The Unobserved IMF*

Kevin A. Clarke[†]
Randall W. Stone[§]

February 5, 2019

*Ioannis Vassiliadis provided research assistance. Errors are our own.

[†]Corresponding author. Associate Professor, Department of Political Science, University of Rochester, Rochester, NY 14627-0146. Email: kevin.clarke@rochester.edu.

[§]Professor of Political Science, Department of Political Science, University of Rochester, Rochester, NY 14627-0146. Email: randall.stone@rochester.edu.

Abstract

We consider the effects of IMF program participation. We begin by estimating an operationalization of Solow's growth model, the results of which indicate a positive effect of program participation. A formal sensitivity analysis, new to the IMF literature, suggests that selection bias is not a major threat to the results. To test these findings, we employ four different instrumental variables configurations including nonlinear instruments and Bartik instruments. The results are remarkably consistent across these specifications. Participating in an IMF program five years ago is associated with 5-6 percentage points of additional growth.

1 Introduction

The International Monetary Fund is the world's premier institution with responsibility for responding to financial crises, and with 189 members, its advice and policy conditionality reach deeply into the economies of most countries. The effects of IMF lending remain controversial. An extensive empirical literature seeks to quantify the average effects of IMF programs on growth, macroeconomic performance, and a variety of other outcomes including conflict and governance. The fundamental question, however, is unresolved: does a country facing a potential financial crisis improve its growth prospects, on average, by participating in an IMF program? If not, the Fund's *raison d'être* is thrown into question. We present compelling evidence that participation in an IMF program does improve growth outcomes particularly after five years.

The central obstacle to inference has been non-random selection into IMF programs, which Goldstein and Montiel (1986) recognized early in the development of the research program. Countries that participate in IMF programs differ systematically from countries that do not because governments are unwilling to submit themselves to intrusive policy surveillance unless facing severe financial constraints (Vreeland, 2003). Program participants have worse outcomes than the population on average, a problem that Bas and Stone (2014) identify as adverse selection. Research strategies that fail to account for endogenous treatments may attribute the effects of the financial crisis to the IMF intervention designed to respond to it. Scholars have taken a variety approaches to the problem of endogenous participation, but we argue that instrumental variables makes the most sense.

Our empirical strategy unfolds in three stages. First, we estimate a series of baseline regressions and defend our use of unit and time fixed effects. Testing suggests no evidence of serial correlation or cross-sectional dependence. Second, we use formal sensitivity analysis to assess how much of a threat endogeneity is to our results. Third, we estimate a series of instrumental variables regressions. We use single and multiple instruments, as well as a nonlinear instrument and two Bartik instruments. We use weak instrument inference to address the lack of strong correlation between our instruments and program participation.

We find compelling evidence that IMF programs are associated with increased growth rates. The estimated coefficient in our baseline model for the 1-year lag is modest and statistically insignificant. The estimated coefficient for the 5-year lag remains modest, but is significant. When we correct for the endogeneity of IMF programs, our best estimates of the effects are 7 percentage points of growth for a program in the previous year, and 6 percentage points for a program five years ago. The substantial increase in the effect size indicates that endogeneity works to suppress the apparent effect, which is consistent with adverse selection: countries that select into IMF programs would likely have below-average growth performance if they had not participated.

2 The effects of IMF program participation

Countries participate in IMF programs to address severe balance-of-payments crises. In a typical case, the country's macroeconomic policies are inconsistent with its exchange-rate commitments or with long-term debt sustainability. A subsequent shock causes capital flight leading to a crisis. During the crisis, capital flight causes a substantial short-term contraction in aggregate demand and a reduction in growth performance. If the crisis forces the country to abandon a pegged exchange rate and domestic banks are leveraged in foreign currency, the result is a balance-of-payments and banking crisis. The IMF responds by providing credit on better terms than the country could get, which eases the pain of adjustment. The IMF insists on policy conditionality intended to put the country back on a sustainable macroeconomic path. These policy reforms are designed to reduce aggregate demand and eliminate barriers to efficient allocation of resources, and they are often painful to implement. Whether the net effect of the liquidity injection and the policy reforms is positive or negative for growth is an open question.

On the negative side, Vreeland (2003) finds little evidence that IMF policy reforms promote long-term growth. Citing evidence that IMF programs depress real wages and government expenditure, he argues that partisan governments use the IMF as a scapegoat to implement policies that serve investors and wealthy industrialists. Aggregate economic growth suffers as a result. It is also possible that IMF programs do not benefit the countries that participate because IMF funds are fully anticipated. That is, the prospect of IMF bail-outs is foreseen and incentivizes unwise policies and risky loans.

On the positive side, IMF programs may trade off short-term adjustment costs for long-term improvements in living standards. The IMF designs the typical program to engineer a period of declining consumption to stabilize the balance of payments and reestablish investor confidence. Once capital flows return to their normal state, consumption and investment can resume, and the economy enjoys a recovery. Thus, we should expect the average program to cause a short-term decline in growth performance, but a long-term improvement as the economy reaps the dividends of economic reform. We use both one- and five-year lags of IMF program participation to address this possibility.

Finally, IMF programs may be beneficial to growth in the short term and in the long-term. These effects can be obscured because program participation may coincide with the deepening of the financial crisis that forced the government to call upon the Fund. Effect masking may occur if, as Bas and Stone (2014) and Chapman et al. (2017) argue, the IMF faces adverse selection of potential borrowers. The observed effects of IMF programs, conditional on covariates, are biased toward zero if countries that apply for loans are privately pessimistic about their growth. Addressing the endogeneity should produce larger effect estimates.

3 The data

Summary statistics for a 1-year lag and a 5-year lag are in Tables 1 and 2. The data span the years 1977 to 2008 and include 138 countries. The dependent variable is the growth of GDP per capita and comes from the Penn World Tables, which has broader historical data coverage than World Development Indicators (WDI). We built the data set with the intent of operationalizing Solow's (1956) growth model. With that in mind, we use WDI for lagged growth rates of gross capital formation and the labor supply. The lagged value of GDP per capita captures the model's expectation that long-run extensive growth converges to a steady state (Solow, 1956). We also include the openness of the economy, measured as exports plus imports divided by GDP, because trade reduces the prices of scarce factors of production and promotes efficiency-enhancing specialization. These covariates generally behave in the expected ways; the growth rates of capital and labor are positively associated with growth, and GDP per capita is negatively associated with growth. Openness is associated with more rapid growth rates.

Table 1: Summary Statistics: 1-year lag

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
GDP per capita growth	2,709	0.021	0.055	−0.456	−0.002	0.046	0.661
Program participation	2,709	0.385	0.487	0	0	1	1
Gross capital formation	2,709	5.696	23.882	−81.772	−3.535	12.564	723.202
Change in labor force	2,709	0.021	0.017	−0.095	0.011	0.032	0.178
Openness	2,709	73.957	47.453	11.087	45.190	92.124	430.392
GDP per capita	2,709	11,371.550	15,858.950	132	1,293.1	14,681.2	111,968
Democracy	2,709	0.229	0.145	0	0.1	0.3	1
Communist	2,709	1.137	10.826	−141.308	−0.255	0.768	204.143
Voting with the U.S.	2,709	0.614	0.487	0	0	1	1
Balance of payments	2,709	0.215	0.411	0	0	0	1
IMF member	2,709	0.006	0.077	0	0	0	1
UN votes	2,709	0.007	0.018	0	0.001	0.005	0.204
Low income states	2,709	-155.38	244.81	-716.3	-336.9	54.53	311.5

We include in the data set two binary covariates, *democracy* and *communist*. Democracy comes from Cheibub et al. (2010) and is coded 1 for electoral systems that feature alternation of power. Przeworski et al. (2000) argue that democracies are associated with lower rates of GDP per capita growth because the public services that they provide promote more rapid population growth by lowering mortality rates. Communist is coded 1 for the countries of the former Soviet bloc, Yugoslavia and Albania prior to 1990. These countries had similar economic systems during the 1980s, but engaged in rapid political and economic transitions in the 1990s. Controlling for this experience reduces heterogeneity in the data. Democracy performs as expected, and the effect of communist is inclusive, which presumably reflects the diverse outcomes of the institutional reforms introduced in the 1990s.

Table 2: Summary Statistics: 5-year lag

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
GDP per capita growth	2,270	0.025	0.056	−0.456	0.002	0.048	0.661
Program participation	2,270	0.406	0.491	0	0	1	1
Gross capital formation	2,270	4.904	25.496	−81.772	−4.627	11.837	723.202
Change in labor force	2,270	0.021	0.018	−0.095	0.011	0.032	0.178
Openness	2,270	71.054	45.808	11.087	43.241	88.831	399.874
GDP per capita	2,270	10,584.830	14,733.750	132	1,220.7	13,242.3	97,678
Democracy	2,270	0.237	0.144	0	0.1	0.3	1
Communist	2,270	0.621	7.988	−141.308	−0.260	0.510	204.143
Voting with the U.S.	2,270	0.636	0.481	0	0	1	1
Balance of payments	2,270	0.213	0.409	0	0	0	1
IMF member	2,270	0.004	0.059	0	0	0	1
UN votes	2,709	0.007	0.018	0.000	0.001	0.005	0.173
Low income states	2,709	-148.35	238.41	-712.43	-325.56	79.56	329.67

We complete the data set with a number of potential instruments for IMF program participation. First, a country‘s UN General Assembly voting pattern is a widely used instrument in the literature (Thacker, 1999; Barro and Lee, 2005; Steinwand and Stone, 2008). As the United States exerts informal influence over the IMF, countries that vote like the United States in the UNGA receive preferential treatment from the Fund. UN voting is an attractive instrument as it is unlikely that UN voting affects growth rates except through access to IMF loans. On the other hand, siding with the United States in the UN often brings with it benefits that might affect growth. Second, balance of payments (in billions of dollars) is negatively associated with IMF program participation. Deficits that are large relative to the world economy pose the danger of international contagion and systemic disruption, which creates strong incentives for the IMF to contain a crisis. Balance of payments is not conclusively linked to growth; a large deficit may be driven by capital inflows that boost domestic investment. Scaling the variable in absolute terms rather than normalizing it by GDP makes it less likely to influence growth, as growth does not depend on country size. Finally, we introduce two Bartik-style variables, UN vote share and low-income status. We discuss these variables at length in Section 6.

4 Empirical Analysis: baseline regressions

Our Solow growth model inspired estimating equation is Equation 1, where s is 1 or 5 depending on the lag.

$$\begin{aligned}
\text{GDP growth}_{it} = &\beta_1 * \text{Program participation}_{it} \\
&+ \beta_2 * \text{Gross capital formation}_{i,t-s} \\
&+ \beta_3 * \text{Change in labor force}_{i,t-s} \\
&+ \beta_4 * \text{Openness}_{i,t-s} \\
&+ \beta_5 * \text{GDP per capita}_{i,t-s} \\
&+ \beta_6 * \text{Democracy}_{i,t-s} \\
&+ \beta_7 * \text{Communist}_{i,t-s} \\
&+ \text{Country fixed effects}_i \\
&+ \text{Year fixed effects}_t \\
&+ \epsilon_{it} \qquad (1)
\end{aligned}$$

Theoretically, unit and time fixed effects make sense. Individual country intercepts control for the variation in development at the outset of our time series. Some states were wealthy, developed countries with well-functioning institutions and urbanized, highly-educated populations. These states also had effective health care systems and extensive infrastructure. Others states were poor with fragile institutions, few services, and were inhabited by rural populations dependent on subsistence agriculture. Some states were rich in natural resources, in human capital, or in capital stock; some had geographical locations that provided easy access to trade routes; some suffered from severe climate challenges. Our country fixed effects control for this country-specific, time-invariant variation.

Time fixed effects control for contemporaneous shocks that influence all countries. A major source of such shocks is the global macroeconomic cycles that emanate from changes in US monetary policy. The policy responses of other countries magnify these changes as they cascade through developed economies and then on to developing countries (Rey, 2016). Examples include the Latin American debt crisis, the expansions of the mid- 1990s and mid- 2000s, and the Great Recession that spread globally in 2009. Another source of contemporaneous shocks is the surge of foreign direct investment that began in the 1990s, and another is the shifts in commodity prices driven by fluctuations in Chinese demand. Technological shifts provide new efficiencies represented by global telecommunications, financial transfers, high-speed computing, and cell phones; and the construction of global supply chains both takes advantage of these changes and spreads them.

The results of our baseline estimations are in Tables 3 and 4. In Table 3, the Solow growth model covariates are lagged one year; in Table 4, they are lagged five years.[1] In each respective table, column one contains estimates from the pooled model. Column two includes only country fixed effect, and column three includes only year fixed effects. Column four contains the two-way model with both country and year fixed effects.[2] For both the 1-year and 5-years lags, the inclusion of country fixed effects increases the magnitude of the estimated effect of program participation. The effect of including country-specific intercepts is strong enough to change the sign on program participation when lagged 1-year.

[1] We multiply the dependent variable by 100 and divide GDP per capita by 1000 to ease interpretation.

[2] We estimate all fixed effects models using the "within" estimator.

Table 3: Baseline Regression Results: 1-year lag

	Dependent variable:			
	Percentage change in GDP			
	Pooled	Unit FE	Time FE	Unit and Time FE
Program participation	−0.354	−0.054	−0.285	0.255
	(0.237)	(0.279)	(0.234)	(0.278)
Gross capital formation	0.030***	0.020***	0.026***	0.019***
	(0.004)	(0.004)	(0.004)	(0.004)
Labor force change	−28.340***	7.457	−25.659***	11.977
	(6.423)	(8.197)	(6.298)	(8.069)
Openness	0.016***	0.062***	0.013***	0.050***
	(0.002)	(0.007)	(0.002)	(0.008)
GDP per capita	−0.014*	−0.101***	−0.018**	−0.244***
	(0.008)	(0.032)	(0.008)	(0.036)
Democracy	0.450*	0.070	0.268	−0.988**
	(0.242)	(0.464)	(0.238)	(0.480)
Communist	−0.020	−0.367	0.833	0.014
	(1.366)	(1.623)	(1.340)	(1.591)
Constant	1.351***			
	(0.334)			
Observations	2,709	2,709	2,709	2,709
R^2	0.050	0.038	0.035	0.039
Adjusted R^2	0.047	-0.016	0.024	-0.026
F Statistic	20.112***	14.640***	14.062***	14.657***

Note: *p<0.1; **p<0.05; ***p<0.01. We estimated all models in R using the plm package (Croissant and Millo, 2008).

The magnitude of the change when including individual intercepts in the model suggests that we need a formal assessment of the model. Testing for fixed versus random effects is complicated by the fact that the simple Hausman test is invalid if the individual intercepts are not independent and identically distributed (iid), which is likely given the expected heteroscedasticity of the data. Instead, we estimate an auxiliary regression suggested by Wooldridge (2010). If the effects are fixed, then the error term of the auxiliary regression is correlated with the regressors, which leads to the statistical significance of additional functions of the regressors. Cluster-robust standard errors are necessary as the error term of the regression is not asymptotically iid (Cameron and Trivedi, 2005). For both lags, the test rejects the null hypothesis of random effects.[3]

We also test for serial correlation and cross-sectional dependence (countries responding to common shocks or spatial dependence). Using fixed effects in a linear model induces negative serial correlation so standard tests routinely reject the null hypothesis of spherical errors. We use Wooldridge's (2010) test for short panels, which regresses the fixed effect residuals on a lagged version of themselves. The test fails to reject the null of no serial correlation with a p-value of 0.46 for the 1-year lag and 0.88 for the 5-year lag. We use Pesaran's (2015) test for cross-sectional dependence, which fails to reject the null of no cross-sectional dependence with p-values of 0.53 and 0.95.

[3] The χ^2-test statistics are 96.511 and 71.42, respectively. We also tested unit and time fixed effects independently from one another, and in each case, the test rejects the null hypothesis of random effects.

Table 4: Baseline Regression Results: 5-year lag

	Dependent variable:			
	Percentage change in GDP			
	Pooled	Unit FE	Time FE	Unit and Time FE
Program participation	0.436^{*}	0.788^{**}	0.236	0.720^{**}
	(0.263)	(0.309)	(0.261)	(0.308)
Gross capital formation	0.0005	-0.008^{*}	0.002	-0.006
	(0.005)	(0.004)	(0.004)	(0.004)
Labor force change	-45.521^{***}	-9.741	-41.823^{***}	-5.591
	(7.001)	(8.936)	(6.899)	(8.810)
Openness	0.013^{***}	-0.001	0.009^{***}	-0.019^{**}
	(0.003)	(0.009)	(0.003)	(0.009)
GDP per capita	-0.013	-0.041	-0.020^{**}	-0.204^{***}
	(0.010)	(0.042)	(0.009)	(0.048)
Democracy	0.115	-0.125	0.027	-0.746
	(0.272)	(0.552)	(0.267)	(0.557)
Communist	0.317	-0.283	0.804	-0.247
	(1.966)	(2.106)	(1.940)	(2.075)
Constant	2.388^{***}			
	(0.383)			
Observations	2,270	2,270	2,270	2,270
R^2	0.030	0.006	0.023	0.017
Adjusted R^2	0.027	-0.060	0.010	-0.060
F Statistic	9.938^{***}	1.869^{*}	7.404^{***}	5.159^{***}

Note: *p<0.1; **p<0.05; ***p<0.01. We estimated all models in R using the plm package (Croissant and Millo, 2008).

5 Sensitivity analysis

The collective results from our baseline regressions suggest that IMF program participation has a positive effect on GDP growth. Program participation is, without much doubt, endogenous. A selection model would address the endogeneity issue by including an estimate of the inverse Mill's ratio in our baseline regressions. Thus, we can see selection bias as omitted variable bias. The goal of our sensitivity analysis is understanding how much selection (omitted variable) bias is required to nullify the effect of having a program on percentage change in growth. Formal sensitivity analysis goes back to Cornfield et al. (1959) and was developed by Rosenbaum and Rubin (1983) and Rosenbaum (1988). The method we use comes from Oster (2017), who builds on Altonjii et al. (2005).

Consider a regression model

$$Y = \beta X + W_1 + W_2 + \epsilon,$$

where X is the scalar treatment, $W_1 = \Gamma\omega$, where ω is a vector of observed controls, and W_2 is unobserved. Let $R_{\max}$ be the R^2 from a hypothetical regression that includes the omitted variable(s), W_2. The proportional selection relationship is[4]

$$\delta\frac{\operatorname{cov}(W_1, X)}{\operatorname{var}(W_1)} = \frac{\operatorname{cov}(W_2, X)}{\operatorname{var}(W_2)}.$$

Oster (2017, 9-10) notes two types of robustness claims that we can

[4] "Omitted variable bias is proportional to coefficient movements, but only if such movements are scaled by movements in R-squared" (Oster, 2017, 3).

make using her method. The first is to assume a value for $R_{\max}$ and then calculate the value of δ for which $\beta = 0$. She argues that a value of "$\delta = 2$, for example, would suggest that the unobservables would need to be twice as important as the observables to produce a treatment effect of zero." The second approach is to use bounds on $R_{\max}$ and δ to develop a set of bounds for β, and then consider whether zero or some other value of interest falls in the bounds. We report results from both approaches in Table 5.

Column 1 of Table 5 lists our treatment variable, program participation, for both a 1-year and 5-year lag. Column 2 lists what Oster refers to as the uncontrolled estimates along with the R^2s associated with the regressions. Column 3 lists the controlled estimates along with their associated R^2s. Column 4 lists the identified sets, which are bounded by the controlled effect and the bias-adjusted effect based on $R_{\max}$ given in the table and $\delta = 1$. Column 5 lists the values of δ that would drive the respective effects to 0 given $R_{\max}$.

Table 5: Selection on Unobservables ($R_{\max} = 0.28, 0.30$).

Treatment Variables	Baseline Effect [R^2]	Controlled Effect [R^2]	Identified Set $R_{\max} = 0.28, 0.30$	δ for $\beta = 0$ Given $R_{\max}$
Program participation 1-year lag	0.208 [0.185]	0.255 [0.217]	[0.255, 0.355]	-3.08
Program participation 5-year lag	0.779 [0.219]	0.720 [0.230]	[0.720, 0.295]	1.587

Note:

Following Oster (2017, 3), we set $R_{\max} = 0.28$ for the 1-year lag and $R_{\max} = 0.30$ for the 5-year lag, which is 1.3 times the observed R^2 from a regression of GDP growth on the full set of covariates as in columns 4 of Tables 3 and 4.[5] Neither identifying set includes 0, and both sets demonstrate

[5]Oster suggests 1.3 because it is the value at which 90% of results from randomized trials are robust.

a reasonable consistency with our previous results. The identified set for the 1-year lag takes the baseline program participation estimate from Table 3 as its lower bound. The reason is that including our six covariates in the regression actually increases our estimate of the program participation effect. The difference between the uncontrolled and controlled also explains why δ is negative.

In the 5-year lag specification, the controlled effect is the upper bound of the identified set. Including the six covariates in the regression, in this case, lowers our program participation effect estimate. Correspondingly, δ is positive with a value of 1.6, which suggests that any omitted variables would have to be 1.6 times more important than the observed variables to drive the effect of program participation to zero. Altonjii et al. (2005) suggest that $\delta = 1$, which means that the observables are at least as important as the unobservables, may be an appropriate cutoff.

6 Instrumental variables analysis

The results above suggest that selection bias is not a major threat to our results. Our priors on the endogeneity of IMF program participation, however, are strong, and additional evidence regarding the participation effect is necessary. We take an instrumental variables (IV) approach to the endogeneity problem as matching methods do not control for unobservables and common selection models have proven severely model dependent and unstable (Cameron and Trivedi, 2005). The challenge with an IV approach is to find appropriate instruments that are both exogenous (given covariates, at least) and correlated with the endogenous variable. To that end, we employ a single instrument, multiple instruments, a nonlinear instrument, and finally two Bartik-style instruments.

Our goal is not to defend a single specification or a single set of assumptions concerning our instruments. Each of our instruments has its strong and weak points. We demonstrate, however, that there is remarkable consistency across these different instruments and specifications. The results that follow make it difficult to believe that IMF program participation has anything other than a positive effect.

Instrumenting for IMF program participation requires a variable that explains participation, but does not affect growth except through participation. We begin with a widely used instrument in the literature to set a baseline: a country‘s pattern of voting in the UN General Assembly (Thacker, 1999; Barro and Lee, 2005; Steinwand and Stone, 2008). The relationship between UNGA voting and program participation is straightforward. The United States exerts informal influence over the IMF, and countries that vote with the United States in the UN General Assembly may receive preferential treatment from the Fund. UN voting, so the argument goes, is unlikely to affect growth rates except through access to IMF loans.

We use two-stage least squares and the "within" estimator. Our first stage estimating equation is equation 2, where program participation is a linear function of UN voting and the same set of covariates from our baseline regression including country and year fixed effects.

$$
\begin{aligned}
\text{Program participation}_{i,t-s} = &\gamma_1 * \text{Voting with the US}_{i,t-s} \\
&+ \gamma_2 * \text{Gross capital formation}_{i,t-s} \\
&+ \gamma_3 * \text{Change in labor force}_{i,t-s} \\
&+ \gamma_4 * \text{Openness}_{i,t-s} \\
&+ \gamma_5 * \text{GDP per capita}_{i,t-s} \\
&+ \gamma_6 * \text{Democracy}_{it} \\
&+ \gamma_7 * \text{Communist}_{it} \\
&+ \text{Country fixed effects}_i \\
&+ \text{Year fixed effects}_t \\
&+ u_{it} \qquad (2)
\end{aligned}
$$

Researchers faced with a binary endogenous variable, such as IMF program participation, are tempted to use a generalized linear model, such as probit or logit, for the first stage of two-stage least squares. Hausman referred to this specification as a forbidden regression in 1975 (Angrist and Pischke, 2008). Only a linear first-stage regression guarantees first-stage residuals that are uncorrelated with fitted values and the other covariates. The correct specification for a binary endogenous variable, then, is the linear probability model in the first stage. We use a nonlinear instrument later in the paper, but the regression, itself, remains linear.

Our results are in Table 6, where column 1 contains the 1-year lag and column 2 contains the 5-year lag. Our estimated coefficients are positive

Table 6: IV Regressions: Voting with the US

	Dependent variable:	
	Percentage change in GDP	
	1-year lag	5-year lag
Program participation	7.163*	6.230*
	(4.294)	(3.257)
Gross capital formation	0.023***	−0.002
	(0.006)	(0.005)
Labor force change	3.256	−10.306
	(10.498)	(9.853)
Openness	0.052***	−0.021**
	(0.009)	(0.010)
GDP per capita	−0.236***	−0.163***
	(0.041)	(0.056)
Democracy	−1.121**	−1.024*
	(0.541)	(0.620)
Communist	5.546	2.635
	(3.861)	(2.798)
Observations	2,709	2,270
Note:	*p<0.1; **p<0.05; ***p<0.01	

and marginally significant. A standard Wald test performed using a cluster-robust variance/covariance matrix returns an F-statistic of 6.2 for the 1-year lag model and 9.3 for the 5-year lag model. These results indicate some weakness in UN voting as an instrument.

The presence of a weak instrument means that we cannot trust the significance tests reported in Table 6. A solution to this problem is to use a test that is unconditionally valid; that is, the test has correct size even when the instrument is weak. One such test was proposed by Anderson and Rubin (1949) and consists of regressing the residuals from baseline regression (equation 1) on the instrument and covariates (equation 2). The Anderson-Rubin test is equivalent to the better known J-test (Davidson and MacKinnon, 1993). Full results are in the Appendix in Table 10 and indicate that the effect remains positive and is marginally significant.

6.1 Multiple instruments

Here we use two instruments for IMF program participation: UN voting and balance of payments deficit. Our first stage estimating equation is equation 3, where program participation is a linear function of UN voting, balance of payments deficit, and the covariates from our baseline regression including country and year fixed effects.

$$
\begin{aligned}
\text{Program participation}_{i,t-s} = {} & \delta_1 * \text{Voting with the US}_{i,t-s} \\
& + \delta_2 * \text{Balance of payments}_{i,t-s} \\
& + \delta_3 * \text{Gross capital formation}_{i,t-s} \\
& + \delta_4 * \text{Change in labor force}_{i,t-s} \\
& + \delta_5 * \text{Openness}_{i,t-s} \\
& + \delta_6 * \text{GDP per capita}_{i,t-s} \\
& + \delta_7 * \text{Democracy}_{it} \\
& + \delta_8 * \text{Communist}_{it} \\
& + \text{Country fixed effects}_i \\
& + \text{Year fixed effects}_t \\
& + \nu_{it} \qquad (3)
\end{aligned}
$$

The results are in Table 7. The estimated coefficients remain positive, and the coefficient on the 5-year lag is marginally significant. Weak instrument testing now requires the use of the Cragg-Donald statistic, which allows a test of multiple instruments (Cragg and Donald, 1993). Critical values of the statistic are given by Stock and Yogo (2005). The test statistics are 7.7 and 11.8 for the 1-year lag and 5-year lag, respectively, which means that we cannot reject the null hypothesis that these instruments are weak (though they do not miss significance by much).[6] Anderson-Rubin re-

[6]The test assumes homoscedasticity, but what test to run under heterscedasticity is an open question.

sults are in Table 11, and we see that UN voting is positive and marginally significant, and balance of payments is not. The robust Wald statistics for these instruments are over 10 for both sets of lags.

6.2 Nonlinear instrument

In our third set of instrumental variables regression, we use a nonlinear instrument. The idea of a nonlinear instrument goes back to Kelejian (1971) and has been revived by Bun and Harrison (2018). The basic idea is to form internal instruments and avoid the use of external instruments. Credible inferences can be therefore be made without a traditional exclusion restriction. Let z_i be an instrument and let w_i and q_i be exogenous variables. The instrument is formed using the second-order polynomial,

$$z_i = \begin{bmatrix} w_i^2 & q_i^2 & w_i * q_i & w_i^2 * q_i & w_i * q_i^2 \end{bmatrix}'.$$

Note that while the first-stage regression using z_i is nonlinear in the parameters, the regression remains linear in terms of its functional form. Thus, we avoid the "forbidden" regression and achieve the proper IV estimates in the second stage. Instruments formed in this way are sometimes weak, and the procedure needs to be combined with robust weak instrument inference as above (Bun and Harrison, 2018).

Table 7: IV Regressions: Voting with the US and Balance of payments

	Dependent variable:	
	Percent change in GDP	
	1-year lag	5-year lag
Program participation	4.809	6.210*
	(3.878)	(3.256)
Gross capital formation	0.022***	−0.002
	(0.005)	(0.005)
Labor force change	6.227	−10.288
	(9.790)	(9.848)
Openness	0.052***	−0.021**
	(0.008)	(0.010)
GDP per capita	−0.239***	−0.164***
	(0.038)	(0.056)
Democracy	−1.076**	−1.023*
	(0.510)	(0.619)
Communist	3.661	2.624
	(3.520)	(2.797)
Observations	2,709	2,270
Note:	*p<0.1; **p<0.05; ***p<0.01	

$$
\begin{aligned}
\text{Program participation}_{i,t-s} = {} & \eta_1 * \text{Gross capital foramation}^2_{i,t-s} \\
& + \eta_2 * \text{Change in labor force}^2_{i,t-s} \\
& + \eta_3 * \text{Openness}^2_{i,t-s} \\
& + \eta_4 * \text{GDP per capita}^2_{i,t-s} \\
& + \eta_5 * \text{Gross capital formation}_{i,t-s} \\
& + \eta_6 * \text{Change in labor force}_{i,t-s} \\
& + \eta_7 * \text{Openness}_{i,t-s} \\
& + \eta_8 * \text{GDP per capita}_{i,t-s} \\
& + \eta_9 * \text{Democracy}_{it} \\
& + \eta_{10} * \text{Communist}_{it} \\
& + \text{Country fixed effects}_i \\
& + \text{Year fixed effects}_t \\
& + \omega_{it}
\end{aligned}
\tag{4}
$$

We use the squares of our continuous exogenous covariates (gross capital formation, change in labor force, openness, and GDP per capita) to form our nonlinear instrument. We omit the cross-product terms to economize on the number of instruments. The first-stage is in equation 4. The results, which are in Table 8, are positive and signfiicant. The Cragg-Donald statistic again suggests that these instruments are somewhat weak (values of 8.5 and 8.6). Anderson-Rubin results are in Table 12, and we see that our nonlinear instruments are mostly significant for the 1-year lag and not for the 5-year lag. However, the robust Wald statistics for these sets of instruments are significant for both sets of lags.

6.3 Bartik instruments

Finally, we turn to our Bartik-style instruments (Bartik, 1991; Goldsmith-Pinkham et al., 2018), sometimes known as shift-share instruments. Bartik estimated labor supply elasticity using two identifies: labor supply equals labor demand in equilibrium and labor demand can be decomposed by sectors and by geographical units. Labor supply shocks are sector-specific and distributed across the units. Therefore, aggregate over-time shocks affect units differently depending upon the concentration of industries in each unit. Interacting national time-series data for employment by sector with cross-sectional data for employment in each sector at $t = 0$ yields an instrument for variations in employment demand in each unit. An attractive feature of this strategy is that the instrument is generally valid if we control for unit and time fixed effects, and in that case, the identification depends only on the differential responses of units to common shocks.

For our application, we construct shift-share instruments based on two distinct intuitions about how shifts over time in the supply and demand for IMF financing differentially affect countries with different starting points. In both cases, the intuition depends on the notion that IMF financing is easier to obtain during years when demand for IMF financing is low, and therefore a larger number of marginal financing candidates are accepted (Vreeland, 2003). For example, there was a serious shortfall in the IMF budget during

Table 8: IV Regressions: Nonlinear instrument

	Dependent variable:	
	Percentage change in GDP	
	1-year lag	5-year lag
Program participation	6.657**	5.426**
	(2.726)	(2.653)
Gross capital formation	0.023***	−0.002
	(0.005)	(0.005)
Labor force change	3.895	−9.617
	(9.509)	(9.554)
Openness	0.052***	−0.021**
	(0.009)	(0.010)
GDP per capita	−0.236***	−0.169***
	(0.040)	(0.054)
Democracy	−1.111**	−0.983
	(0.530)	(0.602)
Communist	5.141*	2.214
	(2.787)	(2.584)
Observations	2,709	2,270
Note:	*p<0.1; **p<0.05; ***p<0.01	

the lull in IMF program participation that preceded the 2008 financial crisis (the Funds income comes from the interest paid by borrowers). IMF staff joked at the time that it was time to "find another Turkey" because Turkey was one of few remaining participants (Stone, 2011). If shocks to the global supply and demand for emergency financing affect the criteria for participation, the participation of similarly-situated countries in IMF programs at different points in time can be regarded as caused by the exogenous over-time shocks. This exclusion criterion is violated, however, if the variable of interest is correlated with the over-time shocks, which is likely the case. Financial crises are driven by the global financial cycle, which in turn is driven by US macroeconomic policy(Rey, 2016). In addition, crises tend to be correlated because of contagion in international financial markets, leading to the waves of financial crises that have spread across Latin America, East Asia, and the Eurozone. Consequently, we cannot rely on exogenous variation in demand to identify IMF program participation. However, instruments that predict differential responses of particular countries to the shocks are valid if we also control for time fixed effects.

The first shift-share instrument is based on the informal politics surrounding the allocation of IMF financing and voting power in the IMF. The IMF has a system of weighted voting, where vote shares depend on contributions to the IMFs resources, or quotas. The formulas that determine the quota sizes depend on GDP, foreign reserves, and foreign trade.[7] Vote shares

[7]Quotas are adjusted infrequently when overall quotas are revised. The original quota formula was designed to generate a politically acceptable distribution of quotas at Bretton Woods, and the formulas have been revised over time as a result of bargaining. European countries tend to be over-represented compared to their shares of global GDP, and emerging market countries tend to be underrepresented.

are important because they underpin formal and informal influence within the institution. Members with sufficient voting power can appoint a member to the Funds governing Executive Board, and those with lower vote shares join coalitions to elect members that represent them collectively. A larger vote share makes ones support more valuable, and guarantees greater attention from ones representative. Other things equal, the IMF staff is likely to be more responsive to countries that enjoy larger shares of voting power. Voting power may not be a valid instrument, however, because GDP, foreign reserves, and foreign trade are likely to be correlated with economic growth. That is, countries that perform well economically increase their voting power over time. Consequently, to construct our instrument, we computed the vote share of each country in 1976, or at the time of membership if the country was not a member in 1976, or at the time when membership was restored, if the countrys membership lapsed. This variable does not vary over time in our dataset, and ranges from .006 to .214. We interact *vote share* with *available funds*, which we calculate as the ratio of untapped funds (total quotas minus outstanding use of Fund resources) to total quotas. This variable does not vary cross-sectionally in our dataset, and ranges from .54 to 1. Controlling for country and year fixed effects, the instrument captures the degree to which the probability of program participation with respect to availability of funds depends on a countrys vote share.

Our second shift-share instrument relies on the fact that the IMF has special lending facilities that are designed for low-income countries, which provide financing at more attractive, concessional terms, for longer maturities, and which generally impose less burdensome conditions on their recipients. All IMF members are eligible to borrow from the General Fund, but eligibility to borrow from these special facilities depends on being below a threshold level of per capita income, which is established each year by the World Bank. (Countries categorized as low-income or lower-middle income by the World Bank are eligible.) Since any country that is eligible to borrow from the IMF would find it more attractive to borrow from the low-income facilities, eligibility should make participation in IMF programs more attractive. These resources are limited by the budget for subsidizing the interest rate, however, so competition for funds depends on the number of countries classified as low- or lower-middle income. Income classification would violate the exclusion restriction because countries that performed well economically would increase in per capita income and graduate from eligibility. Consequently, to create our instrument we interact a measure of a countrys income relative to the IMF membership in 1976 or upon joining the IMF with the number of IMF members. *Relative income* is calculated as the natural logarithm of the ratio of per capita GDP to the average per capita GDP of all IMF members. Since this is coded for each country at a fixed point in time, it does not vary within countries over time, and ranges from -3.9 to 2.5. The number of members does not vary cross-sectionally, and ranges during our data window from 129 to 187. The interaction of *relative income* and *number of members* is correlated with the degree to which eligibility for low-income facilities increases demand to participate, but controlling for time and country fixed effects guarantees that the exclusion restriction holds by eliminating any time- or country-level effects that might be correlated with growth.

Our first stage estimating equation is equation 5, where program participation is a linear function of the interaction of our Bartik instruments (UN vote vote shares and low-income states) and the same set of covariates from our baseline regression including country and year fixed effects.

$$\begin{aligned}
\text{Program participation}_{i,t-s} = & \zeta_1 * \text{Low income}_{i,t-s} \\
& + \zeta_2 * \text{Gross capital formation}_{i,t-s} \\
& + \zeta_3 * \text{Change in labor force}_{i,t-s} \\
& + \zeta_4 * \text{Openness}_{i,t-s} \\
& + \zeta_5 * \text{GDP per capita}_{i,t-s} \\
& + \zeta_6 * \text{Democracy}_{it} \\
& + \zeta_7 * \text{Communist}_{it} \\
& + \text{Country fixed effects}_i \\
& + \text{Year fixed effects}_t \\
& + \epsilon_{it} \qquad (5)
\end{aligned}$$

Our results are in Table 9, where column 1 contains the 1-year lag and column 2 contains the 5-year lag. The estimated coefficient on the 1-year lag Bartik instrument is negligable and insignificant. For the 5-year lag, the estimated coefficient is positive, of the same magnitude as our previous results, and marginally significant.

Table 9: IV Regressions: Bartik instrument (Low income)

	Dependent variable:	
	Percentage change in GDP	
	1-year lag	5-year lag
Program participation	−0.8	6.66*
	(0.095)	(0.031)
Gross capital formation	0.000**	0.000
	(0.000)	(0.000)
Labor force change	0.099	−0.109
	(0.121)	(0.102)
Openness	0.0005***	−0.0002*
	(0.000)	(0.000)
GDP per capita	0.000***	0.000**
	(0.000)	(0.000)
Democracy	−0.009	−0.012
	(0.005)	(0.007)
Communist	−0.006	0.027
	(0.077)	(0.026)
Observations	2,046	2,046
Note:	*p<0.1; **p<0.05; ***p<0.01	

7 Discussion

The effects of IMF programs on economic growth remain controversial despite more than thirty years of scholarly attention. Meanwhile, policymakers continue to behave as if IMF programs effects are well understood and known to be positive. Capital markets appear to be unconvinced; some IMF programs rapidly restore confidence and lead to capital inflows, while others accelerate capital flight (Chapman et al., 2017).

We find compelling evidence that the average effect of an IMF program on per capita GDP growth is indeed positive, and the effect holds after one year and after five years. Diagnostics indicate that pooled growth equations require unit and time fixed effects, and the difference between our results and many other published results is likely due to our use of fixed effects. We interpret the one-year effect on output as capturing the effect of an emergency balance-of-payments loan on macroeconomic variables, and the five-year effect as representing the contribution to growth of efficiency-enhancing reforms. Across a variety of specifications, we consistently find positive coefficients for both effects. Sensitivity analysis indicates that our positive baseline coefficient estimates are reliable. However, we find substantial support for the claim in the literature that estimates of IMF program effects on growth are biased downwards by endogeneity. Instrumental-variables regressions estimate effects that are an order of magnitude larger than our baseline estimates.

We use UN voting agreement with the United States, the balance of payments in billions of dollars, and a non-linear "internal instrument constructed from our exogenous variables as instruments. In the presence of unit and time fixed effects, each of these specifications involves borderline weak instruments. However, the results are broadly consistent across these specifications in terms of the magnitude of the estimated effects, and the results using a non-linear instrument are strongly significant. The effects are substantial, ranging from 4.8 to 7.2 percentage points of additional growth for participating in an IMF program in the previous year and from 5.4 to 6.2 percentage points of additional growth for participating in an IMF program five years ago.

The implication of our findings is that governments that participate in IMF programs are not deluded in their hope that a combination of emergency financing and unpalatable economic reforms will improve their expected growth trajectories. Participating in IMF programs will not allow poor, land-locked, underdeveloped countries to escape their circumstances, but they should take their circumstances into account when they evaluate the effects. The countries that are most likely to participate are the ones that are least likely to grow rapidly under a program or without one, so their growth prospects remain relatively limited. Countries should understand this mechanism when they evaluate their experiences with the IMF. On average, countries that choose to participate in IMF programs grow faster than similarly situated countries that choose not to participate.

Appendix

Table 10: Anderson-Rubin Test

	Dependent variable:	
	Percentage change in GDP residuals	
	1-year lag	5-year lag
UN voting	4.502*	4.832*
	(2.508)	(2.660)
Gross capital formation	−0.0003	−0.0003
	(0.004)	(0.004)
Labor force change	−0.564	−0.278
	(8.063)	(8.801)
Openness	−0.001	−0.002
	(0.008)	(0.009)
GDP per capita	−0.00000	0.00000
	(0.00004)	(0.00005)
Democracy	−0.023	−0.038
	(0.479)	(0.557)
Communist	1.080	0.918
	(1.685)	(2.128)
Observations	2,709	2,270
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 11: Anderson-Rubin Test with Two Instruments

	Dependent variable:	
	Percentage change in GDP residuals	
	1-year lag	5-year lag
UN voting	4.643*	4.787*
	(2.509)	(2.661)
Balance of payments	0.017*	−0.008
	(0.010)	(0.015)
Gross capital formation	−0.0005	−0.0003
	(0.004)	(0.004)
Labor force change	−0.287	−0.363
	(8.062)	(8.804)
Openness	−0.001	−0.002
	(0.008)	(0.009)
GDP per capita	−0.00000	0.00000
	(0.00004)	(0.00005)
Democracy	−0.018	−0.037
	(0.479)	(0.557)
Communist	1.095	0.916
	(1.685)	(2.129)
Observations	2,709	2,270
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 12: Anderson-Rubin Test with Nonlinear Instrument

	Dependent variable:	
	Percentage change in GDP residuals	
	1-year lag	5-year lag
Gross capital formation^2	-0.00004^{***}	0.00001
	(0.00001)	(0.00001)
Labor force change^2	-6.104	-62.982
	(137.561)	(144.543)
Openness^2	-0.0001^{***}	0.00001
	(0.00004)	(0.0001)
GDP per capita^2	0.002^{***}	0.003^{***}
	(0.001)	(0.001)
Gross capital formation	0.010^{*}	-0.003
	(0.006)	(0.006)
Labor force change	0.595	3.900
	(10.200)	(11.200)
Openness	0.037^{***}	-0.0001
	(0.014)	(0.017)
GDP per capita	-0.0002^{**}	-0.0003^{**}
	(0.0001)	(0.0001)
Democracy	-0.113	-0.057
	(0.478)	(0.557)
Communist	0.148	0.055
	(1.569)	(2.069)
Observations	2,709	2,270
Note:		*p<0.1; **p<0.05; ***p<0.01

References

Altonjii, J. G., T. E. Elder, and C. R. Taber (2005, Autumn). An evaluation of instrumental variable strategies for estimating the effects of catholic schooling. *Journal of Human Resources 40*(4), 791–821.

Anderson, T. W. and H. Rubin (1949). Estimation of parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics 20*(1), 46–63.

Angrist, J. D. and J.-S. Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

Barro, R. and J. Lee (2005, October). Mf programs: who is chosen and what are the effects? *Journal of Monetary Economics 52*(7), 1245–1269.

Bartik, T. J. (1991). Who benefits from state and local economic development policies? Technical report, W.E. Upjohn Institute.

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

Bun, M. J. G. and T. D. Harrison (2018). Ols and iv estimation of regression models including endogenous interaction terms. *Econometric Reviews*, 1–14. DOI: 10.1080/07474938.2018.1427486.

Cameron, A. C. and P. K. Trivedi (2005). *Microeconomics: Methods and Applications.* New York: Cambridge University Press.

Chapman, T., S. Fang, X. Li, and R. W. Stone (2017). Mixed signals: Imf lending and capital markets. *British Journal of Political Science 47*(2), 329–349.

Cheibub, J. A., J. Gandhi, and J. R. Vreeland (2010, April). Democracy and dictatorship revisited. *Public Choice 143*(1-2), 67–101.

Cornfield, J., W. Haenszel, E. C. Hammond, A. M. Lilienfeld, M. B. Shimkin, and E. L. Wynder (1959, January). Smoking and lung cancer: Recent evidence and a discussion of some questions. *Journal of the National Cancer Institute 22*(1), 173–203.

Cragg, J. G. and S. G. Donald (1993, June). Testing identifiability and specification in instrumental variable models. *Econometric Theory 9*(2), 222–240.

Croissant, Y. and G. Millo (2008). Panel data econometrics in R: The plm package. *Journal of Statistical Software 27*(2), 1–43.

Davidson, R. and J. G. MacKinnon (1993). *Estimation and Inference in Econometrics*. Oxford University Press.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2018, July). Bartik instruments: What, when, why, and how. Technical Report 24408, NBER Working Paper.

Goldstein, M. and P. Montiel (1986, June). Evaluating fund stabilization programs with multi-country data: Some methodological pitfalls. *IMF Staff Papers 33*(2), 304344.

Kelejian, H. H. (1971). Two-stage least squares and econometric systems linear in parameters but nonlinear in the endogenous variables. *Journal of the American Statistical Association 66*(334), 373–374.

Oster, E. (2017). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business and Economic Statistics 0*(0), 1–18.

Pesaran, M. (2015). Testing weak cross-sectional dependence in large panels. *Econometric Reviews 34*(6-10), 1089–1117.

Przeworski, A., M. E. Alvarez, J. A. Cheibub, and F. Limongi (2000). *Democracy and Development: Political Institutions and Well-being in the World, 1950-1990*. Cambridge University Press.

Rey, H. (2016). International channels of transmission of monetary policy and the mundellian trilemma. *NBER Working Paper* (21852).

Rosenbaum, P. and D. Rubin (1983). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society, Series B 45*(2), 212–218.

Rosenbaum, P. R. (1988, September). Sensitivity analysis for matching with multiple controls. *Biometrika 75*(3), 577–581.

Solow, R. M. (1956, February). A contribution to the theory of economic growth. *Quarterly Journal of Economics 70*(1), 65–94.

Steinwand, M. and R. W. Stone (2008, June). The international monetary fund: A review of the recent evidence. *The Review of International Organizations 3*(2), 123–149.

Stock, J. and M. Yogo (2005). *Testing for Weak Instruments in Linear IV Regression*, pp. 80–108. New York: Cambridge University Press.

Stone, R. W. (2011). *Controlling Institutions: International Organizations and the Global Economy*. New York: Cambridge University Press.

Thacker, S. C. (1999). The high politics of imf lending. *World Politics 52*(1), 38–75.

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

Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data* (2 ed.). Boston: MIT Press.