Structural adjustment and state capacity: Evidence from IMF programs

Bernhard Reinsberg,[1*] Alexander Kentikelenis[1,2], Thomas Stubbs[1,3], Lawrence King[1]

1 – University of Cambridge

2 – University of Oxford

3 – University of Waikato

* Correspondence to: br385@cam.ac.uk, 17 Mill Lane, Cambridge CB2 1RX, United Kingdom

September 7, 2016

Abstract

This paper investigates the impact of International Monetary Fund conditionality on the state capacity of borrowing countries. The empirical analysis draws on a unique dataset of individual conditions between the IMF and its borrowers from 1985 to 2014 to develop new measures of Fund influence that can better account for program heterogeneity. Using multivariate regression analysis corrected for non-random selection into Fund programs, the analysis shows that "structural conditions" – specific conditions that require the overhaul of the state administration and the restructuring of the domestic economy – reduce state capacity. In contrast, "quantitative conditions" – broad targets on macroeconomic indicators – improve state capacity. In addition, results from instrumental-variables estimation show that these effects only exist for countries outside Sub-Saharan Africa, indicating potential endogeneity of conditionality in Sub-Saharan Africa. Overall, these findings suggest that IMF conditionality needs to be carefully designed so that it does not undermine developing countries' institutions.

Keywords: International Monetary Fund; conditionality; structural adjustment; state capacity; bureaucratic quality

JEL codes: F33, F34, F53

Acknowledgements: We would like to thank Timon Forster, Sebastian Ziaja, and members of the Political Economy and Development Policy group at University of Zurich for helpful comments. The authors declare no conflict of interest. Funding by the Institute for New Economic Thinking (INET Grant INO13-00020: 'The Political Economy of Structural Adjustment') to support the data collection and from the Cambridge Political Economy Trust/ Centre for Business Research at Judge Business School ("IMF conditionality and socioeconomic development") is gratefully acknowledged.

1. Introduction

Economists have long agreed that state provision of essential public goods and a legal framework for market exchange are necessary for economic development (Hall and Jones 1999; Easterly and Levine 2003; Rodrik, Subramanian, and Trebbi 2004; Acemoglu 2008). Yet, there is no clear consensus on how capable states arise. The literature emphasizes the key role of a professional bureaucracy (Weber 1978; Evans and Rauch 1999; Cingolani, Thomsson, and de Crombrugghe 2015; Charron, Dahlström, and Lapuente 2016), equating "state capacity" with the ability to devise and implement policies conducive to economic growth.

This article focuses on an international determinant of state capacity that has received little attention in the relevant literature: the role of policy reforms advocated by the International Monetary Fund (IMF). The IMF has long been the most influential agent in the policy environment of developing countries (Stone 2002; Babb 2009; Copelovitch 2010). The organization designs, administers and oversees reforms via loans that come with strings attached, or "conditionalities". That is, borrowing governments – often facing dire economic circumstances – must commit to implement a range of policy reforms.

Debates on the merits of conditionality have focused on identifying appropriate types of policy prescriptions. Some argue that the Fund should focus on macroeconomic targets to reduce excessive aggregate demand (Spraos 1986), while others argue that the Fund should also overhaul the underlying economic structures. In relation to state capacity, IMF programs have advocated policy reforms aimed at dismantling the state and cutting jobs in the public sector (Nooruddin and Vreeland 2010; Dreher and Gassebner 2012; Rickard and Caraway 2014b), or at bolstering revenue collection through strengthened tax inspectorates and customs administration (Crivelli and Gupta 2016).

In this paper, we study the effects of IMF conditionality on state capacity in developing countries over the 1985-2014 period. Using data on bureaucratic quality from the International Country Risk Guide as a proxy for state capacity, we examine the differential impact of structural and quantitative conditions – two types of conditions with different levels of intrusiveness into state affairs (Vreeland 2007; Dreher 2009; Woo 2013). The former refers to specific conditions requiring the overhaul of the state administration and restructuring of the domestic economy; the latter are broad targets on macroeconomic indicators. We find that structural conditions reduce bureaucratic quality, while quantitative conditions improve bureaucratic quality. These results are significant even after accounting for non-random selection into Fund programs and – outside of sub-Saharan Africa – endogeneity of conditionality.

This article also makes two methodological contributions to the literature on the consequences of IMF programs. First, it leverages a new dataset of IMF conditionality to capture program heterogeneity (Kentikelenis, Stubbs, and King 2016), where previous studies relied on a dichotomous indicator of IMF program participation (Dicks-Mireaux, Mecagni, and Schadler 2000), or focused on a particular policy area (Wei and Zhang 2010). Second, it employs a synthetic instrumentation technique to account for endogeneity concerns introduced by deploying conditionality data (i.e., countries with weak state capacity might receive systematically different conditions).

We proceed as follows. Section 2 defines key concepts and reviews related literature. Section 3 discusses the potential mechanisms linking IMF programs and state capacity. The subsequent two sections present our research design and results. Section 6 discusses the findings and concludes.

2. The international determinants of state capacity

State capacity – broadly understood as the ability of the state to provide essential public goods through a professional public bureaucracy[1] – is a key ingredient to economic development. While scholars have long been interested in the macro-historical determinants of state capacity (for comprehensive reviews, see Broich, Szirmai, and Thomsson 2015; Cingolani, Thomsson, and de Crombrugghe 2015; Savoia and Sen 2015), a more recent strand of research has focused on a range of international determinants of state capacity.

First, research shows that foreign aid can deteriorate state capacity, for example by diminishing the recipient government's accountability to domestic constituents (Bräutigam and Knack 2004; Moss, Pettersson, and Van de Walle 2006; Djankov, Montalvo, and Reynal-Querol 2008). However, more recent evidence suggests that aid earmarked for policy and institutional development has a small positive effect on political institutions (Jones and Tarp 2016).

Second, foreign direct investment can increase state capacity. Selowski and Martin (1997) show that foreign firms pressure host governments to improve domestic regulations and bolster administrative capacities. The potential benefits of FDI inflows – including increased capital accumulation and productivity spillovers – can also incentivize governments to provide a more predictable regulatory environment (Bengoa and Sanchez-Robles 2002; Bénassy-Quéré, Coupet, and Mayer 2007; Alfaro, Kalemli-Özcan, and Sayek 2009).

Third, international organizations may impact state capacity through their various interventions (e.g., Levy and Kpundeh 2004; Lake and Fariss 2014). The rationale for such interventions is that aid transfers are more successful when recipients have sufficient absorptive capacity (e.g., Burnside and Dollar 2000; Presbitero 2016). Moreover, a study by Marcoux and Urpelainen (2012) finds that developing countries seek membership in international organizations in part to improve their own administrative capacity.

While institution-building efforts of international organizations like the European Union, the World Bank, and the United Nations are recognized (see Levy and Kpundeh 2004; Lake and Fariss 2014; Moore 2014 for details), little attention has been devoted to the impact of IMF policy reforms, even though they entail conditions that directly affect public bureaucracies.[2] This is a surprising omission, as the sizable literature on the IMF has examined its impacts on a range of factors that interact with bureaucratic quality—from economic growth (Przeworski and Vreeland

[1] This definition builds on a vast body of related literature from sociology (e.g., Weber 1978; Skocpol 1979; Mann 1986; Amsden 1989; Evans and Rauch 1999), political science (e.g., Hendrix 2010; Cingolani, Thomsson, and de Crombrugghe 2015), and economics (e.g., Keefer and Knack 1995; Bräutigam and Knack 2004; Kaufman, Kraay, and Mastruzzi 2009).

[2] Some early qualitative studies linked IMF interventions to reducing the role of the state (Callaghy 1989; Haggard and Kaufman 1989; Biersteker 1990).

2000; Butkiewicz and Yanikkaya 2005; Dreher 2006b) to government spending (Conway 1994; Nooruddin and Simmons 2006; Hamm, King, and Stuckler 2012).

3. Tracing the links between IMF programs and bureaucratic quality

The International Monetary Fund is widely considered the most powerful intergovernmental organization (Stone 2002; Woods 2006; Babb 2009), and has considerable leverage over borrowing-country policies: Faced with economic trouble, developing countries—particularly low-income ones—turn to the IMF for policy advice and financial support, which is conditional on implementing certain policy reforms. This practice of 'conditionality' constitutes a key element of IMF programs. Nonetheless, there is considerable variation in the policy conditions attached to loans across time and space (Kentikelenis et al. 2016).

3.1. The content of IMF conditionality

Soon after its establishment, the IMF adapted its statutes to reflect its de-facto increased use of loan conditionality. These conditions are designed to abate 'moral hazard' and safeguard repayment (Spraos 1986; Polak 1991; Dreher 2009). Until the 1980s, the Fund supervised stabilization programs primarily relying on monetary policies, currency devaluations, and fiscal austerity among its clients. This rationale for these policies is based on the diagnosis that countries in economic crisis suffer from excessive aggregate demand. To reduce excessive demand, the Fund sets specific targets on macroeconomic variables through so-called *quantitative conditions*. In accordance with the IMF mandate, quantitative conditions are meant to be implemented in the short term and thereby restore balance of payments stability in crisis-affected countries. For example, IMF stabilization programs for the former Soviet republics included specific monthly ceilings on the rate of inflation (Stone 2002).

From the mid-1980s, the Fund increasingly embraced so-called *structural conditions* – a different type of conditionality that sought to alter countries' underlying economic architecture. These conditions often required countries to adopt free-market reforms such as privatizing state-owned companies, dismantling the welfare state, and liberalizing financial markets (e.g., Easterly 2005).

A vivid debate centers on the merits and demerits of structural conditions. IMF economists argue that these conditions are necessary in order to tackle the root causes of weak performance (Nowzad 1981; Khan and Knight 1983; Killick et al. 1984). Structural conditions might also send a more credible signal to foreign investors that the government is committed to economic reform (e.g., Bird 2007; Conway 2007; Woo 2013). Critics fear that structural conditions impose unnecessary burdens on borrowing countries and doubt that they are successful in resolving economic crises (e.g., Stiglitz 2001; Chang 2006; Kentikelenis et al. 2016). Even IMF economists have recently considered that the Fund might have gone too far in its pursuit of structural adjustment (Ostry, Loungani, and Furceri 2016).

Our new dataset enables us to explore the differential impact of certain types of conditions, notably quantitative conditions versus structural conditions. These types of conditions are thought

to differ in their "intrusiveness" into state affairs (e.g. Vreeland 2007; Dreher 2009; Woo 2013), which creates a strong *prima facie* claim that state capacity might be affected by these conditions.

3.2. Quantitative conditions and bureaucratic quality

Borrowing states have strong incentives to meet quantitative conditions, which can boost bureaucratic quality in two ways. First, quantitative conditions often entail provisions to increase state revenue, for example by amplifying efforts to collect tax arrears or by levying new taxes. As tax collection is a chief function of states (e.g., Tilly 1990; Bräutigam and Knack 2004; Besley and Persson 2008), they have incentices to invest in building-up administrative capacity. Second, quantitative conditions may indirectly strengthen bureaucratic capacity through increased incentives for compliance with conditionality. By virtue of their quantitative nature, it is easy to verify whether or not a borrowing country has met the agreed targets. Anecdotal evidence suggests that the Fund – mainly for reputational purposes – needs to suspend a program when the government visibly breaches its promises on a chief quantitative target (Stone 2002: 216). However, the government has a vital interest in compliance because its ability to obtain further credit hinges on whether it can implement the agreed target. Assuming that a government is willing to comply, but unable to do so, it has an incentive to redress its capacity deficit in order to increase compliance.

However, quantitative conditions need not improve bureaucratic quality. In particular, expenditure reductions are not administratively complex and are generally easy to accomplish even for low-capacity states (Callaghy 1989: 116). Quantitative conditions can even reduce bureaucratic capacity. For example, IMF programs have often stipulated public sector wage bill reductions, thereby sparking redundancies, hiring freezes or salary cuts (Haggard and Kaufman 1989; Nooruddin and Vreeland 2010; Rickard and Caraway 2014b). Recent empirical evidence shows that the public sector wage bill indeed decreases when the Fund explicitly asks borrowers to do so (Rickard and Caraway 2014a). These policies are not conducive to governments hiring or retaining skilled personnel committed to high-quality public service delivery.

3.3. Structural conditions and bureaucratic quality

Structural conditions can affect bureaucratic quality in different ways. On the one hand, structural conditions can be beneficial for bureaucratic quality. In particular, they may entail reforms to strengthen the independence of the central bank and other institutions related to economic governance. It is widely argued that independent central banks are conducive to sound macroeconomic policies such as low inflation (Alesina and Summers 1993). It might hence be argued that greater independence promotes greater insulation from political pressures, leading to higher bureaucratic quality and a higher predictability of economic policies.

On the other hand, structural conditions can reduce bureaucratic quality. These conditions often require countries to dissolve government departments, to merge ministries, and to dismiss or appoint high-level officials in the administration. There are downside risks of large-scale reorganization, as these measures may instill volatility in the bureaucracy. As a result, policy output may be less predictable and hence levels of bureaucratic quality. In addition, structural adjustment measures often target public sector pay schemes and modalities of social security. For example, agreements have required governments to produce lists of "ghost workers" in the state

administration. As governments may have incentives to protect their loyal supporters inside the bureaucracy, they may not dismiss actual ghost workers, but their most capable bureaucrats. Other agreements have pushed for increased flexibility with respect to hiring, firing, and working hours. At the extreme, these measures may backfire by mounting resistance from public sector workers and failure to attract skilled personnel, thereby reducing bureaucratic quality.

In light of the arguments presented in this section, the effect of conditionality on bureaucratic quality remains ambiguous. For both quantitative conditions and structural conditions, plausible arguments can be found to support either a positive effect or a negative effect. We therefore need to turn to the empirical analysis to adjudicate among these different theoretical expectations.

4. Data and methods

4.1. Operationalization

Dependent variables

Our main dependent variable, BUREAUCRATIC QUALITY, captures a narrow definition of state capacity. The bureaucratic quality indicator from the International Country Risk Guide is widely used in the literature on state capacity (e.g., Van Rijckeghem and Weder 2001; Bäck and Hadenius 2008; Broich, Szirmai, and Thomsson 2015). It scores high in "[…] countries where the bureaucracy has the strength and expertise to govern without drastic changes in policy or interruptions in government services. In these low-risk countries, the bureaucracy tends to be somewhat autonomous from political pressure and to have an established mechanism for recruitment and training" (The PRS Group 2015). To ease interpretation, we rescale the variable to range from 0 to 100.

We argue that meaningful changes in bureaucratic quality most likely occur in the medium term. On the one hand, improvements in bureaucratic quality involve training of staff and related institutional changes that need time to come to bear. Conversely, bureaucratic quality may in principle deteriorate rather quickly, but we did not find evidence of short-term impact of IMF programs. On the other hand, bureaucratic quality as a latent construct cannot be measured without error. We hence would expect short-term variation in the measure of bureaucratic quality to be the result of subjective bias (Kurtz and Schrank 2007). While we want to focus on the medium-term impact of IMF programs, our theoretical discussion does not suggest any particular lag length. The choice of lag length hence remains an open issue that we address through Bayesian model averaging (see below). Following the Bayesian analysis, we average the dependent variable over six-year periods. Averaging also smoothens out short-term fluctuations in BUREAUCRATIC QUALITY that most likely reflect mere noise – an inherent limitation of subjective indicators. We return to this issue in the discussion section. An additional advantage is that multi-year averaging makes the ordinal variable more continuous and hence makes the use of linear regression more appropriate (Dreher and Siemers 2009).

Independent variables

Our expectation is that not all parts of an IMF program necessarily affect state capacity. Different program parts—loan resources, loan conditions, and technical advice—may have different effects. Allowing for such effect heterogeneity, we employ several measures of IMF program participation.

First, we use a binary indicator for an IMF PROGRAM being active in at least five months of a given year, following previous studies using program dummies (e.g., Vreeland 2002; Dreher 2006b; Bas and Stone 2014). If included as the only indicator relating to a Fund program, such a dummy absorbs all of its effects. Therefore, even when assuming that the government complies with all aspects of the IMF program, it remains unclear which part of the program is decisive. Apart from a few exceptions (Dreher 2006b; Dreher and Walter 2010; Woo 2013), most studies use an IMF dummy and hence capture an overall effect (e.g., Vreeland 2003; Nooruddin and Simmons 2006; Casper 2015).

Second, to capture the depth of conditionality, we count the total number of conditions (Dreher and Vaubel 2004; Dreher and Jensen 2007; Copelovitch 2010) – referred to as ALL CONDITIONS in regression outputs. We only consider binding conditions, which include prior actions and performance criteria (Copelovitch 2010; Woo 2013; Rickard and Caraway 2014a). This choice is appropriate because binding conditions directly determine scheduled disbursements of loans, whereas non-binding conditions serve as markers for broader progress assessment (IMF 2001); non-implementation does not automatically suspend the loan and may thus introduce noise to the analysis if included (Stubbs et al. 2016: 3). This measure also is useful to distinguish empirically between lack of compliance and lack of impact in Fund programs because the binding character of these conditions excludes the possibility that a null effect is due to lack of compliance (Vreeland 2003; Dreher 2006b; Vreeland 2006).

Third, we capture the differential impact of conditionality types by using the respective numbers of binding STRUCTURAL CONDITIONS and QUANTITATIVE CONDITIONS in a Fund program. Both conditions must be included jointly because inference from a regression that only includes the count of one type of conditionality might be misleading, as it would also capture the effect of all conditions.[3] We only count binding conditions in each category. All above variables are lagged by one period to allow for some delay in the realization of the effect.

To measure our IMF variables, we use information from a newly constructed database on IMF conditionality available from 1985 to 2014 (see Kentikelenis et al. 2016). This dataset allows for a rich analysis of IMF programs as it includes detailed information regarding both specific types as well as issue areas of conditionality.

Control variables

We rely on a set of control variables drawn from the literature on state capacity. Most studies focus on the macro-historical determinants of state capacity, which include legal origin,

[3] Woo (2013) includes the number of structural conditions along with a program dummy, which may create an inferential problem in the presence of high correlation between structural conditionality and quantitative conditionality. Bulíř and Moon (2004) use the count of structural conditions but do not even control for program participation. This implies that any effect of a Fund program beyond its structural conditionality will be attributed to the latter indicator.

geography, ethno-linguistic fractionalization, and the structure of the economy (e.g., Alonso and Garcimartín 2013; Savoia and Sen 2015). Since none of these determinants are repeated measures, we control for them with a series of country dummies and focus our discussion on time-varying measures (Bräutigam and Knack 2004; Bäck and Hadenius 2008; Charron and Lapuente 2010). Bureaucratic quality is driven by a range of socio-economic factors. First, we consider the natural log of PER CAPITA INCOME, given that wealthier states are better able to sustain capable bureaucracies (e.g., Treisman 2000; Bäck and Hadenius 2008). In addition, we include the incidence of CIVIL WAR, which negatively affects the extractive capacity of the state and thereby reduces its capacity (e.g., Centeno 2002; Besley and Persson 2008). We also control for external war, which is theorized to provide an incentive to invest in state capacity to fend off external competitors (e.g., Tilly 1990; Besley and Persson 2008; Cárdenas, Eslava, and Ramirez 2011). We also include the level of democracy as measured by the POLITY IV INDEX. Citizens in a democracy will hold the government more accountable and increase their demand for public goods, which should increase incentives to invest into a capable bureaucracy (Meltzer and Richard 1981; Bäck and Hadenius 2008; Charron and Lapuente 2010; Fortin-Rittberger 2014). Empirical studies show a positive effect of democracy on governance quality, measured by the control of corruption (Montinola and Jackman 2002), bureaucratic quality (Bäck and Hadenius 2008), and various indices of governance (Adserà, Boix, and Payne 2003; Charron and Lapuente 2010).

Given data availability, we consider three additional control variables to build the extended controls. TRADE OPENNESS is hypothesized to increase pressures for redistribution and hence the need for a more sophisticated bureaucratic apparatus (Rodrik 1997; Bäck and Hadenius 2008; Walter 2010). Aid dependence – measured by FOREIGN AID per capita – could both enhance and diminish state capacity. While it provides resources that can be used to build state capacity (e.g., Jones and Tarp 2016), it also represents non-tax revenue and hence might reduce accountability (e.g., Moss et al. 2006; Djankov et al. 2008). Another source of non-tax revenue that might affect state capacity are natural resources, so we include the natural log of OIL PRODUCTION per capita (e.g., Humphreys 2005; Thies 2010). Detailed variable specifications for the above variables can be found in Appendix D.

4.2. Methods

Our unit of analysis is the country-period observation. Our data comprise 141 developing countries observed from 1985 to 2014. Our analysis proceeds in two steps. First, we compare the effect of IMF programs on state capacity between countries that underwent these programs and the countries that did not. Second, we compare countries with an IMF program but different conditions. This approach enables us to isolate the kinds of conditions that may have the strongest impact on state capacity.

Estimation techniques

We present results using three estimation techniques. We first estimate two-way fixed-effects models that account for both unobserved time-invariant country characteristics and global shocks. Fixed-effects estimation is an initial step to mitigate omitted-variable bias, given that only time-varying confounders may explain any significant finding in these models. All coefficients must be interpreted as within-country effects.

A key methodological challenge is that IMF programs are not randomly assigned. Fixed-effects estimation yields unbiased estimates only when omitted variables do not change over time. Typical approaches to address potential bias due to time-varying omitted variables include instrumental variables estimation and Heckman selection models. Our main concern is with non-random selection into IMF programs, which is captured by a binary indicator. Therefore, a Heckman selection model is the established approach (see also Appendix E for a discussion). In addition to the outcome equation, we also have a selection equation. Prediction errors from this equation yield the so-called non-selection hazard that allows for netting out the effect of unobservable factors of program participation decisions in the outcome equation (e.g., Heckman 1990; Atoyan and Conway 2006; Wei and Zhang 2010).

Based on previous literature, three variables explain the selection of countries into IMF programs (while at the same time being plausibly unrelated to bureaucratic quality). First, past involvement of a country in IMF programs reliably predicts current participation (e.g., Conway 1994; Bird, Hussein, and Joyce 2004; Easterly 2005). We thus include PAST PROGRAM, a dummy variable indicating whether or not a country had a program in the previous period. Second, program participation also is affected by the extent to which the Fund has resources available, which depends on the current number of program countries (Vreeland 2003: 88). Hence, we include the contemporaneous count variable COUNTRIES UNDER PROGRAMS. Third, the literature has repeatedly shown that allies of big powers receive favorable treatment in the international financial institutions (e.g., Thacker 1999; Barro and Lee 2005; Dreher, Sturm, and Vreeland 2009). We hence measure the alignment of voting patterns between the borrowing country and the G7 countries in the United Nations General Assembly (UNGA VOTE ALIGNMENT).; the number of countries under program captures the budget constraint of the Fund (Vreeland 2003: 88); and finally, following common practice, we add all variables from the second stage, regional effects, and time period dummies to the selection model.[4] All variables and their sources are given in Appendix B.

While the two-step procedure helps address concerns about the non-random selection of countries into IMF programs, it does not directly address potential endogeneity of loan conditions (Rickard and Caraway 2014b). Our third method hence employs instrumental variables. We employ a synthetic instrumentation that is employed in aid effectiveness research (e.g, Nunn and Qian 2014; Dreher and Langlotz 2015; Lang 2016). For each type of condition, we construct a synthetic instrument based on the interaction of the within-country average of those conditions and the period-specific budget constraint of the Fund. Appendix E further explains the merits of this approach. While our solution is not perfect, we note that there currently are no other satisfactory solutions to address potential endogeneity of specific types of conditions.[5]

[4] One would think that economic variables should be included, but in fact the prediction performs better without them and their inclusion does not alter our results in the outcome stage. In robustness checks, we permute over a wider range of specifications that also include economic variables (see Appendix D).

[5] Previous studies suggest some excludable instruments for IMF program participation (e.g., Vreeland 2003; Nooruddin and Simmons 2006; Dreher, Sturm, and Vreeland 2009), loan size (e.g., Conway 1994; Bird 1996; Barro and Lee 2005; Copelovitch 2010; Moser and Sturm 2011), and the number of conditions (e.g., Andersen, Harr, and Tarp 2006; Dreher 2006b; Dreher and Jensen 2007), but there is only limited research on the determinants of specific types of conditions that might yield generally valid instruments (e.g., Wei and Zhang 2010; Caraway, Rickard, and Anner 2014).

Model uncertainty and robustness tests

As discussed earlier, our substantive interest lies in the long-term effect of IMF programs, which precludes an annual panel analysis but still leaves room for a range of different lag lengths. To address the remaining model uncertainty regarding the appropriate lag length, we use Bayesian statistics. We find that – if at all – any impact of IMF programs must be expected in a six-year period, and therefore choose to arrange our data in six-year periods (see Appendix C). To untangle the two scenarios – a large number of conditions in a few years *versus* a small number of conditions in almost all years – we divide the number of conditions by the years in which an IMF program was active. This allows us to better capture the depth of conditionality in which we are ultimately interested.

In addition, we consider the possibility that different first-stage models may have implications on the coefficients of interest in the outcome stage. A large body of research on the determinants of IMF programs suggests many plausible specifications (e.g., Vreeland 2003; Nooruddin and Simmons 2006; Moser and Sturm 2011; Kentikelenis et al. 2015). Building on Oberdabernig (2013), we therefore conduct an Extreme Bounds Analysis that generates a distribution of coefficients from a few thousand regressions. Consequently, our coefficients of interest are highly robust to alternative specifications (see Appendix D).

Further robustness checks refer to different ways of operationalizing key variables (see Appendix A). First, we use an alternative measure of conditionality corrected for the number of waivers ceded to a borrower country. A waiver is a formal decision by the Executive Board that the country is freed from the need to fulfill a condition. Waivers can be granted *ad hoc* when circumstances beyond the control of countries have changed such as to justify more lenient treatment, but they can also be ceded as a result of intervention from important shareholders into the bureaucratic decisions of the Fund (Stone 2002: 234; Pop-Eleches 2009: 793). We expect our results to be stronger when using this measure because it only includes the conditions that the country must genuinely implement. Second, we repeat our estimations with a broader measure of state capacity. Using data from the International Country Risk Guide, this measure combines indices of BUREAUCRATIC QUALITY, RULE OF LAW, and CONTROL OF CORRUPTION, and has been used in previous studies on administrative capacity (Bräutigam and Knack 2004; Bäck and Hadenius 2008; Charron and Lapuente 2010).

5. Results

5.1. Descriptive evidence

We first explore the long-term relationship between IMF programs and state capacity through bivariate plots. Figure 1 depicts the evolution of state capacity by IMF program status. It shows the values of state capacity in 1985 plotted against those in 2014, distinguishing countries that had no IMF program during 1985-2014 (left panel) and countries that had at least one program during that period (right panel). Assuming some institutional stickiness over time, we would expect a diagonal line to best represent all observations, absent any policy intervention. If one believes that the Fund as a powerful international institution affects state capacity, there should be a departure

from the diagonal line. This does not seem to be case. Overall, Figure 1 reveals no conclusive evidence that IMF programs in general have an effect on state capacity.

[Figure 1 here]

As a next step, we explore whether certain aspects of IMF programs – such as structural conditions – can be harmful for state capacity. Figure 2 depicts the evolution of state capacity for four groups of countries with different intensities of structural conditionality. The upper left panel pools together non-program countries and IMF countries without structural conditions. The next panels include countries in the first two terciles of structural conditionality. In these panels, the positive relationship between past state capacity and current state capacity as reflected in the diagonal line is intact. If literally interpreted, this suggests that the Fund overall does not seem to have much effect on state capacity. However, the picture is strikingly different for countries in the bottom-right panel: These are the countries with the highest amount of structural conditionality; that is, these countries had to implement between 162 and 422 between 1985 and 2014.

[Figure 2 here]

Of course, these suggestive plots do not account for other confounding factors that could drive the negative relationship between high degrees of structural conditionality and state capacity. We therefore turn to multivariate analysis. We focus our analysis on bureaucratic quality and allow for a time horizon of six years in which the effect of conditionality may unfold (as suggested by the Bayesian analysis).

5.2. Baseline estimates

As our first set of models in the multivariate analysis, we present fixed-effects regressions of the levels of bureaucratic quality on indicators of IMF programs and IMF conditionality along with varying sets of control variables. In Table 1, Models 1 to 3 do not control for any other variables than the fixed effects on borrowing countries and time periods. Models 4 to 6 show regressions with baseline controls, while models 7 to 9 include an extended set of controls.

The results for our variables of interest are robust against different sets of controls and can be summarized as follows. First, IMF programs as such do not have a statistically significant effect on bureaucratic quality. This is in line with the suggestive evidence presented earlier. Second, structural conditions are negatively related to bureaucratic quality. Third, quantitative conditions are positively related to bureaucratic quality. The latter two findings are statistically significant at least at the 5%-level. Substantively, three additional structural conditions (roughly a standard deviation) reduce bureaucratic quality by about three points (equivalent to 8% at its mean). An additional eight quantitative conditions (again roughly a standard deviation) imply an increase of bureaucratic quality by about four points (11% at its mean).

For this set of models, we also present the full estimates for the control variables. Estimates of the control variables are similar in all subsequent models and hence we drop them from the subsequent tables to save space. A one-percent increase in GDP per capita relates to at least a ten-point increase in bureaucratic quality. The effect of income is even stronger when conditioning on countries under an IMF program. The remaining control variables are not statistically significant,

but their direction is consistent with theoretical expectations. For example, civil war tends to reduce bureaucratic quality, while economic growth enhances it. The effect of democracy is ambiguous and depends on the sampling decision.

Table 2 compares the impact of IMF programs in specific regions, notably Sub-Saharan Africa versus all other regions. For several reasons, we believe that results might be different in Sub-Saharan Africa (see also, Kentikelenis, Stubbs, and King 2015). As one of the poorest region and given its colonial legacy, Sub-Saharan Africa receives special attention from the donor community (e.g., Gereffi and Fonda 1992; World Bank 1994; United Nations 2001). Moreover, Sub-Saharan Africa has very low administrative capacity to begin with, due to the high prevalence of civil war (e.g., Centeno 2002; Cárdenas, Eslava, and Ramirez 2011; McBride, Milante, and Skaperdas 2011) and the susceptibility of the public sector to be captured by special interests (e.g., Levy and Kpundeh 2004; Moore 2004; Moore 2014). Indeed, a Hausman test comparing the coefficients for Africa versus all other regions suggests systematic differences in the underlying determinants of bureaucratic quality. Consequently, we produce separate regional estimates for the impact of conditionality on bureaucratic quality. We find that structural conditions only exert a negative effect on bureaucratic quality outside Sub-Saharan Africa. This may be due to floor effects and the fact that Africa is more frequently under IMF programs and hence we might not have much variation left to be explained in fixed-effects estimations.

The results in Tables 1 and 2 are a useful starting point for assessing the impact of IMF programs on bureaucratic quality. This is because the fixed-effects estimation takes into account the potential confounding effects from unobserved time-invariant factors and global trends affecting state capacity. The main weakness of the results, however, is that IMF countries could be different in unobserved ways that also affect bureaucratic quality. For example, a country might become more willing to improve its bureaucratic institutions and – therefore – engage with the Fund (Vreeland 2003). To address such concerns, we conduct estimations that also control for unobserved determinants of selection into programs.

5.3. Selection-corrected estimates

Table 3 presents the results from fixed-effect regressions augmented by a first stage that models the selection into IMF programs. The so-called non-selection hazard – a measure that captures the failure to predict program status based on observable characteristics – features as an additional control variable in our outcome regressions. As before, Model 1 to 3 use no controls, model 4 to 6 the baseline controls, and model 7 to 9 an extended set of controls. We suppress all control variables due to space constraints.

In short, the results become even stronger when taking endogenous selection into IMF programs into account. The coefficients of both structural conditions and quantitative are statistically significant at least at the 5%-level and their magnitudes have further increased. Regardless of the specification, IMF programs as such continue to have no overall effect on bureaucratic quality. When we split the sample by geographic region, we find that structural conditions have a significantly negative effect on bureaucratic quality only outside Sub-Saharan Africa (see Table 4). We also note that unobserved factors of IMF program participation are positively related to bureaucratic quality in Sub-Saharan Africa but not outside it.

The results from this set of regressions further support the earlier evidence that structural conditions reduce bureaucratic quality, whereas quantitative conditions increase it. Note that we can exclude the possibility that our findings are due to differences in the specific nature of countries that submit themselves under IMF programs. However, it may still be the case that within the group of IMF countries, we have not been able to control for factors that determine which types of conditions a given country receives. To address this problem, we turn to instrumental-variable estimations.

5.4. Endogeneity-corrected estimates

Table 5 presents estimates from fixed-effects regressions with non-selection hazard correction in which we also instrument conditionality using synthetic instruments. The rationale for doing so is to address potential endogeneity of conditionality. It might be the case that the Fund imposes more structural conditions on borrowing countries precisely when their bureaucratic quality deteriorates. Conversely, the Fund might also want to avoid overburdening weak-capacity countries with too many structural conditions – given that these conditions are widely considered as hard to implement (e.g., Nelson 1984: 110; Callaghy 1989: 116; Haggard and Kaufman 1989: 246). To our knowledge, there is no study indicating that the Fund allocates conditions on the basis of recipient-country institutional quality (at the exception of not giving out credit if the country is found to be corrupt, see Sandholtz and Gray 2003).

Correcting for potential endogeneity, the coefficients in Table 5 become larger in terms of absolute size but lose their statistical significance. The synthetic instruments are reasonably strong, as suggested by their respective F-statistics. According to the rule of thumb, F-statistics should be above ten (e.g., Staiger and Stock 1997; Stock and Yogo 2005). Structural conditions easily pass this threshold, while the F-statistics for all conditions and quantitative conditions are slightly below the threshold. This is most likely due to fact that almost all programs entail some quantitative conditions, while there is more variation with regards to structural conditions.

Once we split the sample and only consider countries outside Sub-Saharan Africa, the impact of conditionality on bureaucratic quality becomes significant again (see Table 6). Beyond statistical significance, the effects are substantively important: A three-point increase in the number of structural conditions outside Africa reduces bureaucratic quality by about 12 points – this is equivalent to more than 30% at its mean. An empirical example fitting this prediction is Moldova. Comparing the periods from 1997-2002 and 2003-2008, bureaucratic quality dropped by almost 20 points, while the average number of structural conditions per year had increased from 4.20 in 1991-1996 to 7.33 in 1997-2002 (the number of quantitative conditions stayed constant at 23). Another example is El Salvador in the late 1980s and early 1990s. Following a rise in the average annual number of quantitative conditions from 2.33 to 28.33, the country improved its bureaucratic capacity in the subsequent period by almost 40 points; the average number of structural conditions remained constant.

Our theoretical discussion provides some insight into why the two types of conditions have different effects on bureaucratic quality. The positive effect of quantitative conditions is most likely due to their associated incentives for governments to invest into capable bureaucracies that extract more revenue from taxes. For example, taking up the above example of El Salvador, the mid-term review under the Stand-By Agreement prepared by IMF staff praised the government

for making progress on "increasing public sector savings and reducing the overall deficit to the level required to attain the inflation objective," while meeting [its] commitments to raise "social expenditure, security spending, and interest payments on debt owed […]" According to its revised plan, the government would increase tax revenue thanks to an "increase in the VAT rate" and an "improved tax administration." In this context, the IMF review welcomed the improvement of data quality for surveillance purposes specifically due to "the creation of a new analytical unit in the Ministry of Finance" (IMF 1996a: 10-18). This staff review suggests that the El Salvadorian government was remarkably faithful to the quantitative targets of the agreement, while taking measures to build administrative capacity and to cushion against its social consequences. It is not unlikely that these measures led to the subsequent increase in the bureaucratic quality rating.

In contrast, structural conditions often require governments to reorganize the entire administration or to introduce more flexibility in public sector contracts. As an example for the former, consider the Stand-by Arrangement between the Fund and Armenia from 1995, in which the Fund required the government to "finalize a plan to […] eliminat[e] or merge ministries" (IMF 1995). Similarly, in 1996, Georgia was asked to "reduce the number of ministries and committees" (IMF 1996b). As to the latter, the Fund pushed for greater wage dispersion in the bureaucracy in the above Stand-By Arrangement with Armenia; in the case of Bulgaria, the government was required to "[s]ubmit […] amendments to the Labor Code with provisions making hiring, firing, and working hours more flexible" (IMF 2000). In 1998, the Fund asked Benin to "introduce performance-based salaries" as part of a structural adjustment program (IMF 1998).

The key strength of our quantitative analysis is to establish the average pattern linking IMF programs and bureaucratic quality in developing countries. However, the discussion on causal mechanisms remains is based only on anecdotal evidence and our own judgment. Future research may need to have a deeper look into these causal mechanisms, using appropriate methods such as in-depth country case studies.

6. Conclusion

This paper examined the impact of policy reform programs of the International Monetary Fund (IMF) on the bureaucratic quality of borrowing countries over the medium term. Applying three econometric approaches to a sample of all developing countries observed from 1985 to 2014, we find that quantitative conditions are positively related to bureaucratic quality, while structural conditions have a significantly negative relationship. Using instrumental-variable estimation, we demonstrate that the effects even have a causal interpretation when we consider countries outside Sub-Saharan Africa. Substantively, an increase in the number of structural conditions by one standard deviation outside Africa reduces bureaucratic capacity by 30%. Everything else equal, an increase of quantitative conditions by one standard deviation even leads to an increase in bureaucratic quality by 75%.

We note two limitations to our study. First, our results are based on a subjective measure of state capacity. To be sure, subjective measures are widely used in the pertinent literature (Adserà, Boix, and Payne 2003; Bräutigam and Knack 2004; Broich, Szirmai, and Thomsson 2015). Among these indicators, the ICRG index is still the most widely available (covering the period from 1984

to 2014), and its values can be compared over time. Another advantage also is that they seek to measure capacity rather than willingness to provide public goods. However, subjective measures are not uncontroversial, mainly because expert raters often misattribute economic performance to better state capacity because the former is more easily observable and the theoretical link is somewhat appealing (e.g. Stubbs, King, and Stuckler 2014). To mitigate this concern, we include the economic growth rate in all regressions. Nonetheless, we are careful not to over-interpret our findings as reflecting actual changes in state capacity, but changes in a subjective indicator that seeks to capture administrative capacity.

A second limitation pertains to the instrumentation strategy. Scholars have only recently begun using synthetic instruments and more work needs to be done to better understand their behavior and the type of effects that they identify. For our purposes, these instruments work well and we do not see any other possibility to test for different types of conditionality while accounting for potential endogeneity. We hope that our paper helps unleash a research agenda that more carefully considers the individual components of Fund programs and along these lines further develops the available econometric toolkit to study their underlying consequences.

From a theoretical perspective, our findings contribute to the literature on the determinants of state capacity. This literature is imbued by macro-historical determinants of state capacity, which do not admit itself to policy intervention. Only recently scholars have begun looking at the short-term determinants of state capacity that are amenable to purposive intervention (Bäck and Hadenius 2008; Charron et al. 2016). Nonetheless, less work has focused on the international determinants of state capacity. It is widely debated whether foreign aid has adverse effects on state institutions, but we are unaware of studies looking at the specific impact of international organizations such as the International Monetary Fund. In addition, our findings contribute to literature on the effects of IMF programs (e.g., Przeworski and Vreeland 2000; Atoyan and Conway 2006; Nooruddin and Vreeland 2010). Despite noteworthy exceptions (e.g., Rickard and Caraway 2014a), most of the existing research is constrained to a binary indicator of program participation and hence assumes that all programs are alike. Our paper is a useful extension as it addresses the huge variation in program components.

From a policy perspective, our findings imply nuanced conclusions regarding the IMF's role. IMF involvement is neither good nor bad, but conditionality matters. Quantitative conditions strengthen bureaucratic quality, but structural conditions reduce it. This suggests that the IMF is helpful when it stays with its core agenda of providing macroeconomic guidance; its role as guarantor of financial stability can lend countries the credibility to implement reforms that make them more resilient against economic crisis (see also, Stone 2002; Vreeland 2003; Dreher 2009). However, the IMF is harmful when it attempts to micromanage the economy and push for large-scale restructuring in the state apparatus. Recent official statements seem to suggest that the Fund has come to acknowledge that "[d]eveloping economic policy-making capacity in its member states is one of its most critical jobs" (IMF 2016), and its institutional capacity building activities through its regional training institutes are laudable. Yet, this cannot undo the historical legacies of excessive structural conditions and more efforts are needed to improve borrowing-country institutions.

References

Adsera, A., Boix, C., & Payne, M. (2003). Are you being served? Political accountability and quality of government. *Journal of Law, Economics, and organization*, 19(2), 445-490.

Acemoglu, D. (2008). Interactions between governance and growth. In: World Bank (ed.) *Governance, Growth, and Development Decision-making*. Washington D.C.: World Bank, 1-7.

Alesina, A., & Dollar, D. (2000). Who gives foreign aid to whom and why? *Journal of Economic Growth*, 5(1), 33-63.

Alesina, A., & Summers, L. H. (1993). Central bank independence and macroeconomic performance: some comparative evidence. *Journal of Money, Credit and Banking*, 25(2), 151-162.

Alesina, A., & Zhuravskaya, E. (2011). Segregation and the Quality of Government in a Cross Section of Countries. *The American Economic Review*, 101(5), 1872-1911.

Alfaro, L., Kalemli-Özcan, S., & Sayek, S. (2009). FDI, productivity and financial development. *The World Economy*, 32(1), 111-135.

Alonso, J. A., & Garcimartín, C. (2013). The determinants of institutional quality. More on the debate. *Journal of International Development*, 25(2), 206-226.

Amsden, A. H. (1989). *Asia's Next Giant: South Korea and Late Industrialization*. Oxford: Oxford University Press.

Andersen, T. B., Harr, T., & Tarp, F. (2006). On US politics and IMF lending. *European Economic Review*, 50(7), 1843-1862.

Arpac, O., Bird, G., & Mandilaras, A. (2008). Stop interrupting: An empirical analysis of the implementation of IMF programs. *World Development, 36*(9), 1493-1513.

Atoyan, R., & Conway, P. (2006). Evaluating the impact of IMF programs: A comparison of matching and instrumental-variable estimators. *Review of International Organizations*, 1(2), 99-124.

Atwood, J., Joglekar, A., & Smith, V. (2016). Developing consistent estimates of marginal effects in a simultaneous equation model with limited dependent variables. Paper presented at the 2016 Agricultural and Applied Economics Meeting, Boston (MA).

Babb, S. L. (2009). *Behind the development banks: Washington politics, world poverty, and the wealth of nations*. Chicago (IL): Chicago University Press.

Bäck, H., & Hadenius, A. (2008). Democracy and state capacity: exploring a J-shaped relationship. *Governance*, 21(1), 1-24.

Barro, R. J., & Lee, J.-W. (2005). IMF programs: Who is chosen and what are the effects?. *Journal of Monetary Economics*, 52, 1245-1269.

Bas, M. A. and R. W. Stone (2014). Adverse selection and growth under IMF programs. *Review of International Organizations*, 9, 1-28.

Bauer, M. E., Cruz, C., & Graham, B. A. (2012). Democracies only: When do IMF agreements serve as a seal of approval? *Review of International Organizations*, 7(1), 33-58.

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

Bellemare, M. F., Masaki, T., & Pepinsky, T. B. (2015). Lagged explanatory variables and the estimation of causal effects. Available at SSRN 2568724.

Bénassy-Quéré, A., Coupet, M., & Mayer, T. (2007). Institutional determinants of foreign direct investment. *The World Economy*, 30(5), 764-782.

Bengoa, M., & Sanchez-Robles, B. (2003). Foreign direct investment, economic freedom and growth: new evidence from Latin America. *European Journal of Political Economy*, 19(3), 529-545.

Besley, T., & Persson, T. (2008). Wars and state capacity. *Journal of the European Economic Association*, 6(2-3), 522-530.

Biersteker, T. J. (1990). Reducing the role of the state in the economy: a conceptual exploration of IMF and World Bank prescriptions. *International Studies Quarterly, 34*(4), 477-492.

Bird, G. (1996). Borrowing from the IMF: the policy implications of recent empirical research. *World Development*, 24(11), 1753-1760.

Bird, G. (2007). The IMF: A Bird's Eye View of its Role and Operations. *Journal of Economic Surveys*, 21(4), 683-745.

Bird, G. (2009). Reforming IMF Conditionality. *World Economics*, 10(3), 81-104.

Bird, G., Hussain, M., & Joyce, J. P. (2004). Many happy returns? Recidivism and the IMF. *Journal of International Money and Finance*, 23(2), 231-251.

Bräutigam, D. A., & Knack, S. (2004). Foreign aid, institutions, and governance in sub-Saharan Africa. *Economic Development and Cultural Change*, 52(2), 255-285.

Broich, T., Szirmai, A., & Thomsson, K. (2015). Precolonial centralisation, foreign aid and modern state capacity in Africa. UNU-MERIT Working Paper.

Bulír, A., & Moon, S. (2004). Is Fiscal Adjustment More Durable When the IMF is Involved? *Comparative Economic Studies*, 46(3), 373-399.

Burnside, C., & Dollar, D. (2000). Aid, policies, and growth. *American Economic Review*, 847-868.

Butkiewicz, J. L., & Yanikkaya, H. (2005). The Effects of IMF and World Bank Lending on Long-Run Economic Growth: An Empirical Analysis. *World Development*, 33(3), 371-391.

Callaghy, T. M. (1989). Toward state capability and embedded liberalism in the Third World: Lessons for adjustment. In: *Fragile Coalitions: The Politics of Economic Adjustment*, 115-138.

Cárdenas, M., Eslava, M., & Ramirez, S. (2011). State Capacity and Conflict. Unpublished, Department of Economics, Universidad de los Andes.

Casper, B. A. (2015). IMF Programs and the Risk of a Coup d'état. *Journal of Conflict Resolution*, 0022002715600759.

Centeno, M. A. (2002). *Blood and debt: War and the nation-state in Latin America*: Penn State Press.

Chang, H.-J. (2006). Policy Space in Historical Perspective with Special Reference to Trade and Industrial Policies. *Economic and Political Weekly*, 41, 627-633.

Charron, N., Dahlström, C., & Lapuente, V. (2016). Measuring Meritocracy in the Public Sector in Europe: a New National and Sub-National Indicator. *European Journal on Criminal Policy and Research*, 1-25.

Charron, N., & Lapuente, V. (2010). Does democracy produce quality of government? *European Journal of Political Research*, 49(4), 443-470.

Cingolani, L., Thomsson, K., & de Crombrugghe, D. (2015). Minding Weber More Than Ever? The Impacts of State Capacity and Bureaucratic Autonomy on Development Goals. *World Development*, 72, 191-207.

Conway, P. (1994). IMF lending programs: Participation and impact. *Journal of Development Economics*, 45(2), 365-391.

Conway, P. (2007). The Revolving Door: Duration and Recidivism in IMF Programs. *The Review of Economics and Statistics*, 89(2), 205-220.

Copelovitch, M. S. (2010). Master or servant? Common agency and the political economy of IMF lending. *International Studies Quarterly*, 54(1), 49-77.

Cranmer, S. J., Rice, D. R., & Siverson, R. M. (2016). What To Do About Atheoretic Lags. *Political Science Research and Methods*, 1-25. doi:doi:10.1017/psrm.2015.36.

Crivelli, E., & Gupta, S. (2016). Does conditionality in IMF-supported programs promote revenue reform?. *International Tax and Public Finance*, *23*(3), 550-579.

De Boef, S., & Keele, L. (2008). Taking time seriously. *American Journal of Political Science*, 52(1), 184-200.

Dicks-Mireaux, L., Mecagni, M., & Schadler, S. (2000). Evaluating the effect of IMF lending to low-income countries. *Journal of Development Economics*, 61(2), 495-526.

Djankov, S., Montalvo, J. G., & Reynal-Querol, M. (2008). The curse of aid. *Journal of Economic Growth*, 13(3), 169-194.

Dreher, A. (2004). A public choice perspective of IMF and World Bank lending and conditionality. *Public Choice*, 119(445-464).

Dreher, A. (2006a). Does Globalization Affect Growth? Evidence from a new Index of Globalization. *Applied Economics*, 38(10), 1091-1110.

Dreher, A. (2006b). IMF and economic growth: The effects of programs, loans, and compliance with conditionality. *World Development*, 34(5), 769-788.

Dreher, A. (2009). IMF conditionality: theory and evidence. *Public Choice*, 141(1-2), 233-267.

Dreher, A., & Gassebner, M. (2012). Do IMF and World Bank Programs Induce Government Crises? An Empirical Analysis. *International Organization*, 66(02), 329-358.

Dreher, A., & Jensen, N. M. (2007). Independent Actor or Agent? An Empirical Analysis of the Impact of U.S. Interests on International Monetary Fund Conditions. *Journal of Law and Economics*, 50(1), 105-124.

Dreher, A., & Langlotz, S. (2015). Aid and growth. New evidence using an excludable instrument. University of Heidelberg.

Dreher, A., & Siemers, L.-H. (2009). The nexus between corruption and capital account restrictions. *Public Choice*, 140(1-2), 245-265.

Dreher, A., Sturm, J.-E., & Vreeland, J. R. (2009). Global horse trading: IMF loans for votes in the United Nations Security Council. *European Economic Review*, 53(7), 742-757.

Dreher, A., Sturm, J.-E., & Vreeland, J. R. (2015). Politics and IMF Conditionality. *Journal of Conflict Resolution*, 59(1), 120-148.

Dreher, A., & Vaubel, R. (2004). The Causes and Consequences of IMF Conditionality. *Emerging Markets Finance and Trade*, 40(3), 26-54.

Dreher, A., & Walter, S. (2010). Does the IMF Help or Hurt? The Effect of IMF Programs on the Likelihood and Outcome of Currency Crises. *World Development*, 38(1), 1-18.

Easterly, W. (2005). What did structural adjustment adjust? *Journal of Development Economics*, 76(1), 1-22.

Easterly, W., & Levine, R. (2003). Tropics, germs, and crops: how endowments influence economic development. *Journal of Monetary Economics*, 50(1), 3-39.

Evans, P., & Rauch, J. E. (1999). Bureaucracy and growth. *American Sociological Review*, 748-765.

Fortin-Rittberger, J. (2014). Exploring the relationship between infrastructural and coercive state capacity. *Democratization*, 21(7), 1244-1264.

Freedom House (2015). Freedom in the world 2015: Discarding democracy: Return to the iron fist. Retrieved from www.freedomhouse.org (Accessed April 30, 2016).

Garuda, G. (2000). The Distributional Effects of IMF Programs: A Cross-Country Analysis. *World Development*, 28(6), 1031-1051.

Gereffi, G., & Fonda, S. (1992). Regional paths of development. *Annual Review of Sociology*, 18, 419-448.

Gleditsch, N. P., Wallensteen, P., Eriksson, M., Sollenberg, M., & Strand, H. (2002). Armed conflict 1946-2001: A new dataset. *Journal of Peace Research*, 39(5), 615-637.

Grant, T., & Lebo, M. J. (2016). Error correction methods with political time series. *Political Analysis*, 24(1), 3-30.

Haggard, S., & Kaufman, R. R. (1989). The Politics of Stabilization and Structural Adjustment. Developing Country Debt and Economic Performance, Volume 1: The International Financial System, pp. 209-254.

Hall, R. E., & Jones, C. I. (1999). Why do some countries produce so much more output per worker than others? National Bureau of Economic Research Working Paper No. 6564.

Hamm, P., King, L. P., & Stuckler, D. (2012). Mass Privatization, State Capacity, and Economic Growth in Post-Communist Countries. *American Sociological Review*, 77(2), 295-324.

Heckman, J. (1990). Varieties of selection bias. *The American Economic Review*, *80*(2), 313-318.

Hendrix, C. S. (2010). Measuring state capacity: Theoretical and empirical implications for the study of civil conflict. *Journal of Peace Research*, 47(3), 273-285.

Hlavac, M. (2014). ExtremeBounds: Extreme Bounds Analysis in R. Available at SSRN 2393113.

Hulme, D., Savoia, A., & Sen, K. (2015). Governance as a Global Development Goal? Setting, Measuring and Monitoring the Post-2015 Development Agenda. *Global Policy*, 6(2), 85-96.

Humphreys, M. (2005). Natural resources, conflict, and conflict resolution uncovering the mechanisms. *Journal of Conflict Resolution*, 49(4), 508-537.

IMF(1995). Republic of Armenia - Stand-By Arrangement (EBS/95/100). Washington D.C.: International Monetary Fund.

IMF (1996a). El Salvador - Staff Report for the 1995 Article IV Consultation and Midterm Review Under the Stand-By Arrangement (EBS/96/42). Washington D.C.: International Monetary Fund.

IMF (1996b). Republic of Georgia - Enhanced Structural Adjustment Arrangement (EBS/96/21).

IMF (1998). Benin - Request for Second Annual Arrangement Under the Enhanced Structural Adjustment Facility (EBS/98/230). Washington D.C.: International Monetary Fund.

IMF (2000). Bulgaria - Staff Report for the 1999 Article IV Consultation and Third Review Under the Extended Arrangement (EBS/00/45). Washington D.C.: International Monetary Fund.

IMF (2001). Structural Conditionality in Fund-Supported Programs. Washington D.C.: International Monetary Fund.

IMF (2016). Capacity Development. Washington D.C.: International Monetary Fund. Available at http://www.imf.org/external/np/ins/english/ (Accessed July 31, 2016).

Jones, S., & Tarp, F. (2016). Does foreign aid harm political institutions? *Journal of Development Economics*, 118, 266-281.

Khan, M. S., & Knight, M. D. (1983). Determinants of current account balances of non-oil developing countries in the 1970s: an empirical analysis. *Staff Papers*, *30*(4), 819-842.

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2009). Governance matters VIII: aggregate and individual governance indicators, 1996-2008. World Bank Policy Research working paper No. 4978.

Kentikelenis, A. E., Stubbs, T. H., & King, L. P. (2015). Structural adjustment and public spending on health: Evidence from IMF programs in low-income countries. *Social Science and Medicine*, *126*, 169-176.

Kentikelenis, A. E., Stubbs, T. H., & King, L. P. (2016). IMF conditionality and development policy space, 1985-2014. *Review of International Political Economy* (Online first).

Killick, T., Bird, G., Sharpley, J., & Sutton, M. (1984). The IMF: Case for a Change in Emphasis. In: *Adjustment Crisis in the Third World*. New Brunswick: Transactions Books.

Knack, S., & Keefer, P. (1995). Institutions and economic performance: cross-country tests using alternative institutional measures. *Economics & Politics*, 7(3), 207-227.

Kurtz, M. J., & Schrank, A. (2007). Growth and governance: Models, measures, and mechanisms. *Journal of Politics*, *69*(2), 538-554.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. (1999). The quality of government. *Journal of Law, Economics, and Organization*, 15(1), 222-279.

Lake, D. A., & Fariss, C. J. (2014). Why international trusteeship fails: the politics of external authority in areas of limited statehood. *Governance*, *27*(4), 569-587.

Lang, V. F. (2016). The Democratic Deficit and Its Consequences: The Causal Effect of IMF Programs on Income Inequality. Paper presented at the Political Economy of International Organizations, Salt Lake City, January 7-9.

Leamer, E. E. (1983). Let's take the con out of econometrics. *The American Economic Review*, 73(1), 31-43.

Levy, B., & Kpundeh, S. J. (2004). Building state capacity in Africa: new approaches, emerging lessons. Washington D.C.: World Bank.

Mann, M. (1986). *The sources of social power: A history of power from the beginning to A.D. 1760*. Cambridge: Cambridge University Press.

Marcoux, C., & Urpelainen, J. (2012). Capacity, not constraints: A theory of North-South regulatory cooperation. *Review of International Organizations, 7*(4), 399-424.

Marshall, M., Gurr, T., & Jaggers, K. (2015). Polity IV project: Political regime characteristics and transitions 1800−2014. http://www.systemicpeace.org/inscrdata.html (Accessed April 30, 2016).

McBride, M., Milante, G., & Skaperdas, S. (2011). Peace and war with endogenous state capacity. *Journal of Conflict Resolution*, 0022002711400862.

Meltzer, A. H., & Richard, S. F. (1981). A rational theory of the size of government. *The Journal of Political Economy*, 914-927.

Montinola, G. R., & Jackman, R. W. (2002). Sources of corruption: A cross-country study. *British Journal of Political Science*, 32(01), 147-170.

Moore, M. (2004). Revenues, state formation, and the quality of governance in developing countries. *International Political Science Review*, 25(3), 297-319.

Moore, M. (2014). Revenue reform and statebuilding in Anglophone Africa. *World Development*, 60, 99-112.

Moser, C., & Sturm, J.-E. (2011). Explaining IMF lending decisions after the Cold War. *Review of International Organizations*, 6(3-4), 307-340.

Moss, T. J., Pettersson, G., & Van de Walle, N. (2006). An aid-institutions paradox? A review essay on aid dependency and state building in sub-Saharan Africa. Center for Global Development working paper(74), 11-05.

Nelson, F., & Olson, L. (1978). Specification and estimation of a simultaneous-equation model with limited dependent variables. *International Economic Review*, 19(3), 695-709.

Nelson, J. (1984). *The politics of stabilization: Adjustment crisis in the third world*. New Brunswick, NJ: Transaction Books.

Nooruddin, I., & Simmons, J. W. (2006). The politics of hard choices: IMF programs and government spending. *International Organization*, 60(04), 1001-1033.

Nooruddin, I., & Vreeland, J. R. (2010). The Effect of IMF Programs on Public Wages and Salaries. In: *Global Governance, Poverty and Inequality*, 90-116.

Nooruddin, I., & Woo, B. (2015). Heeding the Sirens: The Politics of IMF Program Participation. *Political Science Research and Methods*, 3(01), 73-93.

Nowzad, B. (1981). The IMF and Its Critics. Essays in International Finance No. 146. Princeton (NJ): Princeton University.

Nunn, N., & Qian, N. (2014). US food aid and civil conflict. *The American Economic Review*, 104(6), 1630-1666.

Oberdabernig, D. A. (2013). Revisiting the Effects of IMF Programs on Poverty and Inequality. *World Development*, 46, 113-142.

Ostry, J. D., Loungani, P., & Furceri, D. (2016). Neoliberalism: Oversold? *Finance and Development*, 53(2), 38-41.

Pop-Eleches, G. (2009). Public Goods or Political Pandering: Evidence from IMF Programs in Latin America and Eastern Europe. *International Studies Quarterly*, 53(3), 787-816.

Presbitero, A. F. (2016). Too much and too fast? Public investment scaling-up and absorptive capacity. *Journal of Development Economics*, 120, 17-31.

Przeworski, A., & Vreeland, J. R. (2000). The Effect of IMF Programs on Economic Growth. *Journal of Development Economics*, 62, 385-421.

Raftery, A., Hoeting, J., Volinsky, C., Painter, I., Yeung, K. Y., Sevcikova, M. H., & Suggests, M. (2015). Package 'BMA'.

Rajan, R., & Subramanian, A. (2007). Does aid affect governance? *The American Economic Review*, 97(2), 322-327.

Rickard, S. J., & Caraway, T. L. (2014a). International negotiations in the shadow of national elections. *International Organization*, 68(03), 701-720.

Rickard, S. J., & Caraway, T. L. (2014b). The International Politics of Austerity: The Puzzling Case of Public Sector Reforms. Paper presented at the Political Economy of International Organizations, Berlin, February 12-14.

Rodrik, D. (1997). Sense and nonsense in the globalization debate. *Foreign Policy*, 107, 19-37.

Rodrik, D., Subramanian, A., & Trebbi, F. (2004). Institutions rule: the primacy of institutions over geography and integration in economic development. *Journal of Economic Growth*, *9*(2), 131-165.

Roodman, D. (2012). Fitting fully observed recursive mixed-process models with cmp. *The Stata Journal*, 11(2), 159-206.

Ross, M. L. (2013). Oil and gas data, 1932-2011. Havard University Dataverse.

Rothstein, B., & Teorell, J. (2008). What is quality of government? A theory of impartial government institutions. *Governance*, 21(2), 165-190.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley & Sons.

Sala-i-Martin, X. X. (1997). I just ran two million regressions. *The American Economic Review*, 87(2), 178-183.

Sandholtz, W., & Gray, M. M. (2003). International integration and national corruption. *International Organization*, *57*(04), 761-800.

Savoia, A., & Sen, K. (2015). Measurement, Evolution, Determinants, and Consequences of State Capacity: A Review of Recent Research. *Journal of Economic Surveys*, 29(3), 441-458.

Selowsky, M., & Martin, R. (1997). Policy performance and output growth in the transition economies. *The American Economic Review*, 87(2), 349-353.

Skocpol, T. (1979). *States and Social Revolutions: A Comparative Analysis of France, Russia and China*: Cambridge: Cambridge University Press.

Smith, A., & Vreeland, J. R. (2006). The survival of political leaders and IMF programs: testing the scapegoat hypothesis. In: *Globalization and the nation state: The impact of the IMF and the World Bank*, 263-289.

Spraos, J. (1986). IMF Conditionality: Ineffectual, Misguided, Mistargeted. Essays in International Finance, No. 166. Princeton (NJ): Princeton University.

Stock, J. H., & Staiger, D. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65(3), 557-586.

Stock, J. H., & Yogo, M. (2005). Identification and inference for econometric models. In: *Essays in honor of Thomas Rothenberg*. Cambridge (MA): Cambridge University Press.

Stone, R. W. (2002). *Lending Credibility: The International Monetary Fund and the Post-Communist Transition*: Princeton (NJ): Princeton University Press.

Stone, R. W. (2004). The political economy of IMF lending in Africa. *American Political Science Review*, 98(4), 577-592.

Strezhnev, A., & Voeten, E. (2013). United Nations General Assembly voting data. Harvard University Dataverse.

Stubbs, T., Kentikelenis, A. E., Stuckler, D., & King, L. P. (2016). Health systems and the International Monetary Fund: A cross-national analysis of 16 West African nations, 1995–2014. Mimeo, University of Cambridge.

Stubbs, T., King, L. P., & Stuckler, D. (2014). Economic growth, financial crisis, and property rights: observer bias in perception-based measures. *International Review of Applied Economics*, 28(3), 401-418.

Sturm, J.-E., & De Haan, J. (2005). Determinants of long-term growth: New results applying robust estimation and extreme bounds analysis. *Empirical Economics*, 30(3), 597-617.

Teorell, J., Dahlberg, S., Holmberg, S., Rothstein, B., Khomenko, A., & Svensson, R. (2016). The Quality of Government Standard Dataset, version Jan16. Retrieved from: http://www.qog.pol.gu.se (March 30, 2016).

Thacker, S. C. (1999). The High Politics of IMF Lending. *World Politics*, 52(1), 38-75.

The PRS Group (2015). International Country Risk Guide (ICRG). Retrieved from https://www.prsgroup.com/about-us/our-two-methodologies/icrg (Accessed April 30, 2016).

Thies, C. G. (2010). Of rulers, rebels, and revenue: State capacity, civil war onset, and primary commodities. *Journal of Peace Research*, 47(3), 321-332.

Tilly, C. (1990). *European States, AD 990-1990*. Cambridge and Oxford: Basil Blackwell.

Treisman, D. (2000). The causes of corruption: a cross-national study. *Journal of Public Economics*, *76*(3), 399-457

United Nations (2001). Abuja Declaration on HIV/AIDS, Tuberculosis and Other Related Infectious Diseases.

United Nations (2015). UN Data: A world of information. Available at data.un.org.

Van Rijckeghem, C., & Weder, B. (2001). Sources of contagion: is it finance or trade? *Journal of International Economics*, 54(2), 293-308.

Vreeland, J. R. (2002). The Effect of IMF Programs on Labor. *World Development*, 30(1), 121-139.

Vreeland, J. R. (2003). *The IMF and Economic Development*: Cambridge: Cambridge University Press.

Vreeland, J. R. (2006). IMF program compliance: Aggregate index versus policy specific research strategies. *Review of International Organizations*, 1(4), 359-378.

Vreeland, J. R. (2007). *The International Monetary Fund: Politics of Conditional Lending*: London: Routledge.

Walter, S. (2010). Globalization and the welfare state: Testing the microfoundations of the compensation hypothesis. *International Studies Quarterly*, *54*(2), 403-426.

Weber, M. (1978). *Economy and Society: An outline of interpretive sociology:* University of California Press.

Wei, S.-J., & Zhang, Z. (2010). Do external interventions work? The case of trade reform conditions in IMF supported programs. *Journal of Development Economics*, 92(1), 71-81.

Woo, B. (2013). Conditional on conditionality: IMF program design and foreign direct investment. *International Interactions*, 39(3), 292-315.

Woods, N. (2006). *The Globalizers: The IMF, the World Bank, and Their Borrowers*. Ithaca (NY): Cornell University Press.

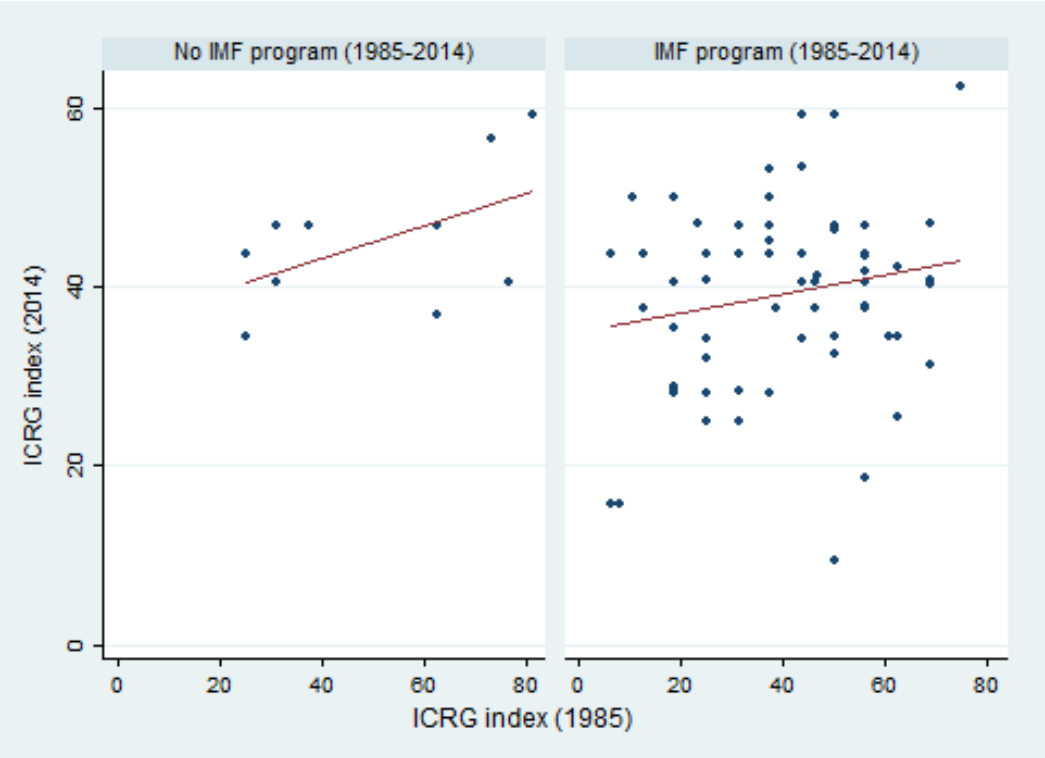
World Bank (1994). Adjustment in Africa: reforms, results, and the road ahead. Washington D.C.: World Bank.

World Bank (2015). World Development Indicators. Retrieved from: http://data.worldbank.org.

Zeugner, S., & Feldkirchner, M. (2015). Bayesian Model Averaging Employing Fixed and Flexible Priors: The BMS Package for R. *Journal of Statistical Software*, 68(4), 1-37.

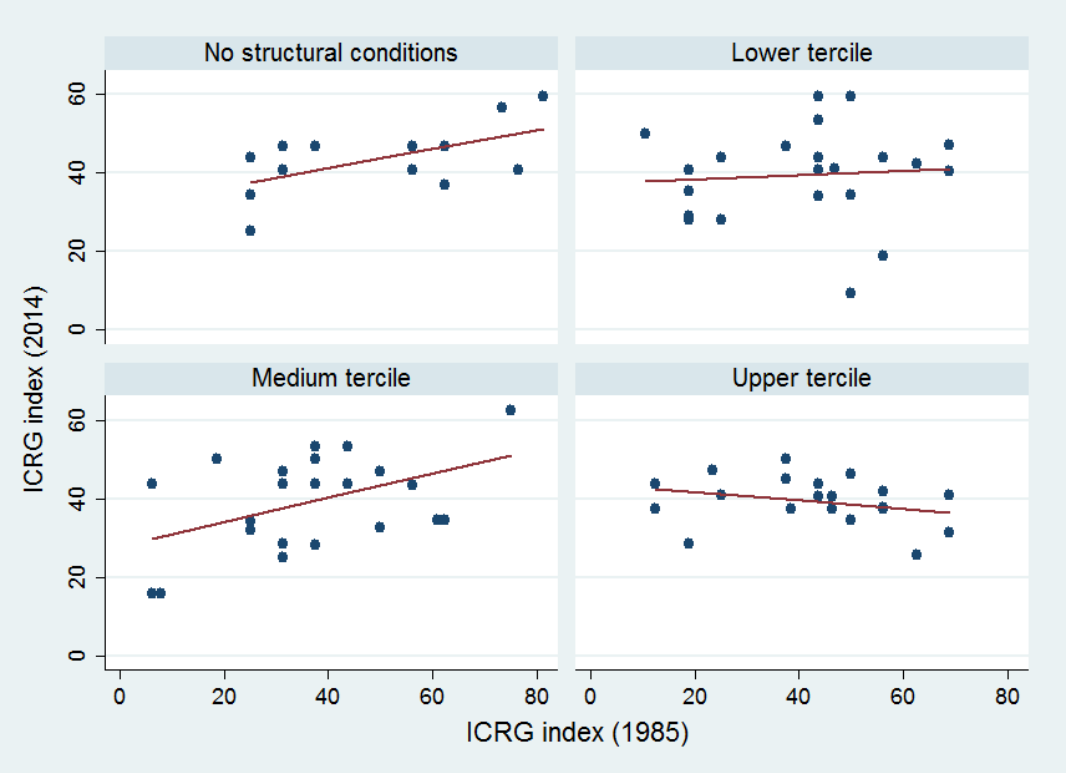
Figures

Figure 1: Before-after plot for non-IMF countries and IMF countries



Notes: The three components of the ICRG index include bureaucratic quality, rule of law, and control of corruption. Plots for the bureaucratic quality only are similar but the fit is less precise.

Figure 2: Before-after plot for non-IMF countries and three groups of structural conditionality



Notes: The three components of the ICRG index include bureaucratic quality, rule of law, and control of corruption. Plots for just the bureaucratic quality indicator are similar but the fit is less precise.

Tables

Table 1: The effect of IMF programs on bureaucratic quality using fixed-effects estimation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
IMF program	-1.59			-3.05			-4.16		
	(2.56)			(2.79)			(2.82)		
All conditions		0.28**			0.16			0.15	
		(0.13)			(0.13)			(0.12)	
Structural conditions			-1.06***			-0.83**			-0.92***
			(0.37)			(0.35)			(0.34)
Quantitative conditions			0.68***			0.48**			0.49**
			(0.21)			(0.20)			(0.19)
Log(GDP per capita)				11.26*	20.79**	17.99*	10.81*	24.20***	21.69**
				(5.90)	(9.20)	(9.14)	(6.20)	(8.60)	(8.38)
Polity IV index				0.41	-0.08	-0.13	0.25	-0.28	-0.32
				(0.35)	(0.44)	(0.46)	(0.34)	(0.44)	(0.45)
Civil war				-1.66	-10.74	-9.35	-1.15	-9.91	-8.20
				(5.32)	(7.27)	(6.99)	(5.29)	(6.67)	(6.46)
GDP growth				0.12	0.18	0.20	0.12*	0.07	0.10
				(0.07)	(0.21)	(0.22)	(0.07)	(0.23)	(0.23)
ODA per capita							0.04	0.09**	0.09**
							(0.03)	(0.04)	(0.04)
Trade openness							0.07	0.14*	0.15*
							(0.06)	(0.08)	(0.08)
Log(Oil per capita)							-4.12	-4.98	-5.84*
							(2.50)	(3.26)	(3.32)
r2_w	0.05	0.06	0.10	0.09	0.13	0.16	0.11	0.19	0.21
N	500	290	290	421	262	262	412	258	258

Country-clustered standard errors in parentheses
* p<.1, ** p<.05, *** p<.01

Table 2: The effect of IMF programs on bureaucratic quality in different regions using fixed-effects estimation

	(1) SSA	(2) Non-SSA	(3) SSA	(4) Non-SSA	(5) SSA	(6) Non-SSA
Structural conditions	-1.49	-0.71**	-1.25	-0.59*	-1.36	-0.68**
	(0.95)	(0.33)	(1.00)	(0.32)	(1.17)	(0.29)
Quantitative conditions	0.66	0.48**	0.37	0.30	0.42	0.28
	(0.40)	(0.18)	(0.41)	(0.19)	(0.46)	(0.18)
Log(GDP per capita)			24.63	-12.59	29.48*	-10.71
			(18.16)	(8.95)	(16.13)	(10.15)
Polity IV index			-0.79	1.56**	-0.92	1.26**
			(0.59)	(0.62)	(0.66)	(0.58)
Civil war			2.10	-12.31***	2.79	-11.55***
			(16.51)	(3.63)	(17.08)	(3.59)
GDP growth			-0.06	0.64**	-0.01	0.59**
			(0.34)	(0.26)	(0.51)	(0.26)
ODA per capita					0.06	0.10***
					(0.11)	(0.03)
Trade openness					0.16	0.04
					(0.15)	(0.08)
Log(Oil per capita)					-3.24	-3.61
					(12.60)	(3.33)
r2_w	0.14	0.28	0.26	0.43	0.29	0.48
N	121	169	104	158	100	158

Notes: SSA is Sub-Saharan Africa. Estimates conditioned on IMF program. Country-clustered standard errors in parentheses

* p<.1, ** p<.05, *** p<.01

Table 3: The effect of IMF programs on bureaucratic quality using fixed-effects estimation and sample selection correction

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
IMF program	-5.06			-5.73			-7.52		
	(4.60)			(5.46)			(5.17)		
All conditions		0.29**			0.17			0.05	
		(0.14)			(0.14)			(0.11)	
Structural conditions			-1.13***			-0.96**			-1.05***
			(0.40)			(0.40)			(0.38)
Quantitative conditions			0.72***			0.54**			0.54***
			(0.23)			(0.20)			(0.20)
Non-selection hazard	2.20	2.00	3.27	1.98	3.40	5.12	2.15	2.49	4.35
	(2.27)	(4.12)	(4.14)	(2.92)	(6.46)	(6.37)	(2.84)	(6.28)	(6.09)
Control variables	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Additional controls	No	No	No	No	No	No	Yes	Yes	Yes
r2_w	0.03	0.06	0.11	0.08	0.14	0.17	0.09	0.09	0.23
N	413	281	281	359	254	254	250	250	250

Country-clustered standard errors in parentheses
* p<.1, ** p<.05, *** p<.01

Table 4: The effect of IMF programs on bureaucratic quality in different regions using fixed-effects estimation and sample selection correction

	(1)	(2)	(3)	(4)	(5)	(6)
	SSA	Non-SSA	SSA	Non-SSA	SSA	Non-SSA
Structural conditions	-1.54	-0.79**	-1.72	-0.60*	-1.96	-0.70**
	(0.99)	(0.38)	(1.09)	(0.33)	(1.27)	(0.29)
Quantitative conditions	0.71	0.52**	0.49	0.32*	0.55	0.30
	(0.42)	(0.19)	(0.42)	(0.18)	(0.43)	(0.18)
Non-selection hazard	4.36	3.96	29.84**	0.21	35.12***	-1.13
	(7.84)	(4.80)	(13.64)	(6.54)	(11.68)	(5.77)
Control variables	No	No	Yes	Yes	Yes	Yes
Additional controls	No	No	No	No	Yes	Yes
r2_w	0.14	0.29	0.33	0.44	0.39	0.49
N	120	161	103	151	99	151

Notes: SSA is Sub-Saharan Africa. Estimates conditioned on IMF program. Country-clustered standard errors in parentheses

* p<.1, ** p<.05, *** p<.01

Table 5: The effect of IMF programs on bureaucratic quality using instrumental-variable estimation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
IMF program	-5.06	-5.75	-31.39	-5.73	-7.80	-34.16	-7.52	-8.14	-31.14
	(4.60)	(9.92)	(66.68)	(5.46)	(9.27)	(36.02)	(5.17)	(8.74)	(29.81)
All conditions		0.04			0.13			0.04	
		(0.56)			(0.50)			(0.48)	
Structural conditions			-2.82			-2.84			-2.76
			(4.96)			(2.85)			(2.49)
Quantitative conditions			2.57			2.61			2.25
			(5.19)			(3.04)			(2.55)
Non-selection hazard	2.20	2.30	4.43	1.98	4.84	4.84	2.15	2.20	4.28
	(2.27)	(2.41)	(5.39)	(2.92)	(4.95)	(4.95)	(2.84)	(2.78)	(4.20)
Control variables	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Additional controls	No	No	No	No	No	No	Yes	Yes	Yes
r2_w	0.03	0.03	0.03	0.08	0.08	0.09	0.09	0.09	0.10
N	413	413	413	359	358	358	351	350	350
F-stat: All conditions		10.82			10.82			11.00	
F-stat: Structural conditions			30.89			22.85			21.21
F-stat: Quantitative conditions			9.16			8.25			8.35

Notes: All estimates are based on two-stage least squares fixed-effects regressions. Country-clustered standard errors in parentheses
* p<.1, ** p<.05, *** p<.01

Table 6: The effect of IMF programs on bureaucratic quality in different regions using instrumental-variable estimation

	(1)	(2)	(3)	(4)	(5)	(6)
	SSA	Non-SSA	SSA	Non-SSA	SSA	Non-SSA
Structural conditions	-2.77	-3.35*	-1.96	-3.77**	-2.46	-3.95**
	(3.29)	(1.86)	(3.25)	(1.84)	(2.93)	(1.88)
Quantitative conditions	3.17*	3.19*	3.49	3.62**	3.74	3.49**
	(1.81)	(1.66)	(3.16)	(1.65)	(3.35)	(1.64)
IMF program	-42.61*	-36.44*	-53.66	-47.45**	-55.61	-46.39**
	(24.02)	(18.68)	(37.37)	(23.04)	(38.98)	(21.85)
Non-selection hazard	10.68*	3.03	19.19	5.38	18.21	4.84
	(6.30)	(3.51)	(11.78)	(4.71)	(13.28)	(4.06)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Additional controls	No	No	No	No	Yes	Yes
r2_w	0.17	0.24	0.30	0.34	0.31	0.37
N	156	257	127	231	123	227
F-stat: Structural conditions	15.19	23.57	14.62	20.11	16.42	17.49
F-stat: Quantitative conditions	2.75	15.47	1.29	14.18	1.14	11.38

Notes: All estimates are based on two-stage least squares fixed-effects regressions. Country-clustered standard errors in parentheses

* p<.1, ** p<.05, *** p<.01

Appendices

Appendix A: Alternative measures

Table A1: Results using implementation-corrected conditionality

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
IMF program	-3.05			-5.73			-5.73	-7.25	-35.60
	(2.79)			(5.46)			(5.46)	(10.86)	(42.56)
All conditions		1.86**			1.68*			0.09	
		(0.84)			(0.91)			(0.62)	
Structural conditions			-2.99			-4.68**			-3.05
			(1.81)			(1.99)			(3.23)
Quantitative conditions			3.44***			3.42***			2.75
			(1.16)			(1.20)			(3.59)
Non-selection hazard				1.98	2.95	3.77	1.98	2.12	4.98
				(2.92)	(6.19)	(5.86)	(2.92)	(2.91)	(5.51)
Non-selection hazard correction	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Endogeneity correction	No	No	No	No	No	No	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional controls	No	No	No	No	No	No	No	No	No
r2_w	0.09	0.15	0.17	0.08	0.15	0.19	0.08	0.09	0.10
N	421	262	262	359	254	254	359	358	358
F-statistics								5.78	11.60
(Excluded instruments)									3.35

Notes: All models include the standard set of control variables. All conditions are net of waivers (see Kentikelenis, Stubbs, and King 2016). Country-clustered standard errors in parentheses

* p<.1, ** p<.05, *** p<.01

Table A2: Results using implementation-corrected conditionality in different regions

	(1)	(2)	(3)	(4)	(5)	(6)
	SSA	Non-SSA	SSA	Non-SSA	SSA	Non-SSA
Structural conditions	-9.09*	-3.33**	-9.70*	-3.23*	-1.96	-3.95*
	(4.75)	(1.61)	(5.16)	(1.68)	(3.25)	(2.02)
Quantitative conditions	3.53	2.43**	3.14	2.20*	3.49	3.80**
	(2.08)	(1.19)	(2.00)	(1.17)	(3.16)	(1.87)
Non-selection hazard			25.65*	0.24	19.19	5.59
			(13.64)	(6.04)	(11.78)	(5.02)
Non-selection hazard correction	No	No	Yes	Yes	Yes	Yes
Endogeneity correction	No	No	No	No	Yes	Yes
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Additional controls	No	No	No	No	No	No
r2_w	0.30	0.44	0.35	0.44	0.28	0.33
N	104	158	103	151	127	231
F-statistics					14.61	13.63
(Excluded instruments)					1.29	9.88

Notes: SSA is Sub-Saharan Africa. All models include the standard set of control variables. All conditions are net of waivers. Country-clustered standard errors in parentheses

* p<.1, ** p<.05, *** p<.01

Table A3: Results using ICRG index of administrative capacity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
IMF program	-2.14			-9.36**			-9.36**	-11.68**	-29.87
	(1.78)			(3.59)			(3.59)	(5.61)	(27.57)
All conditions		0.10			0.11			0.14	
		(0.07)			(0.07)			(0.33)	
Structural conditions			-0.40*			-0.49**			-1.87
			(0.20)			(0.22)			(2.11)
Quantitative conditions			0.26***			0.30***			1.85
			(0.10)			(0.10)			(2.36)
Non-selection hazard				5.05***	3.10	4.00	5.05***	5.27***	7.11**
				(1.87)	(4.27)	(4.17)	(1.87)	(1.77)	(3.43)
Non-selection hazard correction	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Endogeneity correction	No	No	No	No	No	No	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional controls	No	No	No	No	No	No	No	No	No
r2_w	0.17	0.22	0.24	0.19	0.23	0.26	0.19	0.19	0.19
N	421	262	262	359	254	254	359	358	358
F-statistics								8.73	18.01
(Excluded instruments)									5.49

Notes: The dependent variable is the combined index of bureaucratic quality, control of corruption, and rule of law, sourced from the ICRG dataset (The PRS Group 2015). Country-clustered standard errors in parentheses

* p<.1, ** p<.05, *** p<.01

Table A4: Results using ICRG index of administrative capacity in different regions

	(1)	(2)	(3)	(4)	(5)	(6)
	SSA	Non-SSA	SSA	Non-SSA	SSA	Non-SSA
Structural conditions	-0.95	-0.16	-1.29*	-0.19	-0.41	-1.62*
	(0.59)	(0.22)	(0.66)	(0.22)	(1.56)	(0.92)
Quantitative conditions	0.29	0.19*	0.37*	0.22**	0.38	1.68*
	(0.21)	(0.11)	(0.18)	(0.11)	(0.87)	(0.88)
Non-selection hazard			21.41**	-0.41	8.27*	6.43**
			(8.80)	(4.12)	(4.52)	(2.70)
Non-selection hazard correction	No	No	Yes	Yes	Yes	Yes
Endogeneity correction	No	No	No	No	Yes	Yes
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Additional controls	No	No	No	No	No	No
r2_w	0.20	0.43	0.34	0.43	0.21	0.34
N	104	158	103	151	127	231
F-statistics					14.61	13.63
(Excluded instruments)					1.30	9.87

Notes: SSA is Sub-Saharan Africa. The dependent variable is the combined index of bureaucratic quality, control of corruption, and rule of law, sourced from the ICRG dataset (The PRS Group 2015). Country-clustered standard errors in parentheses

* p<.1, ** p<.05, *** p<.01

Appendix B: Descriptive statistics and variable definitions

Table B1: Descriptive statistics

Variable	N	Mean	Sd	Min	Max
Dependent variables					
Bureaucratic quality	506	38.40	22.86	0	100
ICRG index	506	42.84	14.32	6.25	84.24
Covariates for outcome equation					
IMF program	500	0.67	0.47	0	1
All conditions	423	10.02	10.54	0	54.33
Structural conditions	423	1.71	3.15	0	28.67
Quantitative conditions	423	8.31	8.37	0	33.67
Log(GDP per capita)	458	7.13	1.04	4.86	9.31
Polity IV index	468	0.47	6.56	-10	10.00
Civil war	506	0.08	0.27	0	1.00
GDP growth	462	3.11	8.40	-64.05	106.28
ODA per capita	506	37.48	54.26	-14.55	722.66
Trade openness	459	67.09	36.23	0.36	233.94
Log(Oil per capita)	489	0.30	0.58	0	3.39
Covariates for selection stage					
Countries under program	434	57.02	8.71	47.4	68.17
UNGA Vote Alignment with G7	422	0.61	0.08	0.47	0.90
British colonial origin	506	0.30	0.46	0	1
Reserves (in months of imports)	334	3.79	3.88	0.04	34.72
Debt service (% GNI)	356	5.95	4.98	0.17	41.16
Lagged GDP growth	405	3.42	4.20	-16.03	32.25
Current account balance (% GDP)	331	-3.65	7.29	-33.50	30.33
Hyperinflation	375	0.05	0.22	0	1
FDI inflows (% GDP)	443	2.75	6.51	-5.99	89.48
Log(Population)	489	16.40	1.47	12.84	21.02
Lack of civil liberties	496	4.41	1.52	1	7
Political globalization	501	56.09	17.90	21.00	93.45
Executive election	414	0.08	0.27	0	1
Competitive executive election	414	0.43	0.50	0	1
Legislative election	414	0.18	0.39	0	1
Competitive legislative election	414	0.50	0.50	0	1
UNSC Membership	418	0.07	0.26	0	1

Table B2: Correlation table for all second-stage variables (N=290)

	A	B	C	D	E	F	G	H	I	J	K	L	M
A. Bureaucratic quality	1.00												
B. ICRG index	0.66	1.00											
C. IMF program	-0.10	-0.13	1.00										
D. All conditions	-0.04	-0.07	0.63	1.00									
E. Structural conditions	-0.09	-0.09	0.35	0.77	1.00								
F. Quantitative conditions	-0.01	-0.06	0.67	0.97	0.59	1.00							
G. Log(GDP per capita)	0.29	0.24	-0.28	-0.13	-0.11	-0.12	1.00						
H. Polity IV index	0.23	0.06	0.18	0.20	0.13	0.20	0.31	1.00					
I. Civil war	-0.15	-0.18	-0.02	-0.09	-0.09	-0.08	-0.16	-0.05	1.00				
J. GDP growth	0.14	0.09	0.05	0.03	0.01	0.03	-0.16	-0.17	-0.03	1.00			
K. ODA per capita	-0.02	0.05	0.10	0.12	0.07	0.13	-0.13	0.07	-0.09	-0.04	1.00		
L. Trade openness	0.15	0.16	-0.08	0.02	0.08	-0.01	0.19	0.07	-0.24	0.02	0.40	1.00	
M. Log(Oil per capita)	0.02	-0.01	-0.20	-0.16	-0.14	-0.15	0.41	-0.12	-0.06	0.06	-0.27	-0.06	1.00

Table B3: Correlation table for all first-stage variables (N=261)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
A. Bureaucratic quality	1.00																					
B. IMF program	-0.19	1.00																				
C. All conditions	-0.07	0.59	1.00																			
D. Structural conditions	-0.11	0.32	0.78	1.00																		
E. Quantitative conditions	-0.05	0.63	0.96	0.59	1.00																	
F. Countries under program	0.07	0.03	0.21	0.27	0.16	1.00																
G. UNGA Vote Alignment with G7	0.00	0.13	0.40	0.47	0.32	0.56	1.00															
H. British colonial origin	0.12	-0.10	-0.08	-0.04	-0.09	-0.01	-0.27	1.00														
I. Reserves (in months of imports)	0.19	-0.27	-0.11	-0.05	-0.12	0.08	0.08	0.07	1.00													
J. Debt service (% GNI)	0.08	0.05	0.03	-0.06	0.06	-0.01	0.02	-0.13	-0.18	1.00												
K. Lagged GDP growth	0.18	-0.20	-0.01	0.02	-0.02	-0.04	0.01	0.11	0.18	-0.20	1.00											
L. Current account balance (% GDP)	0.16	-0.17	-0.14	-0.07	-0.16	0.04	-0.02	0.02	0.42	-0.24	0.10	1.00										
M. Hyperinflation	-0.11	-0.01	-0.12	-0.10	-0.12	-0.06	-0.09	-0.13	0.01	0.01	-0.22	-0.14	1.00									
N. FDI inflows (% GDP)	-0.04	-0.04	0.17	0.22	0.13	0.21	0.25	-0.11	0.05	0.10	0.28	-0.19	-0.03	1.00								
O. Log(Population)	0.16	-0.12	-0.10	-0.03	-0.11	0.02	-0.13	0.11	0.11	-0.20	0.18	0.24	0.01	-0.16	1.00							
P. Lack of civil liberties	-0.16	-0.04	-0.12	-0.07	-0.13	-0.08	-0.32	0.13	-0.19	-0.08	0.04	0.03	-0.02	-0.06	0.20	1.00						
Q. Political globalization	0.20	-0.03	0.09	0.17	0.05	0.13	0.24	-0.03	0.22	-0.09	0.12	0.23	-0.02	-0.02	0.55	-0.06	1.00					
R. Executive election	-0.08	0.10	-0.01	0.02	-0.02	-0.02	-0.05	-0.08	-0.04	-0.04	0.07	0.04	0.19	-0.01	0.00	0.06	0.02	1.00				
S. Competitive executive election	0.15	0.02	0.13	0.16	0.10	0.22	0.36	-0.09	0.17	-0.16	0.04	0.09	-0.02	0.08	0.00	-0.43	0.26	-0.03	1.00			
T. Legislative election	0.02	0.09	0.06	0.01	0.08	0.08	0.04	-0.02	-0.01	-0.02	-0.12	0.08	0.03	-0.08	0.00	0.00	0.05	0.43	0.05	1.00		
U. Competitive legislative election	0.12	-0.02	0.10	0.13	0.07	0.14	0.33	-0.29	0.15	-0.08	0.10	0.12	0.09	0.13	0.07	-0.40	0.27	0.00	0.66	0.03	1.00	
V. UNSC Membership	0.09	0.05	-0.03	-0.04	-0.02	-0.08	-0.05	0.00	-0.02	-0.03	-0.04	-0.04	0.03	-0.09	0.15	0.01	0.15	0.07	-0.03	0.05	0.01	1.00

Table B4: Variable description and data sources

Variable	Description	Sources
Dependent variables		
Bureaucratic quality	Expert survey-based measure of bureaucratic quality from the International Country Risk Guide (ICRG). According to the methodology, "high points are given to countries where the bureaucracy has the strength and expertise to govern without drastic changes in policy or interruptions in government services. In these low-risk countries, the bureaucracy tends to be somewhat autonomous from political pressure and to have an established mechanism for recruitment and training"	PRS Group 2015
ICRG index	Expert survey-based index measure consisting of bureaucratic quality (L), rule of law (I), and corruption (F) from the International Country Risk Guide (ICRG). Index calculated as 1.5L+I+F and multiplied by 50/9 in order to generate values between 0 and 100	PRS Group 2015
Covariates for outcome equation		
IMF program	Binary variable indicating the presence of an IMF program (any active program in any year during the previous period)	Kentikelenis, Stubbs, and King 2016
All conditions	Number of (binding) conditions in the previous period; a condition is binding if it is either a prior action, a structural performance criterion, or a quantitative performance criterion; number of conditions is divided by number of years in which IMF program was active	
Structural conditions	Number of (binding) structural conditions in the previous period, divided by number of years in which IMF program was active; structural conditions are defined as the sum of prior actions and structural performance criteria	
Quantitative conditions	Number of (binding) quantitative conditions in the previous period, divided by number of years in which IMF program was active; equals the number of quantitative performance criteria	
Log(GDP per capita)	Natural logarithm of GDP per capita; GDP per capita (constant 2005 US$) [NY.GDP.PCAP.KD] at the beginning of the period	World Bank 2015
Polity IV index	Combined index of Polity IV, defined as the democracy score net of the autocracy score, at the beginning of the period	Marshall, Gurr, and Jaggers 2015
Civil war	Binary variable indicating the presence of a civil war according to the UCDP/PRIO definition at the beginning of the period	Gleditsch et al. 2002
GDP growth	GDP growth (annual %) [NY.GDP.MKTP.KD.ZG] at the beginning of the period	World Bank 2015
ODA per capita	Net ODA received per capita (current US$) [DT.ODA.ODAT.PC.ZS] at the beginning of the period	World Bank 2015
Trade openness	Trade (% of GDP) [NE.TRD.GNFS.ZS] at the beginning of the period	World Bank 2015
Log(Oil per capita)	Natural log of one plus oil production in metric tons divided by total population at the beginning of the period	Ross 2013

Covariates for selection stage

Countries under program	Number of countries under any IMF program (lagged by one period as the indicator of IMF program itself)	Kentikelenis, Stubbs, and King 2016
UNGA Vote Alignment with G7	Average vote alignment index of the country with all G7 countries averaged over the previous period	Strezhnev and Voeten 2013
British colonial origin	Binary variable indicating British colonial origin	Alesina and Dollar 2000
Reserves (in months of imports)	Total reserves in months of imports [FI.RES.TOTL.MO] averaged over the previous period, sourced from the QoG database (Teorell et al. 2016)	World Development Indicators (World Bank 2015)
Debt service (% GNI)	Total debt service (% of GNI) [DT.TDS.DECT.GN.ZS] averaged over the previous period, sourced from the QoG database (Teorell et al. 2016)	
Lagged GDP growth	GDP growth (annual %) [NY.GDP.MKTP.KD.ZG] averaged over the previous period, sourced from the QoG database (Teorell et al. 2016)	
Current account balance (% GDP)	Current account balance (% of GDP) [BN.CAB.XOKA.GD.ZS] averaged over the previous period, sourced from the QoG database (Teorell et al. 2016)	
Hyperinflation	Binary variable indicating high inflation above 200% in the previous period, underlying inflation data sourced from the QoG database (Teorell et al. 2016)	
FDI inflows (% GDP)	Foreign direct investment, net inflows (% of GDP) [BX.KLT.DINV.WD.GD.ZS] averaged over the previous period, sourced from the QoG database (Teorell et al. 2016)	
Log(Population)	Natural log of population, sourced from the QoG database (Teorell et al. 2016)	United Nations 2015
Lack of civil liberties	Freedom House indicator: Civil liberties (higher values indicate less free countries), averaged over the previous period, sourced from the QoG database (Teorell et al. 2016)	Freedom House 2015
Political globalization	Political globalization index, averaged over the previous period	Dreher 2006a
Executive election	Binary variable indicating a executive election at the beginning of the previous period, sourced from the QoG database (Teorell et al. 2016)	Database of Political Institutions (see Beck et al. 2001)
Competitive executive election	Binary variable indicating a competitive executive election at the beginning of the previous period (dpi_eipc==7), sourced from the QoG database (Teorell et al. 2016)	
Legislative election	Binary variable indicating a legislative election at the beginning of the previous period, sourced from the QoG database (Teorell et al. 2016)	
Competitive legislative election	Binary variable indicating a competitive legislative election at the beginning of the previous period (dpi_lipc==7) , sourced from the QoG database (Teorell et al. 2016)	
UNSC Membership	Binary variable indicating whether the country was temporary member in the United Nations Security Council at the beginning of the previous period	Dreher, Sturm, and Vreeland 2015

Appendix C: Determining the appropriate lag length

Although the choice of lag length in a time-series cross-section model may be consequential for all subsequent analyses, it can be somewhat arbitrary. To address this uncertainty, we use Bayesian Model Averaging (BMA). BMA has the goal to identify the "best specification" of a statistical model, which maximizes the posterior probability on the basis of model fit and model parsimony. Recent literature suggests using BMA also for the purpose of identifying plausible lag lengths (Cranmer, Rice, and Siverson 2016).[6] Theoretically, we are interested in medium-term effects of IMF programs and therefore define our upper lag limit at seven years. Further expanding the possible lag length severely limits the available observations due to listwise deletion.

As control variables, we use the baseline specification, which includes the natural log of per capita income, democracy, civil war, and economic growth. We add to the predictor space the contemporaneous count of all conditions along with up to seven lags from our database. All models are required to include two-way fixed effects. The results indicate that IMF programs start to matter only after five years, if at all, and the peak effect occurs in the sixth year. We therefore expect IMF programs to have their highest potential impact in a six-year period.

Table C1: Posterior inclusion probabilities for the variable ALL CONDITIONS in annual data

All conditions	PIP (%)	EV	SD	Model 1	Model 2	Model 3
t	0	0.00	0.00	.	.	.
t-1	0	0.00	0.00	.	.	.
t-2	0	0.00	0.00	.	.	.
t-3	4.6	0.0012	0.01	.	.	.
t-4	0	0.00	0.00	.	.	.
t-5	4.9	0.0014	0.01	.	.	0.03
t-6	12.1	0.0043	0.01	.	0.04	.
t-7	3.8	0.0009	0.01	.	.	.
Baseline controls	Yes					
Country-fixed effects	Yes					
Year-fixed effects	Yes					

Notes: The table summarizes the results of regressions of bureaucratic quality on various lags of ALL CONDITIONS and standard control variables using annual panel data. PIP (%) refers to the posterior inclusion probability of a variable measured in percent. EV is the posterior point estimate, SD the posterior standard deviation.

[6] We implement our analysis in R (R Development Core Team 2016) using the BMA package (Raftery, Hoeting, Volinsky, Painter, and Yeung 2015) and BMS package (Zeugner and Feldkirchner 2015).

Appendix D: Extreme Bounds Analysis

Recent advances in computational power have made it possible to explore a potentially vast model space. As discussed in previous appendices, Bayesian Model Averaging can be used to identify the most probable models within that model space. As a pre-processing method, it is useful because it enables the researcher to limit the number of regressions to be shown without falling prey to bias from arbitrary selection of model specifications from the pool of control variables. In contrast, Extreme Bounds Analysis (EBA) does not claim to identify the best model but seeks to provide insights into the robustness of the estimands under different model specifications (for applications in the IMF literature, see, Moser and Sturm 2011; Oberdabernig 2013).

To probe the robustness of our key findings, we conduct an EBA variant in which we permute over a vast set of potential first-stage specifications and examine the distribution of coefficients in the second stage. To keep things manageable, we fix the second-stage specification, using the baseline controls. When is a variable thought to be robust? There is no consensus in the literature on this question (see, e.g., Leamer 1983; Sala-i-Martin 1997; Sturm and de Haan 2005). We follow Sala-i-Martin (1997) and study the entire distribution of coefficients. Following Sturm and de Haan (2005), we consider variables to be robust if at least 95% of the coefficients are either below or above zero. This is a reasonable criterion because some models might be misspecified. Hence, we report the fraction of coefficients with a positive sign (pPos), along with the range of coefficient values (bMin and bMax). The range is a very robust statistic as it is not affected by the weighting that is applied when calculating a weighted-mean coefficient. Nonetheless, we report the mean coefficient, weighted by the plausibility of the model specification as reflected in the associated Bayesian Information Criterion (BIC). This is to ensure that a posteriori more likely models will have more influence on the average coefficient (bMean). We also report the mean standard error, which computes as the square root of the sum of the mean variance and the cross-specification variance of the estimates (e.g., Rubin 1987; Stone 2004; Hlavac 2014). However, we consider the standard errors only as a secondary decision criterion for robustness, following the rule of thumb that robust variables should be significant in at least half of all models. In deviation from Moser and Sturm (2011), we require significance at the five-percent level (pSig95).

An EBA basically requires the researcher to specify two sets of variables – "free variables" that enter each specification, and "doubtful variables" which may or may not be included in any individual specification. As free variables, we use the included instruments from the outcome stage, the excluded instruments of IMF participation – PAST PROGRAM, COUNTRIES UNDER PROGRAMS, and UNGA VOTING ALIGNMENT – and binary indicators for specific regions and time periods. Included instruments from the second stage are the log of per capita income, GDP growth, the log of population, population density, the dependency ratio, democracy, the initial value of the outcome variable, and an indicator for British colonial origin. The latter two variables are included because we estimate pooled outcome regressions in the EBA approach (which is simply a concession to time – least-square dummy variable models are much slower but would yield very similar results).

In addition, we allow for inclusion of up to five so-called "doubtful variables" in each regression. Our doubtful variables include economic variables such as FOREIGN RESERVES in months of imports, CURRENT ACCOUNT BALANCE, ECONOMIC GROWTH, DEBT SERVICE AS OF GNI, HYPERINFLATION, and FOREIGN DIRECT INVESTMENT (Conway 1994; Vreeland 2003; Vreeland

2007; Bauer, Cruz, and Graham 2012), political variables such as EXECUTIVE ELECTIONS (e.g., Przeworski and Vreeland 2000; Dreher and Jensen 2007; Moser and Sturm 2011; Rickard and Caraway 2014; Nooruddin and Woo 2015), CLOSE EXECUTIVE ELECTIONS, LEGISLATIVE ELECTIONS (Dreher 2004; Moser and Sturm 2011), CLOSE LEGISLATIVE ELECTIONS, and temporary membership in the United Nations Security Council (UNSC MEMBER) to capture international horse-trading (Dreher, Sturm, and Vreeland 2009). We also include CIVIL LIBERTIES from Freedom House, POLITICAL GLOBALIZATION from the KOF INDEX, and POPULATION, seeking to capture the political openness of the country to international affairs beyond the already included measure of democracy. We place no restrictions on the possible combinations of doubtful variables except for the requirement that an election indicator and its respective indicator of a competitive election must occur together. Overall, this leads to 3,472 models in the first-stage. The following table shows the distribution of coefficients in the outcome stage for BUREAUCRATIC QUALITY as dependent variable.

We find that past programs, countries under programs, and UNGA voting alignment are robust predictors of current programs. These variables are also almost always significant. Furthermore, the included instruments have meaningful estimates, for example, poorer countries, countries not currently suffering from civil war, and low-growth countries are more likely to have a program with the Fund. Other economic variables found to be robust in terms of their sign are foreign reserves, current account balance, and debt service, while the unexpected negative effect of hyperinflation might be due to reverse causality. Politics also matters. Consistent with prior research, IMF programs are more likely after elections (e.g., Przeworski and Vreeland 2000; Dreher and Vaubel 2004; Moser and Sturm 2011), in either branch of government. Close elections in either branch are associated with a lower likelihood of subsequent IMF programs. Countries that are more outward-oriented and that receive more attention in world politics (for example through membership in the Security Council) are more likely to obtain Fund assistance. Few variables also pass the significance test of robustness (at least 95 percent of the estimates being statistically significant). These are lagged IMF program, GDP per capita, and civil war (British colonial origin does not feature fixed-effect models but is a significant predictor in pooled regressions).

Consider the implications of different first-stage models for the estimands in the outcome equations. Here, we only show the results from the models that include the two types of conditions, but all other tables are available from the authors on request. The main insight is that changes in the first-stage specification are not consequential. All variables (at the exception of democracy) are highly stable across the 3,472 different first-stage models. In particular, structural conditionality always has a negative coefficient. According to the EBA approach, this variable is robust. The same conclusion holds for quantitative conditions, which always have a positive coefficient. In addition, the estimated coefficient is statistically significant at the 5%-level in all models. The control variables (except democracy) are robust determinants of bureaucratic quality, as they are all meaningfully signed, keep that sign throughout all models, and are statistically significant. The non-selection hazard is robustly negative but never reaches statistical significance.

Table D1: Robustness of coefficients for determinants of IMF programs

Variable	bMean	bMin	bMax	pPos	seMean	pSig95	nModels
Lagged IMF program	1.406	1.265	1.644	1	0.29	1	3,472
Countries under program	-0.07	-0.14	0.035	0.022	0.076	0.015	3,472
UNGA Vote Alignment with G7	7.96	4.338	11.288	1	4.486	0.407	3,472
Log(GDP per capita)	-0.917	-1.174	-0.801	0	0.341	1	3,472
Polity IV index	0.004	-0.025	0.034	0.627	0.035	0	3,472
Civil war	-0.948	-1.386	-0.806	0	0.434	0.998	3,472
GDP growth	-0.015	-0.029	0.013	0.021	0.034	0	3,472
British colonial origin	-0.928	-1.306	-0.753	0	0.373	1	3,472
Initial bureaucratic quality	-0.002	-0.003	0.003	0.043	0.006	0	3,472
Reserves (in months of imports)	-0.059	-0.079	-0.043	0	0.044	0	913
Debt service (% GNI)	0.067	0.044	0.078	1	0.038	0.164	900
Lagged GDP growth	-0.115	-0.181	-0.087	0	0.054	0.788	933
Current account balance (% GDP)	0.003	-0.009	0.022	0.592	0.022	0	887
Hyperinflation	-0.435	-1.025	-0.011	0	0.748	0	900
FDI inflows (% GDP)	-0.003	-0.017	0.019	0.314	0.038	0	929
Log(Population)	-0.089	-0.255	-0.006	0	0.12	0.044	908
Lack of civil liberties	-0.006	-0.11	0.062	0.378	0.156	0	885
Political globalization	0.012	0.005	0.03	1	0.012	0.102	896
Executive election	0.528	0.073	1.15	1	0.572	0	325
Competitive executive election	-0.599	-0.781	-0.347	0	0.364	0.058	325
Legislative election	0.426	0.274	0.479	1	0.322	0	311
Competitive legislative election	-0.699	-0.815	-0.488	0	0.355	0.626	311
UNSC Membership	0.369	0.245	0.491	1	0.442	0	918

Notes: First-stage models are probit regressions with individual effects for time periods, income groups, and country regions.

Table D2: Robustness of coefficients in the outcome stage for different first-stage models

Variable	bMean	bMin	bMax	pPos	seMean	pSig95	nModels
Structural conditions	-0.651	-0.666	-0.633	0	0.376	0	3,472
Quantitative conditions	0.394	0.379	0.407	1	0.172	1	3,472
Log(GDP per capita)	6.375	6.064	6.648	1	2.728	1	3,472
Polity IV index	0.004	-0.017	0.022	0.766	0.244	0	3,472
Civil war	-16.262	-16.787	-15.703	0	4.574	1	3,472
GDP growth	0.349	0.342	0.359	1	0.275	0	3,472
British colonial origin	19.442	18.915	19.851	1	3.331	1	3,472
Initial bureaucratic quality	0.289	0.286	0.291	1	0.056	1	3,472
Non-selection hazard	-3.163	-5.704	0.017	0.001	5.615	0	3,472

Notes: All models are pooled ordinary least square regressions with individual effects for time periods, income groups, and country regions, augmented with a non-selection hazard to account for the non-random selection into Fund programs. Standard errors adapted for cross-estimation uncertainty using (see Rubin 1987).

Appendix E: Specification issues

In this appendix, we discuss some further methodological choices in our paper. These choices refer to the use of a level specification rather than a change specification, the issue of atheoretical lags, and the choice of a sample selection correction approach.

Level specification

The first methodological issue pertains to the choice of a level specification versus a change specification of the dependent variable. In fact, both specifications would be equivalent if the level model also contained a lagged dependent variable. De Boef and Keele (2008) formally show the equivalence of the following two models:

(1) $y_{it} = a_0 + a_1\, y_{i,t-1} + ß_0\, x_{it} + ß_1\, x_{i,t-1} + e_{it}$

(2) $\Delta y_{it} = a_0 + a_1\, y_{i,t-1} + ß_0\, \Delta x_{it} + ß_1\, x_{i,t-1} + e_{it}$

They argue that the choice of either model hinges on the preference for obtaining specific interpretable estimands directly from these equations. In a response to their article, Grant and Lebo (2016) warn against the use of the Error Correction Model (equation 2) and emphasize that it can only be used for either co-integrated variables or stationary variables on both sides of the equation.

Other papers dismiss the usefulness of a lagged dependent variable altogether. Achen (2001) reveals that inclusion of a lagged dependent variable shrinks the coefficients of substantive predictors toward negligible values and hence induces bias. He goes on to caution against the inclusion of this variable unless there are strong theoretical reasons. Bellemare, Masaki, and Pepinsky (2015) criticize the widespread practice of "lag identification" and show that the use of lagged dependent variables merely moves the channel through which endogeneity biases causal estimates. Rather than absence of selection-on-observables, the necessary assumption for such a model to be identified is "no dynamics among unobservables", which is not testable.

Researchers often include a lagged dependent variable to mitigate autocorrelation (Achen 2001). In our data, empirical tests indicate autocorrelation regardless of whether we use the dependent variable in levels or in changes. Moreover, tests for stationarity are indecisive in our setting. On the one hand, they reject the null hypothesis that all panels have a unit root. On the other hand, they reject the null hypotheses that all panels are stationary. In the absence of clear guidance from empirical tests, we therefore adopt the specification that is closest to previous research to facilitate the comparability of our results.

Atheoretical lags

Most political science studies adopt a workhorse model of time-series cross-section analysis in which predictors affect outcomes with a one-year lag. We argue that state capacity tends to evolve only over a medium-term period. Short-term fluctuations in state capacity measures may contain rather limited information on the underlying relationships. In the worst case, they may simply reveal the observer biases that plagues many socio-political indicators (e.g., Kurtz and Schrank 2008; Stubbs, King, and Stuckler 2014; Hulme, Savoia, and Sen 2015).

While being able to exclude a one-year lag specification, we are still unclear about the appropriate time frame within which specific policies may affect state capacity. Methodologists advise to start with a general model that includes several distributed lags of a predictor and to test for the significant impact of specific lags (De Boef and Keele 2008). In a similar vein, Bayesian Model Averaging can be used to resolve uncertainty around lag structures (Cranmer, Rice, and Siverson 2016). We acknowledge that there remains a small bit of arbitrariness because even a BMA analysis requires as inputs the predictor space from which it selects the most plausible models. We allow for up to seven lags, arguing that this reflects a useful time horizon for related policy implications. Moreover, it becomes harder to identify any effect of the variables of interests because the set of confounders tends to grow.

Sample selection methods

The need for sample selection correction is apparent in our case. For example, consider political will as one potential source of endogeneity bias: Countries may turn to the Fund for a bailout program because they are willing to reform, and their willingness also has an independent effect on their bureaucratic quality. Any unobserved characteristics that correlate both with the incidence of an IMF program and state capacity mislead the inference from a standard regression.

There are two potential approaches to address endogeneity. One is to instrument program participation and identify a local average treatment effect. For example, if UNGA voting of the borrower with the G7 happens to be a good predictor of IMF programs, it must be assumed that UNGA voting only affects state capacity through its effect on IMF programs (unless other channels are observable and hence be controlled). Moreover, monotonicity is another assumption: UNGA voting should homogeneously affect IMF participation across countries. While these requirements already set a high bar for finding good instruments, an additional complication is that the program indicator is binary, while instrumental variable estimation assumes linearity. Nelson and Olson (1978) provide the first discussion of simultaneous equation models in which at least one endogenous regressor is binary. However, the usefulness of these models is reduced by the difficulty to derive closed-form standard errors and the biasedness of second-stage estimators that require a "Theil-like structural variance correction" (Atwood, Joglekar, and Smith 2016).

In light of these challenges, we opt for the second strategy of addressing the potential endogeneity of IMF programs. Using the so-called Heckman selection approach, we run a selection model, using preferably at least one excludable determinant and other correlates of program participation. The residuals from this model can be used to calculate the non-selection hazard, which can be included as an additional control variable in the outcome equation. Albeit not being perfect, this approach has several advantages. The inclusion of excludable instruments is not necessary but it improves the precision of the second-stage estimates. The Heckman selection correction also is the more natural approach in the presence of a binary selection indicator. In addition, it allows controlling for sample selection when conditioning the outcome equation on the presence of an IMF program. In particular, to evaluate the impact of structural conditions on state capacity, the natural study groups are IMF countries with these specific conditions versus IMF countries without them. However, these groups are both different from countries that do not experience IMF programs. The Heckman correction removes the potential bias from sub-sampling on IMF countries.

We acknowledge that it would be desirable to account for the potential endogeneity of specific types of conditions even among the group of IMF program countries. However, there are no instruments for specific condition types that not also predict IMF programs more generally. The incidence of any type of condition is highly correlated with the presence of an IMF program, simply because the latter is a necessary condition for the former. This beclouds the inferential potential of instrumental variable strategies with more than one endogenous variable.

However, we are able to leverage a recently popularized synthetic instrumentation from the aid effectiveness literature (e.g., Nunn and Qian 2014; Dreher and Langlotz 2015; Lang 2016). Take the example of structural conditions (S_{it}), which we instrument by the interaction between their country average (S_i) and the period-specific budget constraint of the Fund (B_t) (all other terms are to be understood in the usual sense):

(3) $S_{it} = ß_0 (S_i \times B_t) + ß_1 X_{it} + b_i + f_t + u_{it}$

The instrument is relevant, as indicated by the fairly high F-statistics in our results tables. The instrument is also excludable because country-specific changes in structural conditionality deviating from its long-run average are allowed to be brought about only by decisions of the Fund as a whole that do not pertain to the given country. In particular, if the Fund assists more countries in any period, its resources will be more strained, and hence it will ask the given country for more conditions in order to preserve its budget (for the relationship between loan size and depth of conditionality, see Vreeland 2003: 55). However, there is no apparent pathway from the budget constraint to the bureaucratic capacity of the given country – other than through conditionality.

In estimations with two variables on conditionality, we apply the same instrumentation to both types of conditions (both variables are continuous). This implies a two-stage least squares estimation with two potentially endogenous variables; we also allow for robust inference and clustering of standard errors by countries (see Roodman 2012).