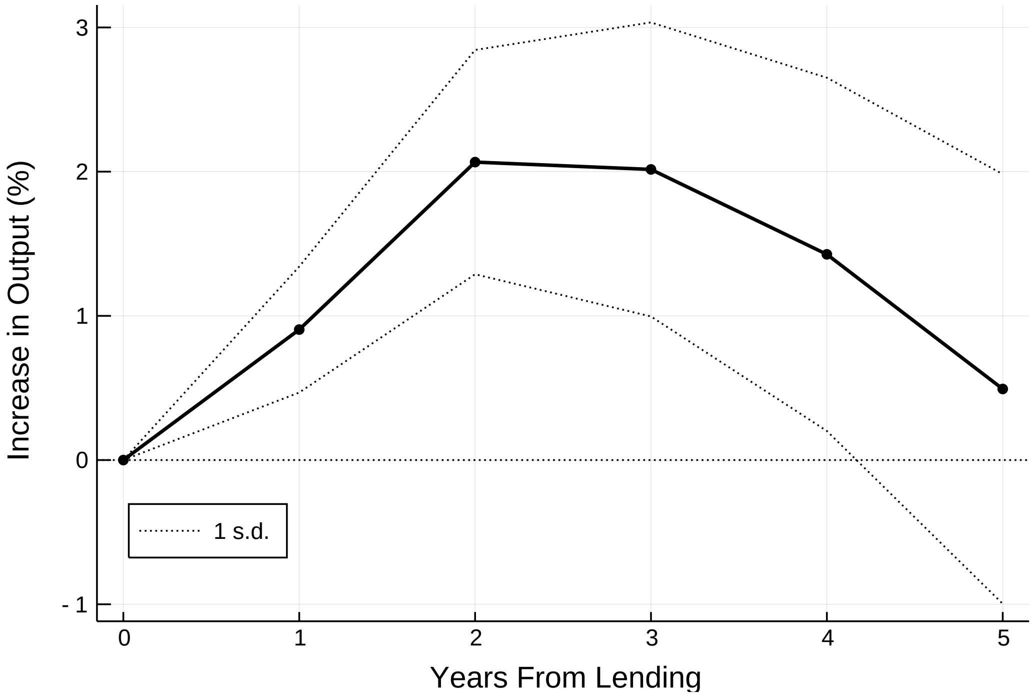
Notes: All *newly initiated* IMF "short-term" programs. The loans included as short-term are: *Stand By Arrangement, Stand By Credit Facility, Rapid Financing Instrument, Rapid Credit Facility, Precautionary Liquidity Line, Flexible Credit Line* and the *Exogenous Shock Facility*.
Source: MONA Database,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

(Carmen M Reinhart and Christoph 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 studies 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 is 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 Fabian Valencia and Luc 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 finds 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 between the treated crisis and its synthetic control in the post-periods of interest.

Using this method I find that IMF involvement in a financial crisis is associated with a significantly faster recovery than would be otherwise anticipated. Figure 1.2 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 that government spending rises faster in treated countries than in 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 are not robust. This pattern is qualitatively consistent with the IMF helping countries through the crisis phase, but not systematically changing long-run potential GDP.

Figure 1.2: 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 1.3.2), joint significance computed as a Hoetelling T^2 test.

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, 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: I only compare 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 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 further mitigate extrapolating from unlike crises 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 confounding: 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 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 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, they should be correlated with the errors arising from the SCM. In practice, this would be reflected in an estimated coefficient on IMF forecasts that is positive in a regression that includes the SCM variables as covariates. This is not 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 side. As an example of a concerning issue, 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 that 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 one could worry about, most seem inconsistent with this pattern.

In fact, the negative correlation between effect sizes and institutional quality is 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 (Craig Burnside and David Dollar, 2000), the negative interaction is not necessarily surprising here. Countries with weak institutions and/or below average policy are likely the 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 (Kevin Kuruc, 2018*a*).

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

Finally, I find that effect sizes are larger for countries with fixed exchange rates than for countries with more flexible regimes. This is in line with theoretical results that spending multipliers—especially from *external* financing—are large under fixed exchange regimes (Emmanuel Farhi and Iván Werning, 2016). Output is estimated to be 5-6 percent larger (cumulatively) in response to IMF loans that are between 1.5-2.5 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 1.2 introduces the empirical setting and formalizes the challenges to overcome. Section 1.3 describes the synthetic control method, both generally and how it is specifically used in this paper. Section 1.4 presents the results under the main specification and shows these findings are robust. Section 1.5 examines the response of other aggregates as well as heterogeneity in the effect sizes to examine the mechanisms underlying the main results.

1.1.1 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 (Angus 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 (Michael D Bordo and Anna J Schwartz, 2000; Michael Hutchison, 2003; James Raymond Vreeland, 2003; Muhammet A Bas and Randall W Stone, 2014; Yasemİn Bal Gündüz, 2016). Methods

[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.

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. This 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." For example, Robert J Barro and Jong-Wha Lee (2005) 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 difficult to interpret 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).

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 Orley 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 (James Heckman, Hidehiko Ichimura, Jeffrey Smith and Petra Todd, 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

[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.

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

[6] Michael A Clemens, Steven Radelet, Rikhil R Bhavnani and Samuel Bazzi (2012) argue for simple estimators in the foreign aid literature precisely for this reason.

this paper, however, is 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 Alberto Abadie and Javier Gardeazabal (2003), it was formally developed by Alberto Abadie, Alexis Diamond and Jens Hainmueller (2010) and extended by Arindrajit Dube and Ben Zipperer (2015). Applications include Giovanni Peri and Vasil Yasenov (2019), Eduardo Cavallo, Sebastian Galiani, Ilan Noy and Juan Pantano (2013), Daron Acemoglu, Simon Johnson, Amir Kermani, James Kwak and Todd Mitton (2016), Alberto Abadie, Alexis Diamond and Jens 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 Andreas Billmeier and Tommaso 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 Aart 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) is consistent with the work of Alan J Auerbach and Yuriy Gorodnichenko (2012). These authors show, using US data, multipliers may be substantially larger in times of recession. Even Valerie A Ramey and Sarah 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

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

in this paper seem consistent with and contribute to the literature on spending multipliers more generally.

1.2 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. 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 with 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.

1.2.1 IMF Programs

The IMF is extremely involved in the global economy both in supplying credit and guiding economic policy. Figure 1.1 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),

---

these settings.

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 roughly understand the empirical regularities of this setting. In the main analysis, since the objective is to measure the differences in outcomes between crises that receive an IMF program and those that do not, I include even those programs that are legally framed as being for medium or long term assistance if the program begins during an acute crisis.[10]

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 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 1.1). 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.

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

[9]Described in the data appendix.

[10]Since only a small fraction of medium/long term instruments go to acute crises, including them in the summary statistics offers a less clear picture of the empirical setting of interest.

Table 1.1: 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)

Even among the subset of IMF programs classified as "short-term" there is a wide range of country situations and loan sizes. Table 1.1 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 Tables, the financial crisis indicators come from Valencia and Laeven (2012) and are 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 1.1 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. For scale, 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 are typically 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 (William Easterly, 2005; Gunes Gokmen, Tommaso Nannicini, Massimiliano Gaetano Onorato and Chris Papageorgiou, 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.

1.2.2 Average Recoveries: An Ashenfelter Dip

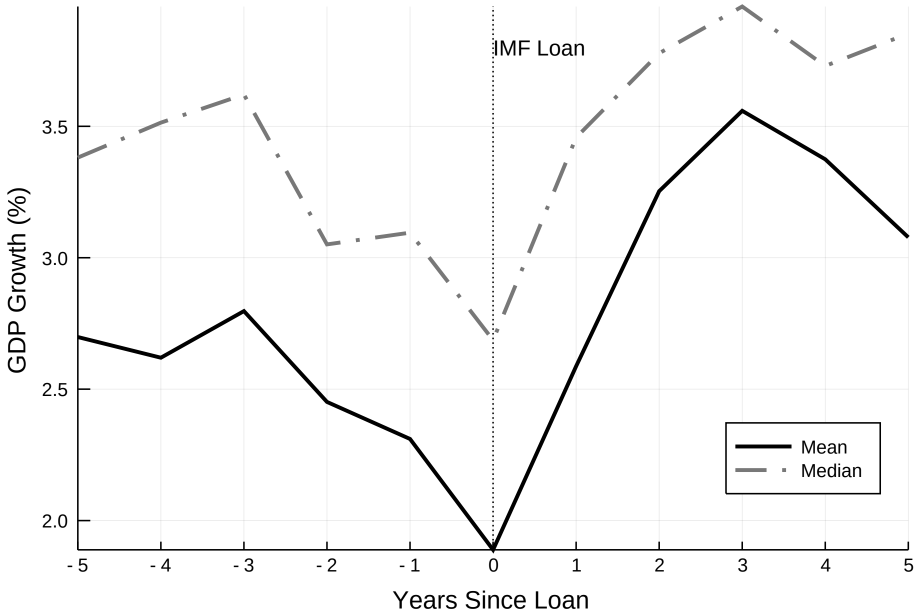
Average growth paths around IMF short-term loans indicate country growth rates recover rapidly following IMF intervention. This is a critical first step towards understanding the empirical regularities of the "treatment" variable and the challenges posed by the setting. Figure 1.3 depicts this pattern in an unconditional event study, following Michael Bruno and William Easterly (1998), Pierre-Olivier Gourinchas and Maurice Obstfeld (2012) and Kuruc (2018*a*). 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).[11] 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, 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 accounts 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

[11] 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*a*) who plot completely unconditional moments surrounding their event of interest in order to formalize interpretable stylized facts.

Figure 1.3: 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 1.1) 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.

inadequate. Country growth slides are not typically associated with such rapid reversals universally (Lant Pritchett, 2000). These appear to be growth slides culminating in some event that marks the bottom of the trough and start to recovery. I verify this concern by documenting a similar pattern surrounding financial crises, and ultimately condition on the occurance of these events.

1.2.3 Financial Crises

Financial crises are taken from 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.[12]

- **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.

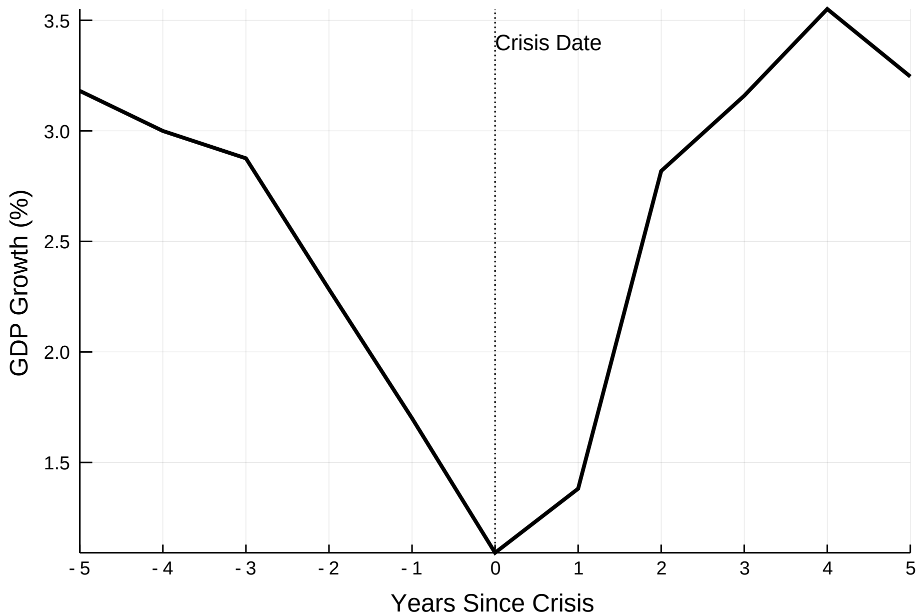
There are many crises of each type in the sample period of interest (1975-2010). These categories are not mutually exclusive, so the number of country-years experiencing a financial crisis is less than the sum of the three crisis types; there are 372 unique crises as I have defined them. This data is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

It is now possible to study whether the pattern observed in Figure 1.3 is likely to be driven by IMF loans going disproportionately into situations that would have recovered regardless. Figure 1.4 runs a similar unconditional event study to Figure 1.3 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 is necessary to construct a plausible counterfactual that shares the recovery properties of the events of

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

Figure 1.4: Growth Path Surrounding Financial Crises



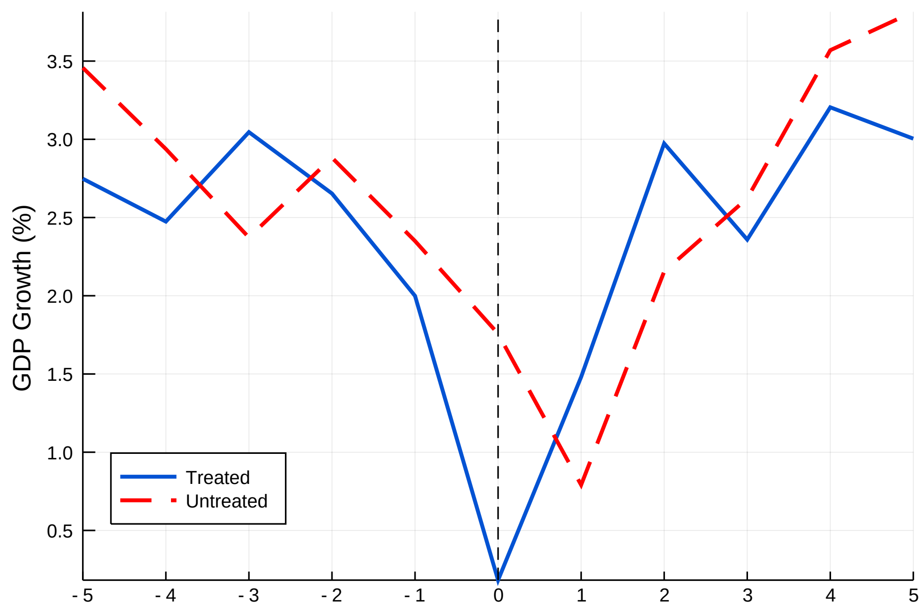
Notes: The average path of IMF short-term programs compared on the same axes as the average path during a financial crisis. Given the large number of financial crises in the data (200+ between 1975 and 2011) it is plausible this large "V" drives the shape of the IMF response.
Source: Author's Calculations

interest. To this end I study the recoveries of financial crises treated by an IMF program and use untreated financial crises to approximate the recovery dynamics that would have arisen otherwise.

1.2.4 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 that began 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.

Figure 1.5: 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.

With these definitions, I can plot the average path of treated versus untreated crises as this is ultimately the comparison of interest. Figure 1.5 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 1.2). 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, attempts to appropriately condition on the pre-crisis growth path in order to correct for this difference and produce a more meaningful comparison.

1.3 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 the 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 very effective.

1.3.1 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 Òscar Jordà (2005); the mean difference between the 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. Andreas Billmeier and Tommaso 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.[13] This trade-off is well-known and originally motivated the use of matching estimators generally (Donald B Rubin, 1977; Rajeev H Dehejia and Sadek Wahba, 2002). I extend this logic to the case of synthetic controls.

[13]The difference with past papers is described in detail in the appendix.

To formalize the discussion, suppose the data generating process can be written as a mean-zero forecasting equation as in (1.1):

$$y_{i,t} = F^t(X_{i,0}, y_{i,0}, y_{i,-1} \dots y_{i,-\infty}) + \theta_t IMF_i + u_{i,t} \tag{1.1}$$

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 $\hat{F}_i^t()$*

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

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

Assumption 1 states that 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, I stray from an infinitesimally small neighborhood so there is an implicit assumption that there is 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 $y_{i,-k}$.[14]

Denote the counterfactual under $IMF_i = 0$ as $y^c_{i,t}$. If these are met then the following result obtains.[15]

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

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 observa-

[14]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.

[15]See the appendix for details.

tion. $Z_{\mathbb{J}}$ 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) are generated by solving the following minimization problem.

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

Prior to solving this problem the potential controls are restricted to some set, J, of qualitatively similar crises. That is why this consideration is absent from the description of the minimization problem.[16]

1.3.2 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 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

[16] 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.

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 ultimately 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 is 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.[17] While this argument provides good reason to choose the procedure that minimizes $\sigma^2_{placebo}$, I include robustness checks to show the main results remain similar when explicitly targeting variables that do not appear to add value in the placebo 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 that 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

[17] 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.

for constructing the likelihood of all possible outcomes under a normal distribution.[18]

1.3.3 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 are 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.

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 1.6 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 discard these observations. Further, crises must fall within the "Growth Boundaries" labelled on

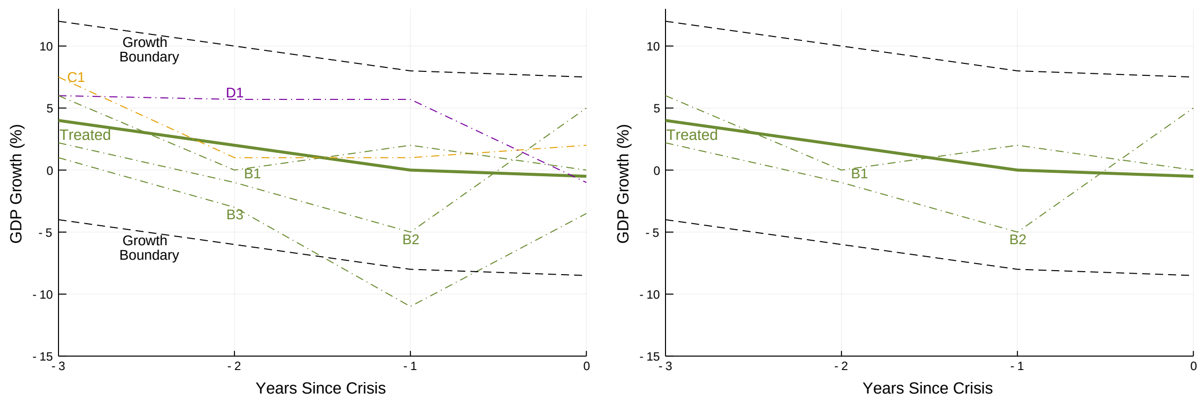
[18] The central limit theorem gives us normality here since the distribution of interest is one of sample means. This is not an assumption about the distribution of errors for individual observations which does in fact have non-normal properties.

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 is discarded as well.

Figure 1.6: 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).

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. Conceptually, severe collapses in growth rates can only continue for so long; once economies shrink there is only so much space to continue shrinking. A convex combination of these episodes is unlikely to serve as a good counterfactual for a mild crash if the severe crashes have unusually large increases in growth rates.[19] Empirically, allowing this to happen makes for bad forecasts in the placebo exercises, substantiating this concern. These placebo forecasts are improved considerably by adding relatively weak restrictions like the ones here.[20] 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.

1.4 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 specification 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 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

[19]See Appendix Figure A3 for a diagram of this concern.

[20]Seven percentage points, the band chosen for the main analysis, is a wide window.

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.

1.4.1 Main Results

Figure 1.2 presents the results from the main specification detailed in the subsection 1.3.3.[21] 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.[22] These level differences come from an underlying comparison of growth rates between the treated and synthetic control directly. Figure 1.8 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 group has faster growth in a period of catch-up. This accounts for the decreasing differences in the later periods of Figure 1.2.

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

[21] 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).

[22] 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.

concern in regressions using cross-country growth rates as these data are known to have thick tails with many extreme values.[23] Figure 1.7 plots the density of level differences after 2 years (chosen because it is the largest point estimate in Figure 1.2).[24] The mean effect size, about 1.9 percent, comes from an underlying distribution with a large fraction of its mass above zero—see Figure 1.7 not from a few large outliers—most crises that are treated with an IMF loan beat their synthetic counterpart. The 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.

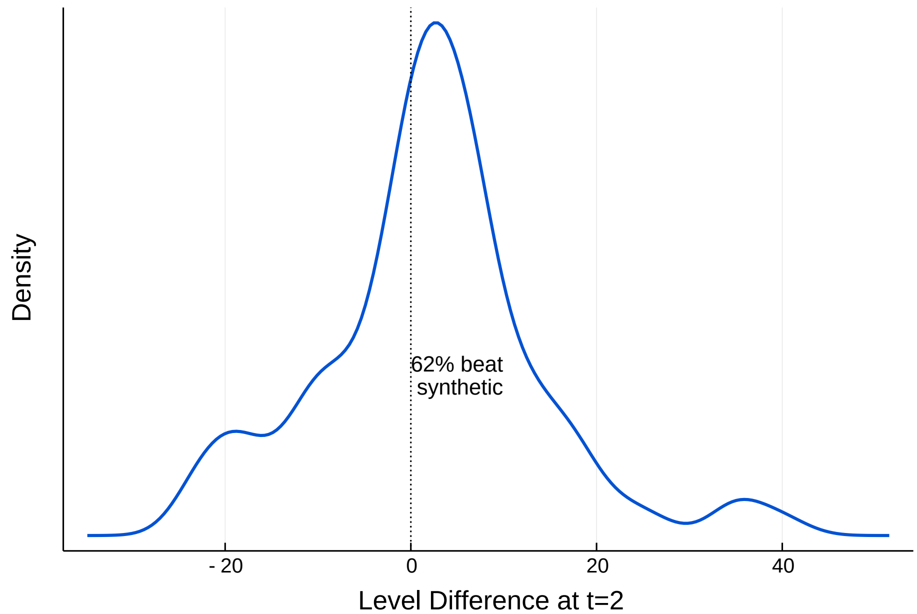
1.4.2 Analyzing the Quality of the Synthetic Controls

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 1.8, the synthetic controls are fairly successful at replicating the pre-crisis growth path. Recall 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, the appendix shows the full distribution of matches achieved by the minimization problem. Some crises and their syn-

[23] 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.

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

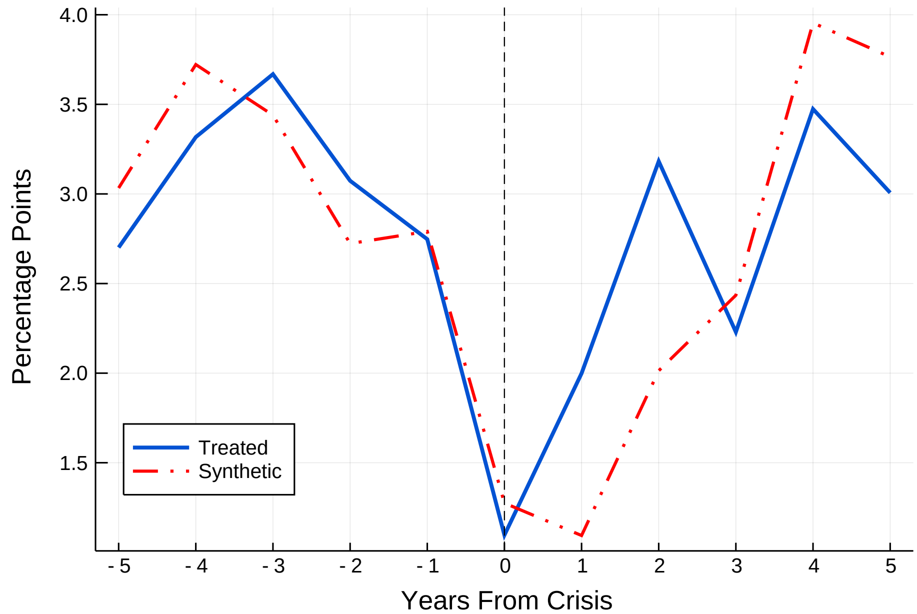
Figure 1.7: Distribution of Effect Sizes



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.

thetic 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 available 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 1.8—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 verify that these badly matched crises do not drive the pattern seen in Figure 1.2.

Figure 1.8: Growth Rates Estimates via Synthetic Control



Notes: Main results for growth rates. The solid line is now the average from the 93 crises with a non-empty set of eligible donors; the dotted red is the average of the synthetic controls constructed for each of these 93 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.

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 0.60 total synthetic controls.[25] Only 30% of untreated observations are not used at all and the maximum contribution of any untreated unit is one with a total weight of 3.1. The appendix 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 1.2 displays such comparisons. Along contemporaneous characteristics in

Table 1.2: 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)

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

[25]There are 157 untreated crises for 93 synthetic controls that need to be generated so each would get a weight less than 1 even in this case.

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 indicate 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 provide evidence in the following subsection that targeting the contemporaneous variables that may be of interest to the IMF does not change the results.

1.4.3 Robustness to Alternative Matching

This subsection performs a battery of specification modifications and shows that the main result is robust. Figure 1.9 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 Penn World Tables as opposed to the World Bank's World Development Indicators. Simon Johnson, William Larson, Chris Papageorgiou and Arvind Subramanian (2013) argue checks like this are necessary when using growth data given the large suspected errors in these poorly measured aggregates.[26] 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.[27]

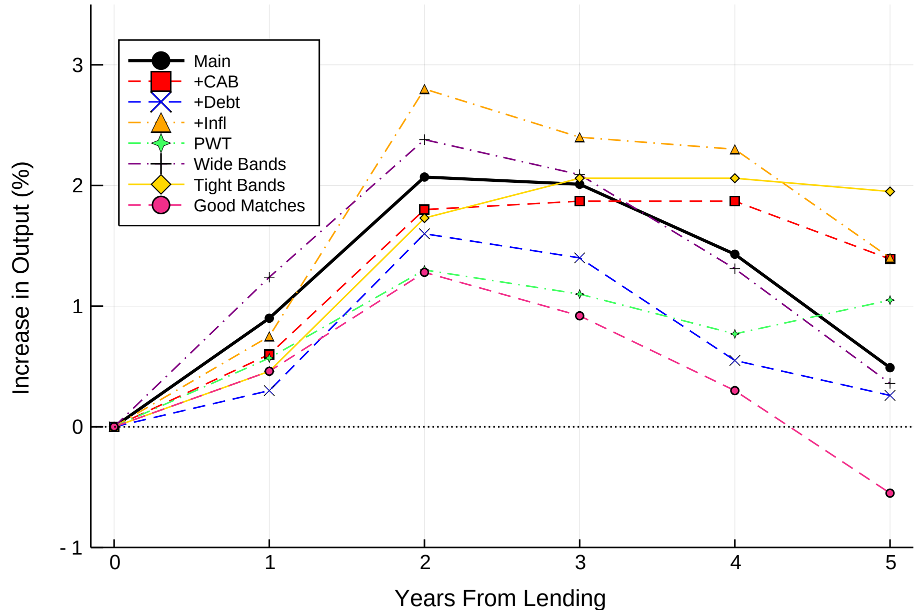
1.4.4 Identification Check Using Historical Forecasts

So far it has been shown that the SCM is an effective method to address selection on observables, this subsection provides evidence that selection on unobservables is unlikely to contaminate

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

[27]These late horizon differences are well within the one standard error bands of Figure 1.2 so I do not want to read into the point estimate differences.

Figure 1.9: Main Effects Are Robust to Alternative Matching Specifications



Notes: Robustness IRFs: each IRF comes from changing the main specification as detailed in section 1.4.3. "+CAB" adds contemporaneous current account balance to the target variables; "+Debt" adds contemporaneous external debt (removing current account balance); "+Infl" adds inflation. "PWT" runs the main specification with the Penn World Tables rather than the World Development Indicators. "Wide Bands" uses ± 8 as the cutoff for qualitatively similar crises; "tight bands" uses ± 6. Finally, "good matches" discards 10% of the observations with the worst match quality in the pre-period.
Source: Author's Calculations from World Development Indicators, World Bank; Valencia and Laeven (2012) and MONA Database, IMF.

the results. Recall this has two sides: countries apply for loans and the IMF chooses whether or not to grant them. I first study the 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 present evidence this is not the case.

I do this by estimating a simple regression model as in Equation (1.2):

$$Y_{i,t} = \gamma_0 + \gamma^f Y_{i,t}^f + \gamma_1 X_i^{SCM} + \xi_{i,1} \tag{1.2}$$

The regression in (1.2) fits a linear model to predict cumulative growth rates at some horizons, t, following a crisis. Here I present 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 (1.2) 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 1.3 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 is 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 ($\approx$ 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.

Table 1.3: 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)

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. For these reasons, the results in Figure 1.2 should be interpretted as causal evidence.

1.5 Transmission & Heterogeneity

In this section I attempt to examine, both directly and indirectly, the transmission mechanisms for the positive output effects that have been estimated in Figure 1.2. Using the SCM specification designed to create counterfactual growth rates is not guaranteed to produce good counterfactual evolutions of other variables.[28] For this reason directly disentangling the effects by comparing the treated and synthetic controls along other dimensions can only reveal so much—the exercise can roughly be thought of as a decomposition. 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. Additionally, heterogeneity in

[28]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.

the treatment effects can be studied both to verify the credibility of the main results and indirectly assess the transmission mechanism.

1.5.1 Government Spending Response

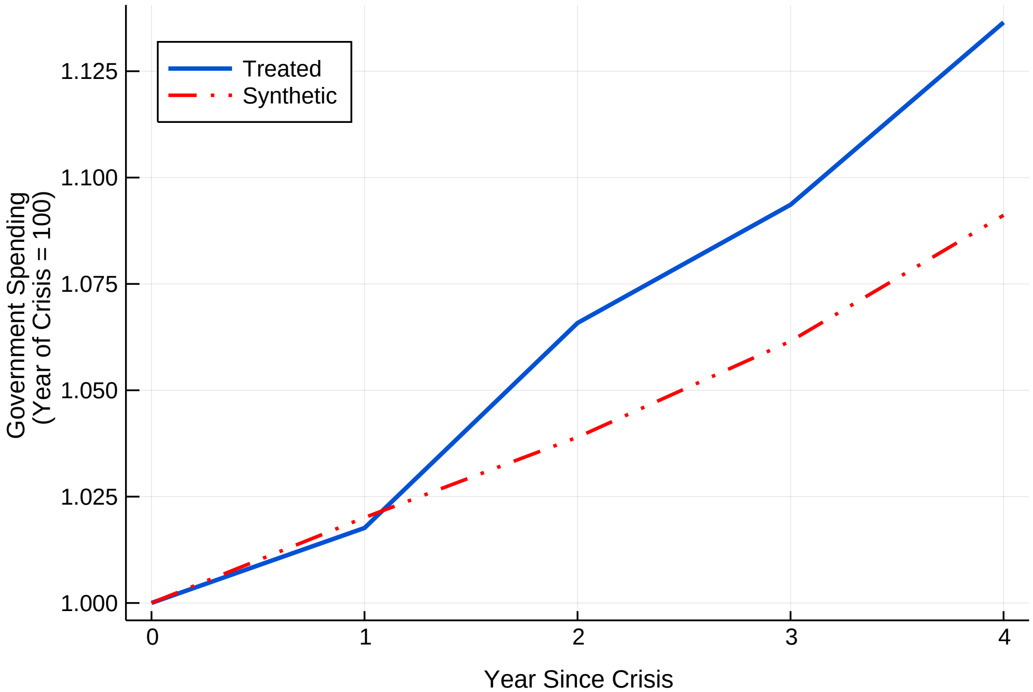
The most obvious place to look for evidence on the transmission of this effect is the evolution of government spending, G. This is analogous to the first-stage in Kraay (2012, 2014). His work runs a similar analysis using World Bank lending decisions to construct an instrument for government spending in order to estimate the associated multiplier. Importantly for the case here, Table 1.2 verified that government spending (as a share of GDP) was nearly identical in the treated and synthetics making the evolution of G for the synthetic group a reasonable counterfactual.

Figure 1.10 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.[29]

While not surprising, this is an important fact in light of typical business cycle policy in low-income countries that is over-whelmingly procyclical (Jeffrey A Frankel, Carlos A Vegh and Guillermo 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

[29]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.

Figure 1.10: 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.

(Ricardo J Caballero and Arvind Krishnamurthy, 2004; Alvaro Riascos and Carlos A 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.

1.5.2 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 is 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 1.7, 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 1.11 (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" (countries with the weakest state capacity) and take the average among the 16 underlying indicators as an 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 1.4 shows the average effect size by region and finds Africa and Small Island Economies to have the largest estimated effects from IMF lending.[30] These are regions that typically suffer from

[30]These numbers are going to be estimated with large error bands that are complicated to compute (recall the placebo inference discussion in Section 1.3). For this reason I just present the averages and read into them with caution.

Table 1.4: 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

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.[31] My own prior work further verifies the possibility of this channel (Kuruc, 2018*a*). In that paper—Chapter 3 of this dissertation—I show that for Fragile States[32] the start of an IMF program is associated with large increases in outside develop-

---

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.

[31] 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.

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

ment financing.[33] 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 Asia (Stiglitz, 2002), and more recently Europe (Olivier J Blanchard and Daniel 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 1.11 (b) shows the average effect size by a measure of exchange rate flexibility from Ethan Ilzetzki, Carmen M Reinhart and Kenneth S 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.

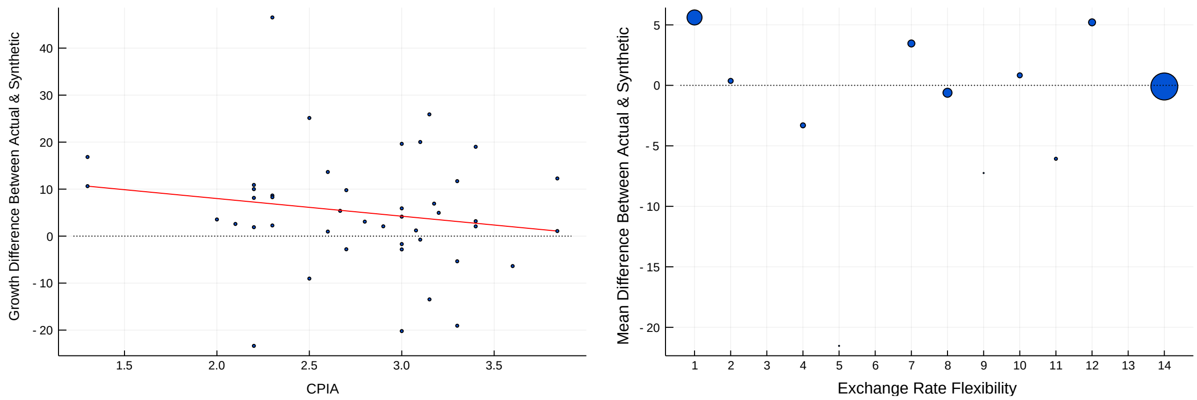
The results in this section help corroborate the main results and further advance a few alternative lines of inquiry with three key pieces of evidence. First, government spending increases more in countries that are treated than would be predicted by their counterparts; an obvious first-

[33] In that paper no distinction between financial crises and “normal times” is made.

Figure 1.11: 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.

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.

1.6 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, are effective 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.

Chapter 2

Quantifying India's Climate Vulnerability (with Melissa LoPalo, Dean Spears and Mark Budolfson)

2.1 Introduction

[1] The Intergovernmental Panel on Climate Change (IPCC) predicts an overall increase in the Earth's temperature over the next century due to climate change caused by human greenhouse gas (GHG) emissions, calling it "virtually certain" that there will be more frequent hot temperature extremes and less frequent cold temperature extremes experienced over most land areas (IPCC, 2007*b*). A large literature from the IPCC and other researchers has estimated or projected economic, health, and other costs of climate change, finding that the net effect on humans will be negative on balance, becoming more negative the more that temperatures rise and the more that other damaging dimensions of climate change affect humanity (IPCC, 2014*a*).

Much of this literature focuses on developed countries. Less is known about the adverse effects of exposure to higher future temperatures on health and economic activity in developing countries and emerging economies. As leading economists recently argued in *Science*, the focus in the prior literature on rich countries is "problematic, both because developing countries currently represent the majority of the world's population and greenhouse gas emissions and because

[1]The dissertator's main contribution to this paper is section 2.3—the planning, modeling, and writing. The dissertator was also involved in the writing of the entire paper and especially in planning which figures ought to be included to highlight the most interesting results in section 2.2.

the nature of impacts and context for policy choice could differ greatly relative to developed regions" (Marshall Burke, Melanie Craxton, CD Kolstad, Chikara Onda, Hunt Allcott, Erin Baker, L Barrage, R Carson, K Gillingham, J Graff-Zivin et al., 2016). Exposure to extreme temperatures is often greater in developing nations, which are disproportionately located in the hotter tropics. Harms conditional on exposure could also be greater: the poor may be less resilient to weather's impacts due to worse overall health. And poor populations may be less able to adapt by reducing exposure to extreme heat and humidity, such as via climate-controlled housing and indoor work.

This paper asks about the climate damages that Indian policy-makers can expect: what is the likely magnitude of climate damages, and how sensitive are they to the level of warming? In other words, how much worse would climate damages be for Indians under, say, 5° of warming rather than 3°? Understanding the magnitude of climate damages and how rapidly they increase as temperature change increases is critical for finding the right climate mitigation policy. Reducing emissions has costs, in part because emissions are a by-product of productive economic activity, and in part because cleaner fuel choices can be more expensive than carbon-emitting fuel choices. These costs are especially salient for a developing country such as India, where many households still lack reliable electricity, and where foregone economic growth implies an important loss of wellbeing for all Indians.

Public economics has a straightforward theoretical answer to externality problems such as climate change, where one decision-maker's action causes external harm to other people. Policy should be chosen so that the marginal social costs of reducing pollution equal the marginal social costs of the harm that is being averted. Still, applying this simple theory is difficult. One difficulty lies in even knowing the quantitative extent of the harm. Because climate change will impact many people—rich and poor; urban and rural; men and women; voting age citizens and their young children and future descendants—understanding the total sum of the harm requires comparing unalike consequences for unalike people (IPCC, 2014*b*; Francis Dennig, Mark B Budolfson, Marc Fleurbaey, Asher Siebert and Robert H Socolow, 2015).

Another well-known difficulty is the politics of collective action: the globally optimal policy package, if it could be enforced for the whole world, may importantly differ from what is in the interest of one country's population, especially the people alive at one time. Under the 2015 Paris Agreement, global mitigation policy will be made through countries' own bottom-up pledges (UNFCC, 2015; Mark Budolfson, Francis Dennig, Kevin Kuruc, Dean Spears and Navroz Dubash, 2019). To know what to pledge, Indian policy-makers need to know the stakes for India. Therefore, our research speaks to the question of what it would be rational for an Indian policy-maker to choose in the self-interest of the Indian population: present and future. As we will detail, when we tally the social costs of climate change, we consider only costs to the population of India.

In short, we find that the cost of climate damages for India is likely to be very large. Although India's climate vulnerability has been widely discussed in the prior literature, quantification is necessary for domestic analysis and policymaking. Moreover, emerging evidence suggests that Indians may bear an even greater share of global climate damages than is been previously understood. For example, because of the combination of heat and humidity of the Indian monsoon months, and because human bodies are more stressed by thermoregulation in humid air than in dry air, India may face a much larger early-life mortality burden from climate change than sub-Saharan Africa (Michael Geruso and Dean Spears, 2018). Among the many tragedies of climate change is the fact that India and other developing countries have not been responsible for much of the world's carbon emissions to date, but Indians nevertheless stand to lose much from climate change. Our quantification of these losses emphasizes the depth of the policy challenge: what is India's best, rational response to this climate injustice?

This paper reviews and integrates microeconomic and macroeconomic literature, in turn. Our analysis emerges from recent collaborative academic research by its authors, especially microeconometric research by Geruso and Spears (2018) and Melissa LoPalo (2019) about the consequences of heat and humidity in combination, and macroeconomic research by Mark Budolfson, Francis Dennig, Marc Fleurbaey, Noah Scovronick, Asher Siebert, Dean Spears and Fabian Wagner (2018) about the dependence of optimal mitigation policy on the unknown future trajectory for economic development of poor, climate-vulnerable countries. But we are far from the first to raise these themes, and we build upon an accomplished literature at the intersection of environmental and development economics (Michael Greenstone and B Kelsey Jack, 2015; Stephane Hallegatte, Mook Bangalore, Laura Bonzanigo, Marianne Fay, Tamaro Kane, Ulf Narloch, Julie Rozenberg, David Treguer and Adrien Vogt-Schilb, 2015; Dennig et al., 2015; IPCC, 2014*b*,*a*).

Section 2.2 considers microeconometric evidence. It considers causally well-identified effect estimates of harms of climate exposure, and uses them to project future damages within India under alternative possible futures for climate policy and outcomes. Section 2.3 presents macroeconomic projections. In this section, we make a novel application of the RICE climate-economy model, which was developed originally by the Yale University economist William Nordhaus (William D Nordhaus, 1992; William D Nordhaus and Joseph Boyer, 2000; William D Nordhaus, 2010). As a global model that explicitly represents different nations, RICE includes assumptions, based on scientific literature, that explicitly represent India's economy and quantify India's climate vulnerability. We use RICE to illustrate India's climate vulnerability by computing the magnitude of hypothetical near-term consumption losses to all Indians that would be an equal-sized loss to social welfare as climate damages. In other words, assuming a method for aggregating social harm across present and future Indians, what size of near-term economic disaster would be comparably bad and compelling for policy-making as climate damages are projected to be? These results will be underestimates, because in using the RICE damage function, we conservatively ignore the new evidence of humidity-based damages in section 2.2.

Section 2.4 briefly builds upon Michael Greenstone, Santosh Harish, Rohini Pande and Anant Sudarshan (2017) *India Policy Forum* study of air pollution. In contrast with climate damages, which will not fully unfold until future decades, India's population is already exposed to hazardous levels of air pollution today. The interaction between air pollution policy and carbon emissions policy is complex because particles in the air that harm human health can also reduce global warming, by reflecting away sunlight. Recent analysis by Noah Scovronick, Mark Budolfson, Francis Dennig, Frank Errickson, Mark Fleurbaey, Wei Peng, Robert H Socolow, Dean Spears and Fabian Wagner (2018) considers the balance between these mechanisms: for India, the health damages from air pollution dominate the computation and offer a compelling reason to simultaneously reduce air pollution and carbon emissions.

Our focus is on understanding and quantifying the damages that India can expect. Only in concluding section 2.5 do we turn to policy implications. What should Indian policy-makers do, in response to these grim facts? Elsewhere we have considered the easier question of what the globally optimal policy would be. In Budolfson et al. (2019), we use the same RICE model to show that the best global emissions policy would take into account inequality in world economic development and the fact that richer countries are more capable of making emissions cuts. So, the globally impartial, welfare-maximizing policy would have the rich countries such as the U.S. very quickly decarbonizing, middle-income countries such as India phasing out carbon emissions more slowly over several decades into the 21st century, and the very poorest countries in sub-Saharan

Africa perhaps continuing to produce some carbon emissions even in the early 22nd century.

But knowing what the globally optimal plan would be may provide little practical guidance to the leaders of India, or any other one developing or middle-income country. Decades of highlighting the injustice of developed countries' emissions policies has done little to change them. Nor, as we show in section 2.3, could India acting alone do much to reduce its own climate damages, even by entirely eliminating its carbon emissions. If India is to escape the climate damages that we project, it will require international policy change. India's leadership must approach the challenge of formulating a best response to climate injustice with an understanding informed by the sober facts of the vulnerability of its population. This may require India to make a strategic concession to protect Indians that is forced upon it by the injustice of rich developed nations who refuse to make equitable emission reductions. It is a moral tragedy if India must make such concessions because of the unethical behavior of others, but making these modest concessions may be the least bad of the tragedies that India realistically must choose between, as refusing to make modest concessions may expose Indians to the worst of the possible future harms from climate change.

2.2 Microeconometric Evidence: The consequences of heat and humidity

In this section, we introduce empirical evidence from microeconometrics about the effects of temperature and humidity on outcomes such as health and productivity. We then compute the implications of these estimate for future Indians, where the combination of heat, humidity, and poverty—especially in the subtropical states of North India—come together to create a unique context of climate vulnerability.

Temperature has been shown to affect many types of relevant outcomes, from human health, to crop yields, to the productivity of workers. Because researchers cannot observe the future climate, the only available empirical strategy is to compare populations and economies exposed to different weather outcomes (or, the same population at different times). But simply comparing countries with hot climates to countries with cold climates to learn about the potential impact of climate change is problematic, because climate may be correlated with other variables that are otherwise correlated with economic outcomes. To overcome these difficulties, an active literature in microeconomics uses short-term fluctuations in weather to make comparisons of hot and cold days (or months) *within a place*. This strategy allows researchers to learn about the impact of temperature and other weather variables separate from other correlated factors (Melissa Dell, Benjamin F Jones and Benjamin A Olken, 2014). This literature has documented impacts of weather fluctuations on outcome variables such as conflict (Solomon M Hsiang, Marshall Burke and Edward Miguel, 2013), health (Olivier Deschênes, Michael Greenstone and Jonathan Guryan, 2009; Alan Barreca, Karen Clay, Olivier Deschenes, Michael Greenstone and Joseph S Shapiro, 2016), and productivity (Marshall Burke, Solomon M Hsiang and Edward Miguel, 2015; Solomon M Hsiang, 2010). We show the implications of estimates from this literature for India by matching data on current temperature distributions and projections of future distributions under various climate change scenarios with effect sizes from these studies.

2.2.1 The Underappreciated Importance of Humidity

From a physiological standpoint, temperature is not the only weather variable that may be important for human well-being. One of the body's main mechanisms for cooling itself is sweating, which lowers temperature through evaporation. Sweating is particularly important at high temperatures. Humidity significantly interferes with evaporative cooling: when the air is saturated with more moisture, sweat evaporates more slowly, meaning that the body is less able to cool itself. The results could be dire: as Steven C Sherwood and Matthew Huber (2010) computed, when exposed to a combination of heat and humidity that is too extreme, the human body cannot cool itself because neither radiative cooling nor evaporative cooling from sweat will be successful. Under feasible bad-case scenarios for climate change, high heat and humidity could make spending several hours outdoors literally deadly in much of the land surface of the world where humans currently live, including much of South Asia.

Recent econometric studies corroborate humidity as an important moderator of the effects of temperature on economic outcomes, even at less extreme levels of exposure. Alan I Barreca (2012) shows that hot and humid days are most dangerous in terms of health impacts in the United States. This has implications for the distribution of health outcomes: these results imply that mortality rates will increase more in hot and humid climates than hot, dry climates as baseline temperatures increase. The literature on temperature and economic outcomes focuses primarily on developed countries, as data on both weather and outcome variables are more readily available in these contexts. However, this evidence suggests that it may be particularly important to understand the impacts of temperature in developing countries: developing countries are more likely to be located in hot and humid areas of the world. In addition, more people in developing countries work outside and fewer have access to adaptive technology such as air conditioning. For these reasons, developing countries are viewed as more vulnerable to the impact of humidity.

2.2.2 Climate Change and Infant Mortality

Motivated by this literature on human thermoregulation, several recent studies estimate the effects for developing-country outcomes of heat and humidity in interaction. Geruso and Spears

(2018) merge Demographic and Health Survey (DHS) data on month of birth and timing of infant deaths with gridded global weather data in four continents. In each country, the DHS collects full reproductive histories from a nationally-representative sample of women of reproductive age. These birth histories include the month of birth (and, when applicable, death) for each child, allowing the authors to match data on weather exposure to births occurring years before the interview. Because many babies are born in the same village in different years and months, their large sample of several million births allows them to identify off of surprise variation in the weather, while controlling for local seasonality, even specific to the village.

Like Sherwood and Huber, Geruso and Spears examine the impact of weather using a variable called "wet bulb temperature," which is a nonlinear function of temperature and humidity that gives a more complete portrait of outdoor conditions than temperature alone. In this literature, ordinary temperature is sometimes called "dry bulb temperature," to distinguish. Wet bulb temperature is the reading that would be given by a thermometer wrapped in a wet cloth; it is always lower than dry bulb temperature for relative humidity less than 100 percent. Geruso and Spears examine the impact of wet bulb temperature semi-parametrically, estimating the impact of replacing a day with a 60-70-degree wet bulb temperature with a day in 9 other bins. They find that hot and humid days in the month of birth predict significant increases in the probability of infant death.

Geruso and Spears estimate that an additional day in a month over 85-degrees wet bulb (approximately 32 degrees Celsius at 80 percent humidity) predicts about half an additional infant death per 1,000 births. In Figure 2.1, we apply the estimate derived from that study to Indian weather data to visually investigate the implications for climate change in India. In Panel A, we first perform a historical decomposition using average counts of experienced wet bulb days above 85 degrees between 2000-2010. This weather data comes from the Princeton Meteorological Forcing

Dataset, which gives information on temperature, humidity, and other weather variables for every 0.25-degree latitude and longitude grid point.[2] We multiply this count by the implied annual effect size. The resulting distribution shows how much lower infant mortality rates per 1,000 births in 2000-2010 would have been in each location if the 85-degree days were replaced by 60-70-degree days. The figure shows that these extremely hot and humid days have thus far been virtually restricted to the northern states of Uttar Pradesh and Bihar in India. Moreover, infant mortality rates would be as much as 3 per 1,000 births lower in some areas if days over 85-degrees wet bulb were replaced with mild days. This accounts for nearly 10 percent of the infant mortality rates in those regions during the period studied, a non-trivial fraction.

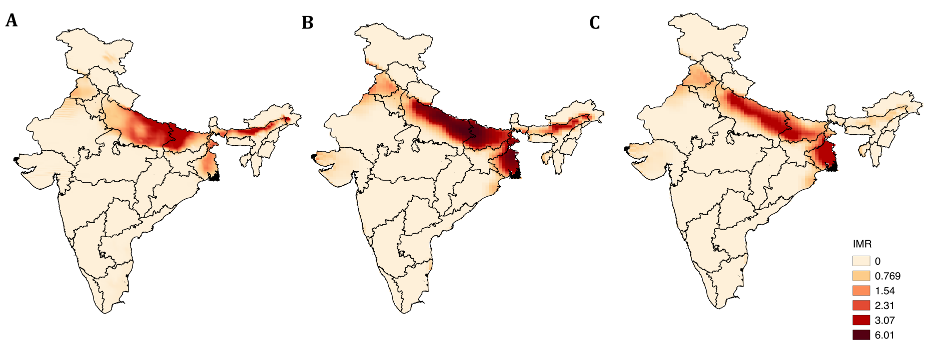
Panels B and C project how climate change may alter the situation depicted in Panel A. Panel B uses projections of heat and humidity obtained from the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP). These data use the Hadley Centre Global Environmental Model (HadGEM2) to predict temperature, humidity, and pressure at a 1-degree latitude/longitude resolution. These types of projections generally categorize predictions into "Representative Concentration Pathways (RCPs)," which characterize different assumptions regarding the trajectory of future greenhouse gas concentration. Panel B uses predictions under RCP 8.5, a pessimistic scenario in which emissions continue to rise throughout the 21st century under assumptions of relatively high population growth and relatively slow income growth, technological change, and energy intensity improvements. Panel B shows the results of these projections for India by 2050.

This map uses the same method as Panel A: it shows the increase in infant mortality rate due to the number of days with wet bulb temperatures above 85 degrees using the same effect

[2]This data is derived from a combination of observational weather data from sources such as satellites, weather balloons, and stations with a physics-based model that extends the data to observationally sparse areas.

sizes from Geruso and Spears (2018). Under this scenario, the ill effects of heat and humidity both spread to new areas in India and worsen in already-affected areas. Under this scenario, in addition to Uttar Pradesh and Bihar, the northwest and Eastern states become severely affected by the types of hot and humid days that have been shown to affect infant mortality. Still, these types of hot and humid days will continue to be concentrated in northern India.

Figure 2.1: Infant Mortality Rate Increases From Extreme Weather



Source: Princeton Meteorological Forcing Dataset; Inter-Sectoral Impact Model Intercomparison Project; Geruso and Spears (2018); Authors' Calculations

Panel C explicitly computes what is at stake when moving from a bad climate outcome (RCP 8.5) to a much better one (RCP 2.6) by showing the differential change in infant mortality under these scenarios. Under RCP 2.6, greenhouse gas concentrations peak mid-century and decline by 2100: an optimistic pathway for emissions. This differential infant mortality increase can be seen as the marginal cost of a bad climate outcome relative to a good climate outcome. Specifically this is calculated as the increase in IMR caused by the yearly count of 85-degree wet bulb days (again, the thought exercise is the excess in IMR over a situation where the 85-degree days are replaced with 60-70-degree days) for RCP 8.5 and RCP 2.6, respectively, and then the difference is taken. The result shows the excess IMR that could be prevented by achieving the RCP 2.6 pathway instead of RCP 8.5 and that the preventable deaths are largely concentrated in Uttar Pradesh, Bihar, and the Eastern states.

All three panels show changes in infant mortality rates, and therefore do not take into consideration the current population numbers or population projections in each place. However, these estimates indicate that these deaths will be taking place in some of the most populous regions in India; as of the 2011 census, Uttar Pradesh was the most populous state while Bihar was the third most populous, together accounting for about a quarter of India's population. These two states also have the highest fertility rates in the country, implying that a large portion of future births will continue to occur in these especially climate-vulnerable regions.[3] Furthermore, Geruso and Spears find that measures of wealth in the DHS do not significantly mediate the impact of wet bulb temperature on mortality, suggesting that even wealthy people in developing countries may be unable to avoid some of the effects of extreme heat and humidity.[4]

2.2.3 Climate Change and Labor Productivity

Infant mortality is an extreme form of climate vulnerability, but it is not the only relevant outcome likely to be affected by the increase in incidence of extremely hot and humid days. Another recent study suggests that this type of weather also significantly impacts labor productivity. LoPalo (2019) examines the impact of weather on a category of workers who are both significantly

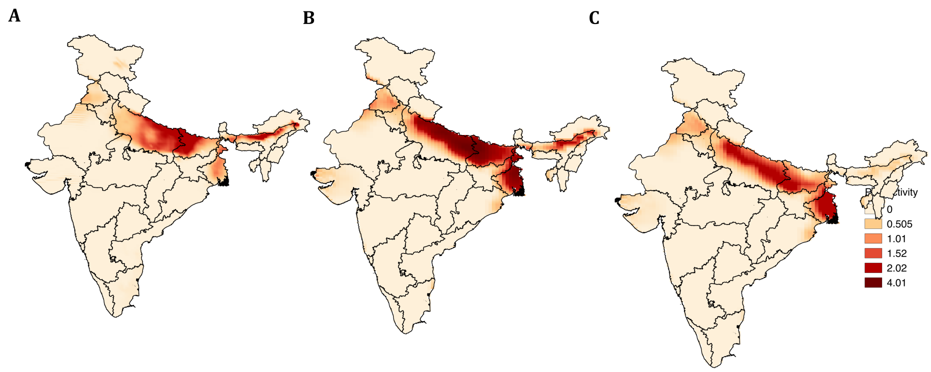
[3]Total fertility rates were 2.74 and 3.41 in Uttar Pradesh and Bihar, respectively, in the 2015-2016 National Family Health Survey, in contrast to 1.83 in Andhra Pradesh.

[4]Some prior literature has found that air conditioners moderated the mortality effects of high temperature in the 20th century United States (Barreca et al., 2016). This is plausible here as well, in part because air conditioners also reduce humidity. Geruso and Spears cannot test for this however, because they study developing countries where air conditioner ownership is sufficiently rare to be not measured in the DHS. In the 2005-6 Indian Human Development Survey, only a small fraction of a percent of households reported owning an air conditioner.

exposed to outdoor temperatures and possible to study using publicly available data: survey interviewers. In other words, LoPalo uses the DHS surveys to study the effects of exposure to the weather on enumerators as workers. She merges data from over 1.1 million interviews conducted in the DHS with data on temperature and humidity in the day of interview and examines the impact of daily average wet bulb temperature on indicators of productivity such as number of interviews completed per hour worked as well as measures of data quality. Her analysis shows that, on days when wet bulb temperature exceeds 85 degrees Fahrenheit, the number of interviews completed per hour declines by approximately 10 percent of the mean. The effects are driven by an increase in working hours rather than a decrease in interviews completed in a day; interviewing teams start earlier in the morning on these hot and humid days but do not complete their work earlier. She also finds that on hot days, the quality of work decreases: data quality problems are more common.

In Figure 2.2, we perform a similar exercise as in Figure 2.1, using the effects from LoPalo (2019). These maps plot the annualized estimate of the effect of temperature on productivity (number of interviews completed per hour in this case), multiplied by the number of high wet bulb days in each grid point. As in Figure 2.1, panel A depicts the impact of 85-degree wet bulb days under current distributions. It shows the impact that replacing each 85-degree day with a 60-70-degree day would have on annual productivity per hour. Panel B shows the same estimates projected on future distributions of temperature under RCP 8.5. Note that the distribution of wet bulb days is precisely the same as in Figure 2.1; what has changed is that the scale is interpretable as an effect on productivity, rather than infant mortality. Finally, Panel C shows the difference in impacts on productivity per hour under the RCP 8.5 vs. RCP 2.6 scenario. Again, these figures show that the greatest impacts will occur in the densely populated areas of Uttar Pradesh and Bihar as well as northeastern India.

Figure 2.2: Labor Productivity Decreases From Extreme Weather

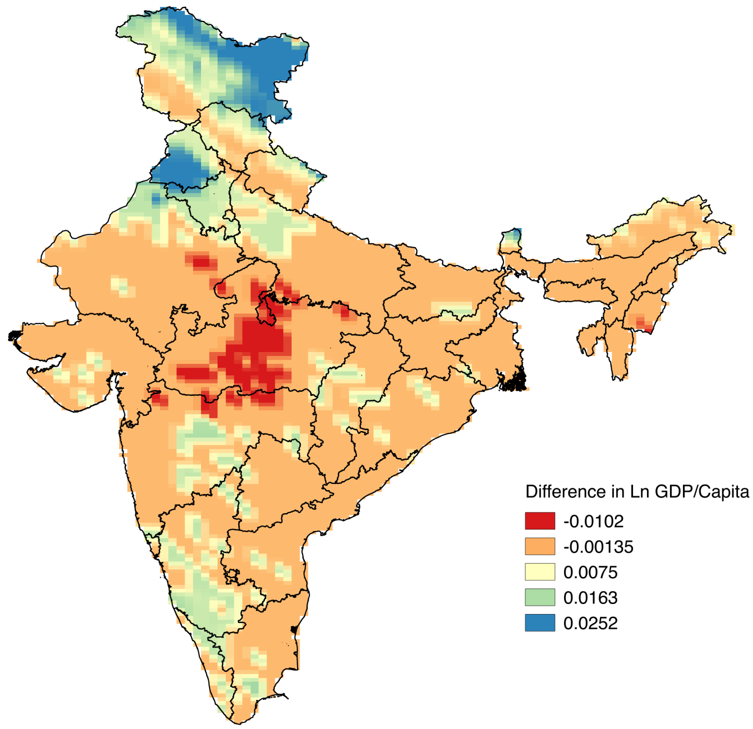


Source: Princeton Meteorological Forcing Dataset; Inter-Sectoral Impact Model Intercomparison Project; LoPalo (2019); Authors' Calculations

Infant mortality and labor productivity are only two examples of the wide range of outcomes that could be impacted by temperature. To get a full picture of the distribution of damages that could be caused by climate change within India, it is also useful to consider the evidence on the impacts of temperature on aggregate production (GDP). Several papers have established correlations between climate and aggregate productivity as well as causal relationships between fluctuations in weather and measures such as GDP. One such paper is Burke, Hsiang and Miguel (2015), which estimates the impact of average annual temperature on GDP per capita using a panel of 166 countries from 1960 to 2010. Similar to other papers in the literature, they make comparisons within countries, and they also difference out country-specific time trends. They find that production per capita is highest at around 13 degrees Celsius and declines sharply at higher temperatures, but also at colder temperatures. They project these estimates to estimate damages under RCP 8.5, finding that colder regions such as Europe may see productivity benefits under climate change, but regions that are warmer on average will see large damages.

Burke et al also conducted a country-by-country exercise to examine the implications of climate change for individual contexts. India is one of the most severely affected countries in the world by their estimates. We use their estimates to conduct an additional exercise to visualize the implied distribution of effects within India. In Figure 2.3, we implement a simplified calculation to show differences in growth rates that might be expected under RCP 8.5 vs. 2.6. We assume that local GDP per capita growth is determined only by annual average temperature, as estimated in Burke et al. We then calculate the growth rates implied by the projected average temperature in 2050 under RCP 8.5 and 2.6, respectively. Finally, we calculate the difference between the two rates, giving an idea of the distribution of impacts on GDP growth under the two emissions scenarios. The results suggest that the most populous areas of India will be significantly negatively affected under RCP 8.5 relative to RCP 2.6. The effects are as large as a decrease in GDP per capita growth of 0.06 percentage points per year. The blue areas in the map signify regions that will be positively impacted by warming: this occurs for areas with an annual average temperature of less than 13 degrees Celsius. A consistent theme across these results is that India is uniquely vulnerable to global warming given the humid climate of South Asia. Within India, climate damages will tend to be greater in the places where the population is already more disadvantaged: we find that Uttar Pradesh, Bihar, Madhya Pradesh and neighboring states tend to show more vulnerability in the projections presented above. Given current inequities, climate damages will not merely reduce the average well-being of the future Indian population; they are also projected to fall disproportionately on the most disadvantaged within India.

Figure 2.3: GDP Changes From Global Warming



Source:Inter-Sectoral Impact Model Intercomparison Project; Burke, Hsiang and Miguel (2015); Author's Calculations

2.3 Macroeconomic Projections: How Much are Climate Damages Worth?

Section 2.2 documented that many important economic and social indicators are vulnerable to temperature and humidity. However, a critical question remains: How does one weigh these costs in total? How should policy-makers aggregate the consequences of climate policy for the full Indian population, including people alive today and people who will not be born for decades to come?

To answer this question, we developed an India-centric Integrated Assessment Model (IAM) by modifying a global IAM in wide use in the climate policy literature—William Nordhaus' RICE (Regional Integrated Climate-Economy) model. Because RICE considers only (dry-bulb) temperature, not humidity, this section does too; the evidence in section 2.2 suggests that these results will therefore underestimate India's climate vulnerability. Our model projects total Indian climate damages to be extremely large. Quantitatively, the damages are as costly as a hypothetical reduction in GDP per capita of 25-30% for each of the next 20 years, which would be widely recognized as a substantial humanitarian disaster. However, as the model shows, these damages cannot be avoided by a reduction in India's emissions alone.

2.3.1 Overview of IAMs

IAMs are macroeconomic growth models with a climate component designed to quantify the economic tradeoffs associated with carbon emissions. The most widely used IAMs (DICE/RICE, PAGE, and FUND) share the same conceptual structure (William D Nordhaus, 2017; Nordhaus, 2010; Chris Hope, 2011; Richard SJ Tol, 1999). In the model, economic consumption generates well-being for the people who consume, but also results in (GHG) emissions. GHG emissions enter a climate module designed to track the stock of CO2 and calculate global temperature dynamics. Higher future temperatures then cause harm to future people according to a relationship called the "damage function."

To measure these trade-offs in a way that assesses the consequences for everyone, we use a standard social welfare function (SWF) that is additive across time. Equation (2.1) formalizes this.

$$W(c;\rho,L)=\sum_{t=0}^{Z}\frac{1}{(1+\rho)^t}L_tU(c_t) \tag{2.1}$$

2.3.2 Social Costs of Emissions in an IAM

Total social welfare is the sum of utility in each period from today ($t = 0$) until some end date ($t = Z$) generated by per-capita consumption, $U(c_t)$, multiplied by the population in that time, L_t, and discounted by $\frac{1}{(1+\rho)^t}$, which is a factor that makes future costs and benefits worth less to the social evaluation than nearer-term costs and benefits.[5]

Temperature, T_t, does not enter (2.1) directly because the models are constructed to deduct climate damages directly from the output available for economic use.

$$Y_t^N = \Big(1 - D(T_t)\Big) Y_t^G \tag{2.2}$$

Equation (2.2) defines net output in each period, Y_t^N, as some fraction of gross output, Y_t^G. The fraction lost, $D(T_t)$, is the damage function. This functional form implies some output is either spent in adaptation efforts (and is therefore unavailable for consumption) or is destroyed from high temperatures. The idea of temperature directly destroying output may be difficult to conceptualize, but it approximates two more realistic interpretations: (i) that more inputs are needed for the same level of output (productivity declines) or (ii) that more output is necessary to retain the same utility level (agents need to be compensated for the higher temperatures).[6]

We are interested in the tradeoffs relevant for an Indian policy-maker, so Equations (2.1) and (2.2) only include Indian inputs. For example, (2.1) is an India-specific social welfare function with projected Indian population and per-capita consumption in each scenario. Climate damages are losses to total Indian welfare from a warmer planet. Costs and benefits for people living outside of India are not counted.

[5] The utility function is assumed to have diminishing marginal returns, specifically of the CRRA form: $\frac{c^{1-\eta}}{1-\eta}$.

[6] This second interpretation is not exact because some fraction is saved rather than consumed, but it is close enough for expositional purposes.

In building towards aggregate damages, we start with decomposition of the social cost of an extra ton of GHG emissions—the social cost of carbon (SCC). This decomposition has a convenient multiplicative form that allows us to highlight each potential channel for damages to increase or decrease. The most uncertain and contested of these channels are the damage function and the social discount rate. We consider these closely below.

Mathematically, the SCC can be shown to be of the form presented in Equation (2.3) (Mikhail Golosov, John Hassler, Per Krusell and Aleh Tsyvinski, 2014).

$$\text{SCC} = \sum_{t=0}^{\infty} \frac{1}{(1+\rho)^t} L_t \frac{\Delta U(c_t)}{\Delta c_t} \frac{\Delta c_t}{\Delta T_t} \frac{\Delta T_t}{E_0} \tag{2.3}$$

The complex economic and atmospheric relationships we hope to capture can be simplified conceptually into 5 multiplicative terms.

i. $\frac{1}{(1+\rho)^t}$: The pure rate of time preference

ii. L_t: The population in time t

iii. $\frac{\Delta U(c_t)}{\Delta c_t}$: Increase in utility that results from an extra unit of per-capita consumption in time t

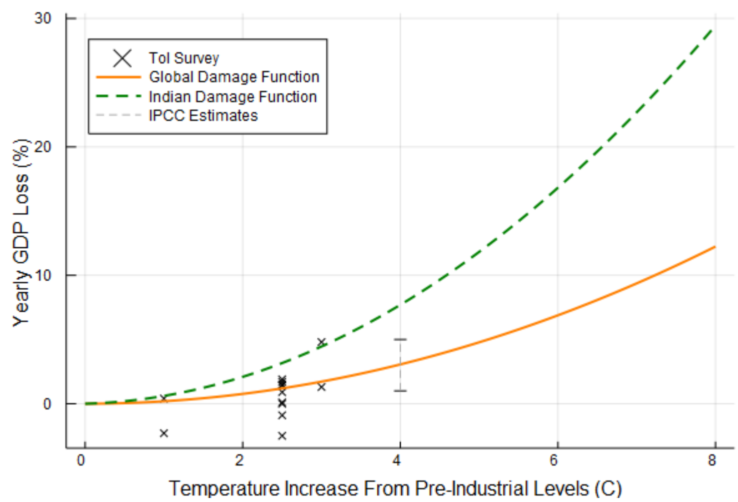
iv. $\frac{\Delta c_t}{\Delta T_t}$: Consumption equivalent losses that result from an increase in temperature at time t

v. $\frac{\Delta T_t}{E_0}$: Increase in temperature at time t from an additional emission *today*

In this paper, we take both (ii) and (v) from other sources. Population projections are taken from the United Nations, and the climate module that relates current emissions to a temperature profile appears to be uncontroversial. In fact, since we present the damages as a function of a temperature change, not emissions, (v) is not directly relevant to our main results.

Our Indian damage function (term iv) is derived by Nordhaus (2010) by scaling up a global damage function to reflect the consensus that India is more vulnerable than a globally-averaged damage function would imply. As documented in William Nordhaus and Paul Sztorc (2013) (and replicated in Figure 2.4) the global estimate is fit to the meta-analysis of Richard SJ Tol (2009).[7] The fitted function is restricted to be quadratic and is calibrated over a range of estimates from 1 to 3 degrees of warming.[8] India's damage function then takes the same functional form, but lies above the global function at all points.

Figure 2.4: Global and Indian Damage Functions



Source: Tol (2009); IPCC (2007*a*); Nordhaus and Sztorc (2013)

A challenge present throughout the IAM literature is that it is especially difficult to know how costly climate damages would be beyond 3°C of warming. We continue to follow Nordhaus

[7]The damage estimates in Tol (2009) are designed to include the monetary costs of optimal adaptation as well as the costs of lost output/well-being. For example, the costs of sea-level rise inclue the cost of building dikes and levees where possible (adaptation) and the cost of damaged/lost landmass where not (residual damages).

[8]While only calibrated on 1-3 degrees, the damage function sits in the IPCC range of estimates for 4 degrees.

(2010) by assuming the calibration at lower temperatures remains informative at higher temperatures. This results in substantial—yet unavoidable—uncertainty over a range of potential outcomes. Subsequent work suggests this uncertainty is one-sided: the DICE/RICE damage function used here is very likely a *lower-bound* for damages at high levels of warming.[9] Specifically, Martin L Weitzman (2012) presents a convincing case that the DICE/RICE implied damages are implausibly low for warming greater than 3 degrees. Likewise, Burke, Hsiang and Miguel (2015) estimate damages using a method less reliant on extrapolation and find a South Asian damage function nearly an order of magnitude larger than what we use here. Nordhaus (2017) himself has even adjusted damages upwards in his most recent work.[10]

Beyond this, no damage function in the IAM literature—including the Nordhaus (2010) specification that we use—considers increases in wet-bulb temperature. As documented in Section 2.2, the importance of humidity makes India more climate-vulnerable (relative to drier developing regions such as sub-Saharan Africa) in a way that has been previously omitted. In order to be grounded in the prior literature, our damage function, too, omits the potentially important role of humidity. Therefore, although the damage function remains a highly uncertain object, we conclude that our results are not driven by unrealistically pessimistic assumptions regarding the damages of climate change.

Terms (i) and (ii) of the SCC quantify the relative importance of damages faced by further-future people compared with damages faced by nearer-future people. These terms reflect the two justifications for discounting over time: (i) merely because damages occur in the future and (ii)

[9]See Delavane Diaz and Frances Moore (2017) for an extensive review of aggregate IAM damage functions.

[10]We use the Nordhaus (2010) version because it is a disaggregated model which allows us to pull India's damage function directly.

because damages are suffered by richer populations. Some combination of these two factors determines how much we ought to value losses to future populations.[11] This is important for our analysis because climate damages will unfold over coming centuries. A large literature in climate economics has recognized that optimal mitigation policy is substantially shaped by the choice of a discount rate: if the social evaluation assumes that the future does not matter, then it is unsurprising that models recommend unaggressive climate mitigation policy. Understanding the respective roles of these parameters is then critical to understanding our results. To reiterate, term (i) plays a simple role of discounting well-being just because it is experienced at a later date. The way term (ii) influences discounting, however, is less obvious.[12]

Term (ii) is the marginal utility of an additional unit of consumption. It is an uncontroversial consensus among social scientists that this changes with income: adding $1 to the budget of a poor person increases his or her well-being more than if we did the same for a richer person.[13] Throughout this literature, economists use functions in which a single parameter, η, controls the importance of extra money to a poorer person, relative to a richer person. This parameter is known as the "inequality aversion" of the model. Inequality aversion is important for discounting in climate policy if we expect future economic growth: because future Indians will be richer than

[11] The exact way these come together to determine the *total discount factor*, δ, under a constant rate of economic growth, g, is represented by the well-known *Ramsey Equation*.

$$\delta = \rho + \eta g$$

[12] Well-being is emphasized because ρ is a discount on utility, not goods. It may be reasonable (as we discuss in the next paragraph) to discount damages to future people because they will be wealthier, but this has nothing to do with ρ.

[13] Nordhaus (2010) and other regionally disaggregated climate-economy models use a solution technique called "Negishi weights" which results in a social welfare function that does not respect this cross-sectionally—$1 to a rich person is as socially valuable as $1 to a poorer person. We interpret Negishi weights as an attempt to solve for the model's equilibrium, rather than a rejection of cross-sectional diminishing returns.

present-day Indians, future money-losses are less important to policy makers than today's money losses to a poorer population.

2.3.3 Social Welfare Parameter Choices

There is a large literature documenting that differences in discount rates drive many of the academic disagreements on climate policy (see, for example, Nicholas Stern (2006); William D Nordhaus (2007); Martin L Weitzman (2007); Partha Dasgupta (2008); John Broome (2012); Hilary Greaves (2017)). After careful review of this past work we have come to agree with the authors who believe that total discounting cannot and should not be inferred from individual economic choices. In our view, ρ reflects the ethical choice of policy-makers: are future Indians as important as present-day Indians? On the other hand, inequality aversion η is, in principle, empirical: it reflects how human well-being increases with increasing levels of consumption. This parameter is unfortunately impossible to estimate in practice.

Therefore—as in essentially every study in the IAM literature—we choose baseline values of ρ and η, and present robustness checks with other values. We believe the appropriate choice of ρ is 0.[14] The list of authors that agree with this choice is long[15], and it follows from a simple argument that in the SWF all Indians, regardless of year of birth, matter equally. Suffering is no less bad whether it occurs 50 or 100 years from now merely because one is further away from us in time.

The parameter that governs the rate of change of marginal utility, η, stands on less firm

[14] In practice some very small positive number is used to follow Stern (2006) who makes an adjustment for the exogenous risk of extinction.

[15] Tyler Cowen and Derek Parfit (1992); Stern (2006); Dasgupta (2008); Broome (2012) are some notable examples.

grounding. We choose a level to match our prior work in Budolfson et al. (2018). To understand the parameter we choose ($\eta = 2$), consider two people, one twice as rich as the other. If the poorer person realizes some consumption gain, our baseline value of η implies that the wealthier person would need to receive four times that gain for it to be as socially good. Zero inequality aversion, in contrast, would imply the richer person would just need the same monetary gain for it to be as socially good, an implication we find implausible. Because any choice is subject to disagreement, we will present robustness checks with additional η values that correspond to the income gains needing to be 2.5 and 5.5 times as large, respectively, rather than the original 4.[16]

2.3.4 Quantitative Results

We can now quantify aggregate damages to India from climate change using the model and parameters just described. These damages are large, even though they do not include the humidity interactions described in section 2.2.

We quantify damages from climate change in terms of consumption-equivalent losses to current people: by what percent would per-capita consumption need to be reduced for the next 20 years to match the welfare losses associated with climate change? What reduction in near-term consumption would be just as bad, from the point of view of the social welfare function, as climate damages will be? Preventing a deep and sustained economic recession would be a top policy priority, so this is a useful way to calibrate the policy importance of climate damages.

[16]The main objection to our resulting discount factor is that individual savings behavior does not match what would be implied by the discount rate on goods we are using. We are not bothered by this. Even if we believed the SWF should be democratically determined (ie, correspond with individual preferences) savings decisions reflect how individuals plan to allocate their resources to their own individual futures. Personal impatience is a different consideration from how society values the lives of future generations.

In particular, the consumption loss that would be equivalent to climate damages is calculated as follows:

1. Exogenously warm the planet to a particular level and compute India's total well-being for all future periods under the resulting level of global warming.

2. Re-run this scenario without climate damages and instead reduce per-capita consumption for the first 20 years until total well-being from step (1) is matched.

3. Repeat (1) and (2) for various possible global warming scenarios.

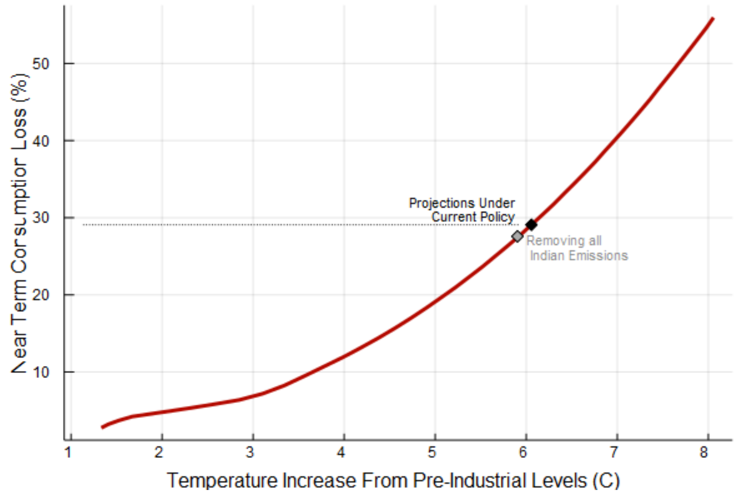
Without any further global mitigation policy, the economic collapse necessary to match projected climate damages is a 29% reduction in GDP per capita for each of the next 20 years. This would be a catastrophic loss. Figure 2.5 presents these near-term consumption equivalent damages under the baseline choices of ρ and η for a wide range of potential climate outcomes.

As Figure 2.5 shows, these damages have the potential to be extremely large. The rightmost point labelled on the curve corresponds to the global "business as usual" (BAU) scenario in DICE: no GHG restrictions are enacted beyond current policy, and mitigation comes only from private sector technological developments.[17] Under this outcome, many decades of the Indian population would experience climate damages amounting to about 15% of GDP.

Perhaps more important than the large level of damages is the slope of this function. At high levels of warming, changes in global temperature cause very large changes in Indian well-being. For instance, the planet is projected to warm by around 3.5 degrees if the national emissions

[17] This corresponds closely to the RCP 8.5 scenario.

Figure 2.5: Near-term Consumption Equivalent Losses



Source:Authors' Calculations

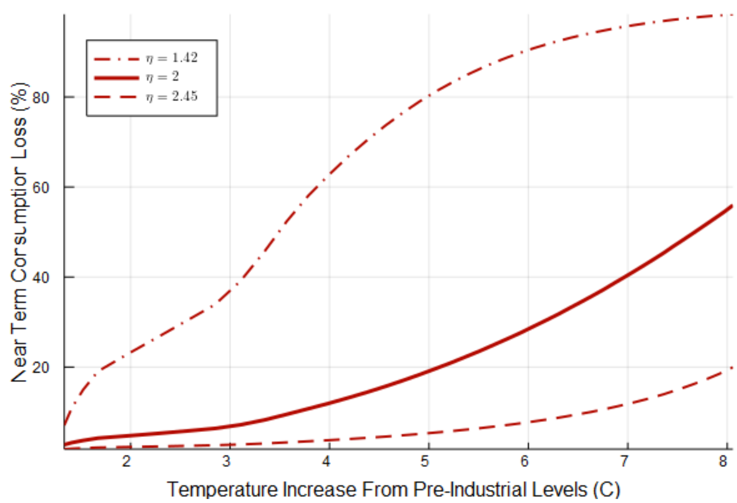
pledges in the Paris Accord are successfully realized (John Reilly, Sergey Paltsev, Erwan Monier, Henry Chen, Andrei Sokolov, Jin Huang, Qudsia Ejaz, Jeffery Scott, Jennifer Morris and Adam Schlosser, 2015). Climate damages would be cut by two-thirds despite warming being reduced by less than one half. Global efforts to reduce warming are especially valuable to India in light of this damage convexity.

A natural reaction to this quantification of India's climate vulnerability may be to suggest that India should quickly and unilaterally decarbonize. The RICE model also allows us to assess the consequences of such a policy. For better or worse, over the coming decades India's emissions are projected to remain a small fraction of the global, historical stock of GHG emissions. The dot to the left in Figure 2.5 shows that the peak global temperature would only decrease slightly if India were to altogether unilaterally eliminate its emissions. As a result, its climate damages would only slightly decrease. Indeed, the global temperature would probably decrease by even less than shown in the picture, because we do not model an endogenous response of other countries: India removing itself from aggregate energy demand would reduce prices and increase other countries' energy use. The message of the RICE model is clear: India is highly vulnerable to climate damages and cannot eliminate the problem by reducing its own emissions.

2.3.5 Robustness

Given the well-known importance and uncertainty over how to discount future costs we report the robustness of our results to alternative choices of the inequality aversion η.[18] Figure 2.6 plots how the results change with these higher and lower values of inequality aversion.[19]

Figure 2.6: Robustness to Inequality Aversion



Source:Authors' Calculations

[18]As we feel much more confident in our choice of ρ, we believe this uncertainty is the result of knowing how fast individual (and social) marginal returns to income diminish.

[19]See the paragraph direclty preceding this "Results" section for the discussion of η values chosen for sensitivity checks.

Because the model assumes that future Indians will be substantially richer than present-day Indians, changes to η are extremely influential for how bad climate damages are perceived to be.[20] Using smaller values of η (1.42 in this case) pushes the damages to very high levels (over 80% for 6°C of warming). But if η is large (2.45 here) total damages become notably smaller. In fact, this graph is conceptually bounded between 0 and 100 so these values span nearly the entire set of feasible outcomes. The fact that the results are heavily shaped by the choice of η is consistent with observations in Dasgupta (2008). However, our choice of η is not low relative to practice in the climate-economy literature so we take little comfort in the low damages associated with an unusually high value of η.[21] This is especially true given the conservative damage function we use.

2.4 Health Co-Benefits

Although the focus of this paper is on climate vulnerability, this section introduces an important near-term vulnerability of the Indian population with impacts for climate policy: air pollution. Reductions in greenhouse gas (GHG) emissions tend to lead to reductions in air pollutants, because both pollutants tend to share common emission sources (e.g. coal-fired power plants). As a result, reductions in GHG emissions are likely to lead to improvements in current human health through improved air quality. These benefits are often called health 'co-benefits' because they

[20] Assuming otherwise—that India will not experience rapid economic growth—would make climate damages even more important to social welfare because a poorer future population would experience the harm.

[21] Although Dasgupta (2008) urges authors to consider larger values for this parameter, the most influential IAM results (Stern, 2006; Nordhaus, 2010, 2017) all use a value less than 2 (some as low as 1). Micro evidence supports our choice as well: Fredrik Carlsson, Dinky Daruvala and Olof Johansson-Stenman (2005) use hypothetical survey questions about the well-being of grandchildren and estimate η to be 2 for intergenerational inequality. Studies directly using governmental behavior in tax policy to infer η in other policy-making spheres find values between 1.3 and 2 (Nicholas Stern, 1977; Frank Cowell and Karen Gardiner, 1999). See Dasgupta (2008) and Greaves (2017) for reviews on total social discounting.

are additional benefits that come alongside the direct climate-related benefits of GHG reductions. Emerging research suggests that these health co-benefits may be large, especially for a nation such as India in which air pollution is one of the nation's leading health problems. For example, according to recent data from the World Health Organization, 14 of the top 20 cities with the highest levels of particulate matter pollution in the world are in India (BBC, 2018). Interestingly, these cities are all located in Northern India, the same region with the highest level of population, fertility, and climate vulnerability in the country: 7 of these cities are located in Uttar Pradesh and Bihar.

Thus, health co-benefits have a critical place within India's climate policy decision-making, and are an additional source of benefits for India from GHG reductions. This is in part because large benefits occur quickly enough to be economically important even with high time discount rates: air pollution is already harming the population alive today (Scovronick et al., 2018). Furthermore—and of particular importance to Indian policymaking—these health co-benefits of GHG reductions can be almost fully captured by a large country such as India through unilateral domestic policy-making, as most co-benefits are realized domestically (in contrast to the fully globally dispersed climate-related benefits of GHG reductions), and co-benefits are not as vulnerable to being negated by the non-cooperative economic and policy response of other nations (in contrast to climate benefits, which are vulnerable to emissions leakage, as discussed below, and can also represent a transfer of GDP from the mitigating nation to other noncooperative nations).

Globally, the benefits from preventing air pollution-related deaths alone may outweigh the mitigation costs of reducing carbon emissions. Drew Shindell, Greg Faluvegi, Karl Seltzer and Cary Shindell (2018) examine the local health impacts of reducing emissions enough in the 21st century to achieve 1.5 degree warming rather than 2 degrees, finding that the drop in air pollution could prevent around 150 million premature deaths, mostly in Asia and Africa. They estimate the health impacts in individual metropolitan areas, showing that Indian metros such as Kolkata, Delhi, Mumbai, and Lucknow will be among the top beneficiaries in terms of number avoided deaths. Similarly, Anil Markandya, Jon Sampedro, Steven J Smith, Rita Van Dingenen, Cristina Pizarro-Irizar, Iñaki Arto and Mikel González-Eguino (2018) find that in some mitigation strategies, co-benefits of carbon emission reductions were almost double the costs in some areas, implying that mitigating enough to achieve 1.5 degree warning would have a net benefit for India, as well as China. Scovronick et al. (2018) find that optimal global mitigation results in immediate net benefits when climate costs, climate benefits, and co-benefits and co-costs are all jointly taken into proper account.

A large literature, recently surveyed by Greenstone et al. (2017), highlights the large costs of air pollution to the health of Indians and people in other emerging nations. Among these, burning coal may be especially important. For example, Aashish Gupta and Dean Spears (2017) estimate the impact of coal plants in India on the health of people living in the same district by studying districts where a new coal plant opened between the 2005 and 2012 waves of the India Human Development Survey. Because the survey visited the same households at the beginning and end of the seven-year interval, Gupta and Spears are able to show that reported respiratory health worsened over time in the districts that acquired a coal plant, relative to the districts that did not. Tellingly, the result is very specific: only respiratory health appears to be harmed by coal plans, not diarrhea or fever. Moreover, other types of new power plants—such as solar or hydroelectric—are not associated with worsening health, which is reassuring that the result is not spuriously due to electrification or economic activity.

One reason that air pollution is so harmful is that the impacts extend to essentially every-body, and are almost impossible to escape. In a recent south Delhi winter Sangita Vyas, Nikhil Srivastav and Dean Spears (2016) conducted an experiment regarding potential avoidance of these harms in an upper-middle-class flat in Green Park. Using air quality monitors the effectiveness of commercially available air filters were tested.[22] Under ideal conditions—never opening room doors, even to the interior of the house—the filters made a difference, but much pollution remained. Under a reasonably normal schedule of opening doors, much of what the filters achieved were erased. Part of the problem—reflected in the fact that indoor air quality remained highly correlated with outdoor air quality—is that even upper-middle-class flats in privileged neighborhoods often do not have window frames and door frames that prevent air from circulating. Perhaps unlike other contexts, such as drinking water solution, even rich Indians have little scope for buying their way out of air pollution.

In recent research that is currently under review, Scovronick et al. (2018) modify the same RICE model that we used in section 2.3: they incorporate an air pollution module, in order to optimize mitigation policy while taking into consideration both climate damages and the near-term harm to health from air pollution. The optimal policy balances countervailing forces: air pollution can be cooling, as particles reflect sunlight away from the earth. They find that the health co-benefits dominate, and recommend more rapid climate mitigation than if air pollution were ignored. Indeed, once health benefits are co-considered, it may be globally economically optimal to limit temperature rise to approximately 2 °C. This finding is especially relevant for India, where severe health costs of pollution are the inverse of the possibility of considerable health co-benefits. Their result suggests that health co-benefits could make aggressive mitigation policy individually rational for India, even if other countries are slower to decarbonize.

[22] These included both a relatively affordable filter and an expensive one.

2.5 Conclusion: India's Best Policy Response to Climate Injustice

Our quantifications show that India is highly vulnerable to climate damages. Our baseline macroeconomic approach suggests that climate change peaking at 5 ° C, rather than 3 ° C would be as detrimental to Indian well-being as a reduction in GDP by 17.5% for each year from 2020-2040. Our microeconomic results suggest that even this may be an underestimate because it ignores the humidity of South Asia. Clearly such a threat to near-term economic outcomes would be an overriding policy priority if political leaders anticipated it. If so, India's climate vulnerability should be a top priority too.

What is India's best response to these facts? As we have argued elsewhere, the Intended Nationally-Determined Contributions that richer polluters (such as the U.S. and the EU) have submitted under the Paris Agreement are inadequate, inequitable, and unjust (Budolfson et al., 2019). We believe that the richer countries should substantially reduce their emissions—quickly and without receiving anything in return—and should substantially fund the climate mitigation and adaption of poorer countries. But what should India do if they do not, as will presumably be the case?

There is no easy answer to this question. Faced with the dilemmas of international cooperation, some analysts suggest that India should "go it alone": either unilaterally eliminate/reduce its GHG emissions, or oppositely pollute as much as necessary to get rich enough to reduce its vulnerability to climate damages. But India cannot go it alone and reduce emissions enough to escape. One reason is limits to state capacity, of the sort that many developing countries face. As Greenstone et al. (2017) summarized in the India Policy Forum:

> A necessary requirement for command-and-control regulation to work is a very well-informed regulator with the willingness and ability to systematically enforce fair

> penalties in cases of non-compliance. In the main, this has been lacking in India. Duflo et al (2013) show how reliable data can be an elusive goal, and Ghosh (2015) identifies severe weaknesses in the enforcement mechanism.

Diane Coffey and Dean Spears (2017) make similar observations about a high-profile rural sanitation program: behavior change is difficult to promote; the small personnel-per-capita size of the Indian state limits capacity; and official statistics can be unreliable even on matters that are routinely measured by straightforward demographic surveys. Developing and promulgating sophisticated and detailed guidelines for the optimal regulation of emissions might, in this context, waste valuable time while having little impact.

The larger reason that India's emissions reductions would be inadequate is that there simply are not enough of them to tip the scales: as we computed in Section 2.3, even if India hypothetically fully eliminated its emissions while the rest of the world did nothing, it would still face almost as many degrees of warming. Worse still, it is unlikely that the rest of the world would be unchanged by India's unilateral decarbonization. Instead, India removing itself from global aggregate demand for fossil fuels might lower the price, so that some of India's emissions reductions would be offset by increases in other countries (this is often called 'emissions leakage').

Nor can India go it alone and escape through unrestrained GHG emissions to accelerate development. That is because the numbers do not realistically add up. Emissions are valuable, but they are not valuable enough to promote the economic growth necessary to enable India to escape via this strategy.

Therefore, India's best response to climate injustice may be first and foremost foreign policy, as well as domestic economic and health policy. The reason the question of what India should do is so challenging is that it depends on India's power to influence other countries' emissions.

One possibility—suggested by the large size of India's climate damages—is that India may have the option of achieving its climate policy goals via strategic international interactions that accept a creative concession in other sectors of policy-making in order to achieve reductions in the emissions of richer countries. We make no suggestions about what sort of non-climate concession (perhaps even a non-economic, symbolic concession) would be effective to offer; we merely note that India's climate vulnerability unfortunately suggests that a Pareto improvement could perhaps be found in the right packaging of a non-emissions concession from India, combined with a large emissions sacrifices from rich countries. How might such a package be invented? Perhaps one desirable feature is to engineer such a package to have time consistency between the concessions India makes and the emissions reductions that developed nations make, with antecedently agreed mechanisms for monitoring and adjustment in light of each side's subsequent compliance. For example, one can imagine trade concessions from India in exchange for deep emissions reductions, where the continuation of those concessions is contingent on reciprocal compliance. Or, perhaps the right package involves a concession in symbolic diplomacy, security policy, or another dimension of international politics—with the concession explicitly linked to and contingent on emissions reductions from China, USA, the EU, and perhaps others. Or perhaps a different package altogether is the best—the current point is merely to illustrate that desirable opportunities may exist for multilateral agreements between India and other nations that have desirable properties.

Inventing the right concession to offer would only be one challenge. Such a strategic concession would only make sense if high-emissions developed countries are sufficiently rational actors in international politics that they could be bargained with; perhaps they are not. The success of such a scheme would require international monitoring of rich-country agreements, so India can be sure it is getting what it bargained for. Efforts to create such monitoring standards should therefore be fully embraced by India. Even in the absence of an agreement between India and high-emission countries, it is to India's benefit that this data be transparently and consistently collected: its vulnerability and low emissions per capita result in it having much to gain and little to lose. Calls for credibility in GHG accounting may constitute a new reason that it would be in the interests of the Indian state to contribute to a norm of accurate official statistics.

It would be a moral tragedy if India must make such a strategic concession to protect Indians from the unjust emissions of rich nations. But climate change involves moral tragedies. If such strategic concession or other action is required and possible, it would be a mistake for India not to do at least what is in the interest of present and future Indians to protect them from the grave threat posed by unbridled climate change.

Chapter 3

The IMF and Fragile States: Assessing Macroeconomic Outcomes

3.1 Introduction

[1] This paper presents a quantitative analysis of the macroeconomic characteristics and performance of countries in fragile and conflict-affected situations (fragile states or FCS), especially in the context of their engagement with the International Monetary Fund (IMF). It provides supporting evidence for the IEO evaluation "The IMF and Fragile States." It investigates, among other things: (i) the persistence and evolution of fragility in individual fragile states; (ii) the macroeconomic performance of fragile states; (iii) the trajectory of economic conditions in fragile states associated with IMF lending; and (iv) the responsiveness of foreign aid flows to IMF program engagement (with or without IMF financing) in fragile states. The study contributes to the ongoing, intense debate on the issue of state fragility, recognizing that 80 percent of the world's needs for humanitarian assistance are driven by conflict and that, by 2030, nearly 50 percent of the world's extremely poor are expected to live in countries characterized by fragility.[2]

At the outset, we must first explain how we identify fragile states for the purpose of this analysis. When identifying such countries, the IMF, along with many others, has broadly relied

[1] This paper was previously published by the International Monetary Fund's Independent Evaluation Office—series number BP/18-01/05 (Kuruc, 2018*a*). Permission to reprint has been granted by the organization.

[2] Fragility, Conflict, and Violence Group, World Bank. Conflict states can be thought of as a subset of fragile states.

on the Harmonized List of Fragile Situations produced by the World Bank.[3] We too adopt this approach, but instead of using the yearly published lists, we use the World Bank's stated criteria to recreate consistent lists going back to the year 2000. This is done for two reasons. First, the Harmonized List is only available from 2010 onwards and we would like to analyze a longer period. Second, some CPIA data were unavailable to us and some appear to have been updated since the creation of the yearly list. For consistency, we remake the FCS list for each year by including countries satisfying the stated criteria based on available data. Details on this procedure are provided in Section 3.2.

Another aspect to keep in mind when analyzing the macroeconomic performance of fragile states is the quality of their national income data, which is generally considered to be questionable (Morten Jerven, 2013). In an attempt to obtain a more reliable indicator of economic activity, we complement the national income data by employing a novel technique of utilizing satellite images of light visible from space as a proxy for economic activity, as recently pioneered in the academic literature (Xi Chen and William Nordhaus, 2011; J. Vernon Henderson, Adam Storeygard and David N. Weil, 2012). The relevant details, as well as the theoretical gains associated with this approach, are discussed in Section 3.3.

The rest of the paper is organized as follows. Section 3.2, after identifying a group of fragile states going back to 2000, analyzes the evolution of fragility in individual fragile states, the macroeconomic characteristics of fragile states as a group, and the extent to which the IMF has been engaged with fragile relative to non-fragile states. Section 3.3 explains the methodology of using satellite images to proxy economic activity, and applies this methodology to quantify the

[3] See http://www.worldbank.org/en/topic/fragilityconflictviolence/brief/harmonized-list-of-fragile-situations.

relationship between the variability of economic growth and a measure of fragility. Section 3.4 uses event-study methodology to assess the impact of IMF lending on economic growth in fragile states and explore how foreign aid flows to fragile states respond to IMF program engagement. Section 3.5 presents conclusions. Finally, the appendix provides details on fragile state classification, the mathematics of the methodology to use satellite images to proxy economic activity, and the event-study methodology.

3.2 Characteristics of Fragile States

3.2.1 Fragility Definition and Persistence

As a first step, a time-consistent means of identifying a fragile state is needed. Our definition broadly follows the criteria used to classify fragile states for the World Bank's harmonized list. We use a dynamic definition, that is, a classification of fragile states that changes from year to year, allowing countries to transition into and out of fragility. In contrast, a static list, which takes a list of fragile states in a given year (say 2015) and holds it fixed for the entire sample period, cannot be used to measure the persistence of fragility or even changes in economic performance over time. To see this, suppose that there is an economic surge associated with leaving fragility.[4] If the only countries classified as fragile over the previous 15 years are the countries that are still fragile at the end of the period, the true underlying distribution of growth outcomes will be severely underestimated, and, as the length of time from the year used to create the static list increases, the misrepresentation will become more severe. Countries that exited fragility must be included for a truly representative average.

[4] It has been observed that an end of conflict is often followed by a surge in economic activity.

This paper defines a country as fragile in a particular year if it meets one of the following two conditions. First, the country's average Country Policy and Institutional Assessment (CPIA) score (provided separately by the World Bank and the Asian Development Bank or the African Development Bank) is below a score of 3.2. The CPIA is the average of subjective rankings of 16 governance indicators intended to capture state capacity. We use the World Bank's CPIA scores if the scores are not available from the Asian Development Bank or the African Development Bank (see the appendix). Second, there has been a peacekeeping or peacebuilding operation in the country in the last three years.[5]

Using this definition, we produce a list of fragile states for each year from 2000 to 2017; 60 countries were classified as fragile at least at some point during these years, with the number fluctuating between 32 and 42 from year to year (Table 3.1). Using the dynamic lists of fragile countries, we estimate a histogram of total years that a fragile state was labeled as fragile between 2000 and 2017 (Figure 3.1). Seventeen countries were labeled as fragile in each of the 18 years; for these countries fragility seems an almost permanent state. However, 24 previously fragile countries are not on the 2017 (or FY 2018) harmonized list. While these data cannot be used to calculate the average length of fragility, the changes highlight the importance of using a dynamic definition for analysis of fragility.[6] Several academic works conclude that the level of state capacity is a highly persistent variable (Lisa Chauvet and Paul Collier, 2008; Matthew Andrews, Lant Pritchett and Michael Woolcock, 2010). The fluid nature of fragility shown in Figure 3.1 is not meant to challenge these claims. Transitions will artificially look more common when using an arbitrary

[5]Missions by the United Nations (UN), African Union (AU), European Union (EU), Organization of American States (OAS), and North Atlantic Treaty Organization (NATO) are listed on the World Bank's website.

[6]Suppose that a country has been classified as fragile for the entire 18-year period. We cannot determine the "true" length of fragility until it ends. Therefore, the observed sample does not allow computation of the true average length.

Table 3.1: Number of Fragile States by Year, 2000-17

Year	Number of Fragile States
2000	41
2001	39
2002	39
2003	36
2004	40
2005	40
2006	37
2007	34
2008	34
2009	32
2010	34
2011	37
2012	37
2013	42
2014	40
2015	39
2016	35
2017	36

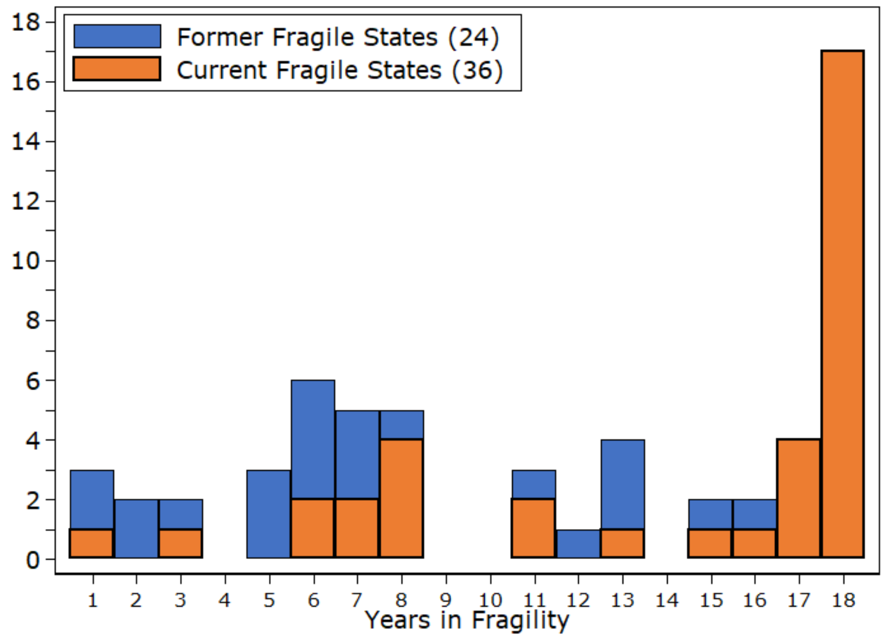
Source: Author's calculations.

cutoff point (such as a CPIA of 3.2) derived from an underlying continuous input, especially if the input variable has high-frequency fluctuations. This is an unfortunate reality for any analysis that compares groups determined by a continuous underlying variable.

3.2.2 Economic Outcomes

Based on the dynamic lists of fragile states, we compare key macroeconomic variables across fragile and non-fragile states. First, Table 3.2 presents aspects of living standards in fragile

Figure 3.1: Persistence of Fragility, 2000-17



Source: Author's calculations based on: CPIA data from the World Bank and African Development Bank; peacekeeping records from the UN, AU, EU, OAS, and NATO.

vs. non-fragile low-income countries (LICs) for 2014.[7] Here, GDP per capita is taken from the IMF's World Economic Outlook (WEO) database, but the averages are qualitatively similar in the Penn World Tables, the leading source among academic researchers for comparisons of purchasing power parity (PPP)-adjusted GDPs. "Access to electricity" and "mortality rate" come from the World Banks' World Development Indicators (WDI). The uncertainty underlying these data will be discussed in more detail later in the paper; for now, these will be regarded as the best available estimates. Two contrasting observations can be made about these estimates for LICs. First, if

[7]The year 2014 is the latest year for which the World Development Indicators provides consistent information on mortality and electricity.

Table 3.2: Standard of Living in Fragile vs Non-Fragile States Low-Income Countries, 2014

	Variable	*Fragile*	*Non-Fragile*
Unweighted	GDP per capita (PPP $)	2241	4535
	Access to electricity (%)	42.4	62.3
	Mortality rate (per 1000)	9.0	7.5
Weighted by Population	GDP per capita (PPP $)	2311	2632
	Access to electricity (%)	42.9	44.7
	Mortality rate (per 1000)	9.1	7.3
	Observations	30	33

Source: Author's estimates based on WEO database; World Development Indicators.

we use a simple average, GDP per capita in fragile states is approximately half as large as that in nonfragile states, and the share of population with access to electricity is two-thirds as high. Second, however, when weighted by population, the gaps between fragile and non-fragile states diminish significantly. In fact, given the uncertainty surrounding these estimates, one cannot be confident that those living in a fragile state are poorer on average than those living in a non-fragile state.

Table 3.3 takes advantage of the dynamic lists of fragile states constructed to report the average macroeconomic performance of a fragile relative to a non-fragile state over the 2000–16 period.[8] Given the dynamic definition, the table captures the trajectory of macroeconomic performance of a yearly cohort of fragile states. Overall, GDP growth seems to be somewhat lower for fragile states, though subject to a large standard deviation.[9] Inflation is higher and external debt larger, while tax revenue is lower. However, fragile states experience smaller current account

[8]As an examle, Cambodia, which is listed as fragile for 2000-7, is included in the "non-fragile" column for 2008-16.

[9]Assuming that the data are perfectly accurate, and that clustering should be done at the country level, this claim is not significant at the 5% level, but it is so at the 10% level. If either of these assumptions does not hold, there is even greater uncertainty about the validity of this claim.

deficits, likely reflecting the fact that they have more limited access to foreign borrowing. These conclusions remain the same (except that external debt looks more comparable) if we use medians rather than means.

Table 3.3: Economic Performance of Fragile vs Non-Fragile States, 2000-2016

	Fragile	*Non-Fragile*
GDP growth (%)	3.7	4.5
Inflation (%)	9.7	6.3
External debt (% of GDP)	74.6	55.6
Tax revenue (% of GDP)	12.1	16.0
Current Account Balance (% of GDP)	-6.1	-9.3
Observations	427	606

Source: WEO database.

One omission from Table 3.3 is the volatility of growth. LICs, and fragile states in particular, are typically thought to suffer from less stable growth. However, this claim is statistically indistinguishable from these countries having "noisier" estimates of GDP. If fragile states, with their lower administrative capacity, are providing less stable estimates of GDP growth, then growth could appear more volatile regardless of the true underlying pattern. The statistical details will be dealt with later, but this paper shows below that fragile states do seem to experience greater growth volatility based on data that are independent of the reliability of measurement of national accounts.

3.3 The IMF's Program Engagement with Fragile States

We now examine the IMF's interaction with the donor community in fragile states. To begin with, it is relevant to know whether fragile states have received financing from the IMF proportionately more or less than non-fragile states. Table 3.4 reports the shares of country-years in which the IMF disbursed any funds, for fragile and non-fragile states, during 2000–16. Of the

Table 3.4: IMF Financial Engagement, 2000-16

	Fragile	Non-Fragile	Fragile	Non-fragile
Has Financing (%)	35	23	38	44
Observations	646	2234	482	688
MICs Included	✓	✓		

Note: "Has Financing" computed as the count of country-years with any IMF disbursement divided by total observations in each category.
Source: Author's estimates based on data from IMF Finance Department.

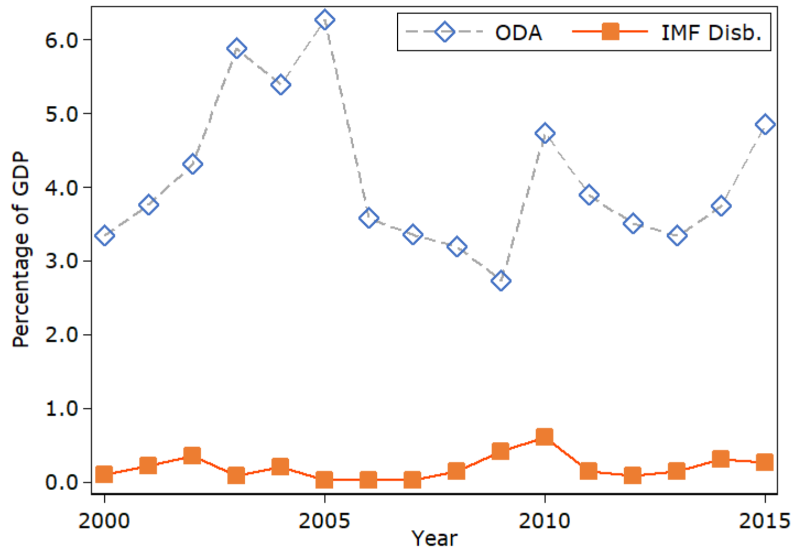
fragile state country-years, 35 percent involved some IMF financing, compared to 23 percent for all non-fragile LIC and MIC country years. However, focusing on LICs only, we find an opposite pattern: 38 percent of fragile LIC country-years involved some IMF financing, compared to 44 percent for non-fragile LICs. Their smaller share may represent the fragile states' greater difficulty in agreeing on a lending arrangement with the IMF or a greater propensity of FCS arrangements to go off-track, given the lack of administrative capacity and the more challenging political environments in these countries. It is also possible that fragile states, especially in a post-conflict situation, received generous official development assistance (ODA) from donors, minimizing the need for IMF financing. This last possibility is consistent with the relatively small financing role the IMF has played in fragile states (Figure 3.2). A second related issue is the importance of IMF financing to fragile states. The most natural indicator for this is how large IMF disbursements are relative to other sources of external financing.

3.4 Use of Satellites to Proxy Economic Activity

3.4.1 Methodology to Use Satellies to Complement National Income Data on Activity

The poor quality of national income data in many developing countries is a constant concern of the international development community (Chen and Nordhaus, 2011; Henderson, Storey-

Figure 3.2: Gross Financing to Fragile States



Note: IMF disbursement defined as any financial resources provided by the IMF to the country concerned. The dynamic definition of fragile states is employed, so the changes over time are a combination of changes per fragile state and a "composition" effect as countries enter and leave fragility.
Source: ODA from the OECD website, IMF disbursements from IMF Finance Department, GDP data from the WEO database.

gard and Weil, 2012; Jerven, 2013; Johnson et al., 2013). In an analysis of the Penn World Tables by Johnson et al. (2013), for instance, revisions between versions 6.1 and 6.2 created such a large variation that the same countries appeared on a list of top-10 and bottom-10 performers in Africa over the same 25-year period. Massive revisions are common in national income data. The most extreme example is for Equatorial Guinea, which was considered a bottom-10 performer in version 6.1 but was the second-highest performer in version 6.2.

Poor data quality is particularly problematic when comparing growth rates. For example, it is substantially less likely that measurement error could lead us to erroneously conclude that the United States is wealthier than Haiti than to erroneously conclude that the United States grew faster than Haiti in a given year. Over a large enough sample, measurement error in growth rates should approximately cancel out, assuming that the mean measurement error is zero. However, when using smaller samples of fewer than 50, as will be done below when assessing the impact of IMF arrangements on fragile states, increasing the precision of the dependent variable can yield substantial improvements in statistical power

In this paper, lights visible from space will be used as an independent source of economic activity estimates. It is now well known that there is a high correlation between how bright a country is in satellite imaging—what will be referred to as "luminosity"—and its level of economic activity (Chen and Nordhaus, 2011; Henderson, Storeygard and Weil, 2012). Not only is long-run development visible from space (e.g. more transport networks, broader access to electricity), but nearly all economic transactions at night require some light. Consider retail shops staying open later when shopping increases, or a manufacturing plant working overtime hours to fill an influx of orders. While this relationship is bound to have its own measurement error, if some new information is provided from a wholly independent source it can be leveraged to reduce total error (even if the new estimate has substantially more measurement error than the original).

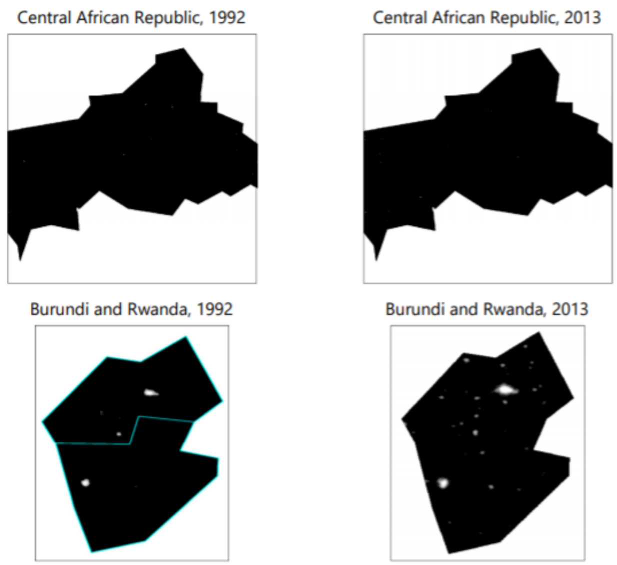
Chen and Nordhaus (2011) and Henderson, Storeygard and Weil (2012) provide useful discussions of what exactly the satellite data are, and the discussion here will closely follow these studies. The data come from the Earth Observation Group in the United States National Oceanic and Atmospheric Administration (NOAA). Each data point is a small pixel with a luminosity score between 0 and 63. To appreciate how fine these pixels are, consider that the surface area of the United States yields more than 16 million pixels. For each country, the pixels are averaged to produce a country-specific luminosity score; this is an appropriate method because the pixel scores are designed to be comparable in proportion (i.e., a pixel with a luminosity score of 60 is approximately twice as bright as a pixel with a score of 30). The country-specific luminosity scores will be the metric used in this paper. Given the obvious role that population density, technology, and culture play in relating luminosity to GDP across countries, only changes in luminosity will be considered. The raw data are yearly images, available from 1992 to 2013, created by the NOAA, which takes averages across all cloud-free images of that year. In some years, more than one satellite performs this operation; for the purposes of this paper their output is averaged.

Before constructing economic activity estimates from satellite images, two visual examples will be useful (Figure 3.3). It is difficult to visually observe small changes in luminosity, so these examples will be over the entire 1992–2013 horizon. Consider first the Central African Republic (CAR). This country is estimated to have had a cumulative GDP growth rate of negative 5 percent and 8 percent, respectively, by the WEO database and the Penn World Tables (PWT). Unfortunately, the CAR is extraordinarily dark to begin with (and end with, given the lack of growth), so the exact luminosity is difficult to see. Fortunately, not many countries experienced such weak growth over this horizon.

Burundi and Rwanda, by contrast, experienced much more robust GDP growth over this period. Burundi's GDP is estimated to have grown by 32 percent and 53 percent, respectively, by the WEO database and by the PWT, and Rwanda's by 100 percent and 87 percent, respectively.[10] As can be seen in the contrasting images, much more light growth has taken place in these countries than in the CAR, and more in Rwanda than in Burundi—which is consistent with a positive relationship between lights and GDP growth.

[10] To be clear, an estimate of 32 percent growth is still poor over a 20-year horizon.

Figure 3.3: Examples of Luminosity Growth



Source: Author production based on stable-average lights data from the NOAA.

3.4.2 Identifying the Economic Growth Characteristics of Fragile States

To construct a proxy for growth in economic activity, by combining national income data on GDP growth with luminosity data (see the appendix for the mathematical details), we:

1. Estimate the lights-GDP relationship;

2. Generate a composite measure of light growth for each year mitigate measurement error from satellites;

3. Take these composite light growth measurements to generate alternative GDP estimates strictly from satellite data, using the relationship from step 1;

4. Use outside sources to estimate the error in national accounts to construct the optimal weight

on predicted growth and reported growth for composite estimates;

5. Feed predicted growth from satellites and growth estimates from national accounts into this weighting for new, better, GDP growth estimates.

While this may seem like a convoluted process, the steps are grounded in statistical and econometric theory and are well known to increase statistical power. In a setting with highly uncertain estimates of GDP and a small sample, any gain in statistical power is critical. It turns out that the resulting revisions are fully consistent with the predictions of the statistical theory.

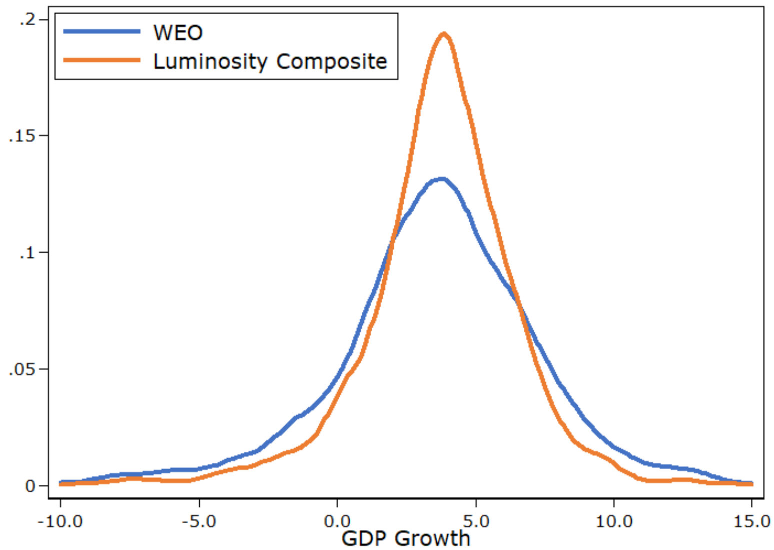
Figure 3.4 depicts the density functions of economic growth as reported by the WEO database and the resulting augmented growth when luminosity data supplements these estimates.[11] In the augmented estimates, there is a larger mass surrounding the mean (somewhere around 4 percent). With measurement error in reported GDP growth, reversion to the mean would be expected; countries reporting deep recessions, for example, are likely to have had a negative reporting error, and vice versa. Satellite data will help correct the cases where luminosity growth does not support that the country suffered a deep recession.[12]

One immediate result that can be confirmed without any econometrics is a difference between countries in the volatility of growth rates. Since there is little reason to believe that measurement error coming from satellite data is substantially worse for fragile states, if the volatility of luminosity growth is larger for fragile states this provides a good check for the results relying on

[11] The tails have been cut off so the distance is visible near the mean. There are extreme growth spurts and recessions recorded in the data that appear less extreme when satellite data are used to augment reported growth data.

[12] To be clear, luminosity does not always revert to the mean. Some growth episodes are revised upwards due to exceptionally strong luminosity growth. Figure 3.4 does not imply the same ordering of countries, but shows that if satellite data do in fact correct measurment error, the expected true distribution should be "tighter" around the mean.

Figure 3.4: Density of Augmented vs. WEO Growth Rates, 1993-2013



Note: Sample period covers 1993-2013 when satellite estimates are available. Details of combining luminosity data and GDP data presented in the appendix.
Source: Autor's calculation based on the WEO database and luminosity data from the NOAA.

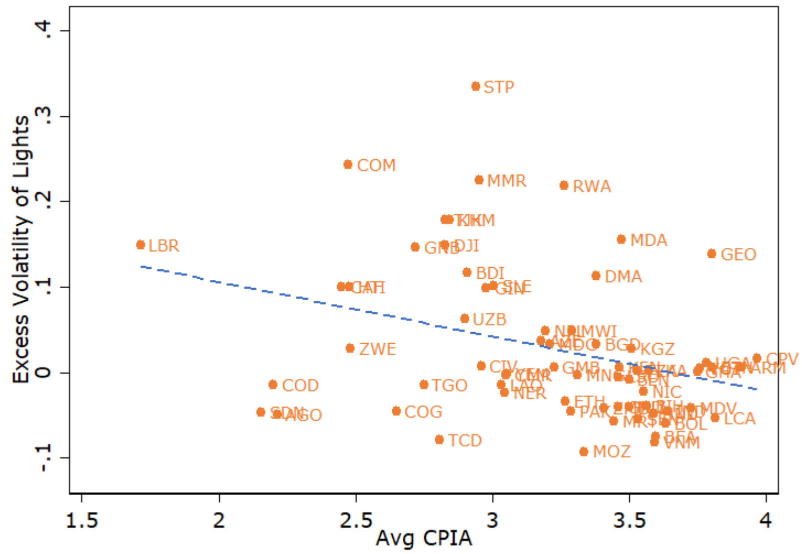
national income data. Figure 3.5 plots the volatility of luminosity growth against CPIA scores (the continuous measure of state capacity), and confirms that a higher CPIA score tends to be associated with larger volatility of luminosity growth.

3.5 Assessing the Impact of an IMF Arrangement on Fragile States

3.5.1 Event-Study Methodology

With revised estimates of GDP in hand, one can investigate the evolution of variables surrounding the approval of an IMF arrangement. The variables of interest will be GDP growth—to investigate the link between IMF financing and basic economic activity—and official developmental assistance—to see if donors respond to IMF program engagement (with or without IMF

Figure 3.5: Volatility of Growth



Note: Average CPIA over the sample against average "excess volatility." Countries with higher growth rates will have a higher variance due to scaling; a more meaningful measure would be relative to average growth. "Excess volatility" here is computed as the residuals of volatility from a regression with volatility as the independent variable and average growth as the dependent variable. Sample covers all years with satellite data, 1992-2013; 3-letter country codes follow IMF convention. *Source*: Stable-average lights from the NOAA; CPIA scores from the World Bank and African Development Bank.

financing). To preview the results, there appears to be an increase in economic growth after an IMF lending arrangement is approved and a strong increase in ODA following the start of IMF program engagement.

The event-study methodology we use to obtain these results does not allow us to make strong causal statements.[13] To see why, consider that to justify a causal statement requires establishing the relevant counterfactual, such as "what would have been the path of GDP in a country had the IMF not provided financing?" If this could be established, it would be straightforward

[13] A "causal" statement being: "IMF engagemet caused growth to increase by 3 percent."

to calculate the average impact of IMF lending by subtracting the observed outcome from the counterfactual. The difficulty in making such a statement becomes immediately obvious: the IMF presumably chooses to lend to a country precisely because the country has a negative projected path that can hopefully be reversed

Past attempts to get around the problem of assigning causality fall into two types. One is "selection correction," which attempts to "control" for the likelihood of the IMF lending to a country with the same observable characteristics as the country of interest. The assumed counterfactual are countries that had similar economic trajectories but did not receive IMF financing. This approach essentially relies on observing two countries with identical projected outcomes, where the IMF has provided financing to only one of them for reasons uncorrelated with these economic projections.[14]

The second approach to "control" for the selection problem and construct a relevant counterfactual is an instrumental variable (IV) strategy (e.g., Barro and Lee (2005)). Deaton (2010) has an excellent discussion on why such strategies are flawed, though in a different context. Applying Deaton's critique to our concerns, we see that the use of an IV strategy would require a variable that is correlated with IMF lending but uncorrelated with economic activity other than through IMF lending. Common instruments in this field are variables such as "political connections to the United States" because these countries are argued to be more likely to receive IMF financing.[15] This may satisfy the first condition, "correlated with IMF decisions," but being tied politically to the U.S. may clearly impact economic activity through other channels.

[14]Technically, the mathematics do not require comparisons of two countries with idential projections because the functional forms imposed allow for extrapolative comparisons. This explanation is useful for thinking through the general assumptions, however.

[15]In fact, this claim passes statistical tests based on conditional correlations.

Once the inherent flaws in these strategies are realized, it is not at all surprising that the statistical findings in the literature vary widely from "IMF financing significantly reduces growth" (Barro and Lee, 2005; Adam Przeworski and James Raymond Vreeland, 2000) to "IMF financing significantly increases growth" (Louis Dicks-Mireaux, Mauro Mecagni and Susan Schadler, 2000). Applying an incorrect counterfactual leaves no promise of recovering anything close to the "true" impact of IMF lending. This paper will avoid making assumptions about the counterfactual path and instead just provide graphical evidence for what it may look like. This conservative approach cannot lead to strong causal claims, but this paper takes the stance that this is a problem with reality, not with any specific statistical technique.

The numerical technique in this methodology will be to compute a (weighted) average path of the change in some variable of interest surrounding the approval of an IMF arrangement for fragile states.[16] This technique normalizes the year of the "event," that is, the approval of an IMF arrangement, to be year $t = 0$ for all countries that have such an event. The average increase in GDP growth across countries in the year following the arrangement, $t + 1$, is then computed. This number can be interpreted, without a counterfactual assumption, as the growth increase that one would expect from an IMF arrangement outside of the sample used for estimation.[17] For this reason, the word "expected" will be used to represent the computed weighted-average path.

Event studies are conducted for economic growth and official aid inflows. What makes an event-study methodology telling is plotting the difference in the variable before and after the

[16]A weighted average is used to mimic feasible generalized least squares (FGLS). The details are presented in the appendix.

[17]This can be thought of as representing the increase in GDP growth rates that would be expected if the IMF engaged in an arrangement with a fragile state today (assuming no further information). It is argued that this is a more informative metric of past "success" than is the simple average of observed outcomes, for reasons presented in the appendix.

approval of an IMF arrangement. If GDP growth was trending upwards prior to an arrangement and continues trending up, it is difficult to argue that the IMF arrangement "caused" the observed increase. If there was no trend prior to an IMF arrangement, but the same increase is observed afterwards, a much stronger case can be made that this growth should not have been anticipated without the IMF arrangement. Hence, readers can judge the plausibility of the counterfactual for themselves by seeing the average pre-trend.

3.5.2 Quantifying the Impact of an IMF Arrangement on Economic Outcomes

Economic Growth

The observations used for GDP growth results are all IMF lending arrangements with fragile states, as defined by the dynamic definition discussed in Section 3.2, that were approved from 2000 to 2012. The observations cover only up to 2012, given the constraints on satellite data, as well as the need for a sufficient "post" period to observe any changes that may or may not have taken place. Only 38 IMF arrangements meet these criteria. The GDP series is the weighted average of growth as observed by satellite data and growth as reported in the Penn World Tables.[18]

Countries that received multiple successive lending arrangements are a complicating factor in the analysis. For instance, Burundi has three separate observations for 2004, 2008, and 2012. Statistically, these three lending arrangements are treated as three independent lending events and GDP outcomes. However, if the IMF is repeating programs for countries that continue to struggle, dropping these observations would increase the average reported change in growth. Likewise, countries with programs that have gone "off-track"—that is, whose loan disbursements stopped

[18] We choose not to use the WEO data to avoid any possible implicit correlation between measurement error in WEO GDP and lending decisions for countries with limited national GDP data. Using an outside source for GDP helps to avoid this.

prior to their originally negotiated terms, because the country was not complying with the agreed program—are left in the analysis. In both cases, the lending arrangements are plausibly less likely to be successful than the average lending arrangement, and hence should not be the cause of a spuriously observed increase in growth outcomes. For the sake of a sufficiently large data series, and with the aim of being objective, these cases are included.

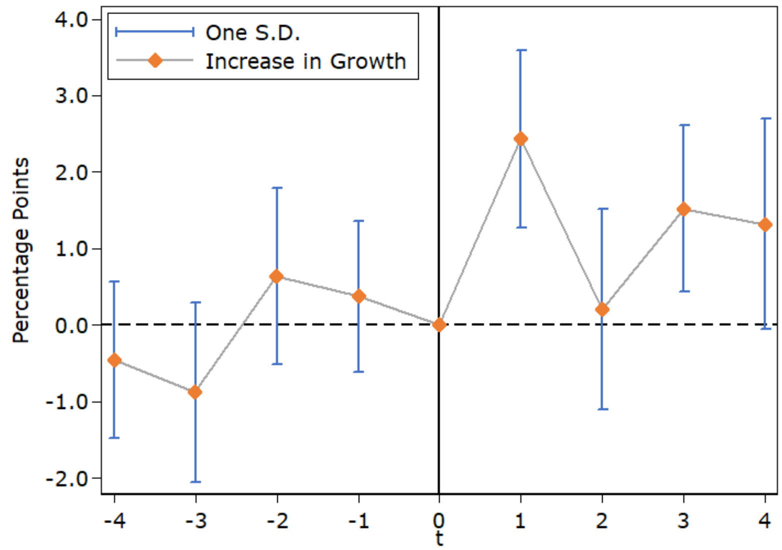
Figure 3.6 depicts the main results, showing the pattern of GDP growth rates before and after the approval of a lending arrangement for fragile states. Notice first that the value is zero at time zero ($t = 0$). This is by construction; the values here are relative to growth at the time of Fund program approval. As shown in Figure 3.6, three years after approval, the expected growth rate is about 1.5 percentage points higher than it was in the year lending began. The blue extensions surrounding these points are standard error bands to provide a measure of statistical confidence.

While the technique does not tell exactly what the average change in growth is, because different countries had different experiences, it provides approximately 70 percent confidence that it is contained within these bands. Given the fact that all confidence intervals overlap with the dotted "zero" line in the "pre-period," the weighted averages suggest that growth does not appear to be systematically lower or higher leading into an IMF arrangement. However, it seems reasonable to believe that after the start of an arrangement the expected change in growth is positive.[19]

Figure 3.6 presents the main results, showing the pattern of GDP growth rates before and after the approval of a lending program for fragile states. Notice first that the value is zero at

[19]While a confidence interval of 70 percent is not typically what is used to assess "statistical significance," it is not necessarily the case that a test would reject a null hypothesis that "growth is not higher following IMF lending." The appropriate test for this claim is an F-test on all four coefficients being zero. Further, the claim "reasonable to believe" should not necessarily rely on a 5 percent level of significance; while the data constraints of course preclude strongly rejecting a null hypothesis, this does not mean that the best guess given the data is that there is no effect.

Figure 3.6: Evolution of GDP Surrounding IMF Financing



Notes: The average change in GDP growth is computed by weighting observations by the inverse of their volatility for growth rates over the sample as would be prescribed by FGLS (see the appendix). Sample covers 2000-12. Countries are included if they are listed as fragile in the year the arrangement began.
Source: GDP from Pen World Tables and satellite imaging from NOAA; dates of financing agreements from IMF Finance Department.

time zero ($t = 0$). This is by construction; the values here are relative to growth at the time of approval. As shown in Figure 3.6, three years after approval the expected growth rate is about 1.5 percentage points higher than it was in the year lending began. The blue extensions surrounding these points are standard error bands to provide a measure of statistical confidence. While the technique is unsure exactly what the average change in growth is since different countries had different experiences, it is approximately 70 percent confident that it is contained within these bands.

While these results are encouraging for the IMF's role in fragile states, they do not necessarily imply that IMF lending per se had a causal impact. For example, if the IMF begins lending at the close of civil conflict, it is likely that growth would increase regardless. This data point would contribute to the dynamic pattern above, without contributing any information about whether the IMF's efforts in fact helped. There are many such possibilities that are consistent with these statistical results. However, the results weigh heavily against the claim that IMF lending to fragile states had a negative impact on economic activity. Taken in context, it is wise to interpret these results with cautious optimism in favor of the IMF's role.

Catalytic effect on aid inflows

A second question of importance for the IMF's role in fragile states concerns what has been termed the "catalytic effect" on donor assistance, the idea that the IMF's financial or program engagement catalyzes additional concessional financing or grants from donors. This effect could result from the signaling the IMF provides to the donor community that a government is sufficiently credible by committing to pursue sound economic policies under the tutelage of IMF monitoring and conditionality.

A similar event study tests the plausibility of this hypothesis by evaluating the evolution of official developmental assistance. In contrast to the analysis using GDP observations, countries that had an IMF arrangement in the prior period are removed. (The hypothesized mechanisms rely on the IMF beginning an arrangement with some country, so the omissions seem necessary in this case.) Likewise, it is not obvious that actual disbursements are necessary for the IMF to be catalytic, so the starts of Staff-Monitored Programs (SMPs), which involve no IMF financing, are included as well. The results are plotted in Figure 3.7, where the normalization is in percentage terms rather than levels: 100 represents ODA flows at the approval of an IMF arrangement (with or without IMF financing), and 150 would represent 50 percent more ODA than at time zero.[20] The results from this exercise strongly support a "catalytic" effect. The confidence bands (expanded to 90 percent) do not come close to zero in the post-period, suggesting high levels of statistical significance in the results. The pre-period does not show any trend of improvement that could suggest that the observed 60 percent increase would have happened in the absence of an IMF arrangement.

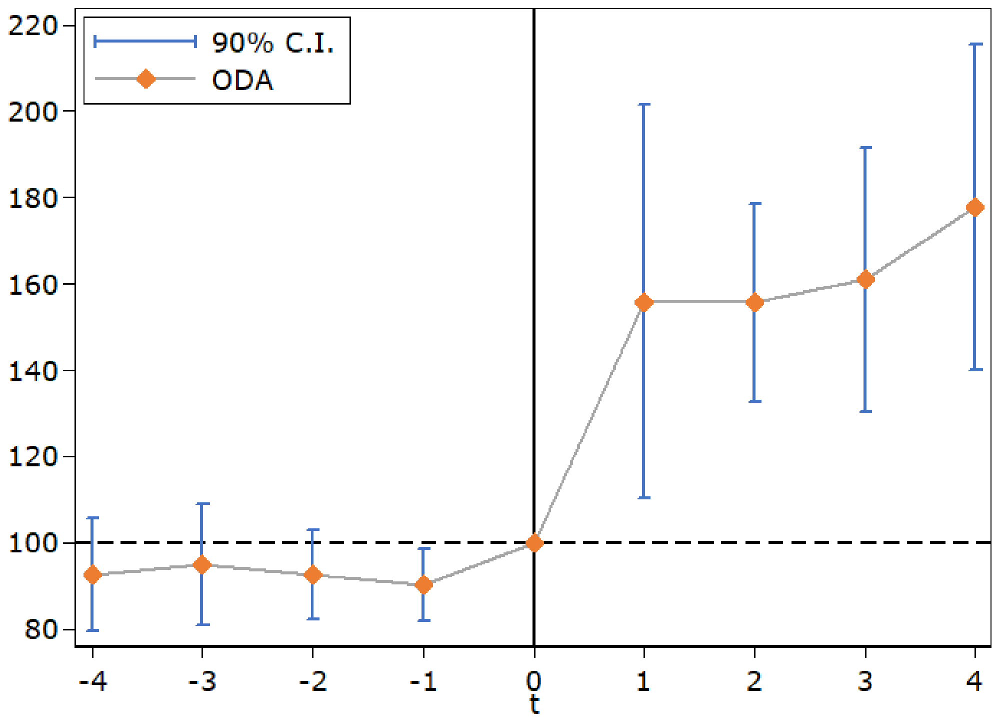
Robustness of the event-study results

Before concluding this section, it is worth commenting on the robustness of the positive GDP growth and ODA results. For GDP growth, it turns out that using weighted averages rather than simple averages changes the results for the $t + 4$ horizon (see the corresponding figures in the appendix). The same is true if we eliminate luminosity measures and use only the reported GDP growth from the Penn World Tables. In Figure 3.6, the $t + 4$ coefficient is approximately 1, but changing either of these assumptions makes this number indistinguishable from zero. There are convincing theoretical reasons to think that one should weight and improve the data with luminosity information—which is why those results are presented and are assumed to be the "better" ones when the results conflict.

The results for ODA are more robust. Figure 3.2 makes clear that there is no upward trend of ODA flows to fragile states over the sample period, so there is no reason to believe that longer-term trends drive this result spuriously. Moreover, using the same sample that was used for the growth analysis (allowing for overlapping programs, and only considering lending arrangements),

[20]For ODA, simple averages are presented, rather than generalized least squares. Volatility in ODA essentially represents political choices. It seems less appropriate to interpret any correction by volatility in these numbers as simply an increase in statistical efficiency.

Figure 3.7: Evolution of ODA Surrounding IMF Financing



Notes: Simple averages of relative ODA computed across all new IMF arrangements (with no arrangement in previous year), including nonlending instruments, 2000-12. Indexed to 100 at year of arrangement's start. Countries are included if they are listed as fragile in the year th arrangement was concluded.
Source: Author's calcuations based on ODA data from OECD; dates of arrangements from IMF Finance Department.

the results become weaker, but the overall pattern is the same. This is exactly what would be expected if only new arrangements matter for "catalyzing" aid.

3.6 Conclusion

This paper has conducted a statistical analysis of the economic performance of fragile states, and of how performance relates to IMF lending or program engagement. Given the poor data quality and selection bias inherent in understanding how IMF engagement interacts with economic performance, much of the paper focused on utilizing improved methodologies for careful assessment.

The results suggest that compared with other states, fragile states grow at a somewhat slower pace, on average, and are more susceptible to growth volatility. Once an IMF arrangement has begun, however, fragile states appear to have significantly higher sustained economic growth rates. Moreover, following the start of IMF program engagement, with or without IMF financing, fragile states experience substantially larger inflows of foreign aid. Overall, these findings suggest that IMF involvement overall has had a positive impact on macroeconomic performance in fragile states.

Appendices

Appendix A

Appendix to Chapter 1

A.1 Data Appendix

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

A.1.1 IMF Loans

IMF loans come from the Monitoring of Fund Arrangements (MONA) database which tracks all IMF programs and their details.The short-term programs are:

- Stand By Arrangements
- Stand By Credit Facility
- Rapid Financing Instrument
- Rapid Credit Facility
- Precautionary Lending Line
- Flexible Credit Line
- Exogenous Shocks Facility

However, the stylized fact of an "Ashenfelter Dip" holds even if only SBAs are used—the Fund's primary lending arrangement for short-term balance of payments problems.

A.1.2 Definition and Sources for Outcomes and Covariates

- Growth Rates: Growth rates are constructed using logged differences of the levels from the World Development Indicators real GDP (in local currency) from national accounts. Robustness is checked using analogous estimates from the Penn World Tables.

- 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

A.2 Details of Synthetic Control Method

Suppose at some horizon, t, following a financial crisis (at time $t = 0$) Equation A.1 determines $y_{i,t}$.

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

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.[1] 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.

[1] 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 (A.1) for constructing a good counterfactual using the SCM is that F^t can be well approximated *locally* by a linear function, $\hat{F}_i^t()$.[2] 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.

$$F^t(X, Y) \approx \hat{F}_i^t(X, Y) = \mathbf{A}_i X + \mathbf{B}_i Y \quad \text{if } (X, Y) \in \mathbb{L}_i \tag{A.2}$$

$\mathbb{L}_i$ 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 A.3).

$$p \in \mathbb{P} \iff (X_p, Y_p) \in \mathbb{L}_i \quad \cap \quad p \in \mathbb{D} \tag{A.3}$$

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 (A.4)-(A.6) hold.

$$Y_i = \sum_{p \in \mathbb{P}} \lambda_p^i Y_p \tag{A.4}$$

$$X_i = \sum_{p \in \mathbb{P}} \lambda_p^i X_p \tag{A.5}$$

$$\sum_{p \in \mathbb{P}} \lambda_p^i = 1 \tag{A.6}$$

Conditions in (A.4)-(A.6) 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

[2] 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.)

$$\begin{aligned}
\sum_{p\in\mathbb{P}} \lambda_p^i y_{p,t} =& \sum_{p\in P} \lambda_p^i F^t(X_{p,0},\ldots,y_{p,0}) + \sum_{p\in\mathbb{P}} \lambda_p^i u_{p,t} \\
=& \sum_{p\in\mathbb{P}} \lambda_p^i (\hat{F}_i^t(X_p, Y_p) + \nu_p) + \sum_{p\in\mathbb{P}} \lambda_p^i u_{p,t} \\
=& \hat{F}_i^t\Big(\sum_{p\in\mathbb{P}} \lambda_p^i X_p, \sum_{p\in\mathbb{P}} \lambda_p^i Y_p\Big) + \sum_{p\in\mathbb{P}} \lambda_p^i (u_{p,t} + \nu_p) \\
=& \hat{F}_i^t(X_i, Y_i) + \sum_{p\in\mathbb{P}} \lambda_p^i (u_{p,t} + \nu_p) \\
=& y_{i,t}(0) - \nu_i - u_{i,t} + \sum_{p\in\mathbb{P}} \lambda_p^i (u_{p,t} + \nu_p) \Rightarrow \\
y_{i,t}(0) =& \sum_{p\in\mathbb{P}} \lambda_p^i y_{p,t} + \underbrace{(u_{i,h} - \sum_{p\in\mathbb{P}} \lambda_p^i u_{p,t})}_{\text{0 in Expecation}} + \underbrace{(\nu_i - \sum_{p\in\mathbb{P}} \lambda_p^i \nu_p)}_{\approx 0 \text{ if locally linear}}
\end{aligned} \tag{A.7}$$

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

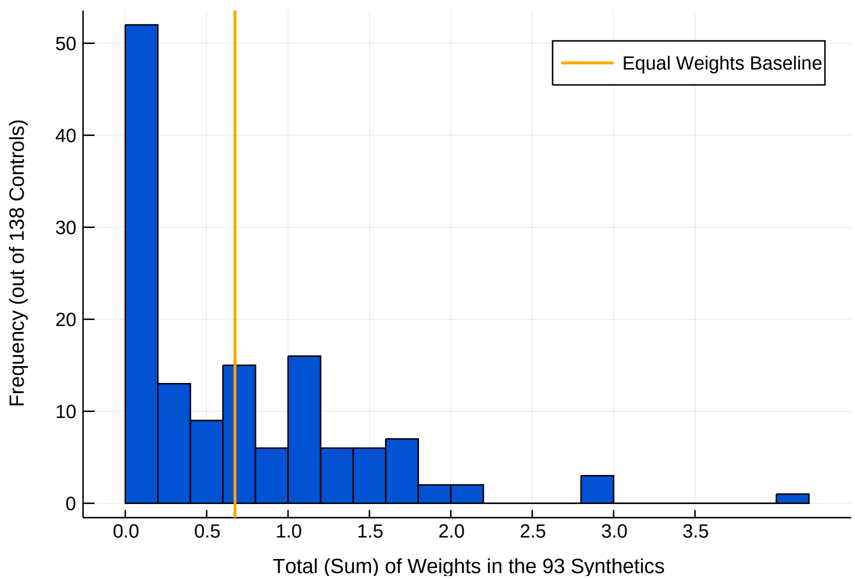
2. The underlying structure only requires *local* linearity rather than the much more restrictive *global* linearity assumption; SCM removes the need for extreme parametric extrapolation

3. For each treated country the counterfactual is directly observable as the convex combination of actual untreated observations making it highly transparent

A.3 Quality of Synthetic Controls

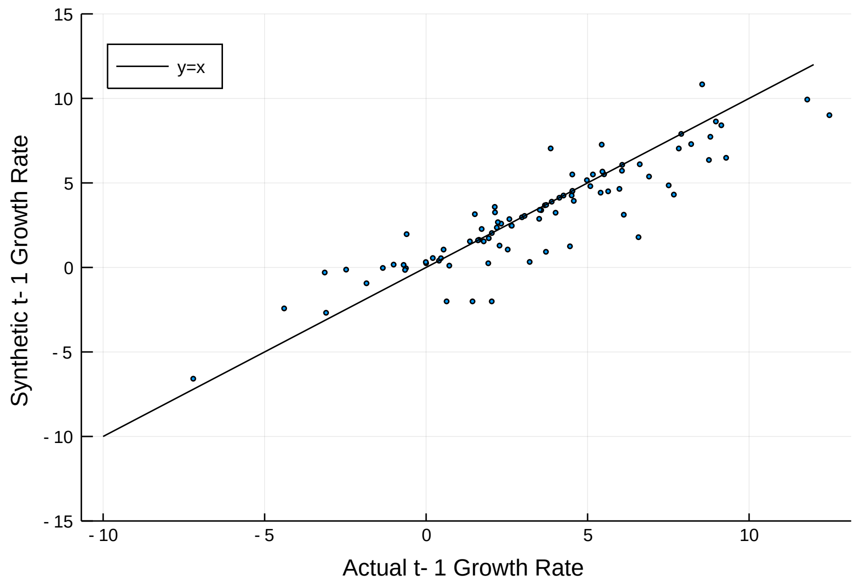
Along with Figure 1.8 there are two additional figures I present here as a way to understand the quality of the synthetic controls. First, Figure A1 shows the distribution of weights on each untreated crisis, measured as the sum of it's contribution to each synthetic. For example, suppose there was an 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 93 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.

Figure A1: Histogram of Synthetic Control Weights



Second, Figure A2 measures 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 "synthetic 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 1.9 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.

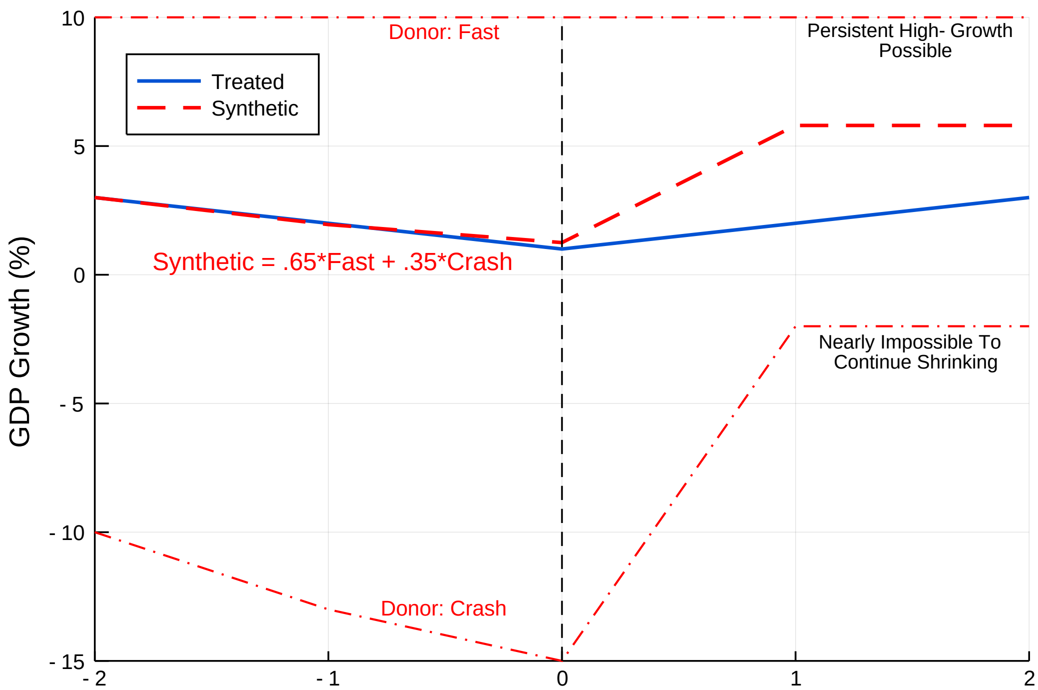
Figure A2: Synthetic vs. Actual for Targeted Growth Rates



A.4 Restricting to Local Crises

Below is a diagramatic representation of what could go wrong if matches were allowed to be drawn from the entire growth space. Extreme crashes that are bound to recover may enter into convex combinations for crises that are qualitatively different.

Figure A3: Restricting to Local Crises



Appendix B

Appendix to Chapter 3

B.1 Fragile Classification

B.1.1 Fragile Classification

While our methodology to classify fragile states is based on the approach taken by the World Bank, the list for any given year does not exactly match the harmonized list, for the following reasons.

First, the Asian Development Bank does not publish its own CPIA scores, although the World Bank publishes the ADB score for a country classified as fragile. This means that ADB's CPIA scores are available only for countries with sufficiently low CPIA scores, preventing us from creating a continuous series of CPIA scores for all low-income countries. Thus, for our analysis the World Bank's CPIA scores are averaged with the scores published by the African Development Bank (AfDB), and for countries not scored by AfDB, the World Bank's CPIA is the only input.

Second, we consistently use the CPIA scores currently available online irrespective of whether those were actually used in classifying countries as fragile on the harmonized list. This occasionally creates an inconsistency between our list and the harmonized list. For example, the World Bank's list for 2010 includes Cameroon, even though Cameroon's CPIA score for that year is 3.43 and the country had no peacekeeping or peacebuilding operation in place as far as we can verify. Our list would not include Cameroon, because the country does not meet either of the two criteria for inclusion as a fragile state.

Our methodology is to use the same criteria consistently to create a list of fragile states year by year, back to 2000, whereas the World Bank started publishing its harmonized list only from 2010. If the objective were only to look as far back as 2010, the harmonized list, as publicly available, could have been sufficient. It would be inconsistent, however, to utilize our approach until 2010 and then switch to the World Bank's published list thereafter. For a statistical analysis, it is important to use a consistent approach and avoid introducing systematic differences by using information selectively.

B.1.2 Mathematical Details of Satellite Use

The positive relationship between luminosity and GDP growth will be used in the following way.[1] First, start by assuming that there is a linear relationship between the percentage growth in true GDP growth, z, and the growth in how bright a country appears from space, l.

$$l_{i,t} = \tau_0 + \tau_1 z_{i,t} + \omega_{i,t}$$

Since true GDP growth is not known and the eventual goal is to feed luminosity growth in to predict GDP growth, an indirect estimate of τ_1 will be obtained by running the following inverted regression:

$$y_{i,t} = \beta_0 + \beta_L l_{i,t} + \mu_{i,t}$$

where $y_{i,t}$ is obserserved GDP growth for country i in year y, $l_{i,t}$ is the percent growth in luminosity, and $\mu_{i,t}$ is a regression error coming from the measurement error in GDP growth as well as the

[1] The discussion presented here is a broad outline of the logic presented in Henderson, Storeygard and Weil (2012), who provide additional details including on mathematical assumptions.

imperfect fit this line has with true GDP growth (coming from ω).

This seemingly circular logic of using observed GDP estimates to derive "independent" estimates of GDP from lights is in fact grounded in econometric theory. Measurement error in the independent variable does not prevent the estimation of the best linear predictor. The measurement error in national income data will inform what weight to place on the data, but this paper will merely take a conservative estimate coming from the academic literature rather than re-estimating the measurement error. With these coefficients estimated, a new, independent series of noisy estimates can be obtained using only information coming from satellite data.

It is less straightforward to mathematically determine the optimal synthesis of these estimates. Let ρ be the weight placed on national income data. Then, the following equation is used to determine our new composite GDP growth estimate, $\hat{z}_{i,t}$:

$$\hat{z}_{i,t} = \rho y_{i,t} + (1 - \rho)\hat{y}_{i,t}$$

Here $\hat{y}$ the growth estimate coming from satellite data and the estimated βs. Letting $z_{i,t}$ represent the true GDP growth of a given country in a given year, the objective is to find the ρ that minimizes the expected mean squared error of the composite from the true value.

$$var(\hat{z} - z) = var(\rho y_{i,t} + (1 - \rho)\hat{y}_{i,t} - z)$$

The equality here just results from substituting in the definition of our composite estimate. The resulting $(1-\rho)$, optimal weight on satellite-data-derived growth, comes from Chen and Nordhaus (2011), and can be worked out with a few steps of calculus and algebra.

$$(1 - \rho) = \frac{\tau_1^2 \sigma_\epsilon^2}{\tau_1^2 \sigma_\epsilon^2 + \sigma_\omega^2}$$

where $\sigma_\epsilon^2, \sigma_\omega^2$ represent the variance in errors for the true lights-growth relationship and the variance over measurement error in national income data, respectively.

It is important to note that the relevant error for lights is in the true relationship, not the observed one. The observed errors in the regression come from both measurement error and noise in the relationship. However, as this equation shows, if the only error came from errors in national income data (i.e. $\sigma_\omega^2 = 0$), all weight should be put on satellite estimates ($1 - \rho = 1$). If errors in national income data are extremely small ($\sigma_\epsilon^2 \rightarrow 0$), there is no relationship between luminosity and GDP ($\tau_1 = 0$) or if the lights–growth errors are very large ($\sigma_\omega^2 \rightarrow \infty$), then $1 - \rho \rightarrow 0$. This equals zero only if it is believed that there is zero error in national income accounts, or there is zero correlation between observable lights and GDP. This is a strong conclusion: the correlation has been verified empirically, so unless national income data are perfect, precision can be increased by incorporating satellite data.[2]

In applying this methodology to actual data, there is one practical issue that has come up in the literature, namely, the non-trivial measurement error in luminosity due to weather or other events that cause variations in the satellite images. While not explicitly addressing this issue, Henderson, Storeygard and Weil (2012) nevertheless conclude that this relationship is not as strong at yearly frequencies. Kevin Kuruc (2018*b*) shows that this weaker relationship is caused by the measurement error and can be corrected by properly accounting for it. In this connection, it should be mentioned that Chen and Nordhaus (2011) are more skeptical about the weight to place on satellite data, given the measurement error in luminosity.

[2]In situations where national income data are trustworthy, lights will optimally be weighted close to zero; this in turn implies that the practical gains will be close to zero. So, while it is theoretically correct that some gain exists in every context, it is not practically relevant for all countries.

For this paper we re-estimate the luminosity–GDP growth relationship at yearly frequencies, accounting for measurement error in satellite data, utilizing the technique introduced in Kuruc (2018*b*), who demonstrates that the measurement error is substantial and provides a strategy for generating superior estimates of yearly luminosity growth for estimating yearly GDP growth. The key insight of this technique is that if measurement error is throwing substantial "noise" into the short-run data, an instrumental-variables technique should recover a coefficient that looks like the result when one uses the long-run data.[3] This is confirmed using an internally constructed instrument inspired by Zvi Griliches and Jerry Hausman (1986). It is then shown that supplementing observed light growth by an average that includes some fraction of the observed cumulative growth rate over the period provides superior estimates.[4]

As there is substantial uncertainty in the literature surrounding these optimal combinations, the weights placed on the supplementing variables (outer growth rates in luminosity, followed by the weight on satellites itself) will be conservative. For nested growth rates, Kuruc (2018*b*) suggests more than half should be placed on the outer difference optimally. This paper uses 55 percent. In the context of statistics produced by low-capacity agencies, Chen and Nordhaus (2011) suggest assigning around one-third weight to luminosity versus GDP, while Henderson, Storeygard and Weil (2012) suggest assigning about half; this paper uses 35 percent. It is important to note that any combination will produce consistent estimates of the true values of interest; using the "optimal" combinations is in hopes of improving accuracy in finite samples. Not believing that

[3] This is a well-known property of instrumental variables. A regression with a noisy independent variable will bias downward the resulting coefficient. A second noisy measure of this independent variable can be used as an instrument to recover the "true" coefficient.

[4] While this technique introduces future and past growth information into the current estimate, this merely results in serial correlation in the errors. If it improves the information on period t growth (which it should), the results are improved, albeit non-independent, estimates of the path of average changes.

these numbers are truly the best combination does not imply that the results are in some way biased.

B.2 Event Study Details

For the event studies, the following regression equation is used:

$$W_{i,t} = C_i + \sum_{j \in J} \alpha_j IMF_{t-j} + u_{i,t}$$

where $W_{i,t}$ is an outcome that varies from its value at the time an IMF arrangement is approved, C_i, non-parametrically depending on how many periods the country is from the start of an arrangement. The set J runs from $\{$-4,4$\}$ excluding 0. IMF_t is an indicator for whether an IMF arrangement began in time t. Finally, α_j is the estimated coefficient that allows this equation to trace out the average path. Suppose, as an example, $\alpha_1 = A$. The prediction of this model is that one year following $(t+1)$ the start of arrangement i, the outcome is $W_{i,t+1} = C_i + A$. When looking at one year following an arrangement, only $IMF_{i,t-1}$ will be non-zero, so only α_1 survives.[5]

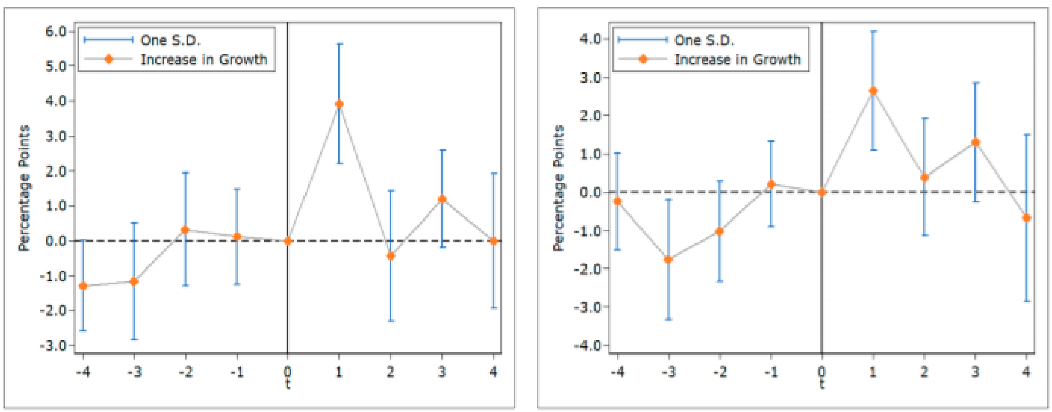
On a technical note, it is well-known that ordinary least squares (OLS) is inefficient when the error, $u_{i,t}$, has different variance across observations (i.e. heteroskedasticity). As mentioned when discussing use of satellite data (see prior appendix), efficiency gains are extraordinarily important, given the data constraints for this problem. Therefore, feasible generalized least squares (FGLS) will be employed. This technique essentially posits that for observations with more "noise" thrown in (large variance for $u_{i,t}$) less can be learned from their observed outcomes. For instance, if a country with very stable GDP growth had an increase following an IMF arrangement, statisti-

[5]Technically, some lending arrangements have windows that overlap. The example presented assumes that this is not the case.

cally there would be more confidence that the growth increase was not driven by randomness than if the same increase were observed for a country with widely fluctuating GDP growth. To put this idea into practice, the observations will be weighted by the inverse of their variance of their yearly growth rates from 1992 to 2013 (the sample for which satellite data are available) to generate the α coefficients.

How robust are the patterns of GDP growth in Figure 3.6 in the text? There are two obvious ways to deviate from the reported results. First (left panel), ignoring the luminosity-implied growth and strictly using Penn World Table growth rates (small changes to the weights make very little difference). Second (right panel), not using the GLS technique and treating every observation with identical weight. The two variations are reported in Figure B1; the results are not quite as strong, but have a similar pattern. Both modifications increase efficiency, so the results being "less strong" is not surprising or particularly concerning, though it would have been reassuring if the results held regardless of the removal of some statistical power. Both panels imply a weaker response four years out, as noted in the text. This is the most troubling disagreement and should give the reader some pause when interpreting the sustained growth implied by the main results.

Figure B1: Robustness Graphs



Source: Author's calculations based on: GDP from Penn World Tables and satellite imaging from NOAA; dates of financing agreements from IMF Finance Department.

Bibliography

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.

Andrews, Matthew, Lant Pritchett, and Michael Woolcock. 2010. "Capability Traps? The Mechanisms of Persistent Implementation Failure." *Center for Global Development, Working Paper*, 234.

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.

Barreca, Alan I. 2012. "Climate Change, Humidity, and Mortality in the United States." *Journal of Environmental Economics and Management*, 63(1): 19–34.

Barreca, Alan, Karen Clay, Olivier Deschenes, Michael Greenstone, and Joseph S Shapiro. 2016. "Adapting to climate change: The remarkable decline in the US temperature-mortality relationship over the twentieth century." *Journal of Political Economy*, 124(1): 105–159.

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.

BBC. 2018. "India cities dominate world air pollution list."

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.

Broome, John. 2012. *Climate matters: Ethics in a warming world.* WW Norton & Company.

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

Budolfson, Mark, Francis Dennig, Kevin Kuruc, Dean Spears, and Navroz Dubash. 2019. "Minimal utilitarian benchmarks for NDCs provide equity, climate and development advantages." Working Paper.

Budolfson, Mark, Francis Dennig, Marc Fleurbaey, Noah Scovronick, Asher Siebert, Dean Spears, and Fabian Wagner. 2018. "Optimal climate policy and the future of world economic development." *The World Bank Economic Review*, 33(1): 21–40.

Burke, Marshall, Melanie Craxton, CD Kolstad, Chikara Onda, Hunt Allcott, Erin Baker, L Barrage, R Carson, K Gillingham, J Graff-Zivin, et al. 2016. "Opportunities for advances in climate change economics." *Science*, 352(6283): 292–293.

Burke, Marshall, Solomon M Hsiang, and Edward Miguel. 2015. "Global non-linear effect of temperature on economic production." *Nature*, 527(7577): 235.

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.

Carlsson, Fredrik, Dinky Daruvala, and Olof Johansson-Stenman. 2005. "Are people inequality-averse, or just risk-averse?" *Economica*, 72(287): 375–396.

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.

Chauvet, Lisa, and Paul Collier. 2008. "What are the preconditions for turnarounds in failing states?" *Conflict Management and Peace Science*, 25(4): 332–348.

Chen, Xi, and William Nordhaus. 2011. "Using luminosity data as a proxy for economic statistics." *Proceedings of the National Academy of Sciences of the United States*, 108(21): 8589–8594.

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.

Coffey, Diane, and Dean Spears. 2017. *Where India Goes: Abandonded toilets, stunted development and the costs of caste.* Harper Collins.

Cowell, Frank, and Karen Gardiner. 1999. *Welfare Weights.* London:Office of Fair Trading.

Cowen, Tyler, and Derek Parfit. 1992. "Against the Social Discount Factor." *Justice Between Age Groups and Generations*, 145.

Dasgupta, Partha. 2008. "Discounting climate change." *Journal of Risk and Uncertainty*, 37(2-3): 141–169.

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.

Dell, Melissa, Benjamin F Jones, and Benjamin A Olken. 2014. "What do we learn from the weather? The new climate-economy literature." *Journal of Economic Literature*, 52(3): 740–98.

Dennig, Francis, Mark B Budolfson, Marc Fleurbaey, Asher Siebert, and Robert H Socolow. 2015. "Inequality, climate impacts on the future poor, and carbon prices." *Proceedings of the National Academy of Sciences*, 112(52): 15827–15832.

Deschênes, Olivier, Michael Greenstone, and Jonathan Guryan. 2009. "Climate change and birth weight." *American Economic Review*, 99(2): 211–17.

Diaz, Delavane, and Frances Moore. 2017. "Quantifying the economic risks of climate change." *Nature Climate Change*, 7(11): 774.

Dicks-Mireaux, Louis, Mauro Mecagni, and Susan Schadler. 2000. "Evaluating the effect of IMF lending to low-income countries." *Journal of Development Economics*, 61(2): 495–526.

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

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan. 2013. "Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India." *The Quarterly Journal of Economics*, 128(4): 1499–1545.

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.

Geruso, Michael, and Dean Spears. 2018. "Neighborhood sanitation and infant mortality." *American Economic Journal: Applied Economics*, 10(2): 125–62.

Ghosh, Shibani. 2015. "Reforming the Liability Regime for Air Pollution in India." *Environmental Law & Practice Review*, 4: 125–46.

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

Golosov, Mikhail, John Hassler, Per Krusell, and Aleh Tsyvinski. 2014. "Optimal taxes on fossil fuel in general equilibrium." *Econometrica*, 82(1): 41–88.

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

Greaves, Hilary. 2017. "Discounting for public policy: A survey." *Economics & Philosophy*, 33(3): 391–439.

Greenstone, Michael, and B Kelsey Jack. 2015. "Envirodevonomics: A research agenda for an emerging field." *Journal of Economic Literature*, 53(1): 5–42.

Greenstone, Michael, Santosh Harish, Rohini Pande, and Anant Sudarshan. 2017. "The Solvable Challenge of Air Pollution in India." *India Policy Forum*.

Griliches, Zvi, and Jerry Hausman. 1986. "Errors in variables in panel data." *Journal of Econometrics*, 31(1): 93–118.

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

Gupta, Aashish, and Dean Spears. 2017. "Health externalities of India's expansion of coal plants: Evidence from a national panel of 40,000 households." *Journal of Environmental Economics and Management*, 86: 262–276.

Hallegatte, Stephane, Mook Bangalore, Laura Bonzanigo, Marianne Fay, Tamaro Kane, Ulf Narloch, Julie Rozenberg, David Treguer, and Adrien Vogt-Schilb. 2015. *Shock waves: managing the impacts of climate change on poverty.* The World Bank.

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

Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review*, 102(2): 994–1028.

Hope, Chris. 2011. "The PAGE09 integrated assessment model: A technical description." *Cambridge Judge Business School Working Paper*, 4(11).

Hsiang, Solomon M. 2010. "Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America." *Proceedings of the National Academy of Sciences*, 107(35): 15367–15372.

Hsiang, Solomon M, Marshall Burke, and Edward Miguel. 2013. "Quantifying the influence of climate on human conflict." *Science*, 341(6151): 1235367.

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.

IPCC. 2007*a*. "Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change."

IPCC. 2007*b*. "Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change."

IPCC. 2014*a*. "Climate Change 2014: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change."

IPCC. 2014*b*. "Climate Change 2014: Mitigation of climate change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change."

Jerven, Morten. 2013. *Poor Numbers: How We are Misled by African Development Statistics and What to do About it.* Cornell University Press.

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*a*. "The IMF and Fragile States: Assessing Macroeconomic Outcomes." *Background Papers: Independent Evaluation Office of the IMF*.

Kuruc, Kevin. 2018*b*. "The Truth Is Out There: Quantifying Errors in National Accounts Using Satellite Data." University of Texas Working Paper.

LoPalo, Melissa. 2019. "Temperature, Productivity, and Adaptation: Evidence from Survey Data Production." Working Paper.

Markandya, Anil, Jon Sampedro, Steven J Smith, Rita Van Dingenen, Cristina Pizarro-Irizar, Iñaki Arto, and Mikel González-Eguino. 2018. "Health co-benefits from air pollution and mitigation costs of the Paris Agreement: A modelling study." *The Lancet Planetary Health*, 2(3): 126–133.

Nordhaus, William, and Paul Sztorc. 2013. "DICE 2013R: Introduction and user's manual." Yale University.

Nordhaus, William D. 1992. "An optimal transition path for controlling greenhouse gases." *Science*, 258(5086): 1315–1319.

Nordhaus, William D. 2007. "A review of the Stern review on the economics of climate change." *Journal of Economic Literature*, 45(3): 686–702.

Nordhaus, William D. 2010. "Economic aspects of global warming in a post-Copenhagen environment." *Proceedings of the National Academy of Sciences*, 107(26): 11721–11726.

Nordhaus, William D. 2017. "Revisiting the social cost of carbon." *Proceedings of the National Academy of Sciences*, 114(7): 1518–1523.

Nordhaus, William D, and Joseph Boyer. 2000. *Warming the world: economic models of global warming.* MIT press.

Peri, Giovanni, and Vasil Yasenov. 2019. "The Labor Market Effects of a Refugee Wave Synthetic Control Method Meets the Mariel Boatlift." *Journal of Human Resources*, 54(2): 267–309.

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.

Przeworski, Adam, and James Raymond Vreeland. 2000. "The effect of IMF programs on economic growth." *Journal of Development Economics*, 62(2): 385–421.

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.

Reilly, John, Sergey Paltsev, Erwan Monier, Henry Chen, Andrei Sokolov, Jin Huang, Qudsia Ejaz, Jeffery Scott, Jennifer Morris, and Adam Schlosser. 2015. "Energy & Climate Outlook: Persepectives from 2015." MIT Joint Program on the Science and Policy of Global Change.

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.

Scovronick, Noah, Mark Budolfson, Francis Dennig, Frank Errickson, Mark Fleurbaey, Wei Peng, Robert H Socolow, Dean Spears, and Fabian Wagner. 2018. "The impact of human health co-benefits on evaluations of global CO2 mitigation policy." Working Paper.

Sherwood, Steven C, and Matthew Huber. 2010. "An adaptability limit to climate change due to heat stress." *Proceedings of the National Academy of Sciences*, 107(21): 9552–9555.

Shindell, Drew, Greg Faluvegi, Karl Seltzer, and Cary Shindell. 2018. "Quantified, localized health benefits of accelerated carbon dioxide emissions reductions." *Nature Climate Change*, 8(4): 291.

Stern, Nicholas. 1977. "Welfare weights and the elasticity of the marginal valuation of income."

Stern, Nicholas. 2006. "Stern Review Report on the Economics of Climate Change." HM Treasury.

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

Tol, Richard SJ. 1999. "The marginal costs of greenhouse gas emissions." *The Energy Journal*, 61–81.

Tol, Richard SJ. 2009. "The economic effects of climate change." *Journal of Economic Perspectives*, 23(2): 29–51.

UNFCC. 2015. "Adoption of the Paris Accord."

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.

Vyas, Sangita, Nikhil Srivastav, and Dean Spears. 2016. "An Experiment with Air Purifiers in Delhi during Winter 2015-2016." *PloS one*, 11(12): e0167999.

Weitzman, Martin L. 2007. "A review of the Stern Review on the economics of climate change." *Journal of Economic Literature*, 45(3): 703–724.

Weitzman, Martin L. 2012. "GHG targets as insurance against catastrophic climate damages." *Journal of Public Economic Theory*, 14(2): 221–244.