CrossMark

# How to evaluate the effects of IMF conditionality

## An extension of quantitative approaches and an empirical application to public education spending

Thomas Stubbs[1,2] ⓘ · Bernhard Reinsberg[2,3] · Alexander Kentikelenis[4] · Lawrence King[5]

## Abstract

Following calls for a more disaggregated approach to studying the consequences of IMF programs, scholars have developed new datasets of IMF-mandated policy reforms, or 'conditionality.' Initial studies have explored how conditions have, inter alia, affected tax revenues, public sector wages, and health systems. Notwithstanding the important contributions of these studies, a methodological quandary arises as to how to quantitatively examine the effects of conditionality, as distinct from other aspects of IMF operations (e.g., credit, technical support, or aid and investment catalysis). In this article, we review and advance these methodological debates by developing an identification strategy for addressing the multiple endogenous components of IMF programs. We begin by surveying the main strategies for studying the effects of IMF programs: matching methods, instrumental variable approaches, system GMM estimation, and variants of Heckman estimators. We then adapt these methods for studying the effects of conditionality per se. Specifically, we utilize a compound instrumental variable design over a system of three equations to address sources of endogeneity related to, first, the IMF participation decision and, second, the conditions included within the program. In Monte Carlo simulations, we demonstrate that our approach is unbiased and performs better than alternatives on standard diagnostics across a range of scenarios. Finally, we apply these methods to investigate how IMF programs impact government education spending as a share of GDP on a sample of 132 developing countries for the period 1990 to 2014, finding exposure to an additional condition results in a 0.05 percentage point decline.

**Keywords** IMF programs · Conditionality · Program evaluation · Selection bias · Education expenditure · International organizations

✉ Thomas Stubbs
thomas.stubbs@rhul.ac.uk

Extended author information available on the last page of the article

🖄 Springer

# 1 Introduction

Established in 1944, the International Monetary Fund (IMF or Fund) is a cornerstone institution of global economic governance. Not only is it central to the functioning of the world economy (Kentikelenis and Seabrooke 2017; Stone 2011; Woods 2006), but it has also played a decisive role in the long-run developmental trajectory of middle- and low-income countries (Babb and Kentikelenis 2018; Dreher 2006; Dreher and Lang 2019; Kentikelenis et al. 2016; Vreeland 2003), affecting the lives of billions in the process (Babb 2005; Kentikelenis 2017). Unsurprisingly, the institution has invited controversies, with a track record rarely praised among scholars.[1] Among its various activities, the most contentious has been its practice of conditional lending. According to its founding charter, the Fund can provide temporary financing under 'adequate safeguards' to countries experiencing balance of payments problems. In exchange for this support, countries must agree to implement IMF-designed policy reform packages—or 'conditionality'—administered through a lending program. These programs typically last from six months to three years, and loan disbursements are phased over the duration in tranches, contingent upon the implementation of policy reforms.

The conditionality apparatus of IMF lending programs has two forms: quantitative and structural conditions (Bird 2009; IMF 2015). The former take the form of quantifiable macroeconomic targets that countries must meet and maintain throughout the program, such as credit aggregates, international reserves, fiscal balances, and external borrowing, and make up the majority of conditionality up to the present (Kentikelenis et al. 2016). Although quantitative conditions may overly restrict governments' fiscal policy space, policymakers can pursue a range of alternative policies to meet them; for example, several types of measures can yield budget deficit reductions. In contrast, structural conditions clearly specify the means that contribute to meeting the macroeconomic targets and other objectives. They concern a wider range of microeconomic reforms and afford governments less flexibility. Such reforms have commonly aimed at altering the underlying structure of an economy; for instance, by privatizing state-owned enterprises, legislating central bank independence, deregulating labor markets, or restructuring tax systems (Kentikelenis et al. 2016).

A large body of quantitative research is devoted to understanding the consequences of IMF programs (Abouharb and Cingranelli 2009; Bas and Stone 2014; Dreher 2006; Nelson and Wallace 2017; Nooruddin and Simmons 2006; Oberdabernig 2013; Stubbs et al. 2016). Most of this work has relied on a broad-brush binary indicator for whether or not a country is under a Fund program in a given year as a measure of the organization's engagement—plugging it into regression models and using it to differentiate control and treatment groups. Notwithstanding the important contributions of these studies, this methodological approach suffers from two major shortcomings. First, it assumes all IMF programs are identical, while—in practice—they entail heterogeneous policy content: some have a wide array of conditions attached, spanning multiple policy areas (e.g., 126 conditions for Romania in 2004); while others require a very limited number of measures (e.g. four conditions for Morocco in 2013). Second, the

---

[1] See, for example, the following: Barro and Lee (2005); Blanton et al. (2015); Dreher (2006, 2009); Dreher and Gassebner (2012); Hartzell et al. (2010); Kentikelenis et al. (2015); Przeworski and Vreeland (2000); Reinsberg et al. (2018); Stiglitz (2002); Stubbs et al. (2017a, b); and Vreeland (2002, 2003, 2007).

technique is unable to differentiate the effects of IMF conditionality from other pathways of program influence; for instance, through credit injections (Dreher 2006), scaled-up technical assistance and policy advice (Broome and Seabrooke 2015; IMF 2016a), aid and investment catalysis (IMF 2004; Stubbs et al. 2016), and moral hazard (Dreher and Walter 2010).

While scholars have developed appropriate methodological solutions to analyze the impact of IMF programs, further advancement of this research agenda has—until recently—been hamstrung by the lack of disaggregated data on conditionality. Prescient of the fact that these methods might soon reach their frontier, Vreeland (2006) called for the adoption of such a disaggregated approach to IMF conditionality over a decade ago. Scholars have responded to this call, assisted by the release of the IMF's Monitoring of Fund Arrangements (MONA) database of conditions and new panel datasets developed by independent scholars (Beazer and Woo 2016; Caraway et al. 2012; Copelovitch 2010a; Kentikelenis et al. 2016; Rickard and Caraway 2018). Given this marked increase in the range and detail of data available on conditionality, along with growing interest in the topic since the controversies surrounding the IMF's handling of the global financial crisis (Grabel 2011), the stage is set to undertake fine-grained quantitative analyses of the impact of conditionality per se. These methodological advances necessitate revisiting established methodologies to examine the impact of IMF programs.

This article elaborates on how to use panel data on conditionality in IMF programs to provide quantitative analyses of their effects, focusing on the impact stemming from the degree of conditionality—that is, the number of conditions applicable either in total, across condition types (i.e., quantitative or structural), or in a given policy area. We consider two endogenous components of treatment: (1) countries may select into IMF programs reflecting, for example, the severity of crises requiring IMF assistance; and (2) once participating in IMF programs, countries may select (or be selected) into greater or lesser degrees of conditionality.

Resolving these methodological quandaries allows scholars to test hypotheses and enrich understandings on the consequences of IMF programs. Does the IMF assist developing countries to improve their economic or social condition? Did changes in IMF policies introduced to address criticisms have the intended effects? Debates about IMF conditionality are, in essence, debates about development and globalization. For borrowing countries, their policy mix of conditionality determines their mode of integration into the world economic system and their ability to provide basic services to their population. Yet, despite such far-ranging implications, much debate has taken place on the basis of haphazard or inadequate empirical data. Accounting for IMF program heterogeneity, as endeavored here, is a step to rectifying this gap in our knowledge.

Our exposition of methodological issues, combined with Monte Carlo simulations, establishes the efficacy of maximum likelihood estimation (MLE) over a system of three simultaneous equations, in effect combining: (a) a compound instrumental variable approach to account for endogeneity of IMF program *participation*, using as an excludable instrument the interaction term of plausibly exogenous time variation in the budget constraint of the IMF and cross-sectional variation in the average of IMF program participation across the period of interest; and (b) a compound instrumental variable approach to account for endogeneity of IMF *conditionality*, using the

interaction term of the budget constraint of the IMF and cross-sectional variation in the average number of conditions a country receives. By including both IMF participation dummy and conditionality variables, we are able to isolate effects of conditions from other aspects of IMF operations. We demonstrate the utility of our approach by applying these methods to examine how IMF programs impact government education expenditure in a sample of 132 developing countries for the period 1990 to 2014. We find that exposure to an additional IMF condition results in a 0.08 percentage point decline in government spending on education as a share of GDP.

The article is structured as follows. The next section reflects on methodological challenges scholars face when studying the effects of IMF programs and provides an overview of four main approaches: matching methods, instrumental variable approaches, system GMM estimation, and variants of Heckman estimators. Section 3 describes our strategy for investigating the effects of IMF conditionality per se. Section 4 illustrates these methods by examining the effects of IMF conditionality on government education expenditure. Section 5 reflects on the limitations and broader relevance of our methodological advances.

## 2 A review of established methods for studying the impact of IMF programs

Scholars interested in estimating the effects of IMF programs typically assemble a time-series cross-section dataset with a large number of countries and years (e.g., all developing countries for a period spanning two decades), where the unit of analysis is the country-year. Studies collapse aspects of the IMF's operations into a binary indicator coded '1' if a country had an active program in a given year, and '0' otherwise (Bas and Stone 2014; Dreher 2006; Vreeland 2002). Existing strategies for estimating the average treatment effect[2] of this IMF program participation variable on some outcome variable—such as economic growth, foreign direct investment, or social expenditures—all confront the issue of selection bias (Steinwand and Stone 2008). This form of endogeneity is introduced because the circumstances of countries participating in IMF programs are systematically different from those not participating, which may in turn affect the outcome of interest. While some of these forces are observable and can thus be included as control variables (e.g., government fiscal balance or international reserves), other factors—such as political willingness to implement reforms—are not directly observable (Przeworski and Vreeland 2000; Stone 2008; Vreeland 2003). Failure to account for factors that are correlated with both IMF participation and the outcome would thus erroneously attribute their effects to IMF participation. Scholars have employed four strategies to overcome this limitation: matching methods, instrumental variables approaches, system GMM estimation, and variants of Heckman estimators.[3] We discuss each in turn.

---

[2] The average treatment effect (ATE) refers to the difference in average outcomes between observations (i.e., country-years) in the treatment group (IMF program participation) and observations in the control group (no IMF participation).

[3] Steinwand and Stone (2008) provide a comprehensive review of the earlier wave of studies on IMF program effects that correct for selection bias.

## 2.1 Matching methods

Matching seeks to address the issue of selection bias arising from observables by pairing observations with similar context but different IMF participation status (Atoyan and Conway 2006). However, it does not offer a solution to selection bias arising from unobservables since it can only be used when variation between participating and non-participating countries can be captured by observed covariates (Hardoy 2003). The advantage of using matching methods is that they do not, in principle, require identification of a valid instrument (discussed later), and reduce dependence on modelling and distributional assumptions that accompany parametric approaches (Copelovitch 2010b).

Matching approaches focus on the impact of IMF participation for countries paired with other countries at a similar likelihood of participation to identify an average treatment effect on the treated (ATET) (Wooldridge 2010). The ATET can be distinguished from an average treatment effect (ATE) insofar as the former identifies the mean effect of those countries that actually participated in an IMF program, whereas the latter refers to the impact of a randomly selected country against a counterfactual non-participation state without considering whether or not the selected country would ever actually qualify for or be interested in participating in an IMF program in the first place (Hardoy 2003).

An initial step in the matching procedure is to calculate the probability of participating in IMF-supported programs conditional on observable economic and political conditions, estimated via a probit model.[4] The next step entails generating matches of similar probabilities, or propensity scores, between pools of participating and non-participating observations, or country-years, to construct a control group (Atoyan and Conway 2006; Bal Gündüz 2016). For instance, if we assume that country selection is driven only by levels of foreign reserves, then we could match Uganda in 1981 with Tanzania in 1983. Both cases had low reserves—0.1 times monthly imports—but only the former country participated in an IMF program. In our hypothetical example, Uganda in 1981 thus enters the treatment group, while Tanzania in 1983 enters the control group, in effect acting as a counterfactual Uganda. This process repeats until all treatment and control country-years are paired. In practice, these matches can be constructed using various tolerance levels and matching techniques, such as nearest-neighbor matching, interval matching, or kernel matching (see Morgan and Winship 2007).[5] The final step involves calculating the ATET as the difference in means between treatment and control groups for matched data.

Several studies deploy matching methods to explore the impact of participation in IMF programs (Atoyan and Conway 2006; Bal Gündüz 2016; Garuda 2000; Hardoy 2003; Nelson and Wallace 2017). For instance, Hardoy (2003) uses nearest-neighbor

---

[4] Matching by propensity score is the traditional and most popular approach, but alternative matching criteria include index scores or Mahalanobis metrics (Augurzky and Kluve 2007; King and Nielsen 2018).

[5] Nearest-neighbor matching—the most commonly deployed matching technique in studies on the effects of IMF—attempts matches in terms of the absolute distance between their propensity scores, subject to the goal of minimizing the sum of all distances over all possible sets of matches (Atoyan and Conway 2006). The choice of tolerance level determines the absolute distance that propensity scores must be equal or less than before two observations are matched; unmatched observations are excluded from the subsequent ATET calculation.

matching to examine the relationship between IMF participation and economic growth, observing no statistically significant difference across treatment and control groups. Atoyan and Conway (2006) also employ nearest-neighbor matching in their study of economic growth, and report no contemporaneous effects of IMF participation. The authors note that their approach excludes 105 of the 181 IMF participation country-years due to a lack of matchable nonparticipants, in effect constraining the analysis to a subsample of countries drawn from the middle of the distribution of propensity scores. More recently, Bal Gündüz (2016) showed IMF participation is positively associated with short-term economic growth for low-income countries in a nearest-neighbor design; and Nelson and Wallace (2017) extend matching procedures in an iterative algorithm—so-called genetic matching (Diamond and Sekhon 2013)—revealing a modest but positive effect of IMF participation on the level of democracy.

Despite the merits of this method, important limitations remain. As mentioned, a key assumption is that all meaningful variation between participating and non-participating country-years can be captured by observed—or pre-treatment—covariates (Hardoy 2003). Since matching relies only on observable determinants of IMF participation to generate propensity scores, it can actually accentuate selection bias (Dreher 2006; Przeworski and Vreeland 2000; Vreeland 2003). Vreeland (2003) explains that matching methods systematically confuse the effect of participation with unobserved factors (e.g., political will), which can affect both selection into IMF programs and the outcome of interest. This selection on unobservables thus contradicts the necessary assumption for matching to provide consistent estimates when comparing means across participating and non-participating countries (Bas and Stone 2014). Another limitation is the inherent trade-off scholars face between minimizing the differences between matches, which may exclude treatment cases due to incomplete matching, or maximizing the number of matches, which may result in poor matches for treatment cases (Atoyan and Conway 2006). With no set benchmark for a suitable level of tolerance, the decision is at the researcher's discretion. Further, matching techniques do not appropriately account for the time-series cross-sectional structure of datasets (Nielsen and Sheffield 2009), which are often used in analyses of the effects of IMF programs. In particular, many matching methods match country-year observations rather than clusters of country-year observations nested within panels, or countries. More generally, King and Nielson (2018) caution against using propensity score matching as a method of causal inference entirely, citing inadequacies in the theoretical justification of its mathematical proof that introduce statistical biases to its results.

## 2.2 Instrumental variable approaches

A second solution to the problem of endogenous explanatory variables is two- or three-stage least squares (2SLS or 3SLS) estimation using one or a series of instrumental variables. To serve as an instrument, a variable must fulfil two criteria: first, the 'exclusion criterion' is that it must not affect the outcome except via IMF participation; second, the 'relevance criterion' is that it must be partially correlated with IMF participation once other exogenous variables have been netted out (Wooldridge 2010). In 2SLS estimation, predicted values are obtained for IMF participation by regressing it on exogenous variables from the outcome equation and the excluded instrumental variables. The outcome variable is then regressed on predicted values of

IMF participation and observed values of exogenous variables. Extending the 2SLS procedure, 3SLS estimation incorporates information from cross-correlations of error terms in a system of simultaneous equations for multiple endogenous variables to produce more efficient parameter estimates (Barro and Lee 2005).[6]

Past research has relied on a range of political economy variables as instruments for IMF participation, which vary depending on the outcome of interest (Barro and Lee 2005; Butkiewicz and Yanikkaya 2005; Dreher 2006; Dreher and Gassebner 2012; Easterly 2005; Moser and Sturm 2011; Oberdabernig 2013; Steinwand and Stone 2008). Most studies rely on United Nations General Assembly (UNGA) voting similarity with the US (Dreher and Gassebner 2012; Steinwand and Stone 2008; Woo 2013); that is, all else equal, countries that vote similarly to the US are more likely to participate in IMF programs. To act as a valid instrument, voting patterns must influence IMF participation (Thacker 1999), but not affect the outcome variable except via IMF participation. Using this instrument, Dreher (2006) shows that IMF participation reduces growth rates even accounting for endogeneity, and Barro and Lee (2005) find that greater participation rates in Fund programs reduce economic growth, democracy, and rule of law.

Nonetheless, identifying valid instruments for all possible outcomes of interest remains a key problem associated with this method. Studies using instruments that proxy the geopolitical importance of a recipient country assume that the Local Average Treatment Effect (LATE) is representative of all IMF programs, not just the politically motivated ones (Dreher et al. 2018). In practice, the LATE might not be generalizable, as politically motivated programs could be less effective. Further, some studies adopting instruments may breach the exclusion criterion. For example, if the outcome is democracy then the UNGA instrument is not excludable (Nelson and Wallace 2017), since democratic states exhibit similar voting patterns to those cast by the US (Carter and Stone 2015). We also observe a clear breach in a study examining the influence of IMF participation on social expenditures by deploying international reserves, bilateral exchange rate, and an exchange rate classification index as instruments (Clements et al. 2013), all of which can affect social spending outside the IMF channel.[7] The challenge of identifying valid instruments is compounded when faced with the additional concern about the endogeneity of IMF conditionality (discussed later).

## 2.3 System GMM estimation

System generalized method-of-moments (GMM) estimators for dynamic panels (Arellano and Bond 1991; Arellano and Bover 1995; Blundell and Bond 1998) have recently been utilized to allay concerns of endogeneity in IMF participation. Unlike standard instrumental variable approaches, this method does not assume that valid

---

[6] 3SLS uses the 2SLS estimates for each equation in a system of simultaneous equations to obtain an estimate of the contemporaneous variance-covariance matrix of the errors across the system. A transformed single-equation representation of the system then yields 3SLS estimates, which are consistent and asymptotically more efficient than 2SLS estimates (Nsouli et al. 2006).

[7] Exchange rates are not excludable because currency depreciation raises the costs of imported drugs and hospital equipment, which can increase government social spending (Kentikelenis et al. 2015). Likewise, governments with greater accumulations of international reserves can draw down on them to safeguard social expenditures during economic downturns (Thomson 2015).

instruments are available outside the immediate dataset, instead employing internally derived instruments based on lagged values of levels and differences of IMF participation. System GMM proceeds by estimating a system of two simultaneous equations: a 'differences' equation—where explanatory variables are first-differences—uses lagged levels of IMF participation from two or more previous time periods to instrument the contemporaneous change in IMF participation; and a 'levels' equation—where explanatory variables are levels—uses lagged first-differences to instrument contemporaneous levels of IMF participation (Roodman 2009a, b).

Studies of the consequences of IMF participation have only infrequently relied on system GMM estimators (Clements et al. 2013; Dreher and Gassebner 2012; Dreher and Walter 2010; Mukherjee and Singer 2010). Dreher and Walter (2010) found a negative association between IMF participation and currency crisis, and a positive association between IMF participation and exchange rate devaluation in response to a crisis, treating IMF participation and a lagged dependent variable as endogenous. IMF staff also adopted a system GMM setup to examine the effects of IMF participation on health and education expenditures, finding a positive effect on spending in low-income countries (Clements et al. 2013). In their approach, GDP per capita and government balance were internally instrumented, whereas IMF participation was externally instrumented. Additional studies have deployed system GMM only in robustness checks (Dreher and Gassebner 2012; Mukherjee and Singer 2010).

Despite its advertised flexibility, system GMM estimation makes strong assumptions about the data generating process. It assumes that the correct model for the outcome is dynamic (i.e., present changes are a function of past trends), that lagged differences can predict contemporaneous levels, and that first differences of instruments are uncorrelated with country fixed effects (Roodman 2009a, b; Stuckler et al. 2012). However, for the latter assumption to hold, country fixed effects and first differences of IMF participation must offset each other across the entire panel. It requires that "throughout the study period, [countries] sampled are not too far from steady states, in the sense that deviations from long-run means are not systematically related to fixed effects" (Roodman 2009b, p. 128). Whether this criterion is fulfilled depends on the sample of countries and time periods included, but is unlikely to be met in the context of IMF interventions. An additional limitation is that system GMM estimation is sensitive to the numerous minutiae—for example, the number of instrument lags, whether they are collapsed, and whether estimation is one- or two-step—none of which have a clear theoretical basis when studying IMF participation (Stuckler et al. 2012). As Roodman (2010) explains, these choices matter: they can make estimates more or less valid, and they can make certain tests of that validity stronger or weaker. There is also a risk of over-fitting endogenous variables by introducing too many instruments, thereby failing to expunge their endogenous components (Roodman 2009a). Roodman (2009a, p. 156) concludes that "the estimators carry a great and under-appreciated risk: the capacity *by default* to generate results that are invalid and appear valid."

## 2.4 Variants of Heckman estimators

Heckman variants correct for selection bias by treating non-random assignment of countries into IMF participating and non-participating groups as an omitted variable problem (Heckman 1979). In effect, the omitted variable is a catch-all term that

captures the qualities that make the entity prone to selection. Like instrumental variable approaches, the appeal of Heckman variants are that they can control for selection on unobservables, such as political will (Vreeland 2003); yet, they are more efficient than instrumental variable approaches when the selection variable is dichotomous—such as IMF participation—rather than continuous (Wooldridge 2015).

Two main variants of Heckman estimators have been deployed in IMF literature: a standard Heckman model; and the control function approach. Both approaches initially employ a probit model to predict a country's IMF participation, thereby generating the 'inverse-Mills ratio.'[8] The participation equation typically requires an 'exclusion restriction'—an excludable instrument that influences selection into IMF programs but not the subsequent outcome of interest (Lang 2016).[9] The inverse-Mills ratio is subsequently added to the vector of controls in an outcome equation estimated with Ordinary Least Squares (OLS) regression. For Heckman models, the outcome equation is limited to observations *only* where the country has selected into the treatment. Although this approach cannot *directly* estimate the effect of participating in an IMF program, it can do so indirectly by estimating another model for observations without IMF participation, and then calculating the weighted difference for the entire sample of selection-corrected parameters for countries participating in IMF programs with selection-corrected parameters for those not participating (Vreeland 2003). Conversely, a control function approach includes *all* observations in the outcome equation (i.e., regardless of whether or not the country selected into treatment) (Wooldridge 2015), and can thereby directly estimate the effect of participating in an IMF program. The approach is frequently mislabeled as a Heckman model in the IMF literature because it draws on insights from Heckman's work vis-à-vis the source of omitted variable bias.

Several studies on the effects of IMF programs use variants of Heckman estimators (Bas and Stone 2014; IEO 2003; Kentikelenis et al. 2015; Mukherjee and Singer 2010; Nooruddin and Simmons 2006; Oberdabernig 2013; Przeworski and Vreeland 2000; Vreeland 2003). For instance, the IMF's Independent Evaluation Office (2003) used a control function approach to test for the effects of IMF participation on social expenditure, identifying a positive association. Employing a similar design, Kentikelenis and colleagues (Kentikelenis et al. 2015) found IMF participation is associated with higher health expenditures in sub-Saharan African low-income countries, and with lower health expenditures in low-income countries elsewhere. Investigating the effects of IMF participation on economic growth, Przeworski and Vreeland (2000) and Vreeland (2003) utilized a Heckman model that corrects both for a country's decision to request an IMF program and for the IMF's decision to approve or reject the request, finding IMF participation lowers growth rates; however, a more recent study reanalyzed their data and found beneficial effects on growth when modelling a country's decision on the *expectation* of the IMF's

---

[8] Przeworski and Vreeland (2000) and Vreeland (2003) elaborate on this method by deploying a bivariate probit design, requiring two participation equations to model the process of IMF participation. This approach corrects both for country selection into IMF participation and IMF selection of countries to lend to, thereby adding two separate Inverse-Mills ratios to the vector of controls in the outcome equation.

[9] What the IMF literature calls an exclusion restriction is typically referred to as an excludable instrument in conventional econometric terminology. We favor the term excludable instrument henceforth. Although Heckman variants that satisfy auxiliary assumptions on the joint distribution of error terms do not strictly require an excludable instrument, we encourage researchers to follow a conservative approach by deploying one.

decision (Bas and Stone 2014). Finally, Oberdabernig (2013) deployed a control function approach and combined it with Bayesian Model Averaging to demonstrate adverse short-term effects of IMF participation on poverty and inequality.

Despite widespread use, Heckman variants are not without limitations. The precision of their estimates depend on the variance of the inverse-Mills ratio, which is determined by the predictive capacity of the first-stage probit model (Winship and Mare 1992). That is to say, it depends on having correctly specified the participation equation. Problems with collinearity of the inverse-Mills ratio may also arise when there is major overlap in the explanatory variables used in the participation equation and the outcome equation (Wooldridge 2012). While these concerns are allayed by introducing an excludable instrument, such a variable may not be readily available (Sartori 2003).[10] A final drawback is that country fixed effects cannot be introduced to the first-stage probit model, due to the well-known incidental parameter problem (Greene 2004).[11]

On balance, while none of the methods are without limitations, we maintain that some clearly perform better than others in dealing with the problem of selection bias. Matching methods are the least palatable option due to their inability to address selection on unobservables. We also discount system GMM estimation because it carries stringent assumptions that are untenable in all but the most exceptional of circumstances; besides, the estimates are too sensitive to arbitrary changes in the model to inspire confidence. We are thus left with instrumental variable approaches and Heckman variants. Both approaches entail the pursuit of a variable that fulfils exclusion and relevance criteria (i.e., an excludable instrument). Here, we prioritize the minimization of potential bias that could be introduced by excluding country fixed effects in the first-stage equation over the gains in efficiency that Heckman variants achieve for dichotomous variables like IMF participation. We thus opt for instrumental variable approaches as our favored strategy. Nonetheless, because researchers may view this potential bias as negligible in their context, we also consider the more efficient control function approach below.

## 3 Adapting methods for studying the effects of conditionality

Notwithstanding voluminous literature on the IMF and its conditional lending, until recently scholars lacked systematic, transparent, and replicable data on the actual policy content of its programs. Assuming the unit of analysis as the country-year, quantitative studies relied on dummy variables that measure the presence of an IMF program in a given year. At a conceptual level, there are two main concerns with this approach. First, such work obscures countries' diverging experiences with IMF programs, which are

---

[10] Sartori (2003) develops an estimator for binary-outcome selection models without excludable instruments, but notes that the rationale for needing an excludable instrument in Heckman approaches applies to both binary and continuous outcomes.

[11] Conditional logit estimation would allow for the inclusion of fixed effects in a binary response model while avoiding incidental parameter bias. However, it cannot be used in a multi-equation framework because its errors have an extreme-value distribution; whereas the multi-equation estimator we introduce in this article assumes a multivariate normal distribution. Conditional logit thus cannot be used unless the researcher wants to extract the endogenous component of *either* IMF participation *or* conditionality, as in previous work (Dreher et al. 2009).

designed ad hoc and thereby entail heterogeneous policy content not accurately captured by a binary variable (Kentikelenis et al. 2016). The empirical approach therefore implicitly sides with critics accusing the IMF of 'one size fits all' policies (Stiglitz 2002), despite the strong rejection of this claim by the organization (Dawson 2002). Second, these studies cannot isolate the effects of conditionality from alternative channels of program influence. These include, inter alia, scaled-up technical assistance and policy advice (Broome and Seabrooke 2015; IMF 2016a), aid catalysis (IMF 2004; Stubbs et al. 2016), and moral hazard (Dreher and Walter 2010). It is possible, for instance, that the impact of conditionality diverges from the impact of other aspects of IMF operations.

Scholars face additional endogeneity concerns when the variable of interest is IMF conditionality and not IMF program participation per se. While existing research has already established that countries select into IMF participation—a form of bias that the methods above attempt to parse out—what is less apparent is whether countries also select into *conditions* (Vreeland 2006). No academic consensus exists concerning whether conditions are requested by countries (Caraway et al. 2012; Rickard and Caraway 2014; Vreeland 2006), or imposed by IMF staff on unwilling borrowers (Chang 2007; Grabel 2011; Simmons et al. 2008; Stiglitz 2002). For proponents of the former argument, certain conditions may be sought by governments to gain leverage over domestic opposition to policy change (Vreeland 2006). The latter line of argument perceives conditionality as a coercive instrument at the disposal of the IMF, used to compel countries into implementing reforms they may not otherwise wish to undertake (Simmons et al. 2008). Where there is agreement is that the circumstances of countries receiving more IMF conditions are systematically different from those receiving fewer conditions. Indeed, several studies find that both domestic political conditions in the borrowing country and international strategic factors influence conditionality (Caraway et al. 2012; Dreher and Jensen 2007; Dreher et al. 2009, 2015; Gould 2003; Stone 2008). Chwieroth (2015) and Nelson (2014) also show that conditionality varies as a function of the professional ties and shared beliefs between IMF staff and borrowing-country officials. Yet, ambiguity remains as to whether these underlying differences would subsequently affect the outcome of interest.

We know of only 11 quantitative studies that examine the effects of IMF conditionality as distinct from IMF program participation, summarized in Table 1. These studies are yet to converge around a single method for addressing possible endogeneity biases.[12] Indeed, three of the 11 studies do not provide any treatment for endogeneity of conditionality (Rickard and Caraway 2018; Stubbs et al. 2017b; Woo 2013). It is also worth noting that these studies vary in the effect they wish to capture: five include an IMF participation variable in addition to the IMF conditionality variable, thereby capturing the total effect—or ATE—of IMF intervention (Bulír and Moon 2004; Chapman et al. 2017; Crivelli and Gupta 2016; Stubbs et al. 2017b; Wei and Zhang 2010); whereas six restrict analyses of conditionality to observations with IMF participation only, thereby capturing the conditioned effect—or ATET—of IMF intervention

---

[12] Although we focus on studies examining effects of IMF conditionality, scholarship on the World Bank and regional development banks face the same methodological quandary surrounding the estimation of effects of conditionality as distinct from program participation (e.g., Smets and Knack 2016). These studies are also yet to converge around a single method for resolving this empirical challenge.

**Table 1** Quantitative studies examining effects of IMF conditionality

| Study | Sample | Conditionality variables | Method to correct for possible endogeneity | Dependent variable | Main findings |
|---|---|---|---|---|---|
| Bulir and Moon (2004) | 112 countries for 1993–1996 (Cross-section) | Dummy variable for existence of fiscal sector structural condition; Total structural condition count | Alternative approach: IMF participation and conditionality variables included in models; estimated parameters for non-participating countries used to simulate macroeconomic policies in participating countries with the impact of participation and conditionality captured residually | Fiscal balance; revenues and grants; expenditure and net lending | No evidence of a statistically significant association of conditionality was found |
| Dreher and Vaubel (2004) | 38 countries for 1997–2003 (Time-series cross-section) | Total condition count; Monetary sector condition count; Public sector condition count | Instrumental variable approach: models restricted to observations with IMF participation; number of conditions instrumented by World Bank loans, real GDP, real per capita GDP growth in OECD countries, and the London Interbank Offered Rate | Monetary growth; budget deficit; current account balance; international reserves; government spending | No evidence of a statistically significant association of conditionality was found |
| Ivanova et al. (2006) | 170 programs covering all countries for 1992–1998 (Pooled program) | Total condition count (logged); share of structural conditions in the total number of conditions | Instrumental variable approach: models restricted to observations with IMF participation; share of structural conditions treated as exogenous; total number of conditions instrumented by share of bilateral aid by G7 prior to program start, approval year, expected program duration, IMF quota, GDP per capita, regional dummies, and population | IMF-program implementation indices | No evidence of a statistically significant association of conditionality was found |
| Wei and Zhang (2010) | ~49,000 country pairings for 1993–2003 (Dyadic) | Dummy variable for existence of trade condition in year t or any year before t during the sample period | Control function approach: IMF participation and conditionality variables included in models; no treatment on selection into IMF participation; treatment on selection into trade conditionality | Bilateral import volume | Existence of trade conditions associated with increase in import volume at high levels of overall condition implementation but no association at low levels of overall condition implementation |
| Woo (2013) | ~90 countries for 1994–2006 (Time-series cross-section) | Total structural condition count (logged); Public sector structural condition count, fiscal | Standard Heckman model: models restricted to observations with IMF participation; treatment on selection into IMF participation; no treatment on selection into conditionality | Inflows of foreign direct investment | Greater number of structural conditions associated with more FDI; same findings for public sector structural |

**Table 1** (continued)

| Study | Sample | Conditionality variables | Method to correct for possible endogeneity | Dependent variable | Main findings |
|---|---|---|---|---|---|
| | | sector structural condition count, financial sector structural condition count (all logged) | Instrumental variable approach: IMF participation and conditionality variables included in models; participation instrumented by UNSC membership; no treatment on selection into conditionality | | conditions; no findings on fiscal and financial structural conditions |
| Crivelli and Gupta (2016) | 126 countries for 1993–2010 (Time-series cross-section) | Dummy variable for existence of revenue conditions | System-GMM approach: IMF participation without revenue conditionality and conditionality variables included in models; IMF participation assumed endogenous and also treated with external instruments; revenue conditionality assumed endogenous | Tax revenues | Existence of revenue conditions associated with greater tax revenues |
| Beazer and Woo (2016) | 79 programs covering 21 post-communist countries for 1994–2010 (Pooled program) | Total structural condition count (logged) Public sector structural condition count (logged) Public sector structural condition ordinal variable | Instrumental variable approach: models restricted to observations with IMF participation; number of structural conditions instrumented by the total IMF disbursement of loans in a given year and the years to IMF governors' scheduled quota review | Composite 'reform progress' measure based on scores assigned by the European Bank of Reconstruction and Development | Greater number of structural conditions associated with more reform progress under left governments but less reform progress under right governments; same findings for public sector structural conditions |
| Casper (2017) | 76 countries for 1992–2004 (Time-series cross-section) | Dummy variable for existence of at least 10 economic conditions | Matching approach: models restricted to observations with IMF participation; genetic matching for conditionality | Dummy variable for coup attempt | Existence of economic conditions associated with higher risk of coup |
| Chapman et al. (2017) | 66 countries for 1992–2002 (Time-series cross-section) | Total condition count | Instrumental variable approach: IMF participation, loan size, and conditionality included in models; no treatment on selection into IMF participation; loan size and number of conditions instrumented by the number of countries participating in an IMF program, the ratio of prior commitments of IMF financing to IMF quota, and an extended program dummy | Sovereign bond yields | Greater number of conditions associated with decreases in sovereign bond yields |

**Table 1** (continued)

| Study | Sample | Conditionality variables | Method to correct for possible endogeneity | Dependent variable | Main findings |
|---|---|---|---|---|---|
| Stubbs et al. (2017b) | 16 West African countries for 1995–2012 (Time-series cross-section) | Total condition count | Control function approach: IMF participation and conditionality variables included in models; treatment on selection into IMF participation; no treatment on selection into conditionality | Government health spending | Greater number of conditions associated with less government health spending |
| Rickard and Caraway (2018) | ~90 countries for 1980–2014 (Time-series cross-section) | Dummy variable for existence of public sector conditions | Standard Heckman model: models restricted to observations with IMF participation; treatment on selection into IMF participation; no treatment on selection into public sector conditionality | Government spending on the compensation of employees | Existence of public sector conditions associated with larger cuts to the wage bill |

(Beazer and Woo 2016; Casper 2017; Dreher and Vaubel 2004; Ivanova et al. 2006; Rickard and Caraway 2018; Woo 2013). The latter option is intuitive, and yields findings that are easier to interpret, since one need only consider the effect of IMF conditionality variables. However, it also means that results can only be interpreted within the context of country-years with an IMF program, in turn offering a more limited set of policy implications surrounding IMF program design more generally. Furthermore, using a restricted sample of the treated still does not absolve the requirement to address selection bias into a program, for instance by including the inverse-Mills ratio (Heckman 1979), nor account for endogeneity of conditionality.

If one wishes to distinguish effects of conditionality from other aspects of IMF programs, but is also interested in how this compares to cases without an IMF program, then *both* a measure of conditionality and a binary indicator for IMF participation should be included in the model. IMF conditionality has been measured on three dimensions: degree, or the number of conditions applicable either in total or within a given policy area; scope, or the total number of policy areas subject to conditionality; and depth, or the relative stringency of each of the conditions (IEO 2007; Kentikelenis et al. 2016; Stone 2008). We limit our analysis to the degree of conditionality, which scholars use as a proxy for the overall burden of conditionality (Copelovitch 2010a; Dreher and Jensen 2007; Dreher and Vaubel 2004; Gould 2003). This measure is admittedly imperfect because it does not capture the difficulty in implementing any individual condition (Dreher et al. 2015). In any case, it may be impossible to measure the difficulty of individual conditions, especially given the vastly different characteristics of IMF borrowers: the same condition in one domestic institutional environment could be easier to implement compared to another environment (Copelovitch 2010a). Coding condition depth entails a substantial—and, in our view, unacceptable—level of subjectivity.[13]

### 3.1 Accounting for multiple endogenous IMF variables

We proceed under the assumption that countries select into both IMF participation and conditionality. Our proposed solution is to utilize maximum likelihood estimation (MLE) over a system of three simultaneous equations, in effect combining an instrumental variable approach to address endogeneity of participation with an instrumental variable approach to address endogeneity of conditionality. If all equations are linear then in theory 3SLS estimation could be used, but since we require greater flexibility due to non-linearity in the IMF program equation, we opt for MLE instead. We subsequently deliberate on plausibly excludable instruments for both IMF variables, before assessing the performance of our strategy in Monte Carlo simulations. Our unit of analysis throughout is the country-year, using time-series cross-sectional data. The methods we propose are adaptable to the inclusion of either a count of all conditions or multiple policy area counts; though, for the sake of parsimony, we focus on the total effect of all conditions.

$$\widehat{IMFPROG}_{it} = \gamma_1 X_{it} + \gamma_2 Z_{it} + \mu_i + \delta_t \qquad (1)$$

---

[13] The IMF's Independent Evaluation Office (2007) attempted to measure the difficulty of implementation for conditions based on whether parliamentary approval was required; however, we found this criterion too insensitive and, regardless, arbitrary in its own right.

$$IMF\widehat{COND}_{it} = \alpha_1 X_{it} + \alpha_2 Y_{it} + \mu_i + \delta_t \qquad (2)$$

$$W_{it} = \beta_1 IMF\widehat{PROG}_{it} + \beta_2 IMF\widehat{COND}_{it} + \beta_3 X_{it} + \mu_i + \delta_t + \varepsilon_{it} \qquad (3)$$

Here, $i$ is country and $t$ is year. Equation (3) is the outcome equation, where $W$ is the outcome of interest; $IMF\widehat{PROG}$ is the fitted value for IMF participation derived from Equation (1); $IMF\widehat{COND}$ is the fitted value for the number of conditions derived from Equation (2); $X$ denotes a vector of controls; $\mu$ is a set of country dummies; $\delta$ is a set of year dummies; and $\varepsilon$ is the error term.[14] Equation (1) is a linear model to obtain predicted values of IMF participation, $IMF\widehat{PROG}$. It is assumed to be a function of $X$, a list of covariates from the outcome equation; $Z$, an excludable instrument; $\mu$, a set of country dummies; and $\delta$, a set of year dummies. Equation (2) obtains the predicted values for IMF conditionality, $IMF\widehat{COND}$. It is also a function of $X$, $\mu$, and $\delta$; as well as, $Y$, another excludable instrument.

As discussed earlier, both instrumental variable approaches and variants of Heckman estimators can control for selection on unobservables into IMF program participation. A variation on Equation (1) is to thus use a control function approach instead, which is more efficient but will not allow for the inclusion of country fixed effects, $\mu_i$.[15] In this setting, the predicted probabilities of IMF participation can be obtained from a probit model.

To implement these analyses, we need a flexible estimator for multi-equation econometric models that can accommodate non-linearity if need be (i.e., if researchers choose to use a control function approach). MLE is suitable for this purpose and can be implemented in the *cmp* module for Stata, which allows us to jointly estimate the covariates of $W$, $IMF\widehat{PROG}$, and $IMF\widehat{COND}$. The procedure produces consistent estimates under the assumption that the system is recursive and errors follow a multivariate normal distribution. It allows for arbitrary cross-equation correlation of errors, and clustered standard errors using the bootstrap. Details on how the model is jointly estimated, including the theoretical properties of the estimator and its distributional assumptions, are available in Roodman (2011) and further detailed in our supplementary online appendices.[16]

## 3.2 Identifying excludable instruments

The perennial challenge of instrumental variable and control function approaches is finding observable variables that affect the endogenous variables the scholar wishes to instrument—here, the number of conditions applicable and the decision to participate in a program in the first place—but not the outcome variable, except via the impact on

---

[14] This model can accommodate varying lag structures for right-hand variables, the appropriateness of which will depend on theoretical expectations and the outcome of interest. For instance, some studies enter IMF variables lagged one year to correspond with the budget cycle (Crivelli and Gupta 2016); while others suggest lags of either zero, one, or two years may be appropriate depending on the effect pathways one purports to measure (Oberdabernig 2013). There may be instances where researchers wish to test on even deeper lags where effects are expected to unfold only after a substantive period of time has elapsed.

[15] Conditional logit estimation would allow for the inclusion of fixed effects in a binary response model while avoiding incidental parameter bias. However, its error distribution is incompatible with the assumptions required for multi-equation MLE. Unlike conditional logit, the latter model affords the flexibility to extract the endogenous components of *both* IMF programs *and* IMF conditionality.

[16] The online appendices are available on *The Review of International Organizations'* website.

conditionality and participation respectively. To overcome this issue, we develop a new instrument for IMF conditionality, before assessing potential instruments for IMF participation used in previous studies.

For conditionality, we repurpose a recently popularized compound instrumentation approach from the aid effectiveness literature (Dreher and Langlotz 2017; Nunn and Qian 2014). Specifically, our instrument is the interaction of the within-country average of the number of conditions across the period of interest with the year-on-year IMF's budget constraint. Formally specified, the predicted values for IMF conditionality specified in Equation (2) are derived as follows:

$$\widehat{IMFCOND}_{it} = \alpha_1 \left( \overline{IMFCOND_i} \times IMFBUDG_t \right) + \alpha_2 X_{it} + \mu_i + \delta_t \qquad (4)$$

Here, $i$ is country and $t$ is year. $\widehat{IMFCOND}$ is the fitted number of IMF conditions; $\overline{IMFCOND}$ is the country-specific average of conditions; $IMFBUDG$ is the budget constraint of the IMF in year $t$; $X$ is a list of covariates from the outcome equation; $\mu$ is a set of country dummies; and $\delta$ is a set of year dummies.

The identifying assumption of Equation (4) is that the outcome of interest in countries with different exposure to conditionality will not be affected differently by changes in the IMF's budget constraint other than through the impact of IMF conditions. This econometric strategy is supported in analytical proofs by Bun and Harrison (2018) and Nizalova and Murtazashvili (2016) that show the interaction of an endogenous variable with an exogenous one can be interpreted as being exogenous under mild assumptions. The approach is akin to a (continuous) difference-in-difference design: the effect of conditionality on the outcome of interest is compared across a group of high-exposure countries to IMF conditions and a group of low-exposure countries as the IMF's budget constraint changes. As in any difference-in-difference design, identification rests on an exogenous treatment and the absence of different pre-trends across groups. Below we discuss these assumptions, along with the usual assumptions on instrumental variables.

The instrument fulfils the relevance criterion because the cross-sectional average of conditionality approximates the general propensity of a country to obtain a specific amount of conditions in any given year, after accounting for observable factors that usually explain such variation. Furthermore, as previous research shows, the number of conditions increases when country demand for IMF loans grows, and decreases or stagnates when country demand for IMF loans is weak (Chapman et al. 2017; Dreher and Vaubel 2004). A plausible rationale for the observed relationship between the number of conditions and country demand is that as the IMF assists more countries, resource scarcity prompts the organization to assign a greater number of conditions to any given country as a safeguard measure for loan repayments (Dreher and Vaubel 2004; Vreeland 2003). The inverse also holds: Lang (2016) shows that the IMF is more generous with its loans when it has high liquidity, implying less conditions in times of resource abundance as the Fund becomes more eager to recruit borrowers, anticipating interest revenues on loans (Babb and Buira 2005).[17] This line of argument is underpinned by the idea that the

---

[17] Vreeland's (2003) conditionality game shows that more conditions decrease country demand for IMF loans, so it would be rational for the IMF to reduce the number of conditions if it wishes to entice countries to borrow.

IMF is, at least to some degree, a self-serving international bureaucracy that seeks to maximize revenues, protect future budgets, and maintain a position of global power (Dreher and Vaubel 2004; Vaubel 1996).

In our analysis, the budget constraint is measured via proxy using the natural log of the IMF's liquidity ratio—calculated as liquid resources divided by liquid liabilities (Lang 2016; Nelson and Wallace 2017). Here, liquid resources is the sum of usable currencies plus Special Drawing Rights contributed; and liquid liabilities is the sum of members' reserve tranche positions plus outstanding IMF borrowing from members (Lang 2016). Figure 1 plots the natural log of IMF liquidity ratio and the mean number of conditions per participating country in a year for 1990 to 2014, showing a significant correlation between the two variables ($r = -.30$). In regressions further below, the instrument consistently satisfies benchmarks for identifying strong instruments, with a Kleibergen-Paap F-statistic above ten (Staiger and Stock 1997).

The instrument fulfils the exclusion criterion because country-specific changes in conditionality that deviate from its long-run average are brought about only by decisions of the IMF that do not pertain to any given country, such as the introduction of social spending floors in the late-1990s or the streamlining initiative of the early-2000s (IMF 2001a; Kentikelenis et al. 2016). While one might be concerned about potential direct effects of the general propensity of a country to obtain a specific amount of conditions in any given year on the outcome variable, we control for this effect through the inclusion of country fixed effects in both conditionality and outcome equations (Dreher and Langlotz 2017). Moreover, there is no apparent pathway from the IMF's own budget constraint to an outcome variable of a given country other than through conditionality, since it is driven by organizational factors that have nothing to do with the characteristics of borrowing countries.

There could be a question on the exogeneity of the budget constraint insofar as wealthy member countries can replenish IMF resources in response to a greater number of countries participating in programs, which would diminish the Fund's risk aversion such
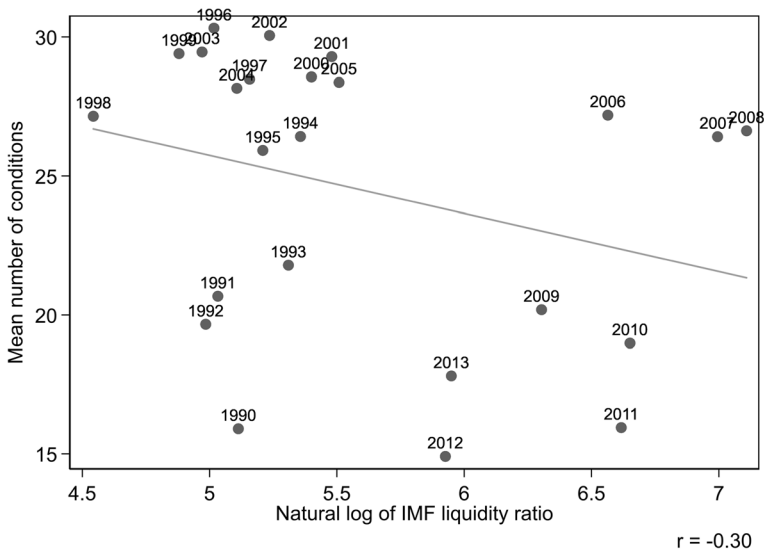


Fig. 1 IMF liquidity ratio and mean number of conditions per year, 1990–2014

that the organization is willing to agree to fewer conditions when bargaining a new program with a recipient country (Dreher and Vaubel 2004). This logic is flawed, however, because the amount of financial resources that members commit to the Fund is determined by exogenous institutional processes: set by the IMF's Board of Governors following quota reviews conducted every five years (IMF 2017a). It is therefore unlikely that IMF resource increases would then be linked to the outcome through unobserved channels (Lang 2016).

A related concern over instrument excludability is that donors may be less willing to spend in times of global financial crisis, resulting in both a reduction of the IMF's concessionary lending budget—which, unlike its non-concessionary account, is replenished through voluntary contributions from member countries rather than quota subscriptions (IMF 2016b)—and deteriorating socio-economic outcomes in aid dependent countries. The inclusion of year dummies in both conditionality and outcome equations can capture common external shocks across all countries, and will ensure that the instrument is not correlated with the error term in the outcome equation (Nunn and Qian 2014). In essence, our identification strategy relies on "the interaction term being exogenous *conditional on the baseline controls*" (Nunn and Qian 2014, p. 1632 emphasis added).

Another potential concern over instrument excludability derives from the underlying bargaining process that determines conditionality (Dreher and Vaubel 2004). It is possible that the within-country average of the number of conditions reflect levels of country 'priority' within the IMF insofar as it relates to geopolitics or risk aversion. For instance, geopolitically important countries have greater bargaining power and may thus receive fewer conditions on average than less geopolitically important countries (Stone, 2002, 2011). If the elasticity of conditionality with regard to the IMF budget constraint also depends on political priority, such that low-priority countries see their conditionality increase faster than high-priority countries when IMF resources become scarce, then the exclusion criterion would be violated through the correlation between the instrument and the unobserved variable 'priority'. To test if this is the case, in Fig. 2 we divide countries into low, medium, and high conditionality groups and graph the IMF liquidity ratio against the number of conditions assigned to a specific country for each year of IMF participation. Countries at different burdens of conditionality have, on average, a similar propensity of receiving more conditions when the amount of countries in IMF programs increases, as indicated by equivalent gradients of the lines of best fit across conditionality groups (all between $r = -.15$ and $r = -.29$). We thus show that the elasticity of conditionality with respect to geopolitical importance is approximately constant; that is to say, the instrument satisfies the homogeneity of treatment assumption necessary for the exclusion criterion to be valid (Hainmueller et al. 2016).

Finally, since the above approach is akin to a difference-in-difference design, non-parallel trends across groups with different exposure would undermine identification (Christian and Barrett 2017). In our case, trends over time in the number of IMF conditions and the outcome variable should be similar across above-mean conditionality exposure and below-mean conditionality exposure groups of countries. In addition, inference would be threatened if there was a non-linear trend in the time-varying part of the compound instrument that is similar to the respective trends in the potentially endogenous variable and the outcome variable in the high-exposure group of countries. This is because such non-linear parallel trends would create spurious correlation not controlled for by year fixed effects unless the same non-linear trends are also present among the low-exposure group of
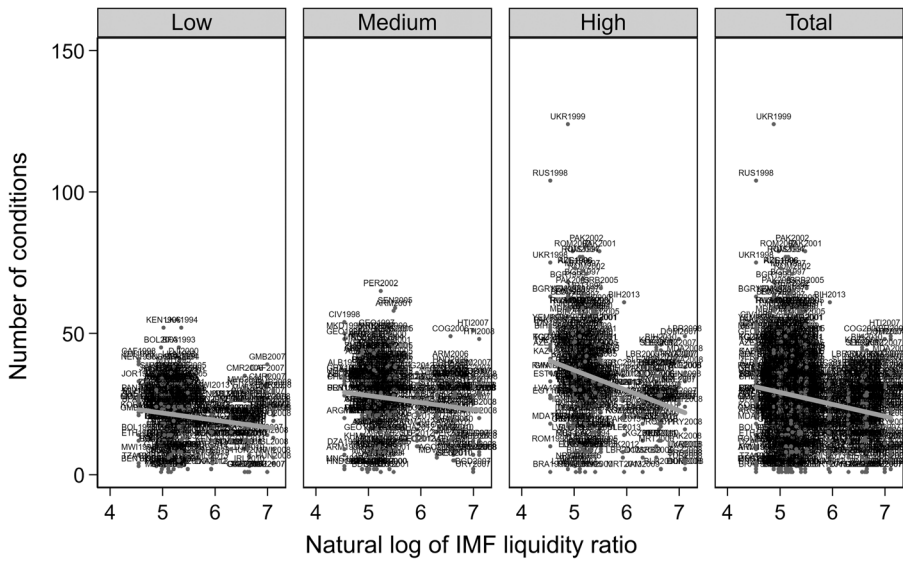
**Fig. 2** IMF liquidity ratio and number of conditions per country-year for low, medium, and high condition-ality countries, 1990–2014. Notes: Excludes country-years without a condition. Countries divided into three equal groups based on a country's average number of conditions per year of IMF participation. Low conditionality (0 to 25.9 conditions) countries includes Antigua and Barbuda, Barbados, Benin, Bolivia, Burkina Faso, Cameroon, Central African Republic, Colombia, Comoros, Czech Republic, Djibouti, El Salvador, Ethiopia, Gambia, Grenada, Guatemala, Guinea-Bissau, Hungary, Iceland, India, Ireland, Jordan, Kenya, Korea Rep., Kosovo, Lesotho, Malawi, Mali, Mozambique, Niger, Nigeria, Panama, Poland, Portugal, Slovak Republic, Sri Lanka, St. Kitts and Nevis, Tanzania, Togo, Tunisia, Zimbabwe. Medium conditionality (25.9 to 30.9) includes Afghanistan, Albania, Algeria, Angola, Argentina, Armenia, Bangladesh, Burundi, Cabo Verde, Cambodia, Congo Rep., Cote d'Ivoire, Cyprus, Dominica, Equatorial Guinea, Gabon, Georgia, Ghana, Guyana, Haiti, Honduras, Macedonia, Madagascar, Maldives, Mexico, Mongolia, Nicaragua, Papua New Guinea, Peru, Philippines, Sao Tome and Principe, Senegal, Seychelles, Solomon Islands, Trinidad and Tobago, Uganda, Uruguay, Zambia. High conditionality (30.9 to 62.7) includes Azerbaijan, Belarus, Bosnia and Herzegovina, Brazil, Bulgaria, Chad, Congo Dem. Rep., Costa Rica, Croatia, Dominican Republic, Ecuador, Egypt, Estonia, Greece, Guinea, Indonesia, Iraq, Jamaica, Kazakhstan, Kyrgyz Republic, Lao, Latvia, Lithuania, Mauritania, Moldova, Nepal, Pakistan, Paraguay, Romania, Russia, Rwanda, Serbia, Sierra Leone, Tajikistan, Turkey, Ukraine, Uzbekistan, Venezuela, Vietnam, Yemen

countries (Christian and Barrett 2017). In our case, we would be concerned about similarly shaped non-linear trends in the IMF liquidity ratio, the number of IMF conditions, and education expenditure if these trends only occur among high-exposure countries but not the low-exposure countries. We graphically assess these two conditions below.

Figure 3 allows us to compare trending behavior across exposure groups for the mean number of conditions (left panel) and our outcome variable of government education spending as a share of GDP (right panel). The trend pattern for the number of conditions is qualitatively similar across groups, although the high-exposure countries face a higher number of conditions by definitional fiat in any given year. We also find similar trends across exposure groups in the outcome variable. In Fig. 4, we present the temporal evolution of the IMF liquidity ratio. Here we observe no similar (non-linear) trends between the IMF liquidity ratio and the mean number of conditions, or between the IMF liquidity ratio and education expenditures. Consequently, there is no apparent violation of the design assumptions of the difference-in-difference approach.
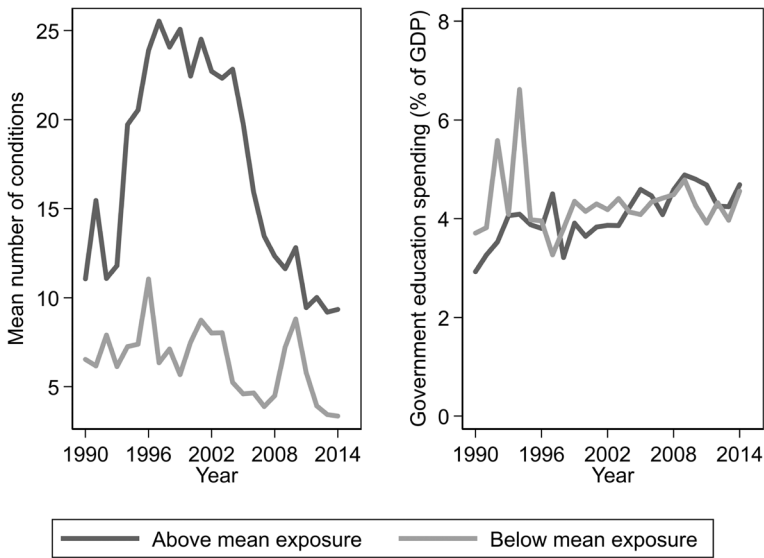
**Fig. 3** Parallel trends in IMF conditionality compound instrument

Overall, we believe that the argument for using the interaction of the within-country average of the number of conditions and the year-on-year IMF budget constraint as an instrument is well grounded. To violate the exogeneity assumption, one would have to find an unobserved variable driving the relationship between the time-variant IMF budget constraint and the outcome of interest; and such an unobserved variable would also need to be correlated with the country-specific average of conditionality after controlling for country and year fixed effects and a vector of controls. Although it is unlikely that such a variable exists, we consider this possibility by incorporating plausible candidate variables for our particular outcome of interest as additional controls in robustness checks, namely education commitments of overseas development assistance and the number of role-equivalent countries participating in IMF programs in the past three years. Furthermore, to violate the parallel and non-overlapping trends assumptions, the contemporaneous trends in the outcome of interest would have to follow different patterns across groups of countries with high compared to low exposure to IMF conditionality, and the trends in the IMF's budget constraint, IMF conditionality, and the outcome of interest would need to overlap (Christian and Barrett 2017).

Excludable instruments are also needed to account for endogeneity of IMF participation. As described in Section 2.2, past studies usually rely on UNGA voting similarity with the US (Dreher and Gassebner 2012; Steinwand and Stone 2008; Woo 2013), but the LATE of the instrument might not be generalizable since politically motivated programs could be less (or more) effective (Dreher et al. 2018). Our preferred approach is thus to use another compound excludable instrument initially proposed by Lang (2016) and adopted more recently by Nelson and Wallace (2017): the interaction of the within-country average of IMF program participation across the period of interest with the year-on-year IMF's budget constraint, again using the natural log of the IMF liquidity ratio as a proxy.

Formally specified, the predicted values for IMF participation specified in Equation (1) are derived as follows:
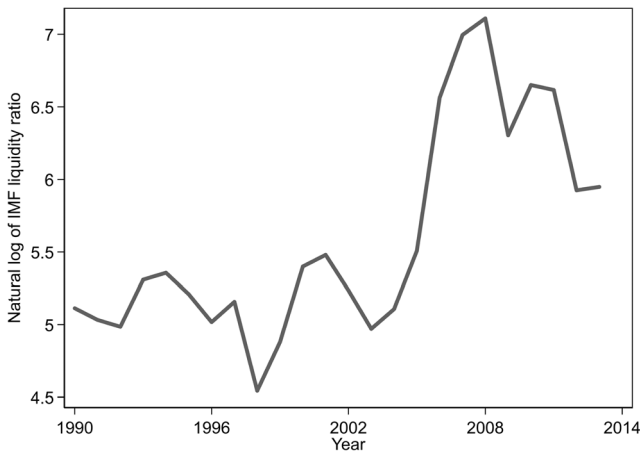
**Fig. 4** IMF liquidity ratio across time, 1990–2014

$$IM\widehat{FPRO}G_{it} = y_1 \left( \overline{IMFPROG_i} \times IMFBUDG_t \right) + y_2 X_{it} + \mu_i + \delta_t \qquad (5)$$

Here, $i$ is country and $t$ is year. $IM\widehat{FPRO}G_{it}$ is the predicted probability of IMF participation; $\overline{IMFPROG}$ is the country-specific average of participation; $IMFBUDG$ is the budget constraint of the IMF in year $t$; $X$ is a list of covariates from the outcome equation; $\mu$ is a set of country dummies; and $\delta$ is a set of year dummies.

Lang (2016) provides a robust defense of the instrument's excludability, which follows the same logic as our IMF conditionality instrument vis-à-vis the exogenous variation of the budget constraint. Again, since such an approach is akin to a (continuous) difference-in-difference design, we check below whether its underlying assumptions hold.

In Fig. 5, we find both exposure groups to be similar in terms of their trending patterns with respect to the outcome variable and the mean probability of IMF program participation. Furthermore, there is no apparent trend similarity between the IMF's budget constraint (see Fig. 4) and the mean probability of IMF program participation and education expenditure, respectively, among above-mean participation exposure countries.

### 3.3 Performance assessment of estimators using Monte Carlo simulation

While the proposed multi-equation approach with correlated errors is preferable on a theoretical basis, its performance relative to alternative approaches needs to be tested. We conduct Monte Carlo simulations, which enable us to control the data generating process and thereby assess how well different estimation methods approximate the true parameter values.

Building on our theoretical discussion, we compare the performance of six estimators: our favored double-instrumental variable maximum-likelihood estimator that essentially linearizes the IMF participation equation to include fixed effects, akin to a 3SLS estimator, which accounts for endogeneity of IMF programs and conditionality (IV/IV/MLE); a conditional mixed process maximum-likelihood estimator that
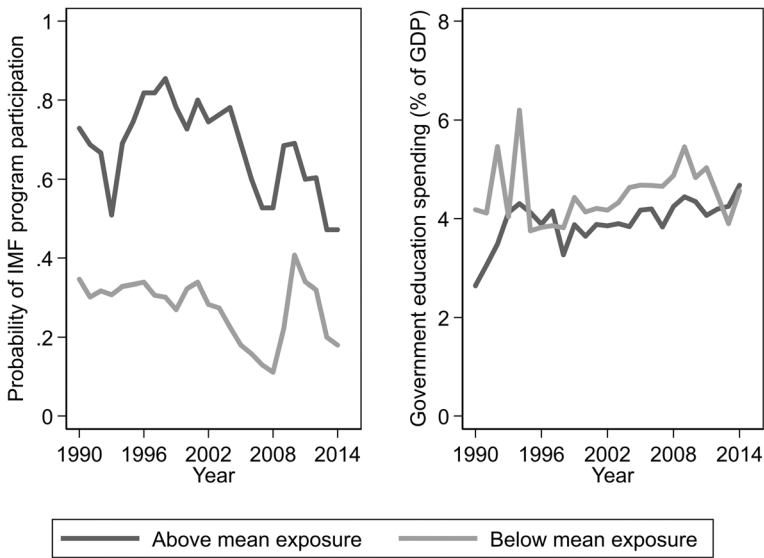
**Fig. 5** Parallel trends in IMF participation compound instrument

accounts for endogeneity of IMF programs and conditionality but does not linearize the IMF participation equation (CFA/IV/MLE); a two-step variant of Heckman estimator using a control function approach without correcting for endogeneity of conditionality (CFA/−/OLS); a 2SLS estimator correcting for endogeneity of IMF programs through a linearized IMF selection equation but with no correction for conditionality (IV/−/OLS); a 2SLS estimator without endogeneity correction for programs but with correction for endogeneity of conditionality through an instrumental variable approach (−/IV/OLS); and a simple OLS estimator without any endogeneity correction (−/−/OLS). We scrutinize these estimators across several scenarios. In the first five scenarios, we match the proposed data-generation process but vary cross-equation correlations and the strength of the instruments. In the last two scenarios, we probe robustness of the estimators to omitted-variable bias that jointly affects IMF treatments and the outcome of interest. For each simulation scenario, the data are generated and models estimated 500 times, and three quantities are calculated: bias, root mean square error, and optimism (Bell and Jones 2015).[18]

We find that our proposed solution—instrumenting for both IMF treatments in a multi-equation framework with joint error structure—is the preferred one unless instruments are weak. Whether CFA/IV/MLE is preferable to IV/IV/MLE, however, depends on the specific parameterization. For example, CFA/IV/MLE performs best when a strong instrument for IMF conditions is available, when there is high residual correlation across equations, and when the model is misspecified by ignoring a confounder

---

[18] We implement the simulation as follows. First, in each replication loop, we create a rectangular dataset of 1000 observations. Second, we generate the normally-distributed predictors, instruments, and errors, and calculate the response variables. The instrument is defined such that it correlates with the main predictor but also includes white noise, which simulates instrument weakness. Third, we run the analysis using six different estimators in each of the seven scenarios, calculating for each coefficient estimate its bias, mean squared error, and optimism with respect to the true parameter. For detailed descriptions and results tables please refer to the online appendices.

that affects both IMF programs and the outcome of interest. Conversely, IV/IV/MLE performs well (and slightly better than CFA/IV/MLE) when both instruments are strong and cross-equation correlation is mild, and in the misspecified case where a third variable causes both IMF conditions and the outcome. Our simulations also suggest conditions under which simpler alternatives underperform. Plain OLS is particularly poor when the model is misspecified. In this case, the bias concentrates in the coefficient of the IMF treatment with unobserved correlation with the outcome. A similar result holds when conditionality is instrumented but the researcher ignores selection into IMF programs. Likewise, when only IMF programs are instrumented but not IMF conditions, results are severely biased under a misspecified model. In this case, a control function approach performs relatively better than a linearized IV model for program selection, but both have unacceptable biases. Only when both instruments are weak, such estimators yield less biased results than the more complex double-instrumentation estimators that our approach uses. In sum, we find that CFA/IV/MLE and IV/IV/MLE are virtually unbiased across a wide range of scenarios and perform best in most cases on root mean squared error and optimism diagnostics. We take this as evidence of the superiority and robustness of our proposed estimator where problems of program selection and endogeneity of conditionality are intertwined. However, a caveat when applying these approaches is that moderately strong instruments are necessary. Failing that, simpler alternatives such as OLS are more robust—as confirmed by recent methodological literature (Young 2018).

Overall, our recommended approach has the advantage of enabling scholars to account for program heterogeneity and to separate the effects of conditionality from other aspects of IMF operations, while addressing potential endogeneity concerns surrounding the IMF variables. Our approach is also flexible enough to accommodate disaggregated counts of conditions (see below).

## 4 An empirical application to government education spending

To illustrate the utility of our proposed method, we present an empirical application where government education expenditure is the outcome variable. This issue has been the subject of sustained controversy, as critics argue that IMF programs result in cuts to funding earmarked for education (Nooruddin and Simmons 2006), while the IMF retorts that its programs safeguard such expenditure (Clements et al. 2013; IEO 2003). Previous studies utilized a dummy variable to measure the presence of a Fund-supported program, thereby masking diverse country experiences (Clements et al. 2013; Huber et al. 2008; IEO 2003; Nooruddin and Simmons 2006). We overcome this limitation by examining whether the number and type of conditions affect spending.

### 4.1 Pathways

We posit three pathways linking IMF conditionality to government education expenditure. These invoke both types of conditions: the quantifiable macroeconomic targets that make up the majority of conditionality ('quantitative conditions'); and the microeconomic reforms aimed at altering the underlying structure of an economy ('structural

conditions'). A further three pathways pertain to IMF operations outside of the conditionality channel.

First, structural conditions that explicitly call for restructuring of the education sector may either increase or decrease education expenditure. For instance, the Nicaraguan program in 1999 required "management by local boards of 95 percent of secondary schools and 65 percent of primary schools" (IMF 2000b); Bolivian officials agreed in 2000 to "develop a reform proposal for higher education in order to reduce the share of public resources for higher education" (IMF 1999); and the Azerbaijani program called for authorities to "establish special government commissions to develop reform plans for the health and education sectors" in 1997 (IMF 1997).

Second, structural conditions that directly stipulate education sector hiring decisions can reduce government education spending. Examples abound: Bulgaria's program in 2006 required "an employment cut of at least 5,500 positions in the education sector" (IMF 2006); Sierra Leone had to "implement concrete measures to control the teachers' payroll" in 2002 (IMF 2002); and in 2004 Tajikistan was required to "reduce the number of employees in the education sector by 5 percent" (IMF 2003). By the same token, quantitative conditions stipulating general—rather than education-specific—wage bill limits on the public sector can indirectly impede a borrowing country's ability to hire and remunerate teachers (Marphatia 2010). In West Africa alone, between 1995 and 2014 a combined 95 of the 211 years with IMF conditions included limits to the wage bill (Stubbs et al. 2017b).

Third, quantitative conditions set by the IMF in the majority of its programs on budget deficit reduction, international reserve holdings, and net domestic borrowing can shrink fiscal space, indirectly forcing a reduction in spending in the education sector (Kentikelenis et al. 2016; Stubbs and Kentikelenis 2018). While IMF staff claim that quantitative conditions stipulating floors on social expenditures reverse this drag on education spending (Clements et al. 2013; IMF 2017b), recent research shows that these floors are observed infrequently, whereas fiscal deficit targets are almost always met (Kentikelenis et al. 2016).

These initial three pathways are represented empirically with a series of IMF conditionality variables (see below). Outside of conditionality channels, the IMF may also bolster government education expenditure via—fourth—the provision of low interest credit provided under its programs, although this additional resource is some-times used to repay external debt instead (Gould 2003). Fifth, the presence of IMF programs may give countries a 'stamp of approval' that could catalyze additional bilateral aid from donor governments, thus boosting education expenditures through donor financing (Clements et al. 2013). Even so, evidence elsewhere shows this aid substitutes—rather than complements—government spending on the social sector, and that aid flows increase for general budget support and debt relief but not for education (Stubbs et al. 2016; Stuckler et al. 2011). Sixth, scaled-up technical assistance can improve budget monitoring and execution, thus increasing the proportion of the education budget commitment that is actually spent on the education sector, rather than being spent elsewhere or going unspent (Stubbs et al. 2017b). For instance, the IMF prioritized assistance to improve the utilization of social sector appropriations in Benin in the late-1990s (IMF 1998), ultimately contributing to higher social spending (IMF 2000a). These subsequent three pathways are captured in our analysis by an IMF participation variable.

## 4.2 Variables

We investigate the effects of IMF conditionality on government education spending as a share of GDP for 132 developing countries over the period 1990 to 2014, as reported by the World Bank (2016). Our IMF conditionality variables are based on the coding of agreements between the Fund and its borrowers (Kentikelenis et al. 2016). The online appendices provide further details on how the dataset was created. We only count binding conditions (known as 'prior actions' or 'performance criteria'), following established procedures in this field of study (Copelovitch 2010a; Rickard and Caraway 2014; Stubbs et al. 2017b; Woo 2013).[19] Binding conditions directly determine scheduled disbursements of loans and must be implemented for the program to continue; whereas non-binding conditions serve as markers for broader progress assessment and non-implementation does not automatically suspend the loan (IMF 2001b), which may thus introduce noise to the analysis if included (Stubbs et al. 2017b). Our chosen measure also allows us to empirically isolate a conditionality effect from an effect—or lack thereof—due to country non-compliance, since the binding character of these conditions precludes the possibility of the latter (Dreher 2006; Vreeland 2003, 2006). Since hypothesized pathways of impact entail both structural and quantitative conditions, our initial measure of conditionality aggregates them. In further analyses, we then test the disaggregated impact of conditions separated into quantitative and structural types, and also explore the effect of conditions in given policy areas. Our IMF participation variable is measured with a binary indicator for whether a country was under an IMF program in a given calendar year for at least five months of the year (Dreher 2006).

Following previous research (Clements et al. 2013; Huber et al. 2008; Stubbs and Kentikelenis 2018), our outcome equation includes control variables for economic and demographic factors that affect both the need for education expenditure and governmental capacity to meet those needs. Entering the model contemporaneously are the natural logarithm of GDP per capita, urban population as a share of total population, population aged under 15 as a share of working-age population, and level of democracy; while government balance as a share of GDP and trade as a share of GDP enter the model lagged one year to correspond with the budget cycle. We expect a positive effect of GDP per capita because, according to 'Wagner's Law', state activities expand to cover new administrative and social functions as economic development takes place, thereby increasing state spending (Brady and Lee 2014; Nooruddin and Simmons 2006; Wagner 1994). Urbanization should have a positive effect since urban dwellers can more easily mobilize to request additional services from governments as well as offering economies of scale (Baqir 2002). The population under 15 accounts for the demographic need for a government to spend on education (Huber et al. 2008). Democracy controls for the well-established finding that democratic governments increase public spending on social programs (Stasavage 2005). The government balance in the previous year controls for a country's fiscal space to increase education spending (Clements et al. 2013). Last, we expect a positive effect of trade openness in

---

[19] In robustness checks, we use alternative measures of conditionality: an implementation-corrected count (which subtracts conditions waived by the IMF); an implementation-discounted count (which discounts conditions during program suspensions); and a combined binding and non-binding condition count.

the previous year because it can lead to increases in social expenditures as governments seek to compensate those adversely affected by it (Rodrik 1998; Rudra 2008). We also introduce country fixed effects to account for time-invariant country-level characteristics, and year fixed effects to control for common external shocks across all countries. Data sources and summary statistics are provided in the online appendices.

As described earlier, our identification strategy entails instrumenting the number of conditions using the interaction of the within-country average of the number of conditions with the natural log of the IMF liquidity ratio; and instrumenting program participation with the interaction of the within-country average of program participation with the natural log of the IMF liquidity ratio. We compare how results differ for instrumental variable and control function approaches to endogeneity of IMF participation. The system of three simultaneous equations is estimated through MLE. Standard errors are calculated using the clustered Sandwich estimator, which adjusts for heteroscedasticity and serial correlation. Analyses are performed using Stata version 15.

### 4.3 Findings

In Table 2, we present the results of our quantitative analyses on government education expenditure as a share of GDP on six variants of our model. The online appendices report results from the first-stage models for IMF conditionality and participation.

To ensure our model specification is appropriate, in Model 1 we initially include only the control variables and conduct simple OLS estimation. Results on controls follow the expected effect direction established by previous studies on government education spending (Clements et al. 2013; Huber et al. 2008; Stubbs and Kentikelenis 2018): positive for urbanization, dependency ratio, democracy, and trade (although only democracy is statistically significant); and negative for GDP per capita and government balance (both statistically significant).[20] For Model 2, we add the IMF condition and participation variables, but again run simple OLS without any endogeneity corrections. Here, none of our IMF variables reach standard thresholds of statistical significance, and results on controls remain stable.

Next, we employ our preferred identification strategy in Models 3 and 4. As described earlier, we correct for endogeneity of the number of conditions and program participation in an instrumental variable approach using compound instruments: the interaction of the within-country average of the number of conditions with the natural log of the IMF liquidity ratio; and the interaction of the within-country average of IMF program participation with the natural log of the IMF liquidity ratio. In Model 3, we exclude potentially endogenous controls that could be affected by IMF intervention, namely democracy, government balance, and trade. In this setting, exposure to an additional IMF condition is associated with a statistically significant decrease of *0.05* percentage points in government education spending as a share of GDP. When we add potentially endogenous controls in Model 4, the headline result is substantively unchanged. The divergent findings from Model 2—which underestimates conditionality

---

[20] Previous studies did not include both country and year fixed effects (Clements et al. 2013; Huber et al. 2008; Stubbs and Kentikelenis 2018). Our specification is thus a more stringent test, which may explain why some of our controls do not reach standard thresholds of significance.

**Table 2** Effect of IMF conditionality on government education spending

| | Dependent variable: Government expenditure on education (% of GDP), 1990–2014 | | | | | |
|---|---|---|---|---|---|---|
| Model specification | 1. Controls only | 2. All conditions and controls | 3. All conditions, exogenous controls | 4. All conditions and controls | 5. All conditions, exogenous controls | 6. All conditions and controls |
| Identification strategy | OLS | OLS | Conditionality IV, Participation IV | Conditionality IV, Participation IV | Conditionality IV, Participation CFA | Conditionality IV, Participation CFA |
| IMF conditions (lagged) | . | −0.0031 | −0.0492* | −0.0500* | −0.0678** | −0.0677*** |
| | . | [0.0031] | [0.0281] | [0.0274] | [0.0264] | [0.0254] |
| IMF participation (lagged) | . | 0.1534 | −0.2135 | −0.2120 | 0.3787 | 0.3700 |
| | . | [0.1152] | [1.3585] | [1.3310] | [0.3034] | [0.3327] |
| Log(GDP per capita) | −0.7197* | −0.7192* | −2.0054*** | −2.1088*** | −1.8654*** | −1.9877*** |
| | [0.4324] | [0.4254] | [0.6226] | [0.6076] | [0.5646] | [0.5772] |
| Urbanisation | 0.0316 | 0.0326 | 0.0562 | 0.0453 | 0.0622 | 0.0520 |
| | [0.0441] | [0.0440] | [0.0414] | [0.0395] | [0.0389] | [0.0384] |
| Dependency ratio | 0.0180 | 0.0184 | 0.0123 | 0.0108 | 0.0161 | 0.0145 |
| | [0.0155] | [0.0153] | [0.0167] | [0.0160] | [0.0150] | [0.0147] |
| Democracy | 0.0889** | 0.0832* | . | 0.1349* | . | 0.0871** |
| | [0.0440] | [0.0436] | . | [0.0729] | | [0.0433] |
| Government balance (lagged) | −0.0195* | −0.0191* | . | −0.0007 | . | 0.0015 |
| | [0.0110] | [0.0112] | . | [0.0130] | | [0.0140] |
| Trade (lagged) | 0.0058 | 0.0058 | . | 0.0083 | . | 0.0081 |
| | [0.0068] | [0.0067] | . | [0.0064] | | [0.0063] |
| Constant | 4.7319 | 4.6291 | 14.2384*** | 14.6120*** | 12.3313*** | 12.8991*** |
| | [3.5733] | [3.4713] | [5.2966] | [4.8583] | [4.3855] | [4.2240] |
| Country fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |

**Table 2** (continued)

| | Dependent variable: Government expenditure on education (% of GDP), 1990–2014 | | | | | |
|---|---|---|---|---|---|---|
| Year fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of observations | 1307 | 1307 | 1307 | 1307 | 1307 | 1307 |
| Number of countries | 134 | 134 | 134 | 134 | 134 | 134 |
| F-statistic for conditionality instrument | n/a | n/a | 33.63 | 35.93 | 29.25 | 31.80 |
| F-statistic for participation instrument | n/a | n/a | 15.03 | 16.61 | n/a | n/a |
| Joint F-statistic | n/a | n/a | 33.64 | 35.96 | n/a | n/a |

F-tests are Kleibergen-Paap statistics. Cluster-robust standard errors in brackets

$* p < 0.10$, $** p < 0.05$, $*** p < 0.01$

effect size and statistical significance—indicates the presence of endogeneity. In particular, the findings suggest that when government education spending is high, either the Fund imposes or a country actively seeks more conditions.

Our finding on the statistically significant negative effect of IMF conditions is also substantively significant. The mean and standard deviation of government education spending as a share of GDP across our sample is *4.41* and *2.15* percent respectively; thus, setting the number of conditions at its mean of *10.61* corresponds to an average decrease of about quarter of a standard deviation in education spending, or *0.53* percentage points, all other factors held constant. Based on findings in Model 4, such an effect is comparable to a four-point decline in democracy, using an ordinal scale from zero (full autocracy) to ten (full democracy) (Teorell et al. 2016).

Outside of the conditionality channel, the IMF effect direction is negative but does not reach standard thresholds of significance. Results on control variables maintain their direction of effect, but government balance is no longer statistically significant in Model 4. Diagnostic statistics indicate that our instruments are strong across Models 3 and 4: Kleibergen-Paap F-statistics of *33.63* and *35.93* respectively for the conditionality compound instrument; and *15.03* and *16.61* for the participation compound instrument.[21] The instruments are also jointly relevant in both models, with F-statistics of *33.64* and *35.96* respectively.

For Models 5 and 6, we maintain an instrumental variable approach to correct for endogeneity of conditionality, but employ a control function approach to correct for the endogeneity of participation. Recall that this approach is more efficient and can still use our compound instrument for IMF participation, but will not allow for the inclusion of country fixed effects in the first-stage equation due to the incidental parameter problem (Greene 2004). Model 5 again excludes potentially endogenous controls, while Model 6 includes the full list of controls. The main results are equivalent across both models: exposure to an additional IMF condition is associated with a statistically significant decrease of *0.07* percentage points in government education spending as a share of GDP. Compared to Models 3 and 4, the effect of IMF conditions is accentuated and reaches the higher significance threshold of $p < 0.01$. The IMF participation variable, while now positive in direction, remains statistically non-significant. Control variables are also qualitatively equivalent with the exception of a (non-significant) reversal of effect on the government balance. The conditionality compound instrument remains strong; and the first-stage probit shows the participation compound instrument is a statistically significant predictor of IMF participation.

To recap our results thus far, we find a consistent negative effect of the total number of IMF conditions on government education spending but no effect for other aspects of IMF operations across multiple model specifications and identification strategies. To demonstrate a potential avenue for future research, we investigate the impact of disaggregated sets of conditions in Table 3. When disaggregating conditionality, care must be taken to ensure all conditions are jointly included in a single model, so that the effect of the residual conditions is not attributed to the condition type or policy area of interest. Thus, we include two IMF conditionality variables in each model. Compound instruments can be constructed for each condition count within a single model based on the interaction of the within-country average of that condition type with the year-on-

---

[21] As a rule of thumb, F-statistics on excluded instruments should be above ten (Staiger and Stock 1997).

**Table 3** Effect of IMF conditionality on government education spending, disaggregated conditions

| | Dependent variable: Government expenditure on education (% of GDP), 1990–2014 | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| Model specification | 7. Structural vs quantitative conditions | 8. Structural vs quantitative conditions | 9. Expenditure vs other conditions | 10. Expenditure vs other conditions | 11. Revenue vs other conditions | 12. Revenue vs other conditions |
| Identification strategy | Conditionality A IV, Conditionality B IV, Participation IV | Conditionality A IV, Conditionality B IV, Participation CFA | Conditionality A IV, Conditionality B IV, Participation IV | Conditionality A IV, Conditionality B IV, Participation CFA | Conditionality A IV, Conditionality B IV, Participation IV | Conditionality A IV, Conditionality B IV, Participation CFA |
| IMF structural conditions (lagged) | −0.0265 [0.0394] | −0.0359 [0.0402] | . | . | . | . |
| IMF quantitative conditions (lagged) | −0.0620 [0.0430] | −0.0852* [0.0476] | . | . | . | . |
| IMF expenditure conditions (lagged) | . | . | −0.4693*** [0.1568] | −0.5407*** [0.1836] | . | . |
| IMF non-expenditure conditions (lagged) | . | . | −0.0005 [0.0302] | −0.0078 [0.0307] | . | . |
| IMF revenue conditions (lagged) | . | . | . | . | 0.0216 [0.2136] | 0.0174 [0.2294] |
| IMF non-revenue conditions (lagged) | . | . | . | . | −0.0649* [0.0393] | −0.0845** [0.0425] |
| IMF participation (lagged) | −0.3958 [1.3234] | 0.0391 [0.5095] | −0.2169 [1.1767] | 0.3133 [0.3947] | −0.2732 [1.4386] | 0.2979 [0.3823] |
| Log(GDP per capita) | −2.2516*** [0.6579] | −2.0521*** [0.6105] | −2.8380*** [0.9000] | −2.5878*** [0.9032] | −2.3593*** [0.7319] | −2.1829*** [0.6726] |
| Urbanisation | 0.0491 [0.0415] | 0.0549 [0.0405] | 0.0365 [0.0477] | 0.0414 [0.0486] | 0.0461 [0.0400] | 0.0534 [0.0389] |
| Dependency ratio | 0.0099 | 0.0149 | −0.0005 | 0.0034 | 0.0090 | 0.0133 |

**Table 3** (continued)

| | | | | | | |
|---|---|---|---|---|---|---|
| | Dependent variable: Government expenditure on education (% of GDP), 1990–2014 | | | | | |
| | [0.0171] | [0.0155] | [0.0197] | [0.0194] | [0.0171] | [0.0152] |
| Democracy | 0.1526** | 0.0980** | 0.1242* | 0.0590 | 0.1516* | 0.0977** |
| | [0.0737] | [0.0459] | [0.0745] | [0.0563] | [0.0810] | [0.0442] |
| Government balance (lagged) | 0.0020 | 0.0051 | 0.0067 | 0.0080 | 0.0018 | 0.0039 |
| | [0.0150] | [0.0172] | [0.0188] | [0.0208] | [0.0144] | [0.0154] |
| Trade (lagged) | 0.0086 | 0.0086 | 0.0057 | 0.0047 | 0.0094 | 0.0092 |
| | [0.0065] | [0.0063] | [0.0065] | [0.0063] | [0.0070] | [0.0071] |
| Constant | 15.6711*** | 13.3314*** | 19.8033*** | 16.9623*** | 16.3623*** | 14.1089*** |
| | [5.1585] | [4.3341] | [6.6400] | [6.2282] | [5.7019] | [4.7014] |
| Country fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Year fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of observations | 1307 | 1307 | 1307 | 1307 | 1307 | 1307 |
| Number of countries | 134 | 134 | 134 | 134 | 134 | 134 |
| F-statistic for conditionality A instrument | 55.17 | 51.42 | 18.17 | 17.06 | 9.52 | 8.90 |
| F-statistic for conditionality B instrument | 32.58 | 24.09 | 26.47 | 20.28 | 29.19 | 26.58 |
| F-statistic for participation instrument | 19.80 | n/a | 14.56 | n/a | 14.48 | n/a |
| Joint F-statistic | 70.45 | 66.47 | 32.07 | 27.85 | 29.98 | 27.30 |

F-tests are Kleibergen-Paap statistics. Cluster-robust standard errors in brackets

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

year IMF budget constraint. The online appendices report first-stage models for IMF participation and both conditionality variables.

In Models 7 and 8, we test the impact of the number of conditions disaggregated into structural and quantitative types using our two alternative identification strategies. Results reveal that quantitative conditions yield a statistically significant negative effect in line with our theoretical expectations, but only when we use a control function approach to correct for endogeneity of IMF participation; however, structural conditions have a non-significant negative association on both models. The IMF participation variable also never reaches standard thresholds of significance and fluctuates in effect direction across the two models. Findings on controls remain consistent with previous models; and diagnostic statistics indicate our compound instruments are strong throughout.

For the next set of models, we investigate the impact of specific policy-area conditions.[22] We examine, in Models 9 and 10, expenditure conditions using the same identification strategies as the previous two models. Predicated on our earlier discussion of pathways linking IMF conditionality to government education expenditure, we expect these will constrain education spending. As a falsification test, in Models 11 and 12 we examine revenue conditions, which we expect will have no effect. Based on our dataset coding, expenditure conditions include all those related to expenditure administration, fiscal transparency, audits, budget preparation, domestic arrears, and fiscal balance; revenue conditions include those related to customs administration, tax policy, tax administration, and audits of private enterprises. The online appendices provide additional details on how conditions are classified into policy areas.

The results confirm our theoretical expectations. Expenditure conditions are negatively associated with government spending on education, and residual policy areas exhibit no effect; whereas revenue conditions yield no association, and residual policy areas are negatively associated with government spending on education. In particular, each additional expenditure condition corresponds to at least a *0.47* percentage point decrease in government education spending as a share of GDP. Setting the number of expenditure conditions at the mean of *1.77* would thus result in a *0.83* percentage point decrease in education spending, all other factors held constant. Outside of the conditionality channel, the IMF effect direction again fluctuates but never reaches standard thresholds of significance; and findings on control variables are also substantively unchanged from previous models. Diagnostic statistics indicate that our compound instruments are strong on all disaggregated condition counts except for those pertaining to IMF revenue conditions in Models 11 and 12, with Kleibergen-Paap F-statistics of *9.52* and *8.90* respectively. It thus cannot be discounted that coefficients in these two models may be biased. Our participation instrument is also strong throughout; and where a control function approach is used, first-stage probit results show the participation compound instrument is a statistically significant predictor of IMF participation.

---

[22] The differential impact of conditionality can be captured across a series of specific policy areas by including a separate variable for the number of conditions requiring adoption of a specific policy area—such as anti-corruption, stabilization, liberalization, deregulation, or privatization—along with a variable for the number of all remaining conditions. We recommend the inclusion of one specific policy area per model to avoid issues of multicollinearity between condition counts of different policy areas. For instance, liberalization conditions tend to be accompanied by privatization conditions, so are highly correlated in analyses.

**Table 4** Robustness checks for effect of IMF conditionality on government education spending

| Model specification | 13. Implementation corrected conditions | 14. Implementation discounted conditions | 15. All conditions, non-binding included | 16. All conditions | 17. All conditions | 18. All conditions | 19. All conditions | 20. All conditions | 21. All conditions |
|---|---|---|---|---|---|---|---|---|---|
| | Dependent variable: Government expenditure on education (% of GDP), 1990–2014 | | | | | | | | |
| Identification strategy | Conditionality IV, Participation IV | Conditionality IV, Participation IV | Conditionality IV, Participation IV | Conditionality IV, Participation IV | Conditionality IV, Participation CFA | Conditionality IV, Participation IV | Conditionality IV, Participation IV (UNGA) | Conditionality IV, Participation IV (UNSC) | Conditionality IV (UNGA & UNSC), Participation IV |
| IMF conditions (lagged) | -0.0510** | -0.0461* | -0.0313 | -0.0426 | -0.1117* | -0.0516* | -0.0722*** | -0.0790*** | 0.1540 |
| | [0.0221] | [0.0238] | [0.0296] | [0.0631] | [0.0678] | [0.0264] | [0.0245] | [0.0278] | [0.1395] |
| IMF participation (lagged) | 0.7511 | 0.3599 | -0.6159 | -1.7435 | 0.0808 | 0.1262 | -3.8708 | 1.6850 | -6.3195** |
| | [0.8008] | [0.8701] | [1.9666] | [2.2149] | [0.3352] | [1.3369] | [2.8524] | [3.2288] | [3.0448] |
| Log(GDP per capita) | -1.0932* | -1.0682 | -2.2177*** | -2.9520*** | -3.1609** | -1.9275*** | -4.5202*** | -1.9031 | -0.0051 |
| | [0.6422] | [0.6628] | [0.7320] | [1.1386] | [1.3763] | [0.6086] | [1.6426] | [1.9414] | [3.9733] |
| Urbanisation | 0.0643* | 0.0642 | 0.0446 | 0.0132 | 0.0423 | 0.0416 | 0.0538 | 0.0562 | -0.0202 |
| | [0.0388] | [0.0402] | [0.0412] | [0.0579] | [0.0474] | [0.0392] | [0.0558] | [0.0376] | [0.0729] |
| Dependency ratio | 0.0109 | 0.0085 | 0.0087 | 0.0006 | 0.0120 | 0.0092 | -0.0089 | 0.0180 | 0.0034 |
| | [0.0154] | [0.0157] | [0.0173] | [0.0264] | [0.0222] | [0.0153] | [0.0342] | [0.0235] | [0.0342] |
| Democracy | 0.1088 | 0.1193* | 0.1566* | 0.2050* | 0.0638 | 0.1193* | 0.3190** | 0.0590 | 0.3007 |
| | [0.0684] | [0.0691] | [0.0903] | [0.1215] | [0.0593] | [0.0708] | [0.1604] | [0.1789] | [0.2058] |
| Government balance (lagged) | -0.0002 | -0.0010 | -0.0016 | 0.0070 | 0.0154 | -0.0027 | 0.0219 | 0.0016 | -0.0458 |
| | [0.0154] | [0.0157] | [0.0140] | [0.0164] | [0.0196] | [0.0131] | [0.0255] | [0.0229] | [0.0551] |
| Trade (lagged) | 0.0048 | 0.0037 | 0.0077 | 0.0112 | 0.0102 | 0.0082 | 0.0125 | 0.0074 | 0.0060 |
| | [0.0065] | [0.0062] | [0.0065] | [0.0077] | [0.0071] | [0.0063] | [0.0095] | [0.0071] | [0.0112] |
| Log(Education commitments of ODA) | . | . | . | 0.1511** | 0.1781** | . | . | . | . |
| | | | | [0.0741] | [0.0866] | | | | |
| Role-equivalent IMF countries | . | . | . | . | . | -0.0089 | . | . | . |
| | | | | | | [0.0085] | | | |
| Constant | 7.8712 | 7.9644 | 15.8164*** | 18.4487** | 16.9320* | 13.7183*** | 33.9482** | 11.7190 | 4.0161 |
| | [5.6730] | [5.9166] | [5.7281] | [8.1397] | [8.7240] | [4.7568] | [13.8999] | [15.8149] | [28.6854] |
| Country fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of observations | 968 | 968 | 1307 | 1040 | 1040 | 1307 | 1285 | 1287 | 1265 |

**Table 4** (continued)

| Dependent variable: Government expenditure on education (% of GDP), 1990–2014 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Number of countries | 129 | 129 | 134 | 120 | 120 | 134 | 132 | 131 | 129 |
| F-statistic for conditionality instrument | 17.14 | 13.21 | 20.74 | 15.96 | 8.55 | 34.16 | 31.75 | 30.61 | 2.76 |
| F-statistic for participation instrument | 17.98 | 14.15 | 13.61 | 8.79 | n/a | 10.84 | 3.89 | 0.08 | 12.88 |
| Joint F-statistic | 21.76 | 17.30 | 20.81 | 16.19 | n/a | 34.66 | 35.5 | 33.10 | 17.13 |

F-tests are Kleibergen-Paap statistics. Cluster-robust standard errors in brackets

* p < 0.10, ** p < 0.05, *** p < 0.01

## 4.4 Robustness checks

We perform a series of robustness checks, presented in Table 4. Results from the first-stage models for IMF participation and conditionality are available in the online appendices. One concern is that countries fail to carry out these conditions, which would mean we erroneously attribute an effect to conditions that were never implemented. While using a binding condition count largely circumvents this issue (as they must be met in order for the program to continue), in some circumstances the IMF's Executive Board can grant waivers that allow countries to miss certain conditions without having their program suspended (Pop-Eleches 2009; Stone 2004). We thus deploy an implementation-corrected measure of conditionality, which deducts the number of waivers from the total number of binding conditions, in order to account for country compliance with conditions. Since the measure is not available beyond 2009, our sample period is reduced. As shown in Model 13, and using our preferred identification strategy, exposure to additional IMF conditions is still associated with decreases in government education spending at standard thresholds of significance, and results on control variables remain consistent. Compound instruments also remain strong. As an alternative, we instead adopt an implementation-discounted measure of conditionality in Model 14. IMF staff conduct reviews of programs regularly, and in case of non-implementation, the conclusion of these reviews are delayed. Thus, we discount conditions during the interruption period to account for non-compliance, as detailed in the online appendices; again, our sample does not extend beyond 2009. We find that accounting for program delays does not substantively alter results.

Second, in the analyses thus far we focus on binding conditions, but it is plausible that non-binding conditions may also have an effect on government education spending. In Model 15, we incorporate the number of binding and non-binding conditions into our measure of conditionality. Results on control variables remain consistent, but exposure to additional IMF conditions has an attenuated effect and no longer holds a statistically significant relationship with decreases in government education spending—consistent with the view that the inclusion of non-binding conditions introduces noise to the analysis (Stubbs et al. 2017b).

Third, to alleviate further concerns over the excludability of our conditionality instrument, we control for additional variables that may be driving the relationship between the IMF budget constraint and education expenditure that are also correlated with the country-specific average of conditionality. One possibility is donors' changing commitment to education via the development of new global initiatives, such as the Global Partnership on Education or Goal 2 of the Millennium Development Goals (i.e., to achieve universal primary education). These kinds of initiatives could prompt member countries to increase voluntary contributions to the IMF's concessionary account, thereby decreasing the IMF budget constraint; and the IMF might also assign more conditions to countries with poor education outcomes. At the same time, donors may also increase funding to countries with poor education performance, which could result in more government education spending. To address this issue, in Model 16 we control for the natural log of total education commitments of overseas development assistance to the recipient country. Our sample is reduced by 20% due to missing data on education aid. In this context, the effect size of IMF conditions is consistent with previous models, but it does not reach standard thresholds of statistical significance

since the loss of observations reduces the precision of our estimates; our compound instrument for IMF participation is also slightly weak. We therefore re-run the analysis in Model 17 using instead the more efficient control function approach to account for endogeneity of IMF participation, finding IMF conditions have a statistically significant negative association with government education spending but that our compound instrument for IMF conditions is slightly weak.

Another possibility is that countries make policy decisions based on a process of social comparison—or 'competitive mimicry'—stemming from the pressure to remain effective and efficient relative to relevant—or 'role equivalent'—others in the global system of states (Henisz et al. 2005). A country's selection decision into an IMF program could be prompted by the number of role-equivalent others currently participating, which will impact the IMF's budget constraint and—in turn—the country-specific average of conditionality. At the same time, a country may expect competition in the global system of states to intensify and expand when a greater number of role-equivalent others are on IMF programs, and will therefore be more compelled to match education spending levels with role-equivalent others. To account for this scenario, we create 18 groups of role-equivalent countries by sub-dividing the six World Bank regional groups (East Asia and Pacific, Europe and Central Asia, Latin America and the Caribbean, Middle East and North Africa, South Asia, and Sub-Saharan Africa) into three World Bank income groups (low income, lower-middle income, and upper-middle income). For each country, we then calculate the number of country-years that role-equivalent countries spent under IMF participation in the past three years. When we include this variable in Model 18, our findings are unchanged.

Fourth, we compare our main results against alternative instruments for IMF participation. Past studies have shown that variables approximating geopolitical importance impact upon the decision to participate in IMF programs without necessarily affecting most domestic economic or social outcomes of interest, though the LATE might not be generalizable (Dreher et al. 2018). For instance, we know that allies of big powers receive favorable treatment by international financial institutions (Barro and Lee 2005; Dreher and Gassebner 2012; Steinwand and Stone 2008; Thacker 1999; Woo 2013). In Model 19, we thus deploy UNGA voting similarities with the US as an instrument for IMF participation instead of our compound instrument. Results remain consistent but the UNGA instrument is relatively weak. An alternative candidate variable for geopolitical importance is temporary membership in the United Nations Security Council (Dreher et al. 2009; Moser and Sturm 2011; Woo 2013), since major shareholders of the IMF may care about how countries vote and some countries are willing to trade their votes for IMF loans (Dreher et al. 2009). Using this variable as an instrument for IMF participation in Model 20 does not substantively alter results—and the instrument does not appear to be valid.

Fifth, we adopt alternative instruments to account for the endogeneity of conditionality in Model 21. As well as being determinants of IMF participation, UNGA voting similarities with the US and UNSC temporary membership are identified as potential determinants of conditionality that are plausibly exogenous with regard to government education spending (Caraway et al. 2012; Chwieroth 2015; Dreher and Jensen 2007; Dreher et al. 2015; Nelson 2014). Countries aligned with the US tend to receive more favorable treatment from the IMF and may thus receive fewer conditions; and major shareholders of the IMF may advocate softer conditionality in return for political

influence over the UNSC. Using this strategy, we find a highly negative effect on government education spending of IMF intervention outside of the conditionality channel, but no statistically significant effect of IMF conditions. Little can be read into this result, however, because our conditionality instruments are weak, with a combined Kleibergen-Paap F-statistic of *2.76*. This reinforces the value of our preferred compound instrumentation approach to examining the effects of IMF conditionality.

## 5 Conclusions

This article offered a new strategy for estimating the effects of IMF programs by incorporating fine-grained data on IMF-mandated policy reforms, or conditionality. The release of such data has enabled scholars to overcome two frequently evoked criticisms of earlier studies: that they treat the policy content of programs as homogenous; and that they do not delineate the effects of conditionality from other pathways of program influence. In so doing, the new data introduces methodological challenges linked to the inclusion of multiple endogenous IMF variables. After a review of existing quantitative approaches, we advocated that future studies utilize MLE over a system of three simultaneous equations, in effect combining the following: (a) an instrumental variable approach to account for endogeneity of IMF program *participation*, using the interaction of the within-country average of IMF program participation with the natural log of the IMF liquidity ratio as an excludable instrument; and (b) an instrumental variable approach to account for endogeneity of IMF *conditionality*, using the interaction of the within-country average of the number of conditions with the natural log of the IMF liquidity ratio as an excludable instrument. Monte Carlo simulations confirmed that our approach to addressing endogeneity is unbiased and performs better than the alternatives, provided that instruments are not weak. Applying our approach, we found that over the last two decades IMF conditions decreased government education spending in developing countries, consistent with expectations from past studies.

We note three main shortcomings of our approach. First, using the number of conditions as a measure of conditionality may not accurately reflect the true burden of conditionality, since it does not capture the relative difficulty of implementing any individual condition. By way of compromise, subsequent studies could consider how to incorporate the scope of conditionality—that is, the number of policy areas subject to conditionality—in addition to the number of conditions. Second, while our method enriches our understanding of the consequences of conditionality, it tells us less about the effect of IMF technical assistance. Our study utilized a dummy variable for the presence of a Fund-supported program to capture the effect of technical assistance, thereby—and in similar vein to criticisms made about the inability of earlier studies to encapsulate heterogeneous policy content—masking diverse country experiences that may, in turn, affect the outcome of interest. This markedly under-researched phenomenon represents the most promising avenue for future research to pursue. Third, compound instrumentation is no panacea. The approach is useful for identification when no obvious (non-compound) instrument is available, but researchers should verify to the extent possible that underlying assumptions hold. The core assumptions include (1) monotonicity of treatment (i.e., changes in the

outcome in countries with different probability of IMF programs and mean number of IMF conditions—in short, IMF treatment variables—are not affected differently by changes in the IMF's budget constraint, other than via IMF programs and IMF conditionality, respectively); (2) parallel trends (i.e., the trends in IMF treatment variables and the outcome are similar across countries with above-mean exposure and below-mean exposure to IMF treatments); and (3) non-overlapping trends (i.e., non-linear trends in these variables for above-mean exposure countries do not overlap with the trend in the IMF's budget constraint).

Why should scholars prefer our approach over others when evaluating IMF programs? Our strategy has the advantage of offering an excludable instrument for conditionality which can—and should—be varied as needed for specific applications to conditionality policy areas, as we demonstrated with IMF expenditure and revenue conditions. Our approach also allows us to isolate where an effect is derived from among the conditions, promising greater nuance on policy advice. For instance, the results of our empirical application suggest that IMF programs should specifically reduce expenditure conditions if government spending on education is to be increased. Taken together, our approach enables scholars to draw more complete interpretations on the consequences of IMF programs. Even so, it is worth bearing in mind that any approach to endogeneity does not absolve scholars from thinking carefully about the underlying data-generating processes and making decisions based on strong theoretical foundations (Chaudoin et al. 2016).

Our methods are also transferable to studies on the effects of other organizations that engage in conditional program lending, such as the World Bank, regional development banks, the European Union, or large bilateral donors. Indeed, studying the effects of these programs is essential if we wish to better understand processes of change. As interest in the IMF—and other international financial institutions—continues to grow, so too will the available data sources and the attendant body of research around it. If we are to make the most of these developments, it is imperative that we revisit the methodologies used to estimate their consequences, paying attention to basic dictums of econometric analysis and the challenges of endogeneity. Until and unless we do, our knowledge of how these international organizations impact social, economic, or political outcomes will suffer.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

# References

Abouharb, M. R., & Cingranelli, D. L. (2009). IMF programs and human rights, 1981–2003. *Review of International Organizations, 4*, 47–72. https://doi.org/10.1007/s11558-008-9050-5.

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies, 58*(2), 277–297. https://doi.org/10.2307/2297968.

Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics, 68*, 29–51. https://doi.org/10.1016/0304-4076(94)01642-D.

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. https://doi.org/10.1007/s11558-006-6612-2.

Augurzky, B., & Kluve, J. (2007). Assessing the performance of matching algorithms when selection into treatment is strong. *Journal of Applied Econometrics, 22*(3), 533–557. https://doi.org/10.1002/jae.919.

Babb, S. (2005). The social consequences of structural adjustment: Recent evidence and current debates. *Annual Review of Sociology, 31*, 199–222. https://doi.org/10.1146/annurev.soc.31.041304.122258.

Babb, S., & Buira, A. (2005). Mission creep, mission push and discretion: The case of IMF conditionality. In A. Buira (Ed.), *The IMF and World Bank at sixty* (pp. 59–83). London: Anthem Press.

Babb, S., & Kentikelenis, A. (2018). International financial institutions as agents of neoliberalism. In D. Cahill, M. Cooper, M. Konings, & D. Primrose (Eds.), *The SAGE handbook of neoliberalism* (pp. 16–27). Thousand Oaks: SAGE Publications.

Bal Gündüz, Y. (2016). The economic impact of short-term IMF engagement in low-income countries. *World Development, 87*, 30–49. https://doi.org/10.1016/j.worlddev.2016.04.012.

Baqir, R. (2002). Social sector spending in a panel of countries. IMF Working Paper No. 35. Washington, DC.

Barro, R. J., & Lee, J.-W. (2005). IMF programs: Who is chosen and what are the effects? *Journal of Monetary Economics, 52*(7), 1245–1269. https://doi.org/10.1016/j.jmoneco.2005.04.003.

Bas, M., & Stone, R. W. (2014). Adverse selection and growth under IMF programs. *Review of International Organizations, 9*, 1–28. https://doi.org/10.1007/s11558-013-9173-1.

Beazer, Q. H., & Woo, B. (2016). IMF conditionality, government partisanship, and the progress of economic reforms. *American Journal of Political Science, 60*(2), 304–321. https://doi.org/10.1111/ajps.12200.

Bell, A., & Jones, K. (2015). Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. *Political Science Research and Methods, 3*, 133–153. https://doi.org/10.1017/psrm.2014.7.

Bird, G. (2009). Reforming IMF conditionality: From "streamlining" to "major overhaul.". *World Economics, 10*(3), 81–104.

Blanton, R. G., Blanton, S. L., & Peksen, D. (2015). The impact of IMF and World Bank programs on labor rights. *Political Research Quarterly, 68*(2), 324–336. https://doi.org/10.1177/1065912915578462.

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics, 87*, 115–143. https://doi.org/10.1016/S0304-4076(98)00009-8.

Brady, D., & Lee, H. Y. (2014). The rise and fall of government spending in affluent democracies, 1971-2008. *Journal of European Social Policy, 24*, 56–79. https://doi.org/10.1177/0958928713511281.

Broome, A., & Seabrooke, L. (2015). Shaping policy curves: Cognitive authority in transnational capacity building. *Public Administration, 93*(4), 956–972. https://doi.org/10.1111/padm.12179.

Bulír, A., & Moon, S. (2004). Is fiscal adjustment more durable when the IMF is involved? *Comparative Economic Studies, 46*(3), 373–399. https://doi.org/10.1057/palgrave.ces.8100051.

Bun, M. J. G., & Harrison, T. D. (2018). OLS and IV estimation of regression models including endogenous interaction terms. *Econometric Reviews*, 1–14. https://doi.org/10.1080/07474938.2018.1427486.

Butkiewicz, J., & 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. https://doi.org/10.1016/j.worlddev.2004.09.006.

Caraway, T. L., Rickard, S. J., & Anner, M. S. (2012). International negotiations and domestic politics: The case of IMF labor market conditionality. *International Organization, 66*(1), 27–61. https://doi.org/10.1017/S0020818311000348.

Carter, D. B., & Stone, R. W. (2015). Democracy and multilateralism: The case of vote buying in the UN general assembly. *International Organization, 69*, 1–33. https://doi.org/10.1017/S0020818314000186.

Casper, B. A. (2017). IMF programs and the risk of a coup d'etat. *Journal of Conflict Resolution, 61*(5), 964–996. https://doi.org/10.1177/0022002715600759.

Chang, H.-J. (2007). *Bad Samaritans: The myth of free trade and the secret history of capitalism*. New York: Bloomsbury Press.

Chapman, T., Fang, S., Li, X., & Stone, R. W. (2017). Mixed signals: IMF lending and capital markets. *British Journal of Political Science, 47*(2), 329–349. https://doi.org/10.1017/S0007123415000216.

Chaudoin, S., Hays, J., & Hicks, R. (2016). Do we really know the WTO cures cancer? *British Journal of Political Science., 48*, 903–928. https://doi.org/10.1017/S000712341600034X.

Christian, P., & Barrett, C. B. (2017). Revisiting the effect of food aid on conflict: A methodological caution. World Bank Policy Research Working Paper No. 8171. Washington, DC.

Chwieroth, J. M. (2015). Professional ties that bind: How normative orientations shape IMF conditionality. *Review of International Political Economy, 22*(4), 757–787. https://doi.org/10.1080/09692290.2014.898214.

Clements, B., Gupta, S., & Nozaki, M. (2013). What happens to social spending in IMF-supported programmes? *Applied Economics, 48*(28), 4022–4033. https://doi.org/10.1080/00036846.2012.744136.

Copelovitch, M. (2010a). Master or servant? Common agency and the political economy of IMF lending. *International Studies Quarterly, 54*, 49–77. https://doi.org/10.1111/j.1468-2478.2009.00577.x.

Copelovitch, M. (2010b). *The International Monetary Fund in the global economy: Banks, Bonds, and Bailouts*. Cambridge: Cambridge University Press.

Crivelli, E., & Gupta, S. (2016). Does conditionality in IMF-supported programs promote revenue reform? *International Tax and Public Finance, 23*(3), 550–579. https://doi.org/10.1007/s10797-015-9379-7.

Dawson, T. C. (2002). Stiglitz, the IMF and globalization. Retrieved December 6, 2016, from http://www.imf.org/external/np/speeches/2002/061302.htm

Diamond, A., & Sekhon, J. S. (2013). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics, 95*(3), 932–945. https://doi.org/10.1162/REST_a_00318.

Dreher, A. (2006). IMF and economic growth: The effects of programs, loans, and compliance with conditionality. *World Development, 34*(5), 769–788. https://doi.org/10.1016/j.worlddev.2005.11.002.

Dreher, A. (2009). IMF conditionality: Theory and evidence. *Public Choice, 141*, 233–267. https://doi.org/10.1007/s11127-009-9486-z.

Dreher, A., & Gassebner, M. (2012). Do IMF and World Bank programs induce government crises? An empirical analysis. *International Organization, 66*(2), 329–358. https://doi.org/10.1017/S0020818312000094.

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*, 105–124. https://doi.org/10.1086/508311.

Dreher, A., & Lang, V. F. (2019). The political economy of international organizations. In R. D. Congleton, B. N. Grofman, & S. Voigt (Eds.), *The Oxford Handbook of public choice, volume 2*. Oxford: Oxford University Press.

Dreher, A., & Langlotz, S. (2017). Aid and growth: New evidence using an excludable instrument. Discussion Paper No. 635. Heidelberg.

Dreher, A., & Vaubel, R. (2004). The causes and consequences of IMF conditionality. *Emerging Markets Finance and Trade, 40*(3), 26–54. https://doi.org/10.1080/1540496X.2004.11052571.

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*, 1–18. https://doi.org/10.1016/j.worlddev.2009.05.007.

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. https://doi.org/10.1016/j.euroecorev.2009.03.002.

Dreher, A., Sturm, J.-E., & Vreeland, J. R. (2015). Politics and IMF conditionality. *Journal of Conflict Resolution, 59*, 120–148. https://doi.org/10.1177/0022002713499723.

Dreher, A., Eichenauer, V. Z., & Gehring, K. (2018). Geopolitics, aid, and growth: The impact of UN security council membership on the effectiveness of aid. *World Bank Economic Review, 32*(2), 268–286. https://doi.org/10.1093/wber/lhw037.

Easterly, W. (2005). What did structural adjustment adjust? The association of policies and growth with repeated IMF and World Bank adjustment loans. *Journal of Development Economics, 76*, 1–22. https://doi.org/10.1016/j.jdeveco.2003.11.005.

Garuda, G. (2000). The distributional effects of IMF programs: A cross-country analysis. *World Development, 28*(6), 1031–1051. https://doi.org/10.1016/S0305-750X(00)00002-4.

Gould, E. (2003). Money talks: Supplementary financiers and International Monetary Fund conditionality. *International Organization, 57*(3), 551–586. https://doi.org/10.1017/S0020818303573039.

Grabel, I. (2011). Not your grandfather's IMF: Global crisis, "productive incoherence" and developmental policy space. *Cambridge Journal of Economics, 35*(5), 805–830. https://doi.org/10.1093/cje/ber012.

Greene, W. (2004). Fixed effects and bias due to the incidental parameters problem in the Tobit model. *Econometric Reviews, 23*(2), 125–147. https://doi.org/10.1081/ETC-120039606.

Hainmueller, J., Mummolo, J., & Xu, Y. (2016). How much should we trust estimates from multiplicative interaction models? Simple tools to improve empirical practice. Working Paper. Palo Alto, CA.

Hardoy, I. (2003). Effect of IMF programmes on growth: A reappraisal using the method of matching. Working Paper. Oslo.

Hartzell, C. A., Hoddie, M., & Bauer, M. E. (2010). Economic liberalization via IMF structural adjustment: Sowing the seeds of civil war? *International Organization, 64*(2), 339–356. https://doi.org/10.1017/S0020818310000068.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica, 47*, 153–161. https://doi.org/10.2307/1912352.

Henisz, W. J., Zelner, B. A., & Guillén, M. F. (2005). The worldwide diffusion of market-oriented infrastructure reform, 1977-1999. *American Sociological Review, 70*(6), 871–897. https://doi.org/10.1177/000312240507000601.

Huber, E., Mustillo, T., & Stephens, J. (2008). Politics and social spending in Latin America. *Journal of Politics, 70*(2), 420–436. https://doi.org/10.1017/S0022381608080407.

IEO (2003). *Fiscal adjustment in IMF-supported programs.* Washington, DC: Independent Evaluation Office of the International Monetary Fund.

IEO (2007). *Structural conditionality in IMF-supported programs.* Washington, DC: Independent Evaluation Office of the International Monetary Fund.

IMF (1997). *Azerbaijan Republic: Request for second annual arrangement under the Enhanced Structural Adjustment Facility, and second review under the Extended Arrangement.* Washington, DC: International Monetary Fund.

IMF (1998). *Benin: Request for the second annual arrangement under the Enhanced Structural Adjustment Facility.* Washington, DC: International Monetary Fund.

IMF (1999). *Bolivia: Staff report for the 1999 article IV consultation and request for second annual arrangement under the Poverty Reduction and Growth Facility.* Washington, DC: International Monetary Fund.

IMF (2000a). *Benin: Request for arrangement under the Poverty Reduction and Growth Facility.* Washington, DC: International Monetary Fund.

IMF (2000b). *Nicaragua: First and second reviews under the second annual arrangemnet under the Poverty Reduction and Growth Facility and request for waiver of performance criteria.* Washington, DC: International Monetary Fund.

IMF (2001a). Managing Directors' report to the International Monetary and Financial Committee—Streamlining conditionality and enhancing ownership. Retrieved January 25, 2017, from https://www.imf.org/external/np/omd/2001/110601.htm

IMF (2001b). *Structural conditionality in Fund-supported programs.* Washington, DC: International Monetary Fund.

IMF (2002). *Sierra Leone: Second review under the Poverty Reduction and Growth Facility and request for a waiver of performance criterion.* Washington, DC: International Monetary Fund.

IMF (2003). *Republic of Tajikistan: Second review under the three-year arrangement under the Poverty Reduction and Growth Facility, and request for a waiver of performance criterion.* Washington, DC: International Monetary Fund.

IMF (2004). *Signaling by the Fund: A historical review.* Washington, DC: International Monetary Fund.

IMF (2006). *Bulgaria: Second review under the Stand-By Arrangement and requests for waiver of performance criteria and postponement of third review.* Washington, DC: International Monetary Fund.

IMF (2015). IMF conditionality. Retrieved May 11, 2015, from http://www.imf.org/external/np/exr/facts/conditio.htm

IMF (2016a). *Corruption: Costs and mitigating strategies*. IMF Staff Discussion Note No. SDN/16/05. Washington, DC.

IMF (2016b). Where the IMF gets its money. Retrieved February 11, 2017, from http://www.imf.org/en/About/Factsheets/Where-the-IMF-Gets-Its-Money

IMF (2017a). IMF Quotas. Retrieved January 11, 2018, from http://www.imf.org/en/About/Factsheets/Sheets/2016/07/14/12/21/IMF-Quotas

IMF (2017b). *Social safeguards and* program *design in PRGT and PSI-supported programs*. Washington, DC: International Monetary Fund.

Ivanova, A., Mayer, W., Mourmouras, A., & Anayiotos, G. (2006). What determines the implementation of IMF-supported programs? In A. Mody & A. Rebucci (Eds.), *IMF-supported programs: Recent staff research* (pp. 160–188). Washington, DC: International Monetary Fund.

Kentikelenis, A. (2017). Structural adjustment and health: A conceptual framework and evidence on pathways. *Social Science & Medicine, 187*, 296–305. https://doi.org/10.1016/j.socscimed.2017.02.021.

Kentikelenis, A., & Seabrooke, L. (2017). The politics of world polity: Script-writing in international organizations. *American Sociological Review, 82*(5), 1065–1092. https://doi.org/10.1177/0003122417728241.

Kentikelenis, A., Stubbs, T., & King, L. (2015). Structural adjustment and public spending on health: Evidence from IMF programs in low-income countries. *Social Science & Medicine, 126*, 169–176. https://doi.org/10.1016/j.socscimed.2014.12.027.

Kentikelenis, A., Stubbs, T., & King, L. (2016). IMF conditionality and development policy space, 1985–2014. *Review of International Political Economy, 23*(4), 543–582. https://doi.org/10.1080/09692290.2016.1174953.

King, G., & Nielsen, R. (2018). Why propensity scores should not be used for matching. Working Paper. Cambridge, MA.

Lang, V. F. (2016). The economics of the democratic deficit: The effect of IMF programs on inequality. Discussion Paper No. 617. Heidelberg.

Marphatia, A. A. (2010). The adverse effects of International Monetary Fund programs on the health and education workforce. *International Journal of Health Services, 40*, 165–178. https://doi.org/10.2190/HS.40.1.j.

Morgan, S., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research*. Cambridge: Cambridge University Press.

Moser, C., & Sturm, J.-E. (2011). Explaining IMF lending decisions after the cold war. *Review of International Organizations, 6*(3), 307–340. https://doi.org/10.1007/s11558-011-9120-y.

Mukherjee, B., & Singer, D. (2010). International institutions and domestic compensation: The IMF and the politics of capital account liberalization. *American Journal of Political Science, 54*, 45–60. https://doi.org/10.1111/j.1540-5907.2009.00417.x.

Nelson, S. C. (2014). Playing favorites: How shared beliefs shape the IMF's lending decisions. *International Organization, 68*(2), 297–328. https://doi.org/10.1017/S0020818313000477.

Nelson, S. C., & Wallace, G. P. R. (2017). Are IMF lending programs good or bad for democracy? *Review of International Organizations, 12*(4), 523–558. https://doi.org/10.1007/s11558-016-9250-3.

Nielsen, R., & Sheffield, J. (2009). Matching with time-series cross-sectional data. Polmeth XXVI Working Paper. New Haven, CT.

Nizalova, O. Y., & Murtazashvili, I. (2016). Exogenous treatment and endogenous factor: Vanishing of omitted variable bias on the interaction term. *Journal of Econometric Methods, 5*(1), 71–77. https://doi.org/10.1515/jem-2013-0012.

Nooruddin, I., & Simmons, J. W. (2006). The politics of hard choices: IMF programs and government spending. *International Organization, 60*(4), 1001–1033. https://doi.org/10.1017/S0020818306060334.

Nsouli, S., Atoyan, R., & Mourmouras, A. (2006). Institutions, program implementation, and macroeconomic performance. In A. Mody & A. Rebucci (Eds.), *IMF-supported programs: Recent staff research* (pp. 140–159). Washington, DC: International Monetary Fund.

Nunn, N., & Qian, N. (2014). US food aid and civil conflict. *American Economic Review, 104*(6), 1630–1666. https://doi.org/10.1257/aer.104.6.1630.

Oberdabernig, D. (2013). Revisiting the effects of IMF programs on poverty and inequality. *World Development, 46*, 113–142. https://doi.org/10.1016/j.worlddev.2013.01.033.

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. https://doi.org/10.1111/j.1468-2478.2009.00556.x.

Przeworski, A., & Vreeland, J. R. (2000). The effect of IMF programs on economic growth. *Journal of Development Economics, 62*(2), 385–421. https://doi.org/10.1016/S0304-3878(00)00090-0.

Reinsberg, B., Kentikelenis, A., Stubbs, T., & King, L. (2018). The world system and the hollowing-out of state capacity in developing countries: How structural adjustment programs impact bureaucratic quality. *American Journal of Sociology, 124*(3).

Rickard, S., & Caraway, T. (2014). International negotiations in the shadow of national elections. *International Organization, 68*(3), 701–720. https://doi.org/10.1017/S0020818314000058.

Rickard, S., & Caraway, T. (2018). International demands for austerity: Examining the impact of the IMF on the public sector. *Review of International Organizations.* https://doi.org/10.1007/s11558-017-9295-y.

Rodrik, D. (1998). Why do more open economies have bigger governments? *Journal of Political Economy, 106*(5), 997–1032. https://doi.org/10.1086/250038.

Roodman, D. (2009a). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics, 71*, 135–158. https://doi.org/10.1111/j.1468-0084.2008.00542.x.

Roodman, D. (2009b). How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal, 9*, 86–136.

Roodman, D. (2010). Squishy findings on aid fungibility. Retrieved June 29, 2015, from http://www.cgdev.org/blog/squishy-findings-aid-fungibility

Roodman, D. (2011). Fitting fully observed recusive mixed-process models with cmp. *Stata Journal, 11*(2), 159–206.

Rudra, N. (2008). *Globalization and the race to the bottom in developing countries: Who really gets hurt?* Cambridge: Cambridge University Press.

Sartori, A. E. (2003). An estimator for some binary-outcome selection models without exclusion restrictions. *Political Analysis, 11*(2), 111–138. https://doi.org/10.1093/pan/mpg001.

Simmons, B., Dobbin, F., & Garrett, G. (2008). *The global diffusion of markets and democracy.* Cambridge: Cambridge University Press.

Smets, L., & Knack, S. (2016). World Bank lending and the quality of economic policy. *Journal of Development Studies, 52*, 72–91. https://doi.org/10.1080/00220388.2015.1068290.

Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica, 65*(3), 557. https://doi.org/10.2307/2171753.

Stasavage, D. (2005). Democracy and education spending in Africa. *American Journal of Political Science, 49*(2), 343–358. https://doi.org/10.1111/j.0092-5853.2005.00127.x.

Steinwand, M. C., & Stone, R. W. (2008). The International Monetary Fund: A review of the recent evidence. *Review of International Organizations, 3*(2), 123–149. https://doi.org/10.1007/s11558-007-9026-x.

Stiglitz, J. (2002). *Globalization and its discontents.* New York: W. W. Norton.

Stone, R. W. (2002). *Lending credibility: The International Monetary Fund and the post-communist transition.* Princeton: Princeton University Press.

Stone, R. W. (2004). The political economy of IMF lending in Africa. *American Political Science Review, 98*(4), 577–591. https://doi.org/10.1017/S000305540404136X.

Stone, R. W. (2008). The scope of IMF conditionality. *International Organization, 62*(4), 589–620. https://doi.org/10.1017/S0020818308080211.

Stone, R. W. (2011). *Controlling institutions: International organizations and the global economy.* New York: Cambridge University Press.

Stubbs, T., & Kentikelenis, A. (2018). Targeted social safeguards in the age of universal social protection: The IMF and health systems of low-income countries. *Critical Public Health, 28*(2), 132–139. https://doi.org/10.1080/09581596.2017.1340589.

Stubbs, T., Kentikelenis, A., & King, L. (2016). Catalyzing aid? The IMF and donor behavior in aid allocation. *World Development, 78*, 511–528. https://doi.org/10.1016/j.worlddev.2015.10.010.

Stubbs, T., Kentikelenis, A., Stuckler, D., McKee, M., & King, L. (2017a). The IMF and government health expenditure: A response to Sanjeev Gupta. *Social Science & Medicine, 181*, 202–204. https://doi.org/10.1016/j.socscimed.2017.03.050.

Stubbs, T., Kentikelenis, A., Stuckler, D., McKee, M., & King, L. (2017b). The impact of IMF conditionality on government health expenditure: A cross-national analysis of 16 west African nations. *Social Science & Medicine, 174*, 220–227. https://doi.org/10.1016/j.socscimed.2016.12.016.

Stuckler, D., Basu, S., & McKee, M. (2011). International Monetary Fund and aid displacement. *International Journal of Health Services, 41*, 67–76. https://doi.org/10.2190/HS.41.1.e.

Stuckler, D., King, L., & McKee, M. (2012). The disappearing health effects of rapid privatisation: A case of statistical obscurantism? *Social Science & Medicine, 75*, 23–31. https://doi.org/10.1016/j.socscimed.2011.12.054.

Teorell, J., Dahlberg, S., Holmberg, S., Rothstein, B., Khomenko, A., & Svensson, R. (2016). The Quality of Government Standard Data: Version January 2016. Retrieved February 5, 2016, from http://qog.pol.gu.se/data/datadownloads/qogstandarddata

Thacker, S. (1999). The high politics of IMF lending. *World Politics, 52*, 38–75. https://doi.org/10.1017/S0043887100020025.

Thomson, S. (2015). *Economic crisis, health systems and health in Europe: Impact and implications for policy.* Maidenhead: Open University Press.

Vaubel, R. (1996). Bureaucracy at the IMF and the World Bank: A comparison of the evidence. *The World Economy, 19*(2), 195–210. https://doi.org/10.1111/j.1467-9701.1996.tb00672.x.

Vreeland, J. R. (2002). The effect of IMF programs on labor. *World Development, 30*, 121–139. https://doi.org/10.1016/S0305-750X(01)00101-2.

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. https://doi.org/10.1007/s11558-006-0161-6.

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

Wagner, A. (1994). Three extracts on public finance. In R. A. Musgrave & A. T. Peacock (Eds.), *Classics in the theory of public finance* (pp. 1–15). London: Palgrave Macmillan.

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*, 71–81. https://doi.org/10.1016/j.jdeveco.2008.11.008.

Winship, C., & Mare, R. D. (1992). Models for sample selection bias. *Annual Review of Sociology, 18*(1), 327–350. https://doi.org/10.1146/annurev.so.18.080192.001551.

Woo, B. (2013). Conditional on conditionality: IMF program design and foreign direct investment. *International Interactions, 39*(3), 292–315. https://doi.org/10.1080/03050629.2013.782303.

Woods, N. (2006). *The globalizers: The IMF, the World Bank, and their borrowers*. Ithaca: Cornell University Press.

Wooldridge, J. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). Cambridge: MIT Press.

Wooldridge, J. (2012). *Introductory econometrics: A modern approach* (5th ed.). Mason, Ohio: Cengage Learning.

Wooldridge, J. (2015). Control function methods in applied econometrics. *Journal of Human Resources, 50*(2), 420–445. https://doi.org/10.3368/jhr.50.2.420.

World Bank. (2016). World Development Indicators. Retrieved February 5, 2016, from http://data.worldbank.org

Young, A. (2018). Consistency without inference: Instrumental variables in practical application. Working Paper. London.

## Affiliations

**Thomas Stubbs**[1,2] · **Bernhard Reinsberg**[2,3] · **Alexander Kentikelenis**[4] · **Lawrence King**[5]

Bernhard Reinsberg
bernhard.reinsberg@glasgow.ac.uk

Alexander Kentikelenis
alexander.kentikelenis@unibocconi.it

Lawrence King
lpking@econs.umass.edu

[1]  Department of Politics and International Relations, Royal Holloway, University of London, London, UK

[2]  Centre for Business Research, University of Cambridge, Cambridge, UK

[3]  School of Social and Political Sciences, University of Glasgow, Glasgow, UK

[4]  Department of Social and Political Sciences, Bocconi University, Milan, Italy

[5]  Department of Economics, University of Massachusetts Amherst, Amherst, USA