

World Bank Policy Lending and the Quality of Public Sector Governance

Lodewijk Smets
Stephen Knack



WORLD BANK GROUP

Development Research Group

Human Development and Public Services Team

May 2015

Abstract

This study investigates the impact of World Bank development policy lending for public sector governance on the quality of public sector management and institutions. The World Bank's Country Policy and Institutional Assessments (CPIA) are used to measure the latter, the study considers only policy conditions targeted at improvements in those areas. The analysis uses a comprehensive country-year panel data set of aid receiving-countries and finds a significant and inverse U-shaped effect of public sector conditions on the quality of public sector governance. For most observed values in the data, the impact is positive, but it turns negative beyond a value of 80 conditions. At that point, the predicted CPIA score is about 0.25 point (0.3 standard deviation) higher than with zero conditions. For most observations, the number of cumulative conditions is below

80, so the estimated effect of more conditions is generally positive. The analysis corrects for potential endogeneity and shows that the results are robust to sample restrictions, the use of an alternative governance measure, and the inclusion of an extended set of control variables. Falsification tests are also consistent with a causal interpretation from conditions to quality of public sector governance. The paper shows that conditions related to public financial management and tax reforms are more effective than those related to anti-corruption or civil service and administrative reform, where progress requires changing the behavior of a larger set of "deconcentrated" actors. The paper concludes by describing some innovative ideas in the Bank's ambitious new public sector management strategy that could improve the effectiveness of its support for public sector governance reform.

This paper is a product of the Human Development and Public Services Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at sknack@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

World Bank Policy Lending and the Quality of Public Sector Governance

Lodewijk Smets^{a,*}, Stephen Knack^b

^a*KU Leuven, LICOS Centre for Institutions and Economic Performance, Belgium*

^b*World Bank, Washington DC*

Keywords: policy lending, public sector governance, Aid effectiveness, World Bank

JEL codes: O10, O19

1. Introduction

Since 1980 the World Bank has been providing conditional financing to recipient governments to support specific policy and institutional reforms. Providing such policy support has become an important component in the financing of development operations. For example, during 2009-2012 World Bank policy lending reached around 45 billion USD. Furthermore, in contrast to early adjustment lending, the World Bank development policy loans (DPLs) now seek to improve policy in many different sectors, from agricultural policy to public sector governance (see table 1).

Although there is an extensive empirical literature evaluating the effects of adjustment lending on growth and macroeconomic stability,¹ only a limited number of studies investigate the impact of donor support on a recipient country's public sector. In this study we add to this line of research by examining the association of conditions pertaining to public sector reform in World Bank DPLs with the quality of public sector governance (PSG).

*Corresponding author

Email addresses: lode.smets@kuleuven.be (Lodewijk Smets), sknack@worldbank.org (Stephen Knack)

¹See Smets and Knack (2014) for a recent survey of that literature.

The main dependent variable in our study, from the World Bank’s Country Policy and Institutional Assessment (CPIA), measures what World Bank country teams are attempting to achieve when they design public sector reform conditions to include in DPLs, i.e., an increase in the quality of public sector management. In contrast with the existing literature, we consider the number of cumulative conditions related to public sector governance in DPLs as the main variable of interest. We look at the accumulation of conditions since we believe supporting policy reform in the public sector is a multistage and long term process (see, e.g., Pritchett and de Weijer, 2010).

Results from panel estimations reveal a significant quadratic relationship. These findings are robust to sample restrictions, additional controls, the use of alternative measures of public sector governance, and correction for endogeneity with system GMM. Further evidence is provided by instrumenting our variable of interest in a cross-sectional design. Falsification tests are also consistent with a causal interpretation from conditions to quality of public sector governance. Although results are statistically significant, the impact of PSG conditionality is modest. For instance, the predicted turning point in our preferred GMM specification is at 80 conditions, where the CPIA score is increased by about 0.247 points (or 0.3 standard deviations) higher on average, compared to the case of 0 conditions.

Next, we replicate the analyses after disaggregating the conditions and CPIA indicators by sub-sector. There are four CPIA indicators on public sector governance: (1) quality of budgetary and financial management, (2) efficiency of revenue mobilization, (3) quality of public administration and (4) transparency, accountability and corruption in the public sector. The PSG-related conditions in DPLs can be classified along similar lines. These disaggregated tests find that conditions are effective in supporting public financial management (PFM) and revenue systems, but not for promoting accountability and combating corruption. For quality of public administration, there is a significant quadratic relationship in our preferred GMM specification, but it is not robust to changes in sample

or specification.

The remainder of this paper is structured as follows. In the next section we discuss the related empirical literature. In section 3 we review the data and elaborate on the methodology used. Section 4 presents the empirical findings and section 5 concludes.

2. Related Literature

In its early years adjustment lending mainly emphasized economic stabilization and correction of balance of payments distortions (Kapur et al., 1997). At the beginning of the 1990s however, more emphasis was put on protecting the poor from the adverse effects of adjustment (Dreher, 2002). Furthermore, the development community came to realize that social, political and economic institutions matter for sustained implementation of sound macroeconomic policies, economic growth and poverty reduction (see, e.g., World Bank, 1998). Reflecting these concerns, the Bank's adjustment operations changed along a number of dimensions with an increased attention to public sector governance (PSG). In the 1980s only 24 percent of all conditions were related to public sector governance. By fiscal year 2007, the share of conditionality aimed at improving the public sector had doubled to 50 percent (see figure 1).²

The available empirical evidence seems to indicate however that supporting reforms in the public sector is difficult and success uneven. Cashel-Cordo and Craig (1990) provide an early analysis covering the period 1975-1980. Estimating a fixed-effects model with lagged independent variables, the authors show that the impact of aid on tax collection varies considerably among donors. For instance, Cashel-Cordo and Craig (1990) find that IMF lending reduces government revenues, while loans and grants from bilateral donors do not influence government behavior. Analyzing cross-country data, Knack (2001) finds

²Other donors are also heavily involved in supporting public sector reform. For instance, Andrews (2013) notes that, for the period 2004-2010, more than half of the projects carried out by Britain's Department for International Development (DfID) contain a PSG component.

that higher levels of official development assistance (ODA) diminish the quality of public sector governance – as measured by an index of governance indicators from the International Country Risk Guide (ICRG).³ His findings are robust to several sample restrictions, alternative model specifications and correction for endogeneity with 2SLS. Focusing on Sub-Saharan Africa, Brautigam and Knack (2004) repeat this exercise and come to similar conclusions. Furthermore, they find that higher aid levels are associated with a lower effort in tax collection. Alesina and Weder (2002) analyze the effect of aid on corruption. Employing OLS on a panel of 84 countries for the period 1984-1995, Alesina and Weder (2002) find tentative evidence that aid increases corruption. They relate their findings to the “voracity effect”, i.e., the tendency of foreign aid to foster corruption by increasing the size of resources available (see Svensson, 2000). However, the authors note that they ‘cannot fully resolve the question of causality in the relationship between changes in aid and corruption [...] Therefore, these results on the dynamic relationship between aid and corruption have to be taken very cautiously’ (Alesina and Weder, 2002, p. 1136). Estimating an error-correction model on a dataset covering the period 1970-1999, Remmer (2004) finds partial evidence that foreign aid reduces the revenue generating effort of low and middle-income countries. Using fixed effects analysis and IV estimation, Rajan and Subramanian (2007) find that, during the period 1981-1990, foreign aid slowed down industry growth in sectors that rely on contract enforcement and government regulation. The authors argue that aid reduces the incentive for governments to invest in rule of law and good governance. In a cross-country regression, de Renzio et al. (2011) investigate the impact of donor support for PFM reform on quality of PFM systems, as measured by Public Expenditure and Financial Accountability (PEFA) assessments. Running OLS on a dataset of 93 countries, the authors find a significantly positive, but economically

³The index from ICRG is an 18-point scale, created by summing the following three six-point scales: corruption in government, bureaucratic quality, and the rule of law.

small effect of donor support for PFM reform on PEFA scores, a result robust to different estimation methods and several sample restrictions. de Renzio et al. (2011) acknowledge that causation could run in both directions, but report that attempts to find “adequate instrumental variables” were unsuccessful. Therefore the authors note that their findings should be treated with caution.

A few studies specifically focus on World Bank support. A large-scale evaluation of PSG lending during 1999-2006 by IEG (2008) – combining a with-without approach, comparison to objectives, before-after analysis with case study findings – suggests that the Bank has been successful in supporting changes in public financial management and tax administration, without much progress however in controlling corruption and improving the functioning of civil service. Cruz and Keefer (2013) show that World Bank public sector reform projects are more likely to be rated as successful in countries with more programmatic political parties. Their explanation is that politicians in those countries have stronger incentives to pursue public policies that require well-functioning public administrations, and therefore to assist rather than block the Bank’s reform efforts. Blum (2014) similarly investigates the country and project-level factors associated with higher IEG ratings for PSG projects, finding that they perform better in more democratic and aid-dependent countries. Moloney (2009) estimates the impact of public sector projects on the quality of governance – as measured by Kaufmann et al. (2006) – covering the period 1996-2005. Controlling for GDP per capita and aid over GNI in a random effects model, the author claims that World Bank projects have a positive effect on regulatory quality, but a negative impact on control of corruption. Moloney (2009) does not correct for endogeneity, however, and acknowledges that PSG projects may target the most problematic countries. Moreover, due to limited availability of the CPIA data we use in this study, her study relied on other governance indicators “that the Bank itself does not use . . . to evaluate its clients.” In all of these studies the majority of PSG projects considered are

investment loans, not DPLs; Blum (2014) excludes DPLs entirely from his sample.

Our study contributes to this literature in several important ways. First, the majority of the abovementioned studies do not convincingly identify causal effects of aid on public sector performance. We attempt to correct for endogeneity through the use of instrumental variables techniques, i.e., system GMM and cross-sectional IV. Second, instead of considering aggregate measures such as total aid over GDP and overall public spending, our analysis provides a closer match between dependent and independent variables. Namely, our dependent variables measure more narrowly what World Bank teams are attempting to achieve when they include public sector conditions in DPLs, an improvement in the quality of public sector management and institutions. In this respect, our study is similar in spirit to Clemens et al. (2012), who show that aid's estimated impact on short-run growth strengthens when humanitarian and other components of aid are excluded that are not intended to further short-run growth. Third, we analyze the impact of conditions on performance for several PSG subcomponents – budgetary management, revenue mobilization, public administration, and transparency, accountability and corruption. Finally, we consider only PSG conditionality in DPLs, while estimates from other studies reflect a mixture of DPLs and investment loans with public sector components.

3. Data and Methodology

In its first 30 years of policy lending (1980-2010) the World Bank extended 1,002 DPLs⁴ to 124 different countries. Table 1 indicates that public sector governance was the primary sector targeted in 13 % of these loans. It is important to note, however, that these loans generally include some conditions pertaining to reforms in other sectors; similarly, the other 87 % typically include some PSG conditions. We use cumulative PSG conditions, rather than PSG loans, as our key independent variable because they potentially affect

⁴We have counted DPLs that became effective in 1980 or later.

the quality of public sector governance even when they are attached to a loan that mostly targets another sector.⁵

There is considerable variation in PSG lending across countries as figure 2 shows. It presents the distribution of the total number of cumulative PSG-related conditions countries received. Argentina easily tops the list with a total with a total of 210 conditions related to public sector governance, in large part due to the World Bank’s involvement in the large-scale structural reforms in Argentina during the 1990s and early 2000s (see e.g. Bambaci et al., 2002).

To examine the relationship between cumulative PSG lending and the quality of public sector governance we estimate the following equation:

$$y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 (X_{it})^2 + \beta_3 Z_{it} + \delta_i + \epsilon_{it} \quad (3.1)$$

where y_{it} is the CPIA “Cluster D” score for country i in year t .⁶ CPIA cluster D on public sector management and institutions consists of five components – rule-based governance, quality of budgetary management, efficiency of revenue mobilization, quality of public administration and transparency, accountability and corruption in the public sector – each rated on a one to six scale. The resulting cluster score represents the arithmetic mean of the five components.⁷ X_{it} represents our variable of interest for country i in year t . As mentioned above, we test our model with the number of cumulative prior actions related to public sector governance from all DPLs approved before or during year t . This variable can be decomposed into cumulative conditions related to civil service reform,

⁵We replicated the tests reported below, substituting the number of cumulative PSG *loans* for cumulative PSG conditions, and results were not statistically significant. This finding could be interpreted as a product of measurement error, however, as many PSG-related conditions are missed by this measure, and many non-PSG conditions (e.g., on agricultural policy) are included.

⁶We also show that our main results are robust to using an alternative governance measure.

⁷See the appendix for a detailed description of the components and the assessment procedure used to generate them.

decentralization, public expenditure and financial management (PFM), anti-corruption, tax policy and administration and other public sector governance issues. This breakdown allows us to estimate the effect of PSG conditionality on four CPIA cluster D components. The following dependent-independent variable pairs are tested: public expenditure and financial management conditions on CPIA's quality of budgetary management, tax policy and administration conditions on CPIA's efficiency of revenue mobilization, anti-corruption conditions on CPIA's accountability and corruption in the public sector and civil service reform conditions on CPIA's quality of public administration.

Z_{it} is a vector of control variables. As other donors also might be engaged in supporting policy change, we include total aid over GDP as a control variable. Next, we also add the logarithm of GDP per capita as a regressor. High levels of per capita income could improve the quality of governance by increasing tax revenues if government funds are a binding constraint. Higher income levels could also reflect a greater volume and size of transactions, increasing the benefits of developing policies for contract enforcement. Furthermore, as many factors influence a country's welfare, GDP per capita also serves as a general control. Following Besley and Persson (2011) among other studies, we also include a measure of democracy, i.e., the Freedom House index of political freedoms. Finally, country and year fixed effects are added as regressors. Cumulative conditions are positively correlated with time, so the inclusion of year effects is important to control for any time trend or transitory events that could otherwise bias the coefficient on cumulative conditions. In section 4.3, we also test whether results are robust to including several additional controls. We have drawn upon the public sector reform literature to select the additional covariates. It is often argued that fiscal pressures and crises are engines for reform (see, e.g., Krueger, 1993; Ranis and Mahmood, 1992). On the other hand, crises in and by themselves do not necessarily entail deep, structural change, and could actually decrease the political capital for engaging in costly reform (Brinkerhoff, 2000; Tommasi, 2004). To control for any crisis-

related effects, we include a dummy coded 1 if a debt crisis occurred at year $t - 1$. As several country cases – India, Brazil, China – illustrate that economic growth is conducive to public sector reform, we add annual GDP growth to the model.⁸ Schneider and Heredia (2003) argue that integration into the global economy strengthens incentives for governments to engage in PSG reform, so we include trade openness as an additional control. In addition to these economic variables, we add several political controls. Chauvet and Collier (2009) show that elections matter for policy and distinguish between a positive frequency effect of elections and a non-linear cyclical effect of elections. We therefore add a measure of elections frequency and a variable that captures the stage of the political business cycle. Keefer (2011) and Cruz and Keefer (2013) find that governments with institutionalized parties, i.e., parties with internal accountability and sanctioning mechanisms, are more likely to pursue pro-development policies. Hence, we include a dummy for institutionalized parties in the analysis. As a fourth political variable, we test whether electoral systems with proportional representation affect the quality of public sector governance (Geddes, 1991). Next, we add two political variables related to regime type: the age of democracy and a dummy whether a country transitioned to democracy. Ideally, democratization creates electoral competition and demand for a merit-based bureaucracy. On the other hand, in young democracies, the bureaucracy could be used for electoral ends, which might hamper reform (Bunse and Fritz, 2012). Finally, we include the logarithm of population and gross IDA disbursements as additional controls. Concerning the latter, countries with higher CPIA ratings receive higher allocations of IDA aid, other things equal, which in turn may increase the likelihood of receiving a DPL with its associated conditions. Because any causal effect of CPIA ratings on DPLs (and conditions) is mediated by IDA disbursements, controlling for the latter will effectively correct for this potential source of endogeneity bias. Descriptive statistics for the panel data variables are found in table 2.

⁸To avoid multicollinearity with the crisis dummy, we lag growth two periods.

We estimated our model with linear, logarithmic and quadratic specifications and retain the quadratic one as it produced the best fit. The graphical output of a semi-parametric estimation further justifies the choice of the quadratic model (see figure 3).

Model 3.1 is tested using data from a comprehensive country-year panel of aid receiving countries from 1996 to 2008. In robustness tests, however, we employ three alternative sample restrictions. First, we drop all observations for a country after its last DPL containing PSG conditions has closed. Second, we follow Easterly (2005) in limiting the sample to countries that received at least one DPL with PSG-related conditions. Third, we estimate the model excluding Argentina (a strong outlier) from the sample.

3.1. correction for endogeneity: two approaches

When analyzing the association of World Bank lending with the quality of public sector governance, we have to take into account a potential selection bias problem. That is, countries often receive policy loans because of governance deficiencies, so the coefficient on DPL conditions may be biased downward when examining its impact on policy outcomes (Easterly, 2005). On the other hand, the coefficient may be biased upward, if loans (and their associated conditions) tend to go to motivated governments that would have reformed even in the absence of support. The basic rationale of a DPL is that the prospect of receiving a loan motivates a government to implement a set of “prior actions” or policy conditions negotiated with the Bank, and funds are then disbursed in anticipation of further reforms. Conceivably, however, governments might have implemented a similar set of measures whether or not a DPL with PSG conditions had been negotiated with the Bank.

Correcting for this problem calls for a robust identification strategy, which we implement in two alternative ways. First we estimate equation 3.1 with system GMM (Arellano and Bover, 1995; Blundell and Bond, 1998) and instrument for number of conditions with its lagged differenced values. The Arellano and Bond (1991) tests indicate $AR(3)$ autocorrelation. Hence, we lag number of conditions by four periods. Furthermore, as the number

of time periods grows large, the instrument count increases exponentially, making results about estimators and related specification tests invalid (Roodman, 2009). One solution to this problem is to use only certain lags. We implement this by using only lags 4 through 6. In order to minimize correlation across countries in the idiosyncratic errors, we also include time dummies. Other configurations – e.g., collapsing the instrument matrix, increasing the number of lags per time period, including different lags – generate equally significant coefficient estimates for number of DPL conditions, with acceptable test statistics for over-identification. (See the appendix for a regression with a collapsed instrument matrix, using lags five to ten.)

The validity of this identification strategy rests on the assumption that $E(\Delta x_{i,t-j}(\delta_i + \epsilon_{i,t})) = 0$, which implies that changes in the number of DPL conditions is uncorrelated with the fixed effects. According to the Diff-in-Hansen test reported in table 5, these additional moment conditions are met. Next, system GMM estimators may suffer from weak instrument biases (Bun and Windmeijer, 2010). Unfortunately, there are no formal tests available to evaluate instrument strength. Hence, we assume that the internal instruments used here are sufficiently strong to identify the effect of DPL conditions on the quality of public sector governance.

Given the restricted assumptions required for system GMM, we use a cross-sectional version of the dataset and employ 2SLS as a second correction for possible selection bias.⁹ We estimate the following cross-sectional equation:¹⁰

⁹With the panel dataset, we are limited to using mechanical instruments in GMM, because substantive instruments that significantly predict DPL conditions exhibit little or no time series variation. Moving to cross-section data allows us to avoid that problem as well as complications associated with serial correlation in the dependent variable.

¹⁰Instead of estimating the quadratic specification, we test the linear model which enables us to employ specification tests. Note that the number of conditions exceeds the estimated turning point for only a tiny minority of observations in our data (see figure 3). Results of this test are relevant only to establishing a positive causal relation between PSG conditions and quality of governance for the vast majority of observations to the left of the turning point.

$$\Delta y_{i,t} = \gamma_0 + \gamma_1 y_{i,t_0} + \gamma_2 \Delta \hat{X}_{i,t} + \gamma_3 Z_{i,\cdot} + v_i \quad (3.2)$$

The dependent variable here is the change in quality of governance, measured over the period 1996-2008. A convenient implication of using the change in quality of governance as the dependent variable is that time-invariant sources of heterogeneity between countries – such as colonial heritage, legal tradition, and cultural norms – are not likely to matter much, because their effects arguably will be captured mostly by controlling for initial-year (1996) quality of governance.

The key independent variable is the number of cumulative conditions from 1996 through 2008. In the first stage we instrument for the number of conditions with the logarithm of population (in 1996) and the number of cumulative agricultural conditions from 1996 through 2008. Population size is used as an instrument for aid in many other studies (e.g., Boone, 1996; Burnside and Dollar, 2000; Djankov et al., 2008) and has been shown to be statistically unrelated to a wide variety of institutional indicators (Rose, 2006). However, its use is not without problems, especially where economic growth is the dependent variable. Most notably, Bazzi and Clemens (2013) argue that country size may affect growth through multiple (endogenous) channels – trade, FDI, aid, . . . – which makes population an invalid instrument when those factors are omitted or not properly instrumented for. The fact that many empirical studies have used population as an instrument for some endogenous variable without controlling for other channels through which country size could affect the dependent variable confirms their point. However, with the quality of governance (in contrast to growth) as the dependent variable, theory and evidence about any effects of country size are more ambiguous and scarce (Knack and Azfar, 2003). Obviously, we cannot exclude the possibility that orthogonality conditions are not met, even though test statistics for overidentification are reassuring. For our second instrument – the number

of agricultural conditions – we exploit the fact that the reform programs countries receive often contain conditionalities that target several sectors. That is, development policy loans that contain PSG conditions may also contain conditions targeting agricultural policy. As the results chain from inputs to outputs in public sector governance is often long and hard to influence (World Bank, 2013), we do not expect sectoral conditions targeting agricultural productivity to have a direct effect on the quality of public sector governance. Specification tests provided in table 6 are consistent with this reasoning.

As controls we include the initial level of the quality of governance, average annual aid as a share of GDP, and average annual growth in GDP per capita over the period 1996-2008, the logarithm of initial income per capita, a measure for ethnic fractionalization, initial political freedom and the change in political freedom over the period 1996-2008. The coefficients of equation 3.2 are estimated with robust standard errors using 126 observations. See table 3 for descriptive statistics. In the next section we discuss our empirical findings.

4. Empirical Findings

4.1. OLS and alternative governance indicator

Table 4 presents the results of estimating the base model using OLS. We find a significant quadratic relation between PSG-related lending conditionality and the quality of public sector governance. According to the results of table 4, the predicted turning point lies between 54 and 55 conditions, with the CPIA cluster D score increasing by about 0.129 points (or about 0.15 standard deviations) on average at that point, compared to the case of 0 PSG conditions. Ceteris paribus, beyond 110 PSG conditions PSG lending becomes detrimental, relative to the case of 0. In our dataset five countries received more than 110 PSG conditions: Morocco (113), Pakistan (121), Ghana (136), Bangladesh (166) and Argentina (210). Table 4 also indicates that a well governed public sector is associated with

high income levels, corresponding with earlier findings of, e.g., Hesse (2000) and de Renzio (2009). Finally, we also find a positive partial correlation between aid and the quality of public sector governance. The positive coefficient (partly) incorporates the endogenous effect due to reverse causality going from good governance to aid disbursements. On the other hand, the availability of aid funds may provide a financial cushion for implementing politically-costly PSG reforms.

As CPIA ratings are produced within the World Bank, one might argue that results could be driven by spurious correlation or tautological relationships. As a first possibility, CPIA scores for a country might be inflated to justify a shift toward development policy lending (budget support) and away from investment lending (project aid). Second, over-optimistic beliefs about the efficacy of PSG conditions could bias CPIA scores upward. Third, it is possible that PSG conditions are selected to match the CPIA criteria, so that their implementation virtually guarantees an improvement in the CPIA ratings. Regarding this last possibility, however, prior actions tend to be modest “de jure” reforms – such as cabinet approval of an access to information law – that would rarely be significant enough to warrant an increase in a CPIA rating. The intent of reforms in this area is to make government more responsive to civil society’s demands for information and thereby improve accountability and performance, and the CPIA content reflects these larger objectives rather than the mere presence of an access to information law.¹¹ Nevertheless, for the above three reasons it is useful to show that the main result in equation 1 is robust to using an alternative dependent variable from the Heritage Foundation, that is not influenced in any way by the judgments of World Bank staff.¹² Equation 2 reveals a similar concave relationship with, again, a limited impact of PSG conditionality in DPLs on the quality

¹¹Note that DPLs each specify intended results that are distinct from and broader than the prior actions contained in the loan (World Bank, 2013).

¹²Since 1995 the Heritage Foundation creates an index of economic freedom, based on 10 subcomponents. We have used the average of two of those components, property rights and freedom from corruption, as an (incomplete) alternative for the CPIA cluster D score.

of governance: the turning point lies at 63 conditions, where the Heritage score (ranging from 0 to 100) is predicted to be 4 points higher on average than with 0 conditions.

The use of the CPIA ratings thus does not appear to create an upward bias in the measured relationship between PSG-related conditions and quality of public sector governance; nor is it merely reflecting a tautological relationship. Because the CPIA – in comparison with the Heritage Foundation measures or other governance indicators – better reflects the Bank’s objectives in strengthening public sector governance, we use it as our preferred dependent variable in the remainder of the study. Note that we use GMM and 2SLS methods in all tests reported below, as a general correction for potential sample selection and endogeneity bias, including any bias that might be produced by CPIA ratings inflation.

4.2. Endogeneity of PSG conditionality

Results for our base GMM specification are presented in table 5. A similar quadratic relationship appears as in table 4, with a stronger effect on the quality of public sector governance: the maximum impact occurs at 80 cumulative conditions, with an average increase in the CPIA cluster D score of 0.247 points (about 0.3 standard deviations) relative to 0 conditions. Note also that here, in contrast to OLS, democratic regimes are significantly associated with higher public sector quality.¹³ The test statistics presented at the bottom of table 5 are reassuring. The p-value of the Hansen J statistic does not indicate rejection of the null that instruments are exogenous. The value reported for the Diff-in-Hansen test provides an indication whether the additional moment restrictions necessary for system GMM are met (Bond et al., 2001). With a p-value of 0.625, we do not reject the null that the additional moment conditions are valid.

For about 95 % of the observed cases in the sample, the number of cumulative conditions is below the turning point of 80. For most but not all of the range of observed

¹³‘Political Rights’ varies from 1 (most democratic) to 7 (least democratic), so a negative coefficient implies that more political freedoms are associated with higher CPIA ratings.

values, therefore, the estimated effect of PSG conditions on the quality of governance is positive. A possible explanation for the negative relationship estimated for the upper tail is selection bias: countries with the most severe governance problems or those most at risk of deterioration may receive loans with more stringent conditionality. If our GMM procedure fully corrected for such a bias, then few if any observations should lie above the turning point. If it is even somewhat effective in correcting for it, however, we should observe a lower turning point using OLS than GMM. In fact, that is what we observed above: the estimated turning point in OLS was 55 conditions, and 11 % of observations lie beyond that point, compared to only 5 % lying above the turning point estimated using GMM.

4.3. Sample restrictions and additional controls

The remaining regressions in table 5 report results of several robustness tests. First, we follow Easterly (2005) and limit the sample to include only countries that have received at least one DPL with PSG-related conditions over the period 1980-2010. As shown in equation 2, results remain significant, and quantitatively slightly larger than in equation 1. At 88 conditions, the turning point, the predicted CPIA cluster D score is 0.31 points higher on average than in the case of zero conditions.

As a second sample restriction, we drop all observations for a country after the last DPL with PSG conditions to that country has closed.¹⁴ If reforms associated with DPLs are often not sustained following completion of the loan, then the estimated effects should be larger when the years following loan closing are dropped (about one-seventh of all observations). However, the effects of cumulative conditions in equation 3, where those years are dropped, are very similar to those in the base model.

As a third sample restriction we test our model excluding Argentina from the sample. With 210 PSG conditions, Argentina can be considered an outlying observation.¹⁵ Equa-

¹⁴Data on closing years were extracted from a less comprehensive dataset.

¹⁵However, based on a visual inspection one could argue that the data on cumulative PSG lending are

tion 4 shows that Argentina is not particularly influential, as results are similar to those in equation 1, though not significant at conventional levels.

Finally, we test whether results are robust to including additional controls. Equation 5 of table 5 shows that coefficient estimates for the number of conditions variable remain significant, even beyond the 1% significance level, and their magnitudes are unaffected by the inclusion of the added regressors. As expected, equation 5 indicates a positive association between IDA disbursements and quality of governance. Furthermore, we find that a debt crisis negatively affects the quality of public sector governance, while growth spurts have a positive effect, possibly relaxing the political constraints by providing resources and credibility for reform. The coefficient estimate for trade openness is significantly positive, as such, we find empirical support for the hypothesis that integration into the global economy pushes governments to reform (although causality might run in both directions). Concerning the political variables, transitioning to a democracy negatively affects public sector functioning, which might be due to the lack of complementary institutions that support a smooth transition (Collier and Vicente, 2008, 2012). Finally, over and above a rough transition year, young democracies are able to reap the benefits of increased electoral competition, as the coefficient on age of democracy is negative and significant (although small in magnitude).

4.4. Cross sectional 2SLS tests

As an alternative to GMM, we correct for possible endogeneity by employing 2SLS and estimating equation 3.2 in a cross-sectional version of the data. The two methods rely on independent sources of exogenous variation in the endogenous variable (PSG conditions), but they both point to a similar conclusion, that PSG conditions in DPLs produce significant improvements in the quality of public sector governance.

drawn from an exponential distribution with the PSG loans to Argentina as part of such a data generating process. A Kolgomorov-Smirnov test does not allow to reject this hypothesis.

In the 2SLS tests the dependent variable is the change in CPIA cluster D ratings between 1996 and 2008, and we control for the initial (1996) level. The variable of interest is the number of cumulative PSG conditions from 1996 through 2008. In the first stage we instrument for number of PSG conditions with the logarithm of population (in 1996) and the number of cumulative agricultural conditions from 1996 through 2008. Results for OLS and 2SLS regressions are reported in table 6. Equation 1, table 6 shows that the effect of conditions on changes in the quality of governance is positive but not statistically significant. Furthermore, the coefficient for initial level of governance quality is significantly negative, implying a regression toward the mean effect. Both the initial level of political rights and its change over the period are associated with improved quality of public sector governance. Also, higher income countries have a better functioning public sector. Equations 2 and 3 present the results from a 2SLS regression. As shown in the first-stage regression (equation 2), both of our instruments are significantly and positively related to the number of conditions. The F-statistic of excluded instruments is 10.61, indicating a sufficiently strong association of our instruments with the receipt of (repeated) World Bank PSG conditions. Furthermore, Wooldridge (1995)'s robust score test of overidentifying restrictions does not reject the null that the excluded instruments are exogenous to the quality of governance (test score = 0.006, p-value = 0.935). In equation 3, the exogenous effect of PSG conditionality is reported. The coefficient on conditions is positive and significant: 50 PSG conditions over the period 1996-2008 are predicted to increase the CPIA cluster D score by 0.56 points on average. The coefficient substantially increases in magnitude in comparison with its OLS counterpart, consistent with a negative endogeneity bias (i.e., more PSG conditions are attached to loans for countries that are making slower progress on quality of governance). Furthermore, the 2SLS regression confirms the regression toward the mean effect and the positive relationship between both political rights (and income) on the quality of governance.

4.5. Impact of PSG conditionality on cluster D components

Prior conditions related to PSG can be disaggregated and matched with the most relevant component of CPIA Cluster D, to test for differences in their impacts. Table 7 reports results for these tests, which in all other respects replicate those from tables 5 and 6.¹⁶ For brevity, the table reports only the coefficient estimates and standard errors for the conditions variables.

Using the base GMM specification (corresponding to equation 1 from table 5), a significant quadratic relationship between conditions and CPIA scores is found for three of the four Cluster D components, with transparency and corruption as the only exception. Among the other three components, results for budget management and revenue mobilization (see middle two columns of table 7) are relatively robust to changes in the sample, inclusion of more control variables, and to correcting for endogeneity using the cross-sectional 2SLS design. As seen in the first column of table 7, however, results for the component on quality of public administration (which focuses on human resource management and sustainability of the public sector wage bill) turn out to be less robust to these variations in sample, specification and method. Vested interests, clientelism and rent-seeking behavior of public officials may partly explain the limited impact of PSG support in this sub-sector (see, e.g., Kelsall, 2011; van de Walle, 2003). IEG (2008, p. 54) notes that the Bank's analytical tools, including diagnostic instruments and monitoring indicators, are weaker in the area of administrative and civil service reform (ACSR) than in other areas such as PFM, with adverse consequences for program design.

The increased focus of the international community on public financial management – half of all PSG conditions are related to PFM reform – may help explain the positive results in this sub-sector. Compared to ACSR, there appears to be a stronger consensus

¹⁶Consistent measurement of separate CPIA criteria are available only from 1998 onwards, reducing the sample to 1430 observations.

among PFM specialists regarding what a well-performing system looks like, and stronger diagnostic and measurement tools. Base model estimations indicate that, on average, the maximum predicted increase in the CPIA rating on Quality of Budgetary and Financial Management PFM component is 0.44 points, occurring at 30 PFM conditions.

Andrews (2009) distinguishes between public sector reforms involving relatively “concentrated” and “deconcentrated” sets of actors. Successful implementation of reforms is more difficult when it requires changing the behavior and norms of larger and more disparate groups of public officials or other agents. Our findings are consistent with this distinction. Among the four sub-sectors analyzed here, PFM and tax conditions appear to be the most effective in improving the quality of governance (as measured by the respective CPIA criteria), and reforms in these areas arguably involve more concentrated actors than in civil service reform and anti-corruption.

4.6. A falsification test

In this subsection we report results from a falsification test that further supports the interpretation that PSG conditions cause improvements in the quality of public sector governance, as measured by the CPIA. This test exploits the presence of three other “clusters” in the CPIA: macroeconomic policy (CPIA A), structural policies (CPIA B) and social policies (CPIA C). If DPLs (with their conditions) and higher CPIA ratings are both the product of some positive shock that is not cluster specific (e.g., a democratic opening, or increased geopolitical importance of a country, that yields more favorable treatment by donors), then PSG conditions should be significantly related to CPIA ratings not only in cluster D, but also in clusters A-C. As shown in table 8, however, PSG conditions turn out to be unrelated to higher ratings in each of the other three clusters, controlling for conditions pertaining to the respective clusters. The conditions that “should” matter for each of those other CPIA clusters are significant in each case; e.g., conditions related to structural policies have a significant concave relationship with cluster C ratings. These

results indicate it is very unlikely that our main results are plagued by certain namely, non-cluster specific, sources of omitted variable bias. Any such bias would have to be specific to PSG conditions and CPIA cluster D ratings. It would also have to vary over time, as the country fixed effects control for time-invariant country characteristics.

5. Summary and discussion

In this study we investigate the impact of World Bank lending on the quality of public sector governance. We find a significant concave relationship between PSG conditions and quality of public sector governance. Coefficient estimates from the base GMM specification indicate that 80 PSG conditions – which is around the maximum of the predicted curve – increases the CPIA cluster D score by about 0.247 points on average, or 0.3 standard deviation units. The number of cumulative conditions is below 80 for the vast majority of observations, so the predicted effect of additional conditions is positive in most cases. This result is robust not only to alternative approaches to correcting for potential endogeneity, but also to sample restrictions, the use of an alternative governance indicator, and the inclusion of additional controls. Replication of findings using the alternative governance measure (from the Heritage Foundation) ensures our key results do not merely reflect correlated errors in Bank staff’s judgments regarding which countries should receive DPLs (with their accompanying conditions) and higher CPIA ratings. Coefficients on the conditions variables are unchanged by controlling for IDA inflows, ruling out any upward bias from the fact that countries with higher CPIA ratings receive higher IDA allocations, and potentially more DPLs. Falsification tests further support a causal interpretation, by showing that omitted variable bias could not be driving our main findings, unless the bias is specific to public sector governance and does not affect the CPIA ratings for macro, structural, or social policies.

When we disaggregate the analysis by the CPIA cluster D components, we find that

World Bank lending appears to be less successful in improving the quality of public administration and fighting corruption. On the other hand, conditions targeting PFM and tax system reforms appear to be effective. E.g., on average, 30 PFM conditions increase the CPIA PFM rating by 0.44 points. These sub-sector differences in performance generally accord with findings from other studies on World Bank public sector reform (IEG, 2008; Blum, 2014).

These findings stand in contrast with the results of Smets and Knack (2014), who investigate the impact of World Bank lending on the quality of economic policy (as measured by the CPIA cluster A and B score). For comparability, Smets and Knack (2014) find that 62 economic policy conditions increase the CPIA cluster A score by more than half a point on average.¹⁷

How then to explain the mixed findings of PSG support? Factors on both the donor and recipient side matter. On the recipient side, political economy challenges are often cited (Devarajan et al., 2011; World Bank, 2012; Blum, 2014). Implementing reforms is (politically) costly in the short term, especially in the public sector, which is often used as an instrument for patronage and clientelism. On the other hand, building a well functioning state apparatus might take several years, if not decades (Pritchett and de Weijer, 2010). This time inconsistency problem makes it for many governments unattractive to engage in PSG reform. In other words, PSG reforms might violate the government's incentive compatibility constraints.¹⁸ Related to this, IEG (2008, p. 56) notes that for certain PSG programs – especially in the civil service – it is not obvious to identify tangible benefits.

¹⁷The estimate of a 0.504 point increase is based on a quadratic model specification including Argentina in the sample.

¹⁸One way recipient governments deal with adversarial conditionality is what Andrews et al. (2013) call “isomorphic mimicry”, i.e., the tendency of governments to superficially implement (donor-driven) reforms as a way of windowdressing without a structural change in the functioning of the public sector. As an example the authors mention procurement reform in Liberia and Afghanistan, where the introduction of new laws and implementation procedures has not done much for public sector efficiency. Applied to World Bank lending, Bunse and Fritz (2012) note that PSG programs often contain technical, de jure reforms that do not change actual practice in the public sector.

Consequently, political leaders are more likely to undervalue such programs, resulting in a low commitment to reform.

In contrast with economic policies, which are often characterized by a short chain from inputs to outputs - e.g., “stroke-of-the-pen” reforms such as reduction in trade tariffs - the governance results chain in the public sector is much longer and thus harder to influence (World Bank, 2013). Implementing PSG reforms may be ineffective if other constraints in the results chain are binding. As an example World Bank (2012) notes that the introduction of a school-based management policy may improve school autonomy, but will have little impact on learning outcomes without complementary policies that increase the supply of high-quality and motivated teaching staff. In the same vein, Reinikka and Svensson (2006) argue that traditional supply-side mechanisms to reduce corruption – such as anti-corruption commissions, audits and legislation – may not be sufficient for bringing about change, especially in developing countries.

On the donor side, Pritchett et al. (2013) note that development organizations (still) push for the adoption of best practice public policies. However, there is relatively little evidence-based knowledge about what matters most in improving public sector governance (World Bank, 2012). What is more, as many aspects in public sector governance entail discretionary face-to-face transactions (Pritchett and Woolcock, 2004), customized program designs are required based on deep knowledge of local context and behavior. This is not straightforward. For instance, research from behavioral economics has shown that mental models are sometimes hard to change. Harun (2007) notes that senior civil service officials in Indonesia are reluctant to embrace new accounting systems since they share the opinion that ‘what worked before will continue to work in the future’.

World Bank (2012) points to two additional organizational issues that may raise particular challenges to PSG lending. First, in order to design effective reforms that are politically feasible, continuous engagement and policy dialogue are required. However,

World Bank (2012) notes that World Bank PSG lending is rather episodic and dependent on identifying a feasible lending opportunity. Second, public sector reform outcomes are inherently uncertain, and the Bank has tended to downplay the consequent risks (World Bank, 2012). Blum (2014) finds that task team leaders (TTLs) that manage PSG projects do a good job at assessing the risks, but that prevailing incentives encourage TTLs to underreport them.

To conclude, how can the World Bank become more effective in supporting public sector governance reforms? Its most recent public sector management strategy (World Bank, 2012) proposes several new actions and shifts in emphasis, as influenced by recent academic thinking on public sector reform, e.g. Andrews et al. (2013); Grindle (2007); Pritchett and Woolcock (2004). According to World Bank (2012), the Bank should embrace an agnostic approach towards PSG lending, letting go strong priors about the nature of the issue at hand and the appropriate ways to fix it. This more diagnostic, problem-solving approach would mean moving away from “best practice” towards “best fit,” being more flexible in implementation, and experimenting more with novel approaches (Rodrik, 2008). As one example, Kelsall (2011) provocatively argues that instead of “going against the grain” in Africa, policy and institutional reform should rather build on local patterns of thought and governance. The author suggests that, instead of imposing an individualistic, merit-based bureaucracy, public sector reform should exploit the strong kinship ties in African societies and organize the civil service accordingly. However, Xavier (2013) cautions – while supporting most elements of the Bank’s new strategy – that in some cases imported “best practice” models can provide a useful narrative that helps legitimize reforms. But even in those cases, as Xavier (2013) acknowledges, reforms may fail if they are not sufficiently sensitive to local contexts and focused on solving specific problems.

The Bank’s new strategy also acknowledges the need for improving risk management at the project and portfolio level, and to encourage more candid assessments of project

risk by TTLs (World Bank, 2012). Next, as explicit knowledge and evidence are lacking, the Bank and other donors should invest in learning about the dynamics of public sector reform and the impact of PSG interventions.¹⁹ To achieve this, more and better data on the functioning of public sector institutions are required.

Finally, the Bank’s new diagnostic, “best fit” approach envisions a more continuous dialogue with recipient country counterparts in government, and greater participation of other stakeholders including the public. A more continuous and broadened engagement of this sort by Bank task teams can generate mutual trust and help build agreement on a convincing diagnostic story and possible solutions that take into account incentives of key actors including politicians (World Bank, 2012).

Acknowledgement

We would like to thank Vincenzo Verardi, Adam Wagstaff, Peter Moll and Deon Filmer for useful comments and suggestions. Lodewijk is also indebted to the Institute of Development Policy and Management (University of Antwerp) and the Research Foundation Flanders (FWO) for financial support.

References

- Alesina, A., Weder, B., 2002. Do corrupt governments receive less foreign aid? *The American Economic Review* 92 (4), pp. 1126–1137.
- Andrews, M., 2009. Isomorphism and the limits to african public financial management reform. Harvard Kennedy School of Government Working Paper Series RWP09-012.
- Andrews, M., 2013. Explaining positive deviance in public sector reforms in development. Harvard Kennedy School of Government Working Paper Series RWP13-040.

¹⁹Already several initiatives are set up to improve the effectiveness of PSG lending. For instance, in 2005 the development impact evaluation initiative (DIME) was created to generate knowledge on policy effectiveness. Currently, new impact evaluations are being developed to address specific PSG bottlenecks in civil service and PFM reform.

- Andrews, M., Pritchett, L., Woolcock, M., 2013. Escaping capability traps through problem driven iterative adaptation (PDIA). *World Development* 51, 234 – 244.
- 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), 277–97.
- Arellano, M., Bover, O., 1995. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68 (1), 29–51.
- Bambaci, J., Saront, T., Tommasi, M., 2002. The political economy of economic reforms in Argentina. *Journal of Policy Reform* 2 (2), 75–88.
- Bazzi, S., Clemens, M. A., 2013. Blunt instruments: Avoiding common pitfalls in identifying the causes of economic growth. *American Economic Journal: Macroeconomics* 5 (2), 152–86.
- Besley, T., Persson, T., 2011. *Pillars of Prosperity*. Princeton University Press, Princeton and Oxford.
- Blum, J. R., 2014. What Factors Predict How Public Sector Project Perform? A Review of the World Bank’s Public Sector Management Portfolio. World Bank Policy Research Working Paper 6798, World Bank, Washington DC.
- Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87 (1), 115–143.
- Bond, S. R., Hoeffler, A., Temple, J., 2001. GMM estimation of empirical growth models. C.E.P.R. Discussion Papers 3048, Center for Economic Policy Research.
- Boone, P., 1996. Politics and the effectiveness of foreign aid. *European Economic Review* 40 (2), 289 – 329.
- Brautigam, D. A., Knack, S., 2004. Foreign aid, institutions, and governance in Sub-Saharan Africa. *Economic Development and Cultural Change* 52 (2), pp. 255–285.
- Brinkerhoff, D. W., 2000. Assessing political will for anti-corruption efforts: An analytic framework. *Public Administration and Development* 20 (3), 239–252.
- Bun, M. J. G., Windmeijer, F., 2010. The weak instrument problem of the system GMM estimator in dynamic panel data models. *Econometrics Journal* 13 (1), 95–126.
- Bunse, S., Fritz, V., 2012. Making public sector reforms work. Policy Research Working Paper 6174, The World Bank.
- Burnside, C., Dollar, D., 2000. Aid, policies, and growth. *The American Economic Review* 90 (4), 847–868.
- Cashel-Cordo, P., Craig, S. G., 1990. The public sector impact of international resource transfers. *Journal of Development Economics* 32 (1), 17 – 42.

- Chauvet, L., Collier, P., 2009. Elections and economic policy in developing countries. *Economic Policy* 24, 509–550.
- Clemens, M. A., Radelet, S., Bhavnani, R. R., Bazzi, S., 2012. Counting chickens when they hatch: Timing and the effects of aid on growth. *The Economic Journal* 122 (561), 590–617.
- Collier, P., Vicente, P. C., 2008. Votes and violence: Evidence from a field experiment in Nigeria. CSAE Working Paper Series 2008-16, University of Oxford.
- Collier, P., Vicente, P. C., 2012. Violence, bribery, and fraud: The political economy of elections in Sub-Saharan Africa. *Public Choice* 153 (1-2), 117–147.
- Cruz, C., Keefer, P., 2013. The organization of political parties and the politics of bureaucratic reform. Policy Research Working Paper Series 6686, The World Bank, Washington D.C.
- de Renzio, P., 2009. Taking stock: What do PEFA assessments tell us about PFM systems across countries? ODI Working Papers 302.
- de Renzio, P., Andrews, M., Mills, Z., 2011. Does donor support to public financial management reforms in developing countries work? ODI Working Paper 329, London.
- Devarajan, S., Khemani, S., Walton, M., 2011. Civil society, public action and accountability in Africa. Policy Research Working Paper Series 5733, The World Bank.
- Djankov, S., Montalvo, J., Reynal-Querol, M., September 2008. The curse of aid. *Journal of Economic Growth* 13 (3), 169–194.
- Dreher, A., 2002. The development and implementation of IMF and World Bank conditionality. Hamburg Institute of International Economics.
- 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.
- Geddes, B., 1991. A game theoretic model of reform in Latin American democracies. *The American Political Science Review* 85 (2), pp. 371–392.
- Grindle, M. S., 2007. *Going Local: Decentralization, Democratization, and the Promise of Good Governance*. Princeton University Press.
- Harun, 2007. Obstacles to public sector accounting reform in Indonesia. *Bulletin of Indonesian Economic Studies* 43 (3), 365–376.
- Hesse, J. J., 2000. *Rebuilding the State: Public Sector Reform in Central and Eastern Europe*. Nomos, Sinzheim.
- IEG, 2008. *Public Sector Reform: What Works and Why? An IEG evaluation of World Bank Support*.

- No. 6484 in World Bank Publications. The World Bank.
- Kapur, D., Lewis, J. P., Webb, R., 1997. *The World Bank: Its first half century. Volume 1. History.* Brookings Institution Press, Washington, D.C.
- Kaufmann, D., Kraay, A., Mastruzzi, M., 2006. *Governance matters V: Aggregate and individual governance indicators for 1996 – 2005.* World Bank Policy Research Working Paper Series 4012, Washington DC.
- Keefer, P., 2011. Collective action, political parties and pro-development public policy. *Asian Development Review* 28 (1), 94–118.
- Kelsall, T., 2011. Going with the grain in African development? *Development Policy Review* 29 (s1), s223–s251.
- Knack, S., 2001. Aid dependence and the quality of governance: Cross-country empirical tests. *Southern Economic Journal* 68 (2), 310–329.
- Knack, S., Azfar, O., 2003. Trade intensity, country size and corruption. *Economics of Governance* 4 (1), 1–18.
- Krueger, A. O., 1993. *Political economy of policy reform in developing countries.* The MIT Press, Cambridge, MA and London.
- Moloney, K., 2009. Public administration and governance: A sector-level analysis of World Bank aid. *International Review of Administrative Sciences* 75 (4), 609–627.
- OPCS, 2009. *Country policy and institutional assessments: 2009 assessment questionnaire.* Operations Policy and Country Services, World Bank.
- Pritchett, L., de Weijer, F., 2010. *Fragile states: Stuck in a capability trap?* Background paper for the 2011 World Development Report.
- Pritchett, L., Woolcock, M., 2004. Solutions when the solution is the problem: Arraying the disarray in development. *World Development* 32 (2), 191 – 212.
- Pritchett, L., Woolcock, M., Andrews, M., 2013. Looking like a state: Techniques of persistent failure in state capability for implementation. *Journal of Development Studies* 49 (1), 1–18.
- Rajan, R., Subramanian, A., 2007. Does Aid Affect Governance? *The American Economic Review* 97 (2), 322–327.
- Ranis, G., Mahmood, S. A., 1992. *The political economy of development policy change.* Basil Blackwell, Cambridge, MA and Oxford.
- Reinikka, R., Svensson, J., 2006. Using micro-surveys to measure and explain corruption. *World Develop-*

- ment 34 (2), 359–370.
- Remmer, K. L., 2004. Does foreign aid promote the expansion of government? *American Journal of Political Science* 48 (1), 77–92.
- Rodrik, D., 2008. The New Development Economics: We Shall Experiment, but How Shall We Learn? Harvard Kennedy School Working Paper Series rwp08-055.
- Roodman, D., 2009. A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics* 71 (1), 135–158.
- Rose, A. K., 2006. Size really doesn't matter: In search of a national scale effect. *Journal of the Japanese and International Economies* 20 (4), 482 – 507.
- Schneider, B. R., Heredia, B. (Eds.), 2003. *Reinventing Leviathan: the Politics of Administrative Reform in Developing Countries*. North-South Center Press, University of Miami, Miami.
- Smets, L., Knack, S., 2014. *World Bank lending and the Quality of Economic Policy*. World Bank Policy Research Working Paper 6924, World Bank, Washington DC.
- Svensson, J., 2000. Foreign aid and rent-seeking. *Journal of International Economics* 51 (2), 437–461.
- Tommasi, M., 2004. Crisis, political institutions and policy reform: The good, the bad, and the ugly. In: Tungodden, Bertil, N. S., Kolstad, I. (Eds.), *Annual World Bank Conference on Development Economics*. World Bank and Oxford University Press.
- van de Walle, N., 2003. Presidentialism and clientelism in Africa's emerging party systems. *The Journal of Modern African Studies* 41 (02), 297–321.
- Wooldridge, J. M., 1995. Score diagnostics for linear models estimated by two stage least squares. In: Maddala, G. S., Phillips, P. C. B., Srinivasan, T. N. (Eds.), *Advances in Econometrics and Quantitative Economics: Essays in Honor of Professor C. R. Rao*. Blackwell, Oxford, pp. 66 – 87.
- World Bank, 1998. *Assessing aid: What works, what doesn't, and why*. Oxford University Press for the World Bank, Oxford and New York.
- World Bank, 2007. *Conditionality in Development Policy Lending*. World Bank, Washington, D.C.
- World Bank, 2012. *The World Bank's Approach to Public Sector Management 2011–2020: Better Results from Public Sector Institutions*. Poverty Reduction and Economic Management, World Bank, Washington, D.C.
- World Bank, 2013. *Development Policy Retrospective 2012: Results, Risks, and Reforms*. World Bank, Washington, D.C.
- Xavier, J. A., 2013. *The world bank approach to public sector management 20112020: lessons from the*

malaysian experience. *International Review of Administrative Sciences* 79 (3), 426–432.

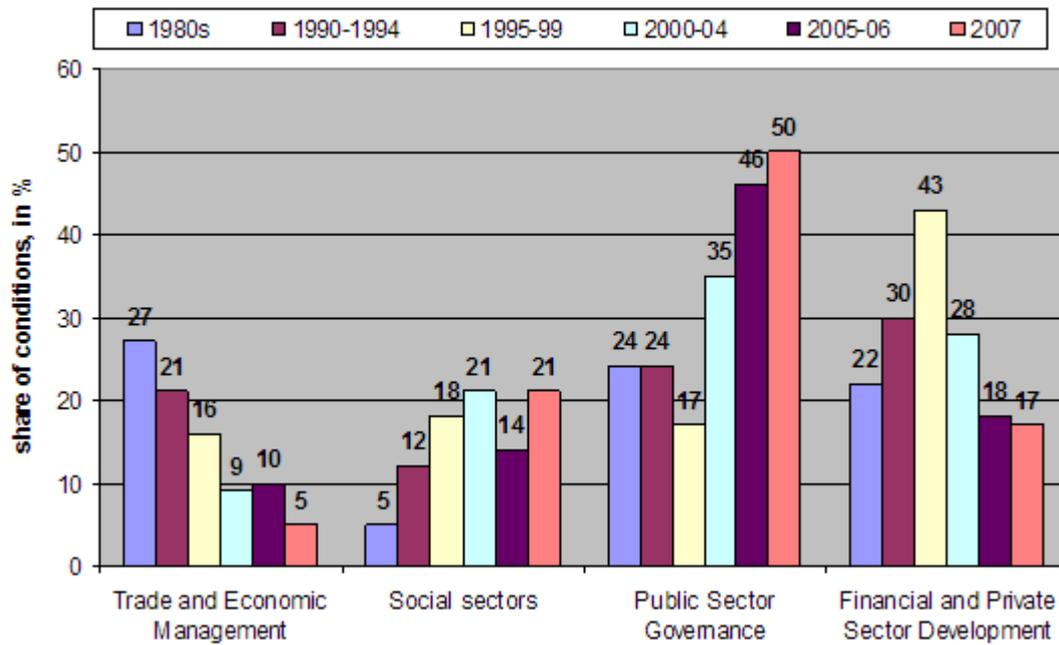


Figure 1: Sectoral coverage of conditionality in policy based lending (Source: World Bank (2007))

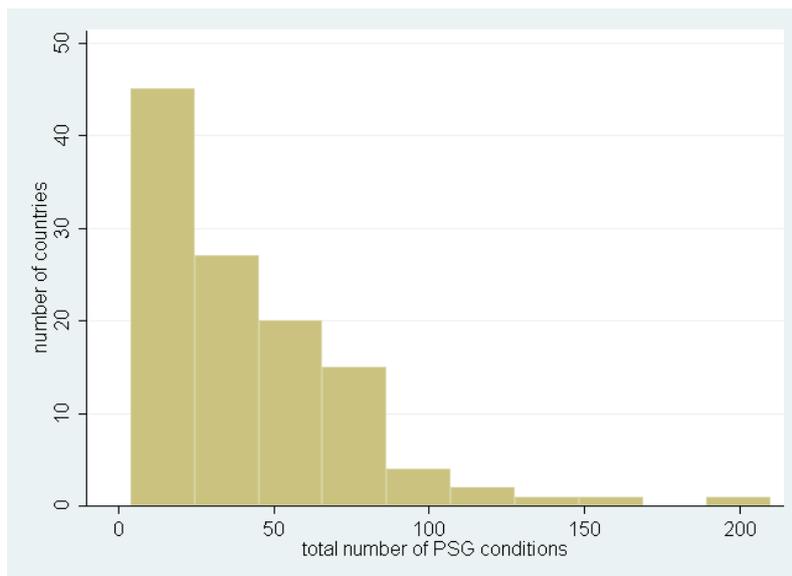
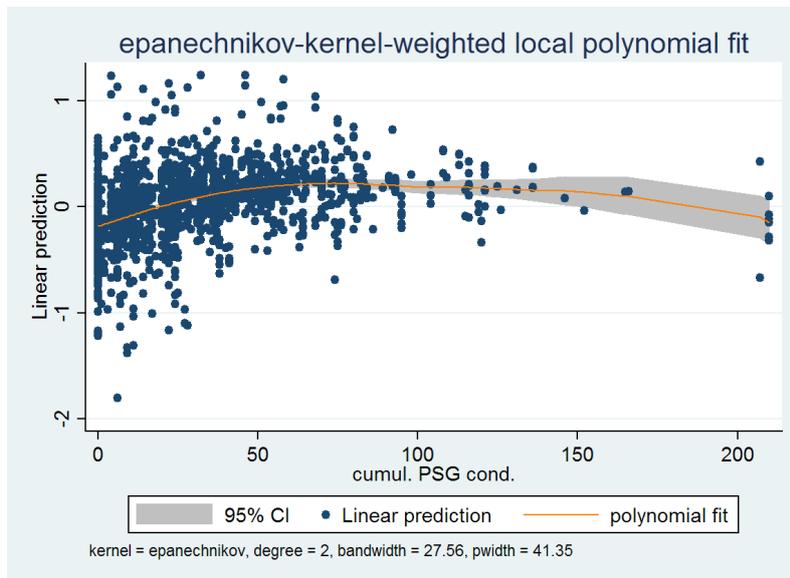


Figure 2: Distribution of cumulative PSG conditions for the period 1980-2010

Figure 3: Non-parametric fit of cumulative conditions



Note: semiparametric fixed-effects regression using STATA's xtsemipar command with CPIA cluster D as dependent variable, log of per capita GDP, aid over GDP, political rights and a time trend as parameterized variables and cumulative conditions as non parameterized variable. Polynomial of degree two fitted. Standard errors clustered by country.

Table 1: sectoral distribution of all effective adjustment loans for the period 1980-2010

| sector | frequency | percentage |
|--|-----------|------------|
| Agriculture and Rural Development | 62 | 6.19 |
| Economic Policy | 450 | 44.91 |
| Education | 29 | 2.89 |
| Energy and Mining | 46 | 4.59 |
| Environment | 14 | 1.4 |
| Financial Management | 1 | 0.1 |
| Financial Sector | 12 | 1.2 |
| Financial and Private Sector Development | 121 | 12.08 |
| Global Information/Communications Techn | 2 | 0.2 |
| Health, Nutrition and Population | 8 | 0.8 |
| Poverty Reduction | 51 | 5.09 |
| Private Sector Development | 7 | 0.7 |
| Public Sector Governance | 127 | 12.67 |
| Social Development | 2 | 0.2 |
| Social Protection | 49 | 4.89 |
| Transport | 5 | 0.5 |
| Urban Development | 14 | 1.4 |
| Water | 2 | 0.2 |
| Total | 1,002 | 100 |

Table 2: Summary statistics panel data models

| Variable | Mean | Std. Dev. | Min. | Max. |
|----------------------------------|-------------|------------------|-------------|-------------|
| CPIA cluster A | 3.713 | 0.869 | 1 | 6 |
| CPIA cluster B | 3.513 | 0.706 | 1 | 5.7 |
| CPIA cluster C | 3.422 | 0.700 | 1 | 6 |
| CPIA cluster D | 3.226 | 0.717 | 1 | 5.5 |
| Heritage | 37.959 | 16.318 | 10 | 91.5 |
| quality of public administration | 3.148 | 0.683 | 1 | 6 |
| quality of budget management | 3.345 | 0.79 | 1 | 6 |
| efficacy of revenue mobilization | 3.477 | 0.711 | 1 | 6 |
| transparency and corruption | 3.043 | 0.796 | 1 | 6 |
| cumul. PSG cond. | 27.562 | 30.897 | 0 | 210 |
| cumul. civil service cond. | 4.576 | 6.68 | 0 | 39 |
| cumul. financial mgmt. cond. | 8.664 | 11.008 | 0 | 66 |
| cumul. tax policy cond. | 5.661 | 8.577 | 0 | 73 |
| cumul. anti-corruption cond. | 0.448 | 1.302 | 0 | 10 |
| cumul. econ. mgmt. cond. | 11.92 | 14.365 | 0 | 116 |
| cumul. struc. pol. cond. | 40.341 | 42.469 | 0 | 220 |
| cumul. social. pol. cond. | 8.418 | 12.971 | 0 | 83 |
| log GDP per capita (PPP) | 8.012 | 1.017 | 5.076 | 10.352 |
| aid over GDP | 0.04 | 0.063 | -0.019 | 0.806 |
| Political Rights | 3.755 | 1.987 | 1 | 7 |
| log gross IDA | 2.102 | 2.224 | 0 | 8.311 |
| debt crisis t-1 | 0.007 | 0.085 | 0 | 1 |
| growth t-2 | 4.474 | 6.679 | -50.248 | 106.28 |
| trade openness | 0.37 | 0.212 | 0.06 | 1.162 |
| programmatic party | 0.553 | 0.497 | 0 | 1 |
| transition to democracy | 0.013 | 0.114 | 0 | 1 |
| age of democracy | 27.318 | 25.316 | 1 | 139 |

Table 3: Summary statistics cross-sectional model

| Variable | Mean | Std. Dev. | Min. | Max. |
|---------------------------------------|-------------|------------------|-------------|-------------|
| change in CPIA cluster D | 0.089 | 0.709 | -1.578 | 2.081 |
| cumulative PSG conditions 1996-2008 | 17.69 | 22.074 | 0 | 130 |
| CPIA cluster D 1996 | 3.262 | 0.956 | 1 | 5.478 |
| average annual GDP per capita growth | 4.770 | 8.168 | -2.486 | 82.035 |
| log of GDP per capita 1996 | 6.834 | 1.168 | 4.191 | 9.031 |
| Political Rights | 3.738 | 2.06 | 1 | 7 |
| change in Political Rights | -0.079 | 1.312 | -5 | 3 |
| ethnic fractionalization | 0.473 | 0.252 | 0 | 0.930 |
| average annual aid over GDP | 0.035 | 0.045 | 0 | 0.277 |
| cumulative agri. conditions 1996-2008 | 1.706 | 3.596 | 0 | 22 |
| log of population | 15.588 | 1.958 | 10.618 | 20.92 |

Table 4: Panel regression, OLS

| Dependent variable | CPIA cluster D | Heritage |
|--|-------------------------|---------------------|
| number of cumulative PSG conditions | .005 (.002)** | .132 (.042)*** |
| number of cumulative PSG conditions, squared | -.00004 (0,00001)*** | -.001 (.0002)*** |
| log GDP per capita (PPP) | .498 (.139)*** | 7.392 (2.539)*** |
| aid over GDP | 1.525 (.339)*** | -9.018 (12.151) |
| Political Rights | -.005 (.021) | -.302 (.404) |
| <i>country fixed effects</i> | <i>yes</i> | <i>yes</i> |
| <i>year fixed effects</i> | <i>yes</i> | <i>yes</i> |
| Observations | 1690 | 1607 |
| Countries | 139 | 132 |
| R^2 | .144 | .169 |

* significance at 10%; ** significance at 5%; *** significance at 1%; constant not reported. Standard errors reported in brackets are adjusted for country clustering of observations.

Table 5: panel regression, GMM

| equation no. | (1) | (2) | (3) | (4) | (5) |
|------------------------------|------------------------|------------------------|-----------------------|---------------------|-----------------------|
| variation | GMM | PSG > 0 | closing year | excl. ARG | controls |
| cumul. cond. | .006 (.003)** | .007 (.003)** | .006 (.003)* | .005 (.004) | .006 (.003)** |
| cumul. cond, sq. | -.00004 (.00001)*** | -.00004 (.00001)*** | -.00004 (.00002)** | -.00003 (.00003) | -.00003 (.00001)** |
| log GDP per capita | .243 (.059)*** | .223 (.065)*** | .233 (.064)*** | .240 (.060)*** | .394 (.069)*** |
| aid over GDP | -.260 (.552) | -.312 (.691) | -.312 (.590) | -.287 (.556) | 1.024 (.764) |
| Political Rights | -.135 (.028)*** | -.081 (.026)*** | -.139 (.031)*** | -.136 (.028)*** | -.139 (.029)*** |
| log of gross IDA | . | . | . | . | .088 (.021)*** |
| debt crisis t-1 | . | . | . | . | -.574 (.140)*** |
| growth t-2 | . | . | . | . | .001 (.006) |
| trade openness | . | . | . | . | .353 (.250) |
| prog. party | . | . | . | . | .030 (.083) |
| trans. to democ. | . | . | . | . | .020 (.117) |
| age of democracy | . | . | . | . | .004 (.002)* |
| <i>country fixed effects</i> | <i>yes</i> | <i>yes</i> | <i>yes</i> | <i>yes</i> | <i>yes</i> |
| <i>year fixed effects</i> | <i>yes</i> | <i>yes</i> | <i>yes</i> | <i>yes</i> | <i>yes</i> |
| Observations | 1690 | 753 | 1456 | 1677 | 1526 |
| Number of countries | 139 | 61 | 137 | 138 | 127 |
| Number of instruments | 114 | 114 | 114 | 114 | 121 |
| Wald statistic | 197.46 | 139.46 | 164.03 | 198.21 | 292.06 |
| <i>p-value</i> | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 |
| Hansen J-test | 99.84 | 51.42 | 92.34 | 99.78 | 92.94 |
| <i>p-value</i> | 0.374 | 0.999 | 0.587 | 0.376 | 0.570 |
| Diff-in-Hansen test | 23.14 | 2.67 | 25.43 | 27.48 | 25.42 |
| <i>p-value</i> | 0.625 | 0.999 | 0.495 | 0.385 | 0.495 |

Note: Dependent variable: CPIA cluster D average. Cluster-robust standard errors are reported. Coefficients estimated with forward orthogonal deviations and level equations for IV style instruments. * significance at 10%; ** significance at 5%; *** significance at 1%.

Table 6: cross-sectional 2SLS

| equation no. | (1) | (2) | (3) |
|---------------------------------------|--------------------|---------------------|--------------------|
| | OLS | First stage | Second Stage |
| cumulative PSG conditions 1996-2008 | .0002 (.002) | . | .011 (.005)** |
| CPIA 1996 | -.710 (.044)*** | 1.388 (1.874) | -.739 (.049)*** |
| average annual GDP per capita growth | .0009 (.010) | .064 (.108) | .002 (.009) |
| log of GDP per capita 1996 | .100 (.049)** | -1.601 (3.410) | .173 (.074)** |
| Political Rights 1996 | -.111 (.024)*** | -1.209 (1.059) | -.100 (.026)*** |
| change in Political Rights | -.122 (.027)*** | .419 (1.495) | -.115 (.029)*** |
| ethnic fractionalization | -.043 (.196) | 2.261 (6.240) | -.125 (.190) |
| average annual aid over GDP | -1.231 (1.101) | 38.570 (60.350) | -.009 (1.219) |
| cumulative agri. conditions 1996-2008 | . | 1.194 (.536)** | . |
| log of 1996 population | . | 4.721 (1.696)*** | . |
| No. observations | 126 | 126 | 126 |
| R^2 | .681 | .247 | .575 |
| F test of excluded instruments | . | 10.613 | . |
| <i>p-value</i> | . | 0.0001 | . |
| test of endogeneity | . | . | 9.56685 |
| <i>p-value</i> | . | . | 0.002 |
| Overidentification test | . | . | .006629 |
| <i>p-value</i> | . | . | 0.9351 |

Note: Dependent variable is the change in policy quality over the period 1996-2008, as measured by the CPIA cluster D average. Robust standard errors in parentheses. * significance at 10%; ** significance at 5%; *** significance at 1%.

Table 7: CPIA components

| component | public admin. | budget mgt. | revenue mobil. | transp. & corrupt |
|--------------------|---------------|-------------|----------------|-------------------|
| GMM | .038 | .029 | .027 | -.027 |
| | (.019)** | (.008)*** | (.009)*** | (.080) |
| PSG > 0 | -.001 | -.0005 | -.0005 | .008 |
| | (.0006)** | (.0001)*** | (.0001)*** | (.010) |
| Closing year | .00009 | -.0001 | -.0004 | .00007 |
| | (.0007) | (.0002) | (.0001)*** | (.006) |
| Excl. ARG | .010 | .034 | .016 | .068 |
| | (.025) | (.009)*** | (.011) | (.118) |
| Controls | -.0004 | -.0006 | -.0003 | .004 |
| | (.0007) | (.0002)*** | (.0001)** | (.020) |
| Cross-sectional IV | -.001 | .028 | .038 | -.021 |
| | (.022) | (.014)** | (.018)** | (.082) |
| Controls | .0003 | -.0005 | -.001 | .007 |
| | (.0008) | (.0003) | (.0007) | (.010) |
| Controls | .015 | .017 | .027 | .043 |
| | (.018) | (.009)* | (.009)*** | (.056) |
| Cross-sectional IV | -.0005 | -.0003 | -.0005 | -.007 |
| | (.0005) | (.0001)** | (.0001)*** | (.006) |
| Cross-sectional IV | .025 | .065 | .059 | .089 |
| | (.018) | (.024)** | (.032)* | (.097) |

Note: Results from estimating regression (GMM) models on CPIA cluster D components: Quality of Public Administration, Quality of Budgetary and Financial Management, Efficiency of Revenue Mobilization and Transparency, Accountability, and Corruption in the Public Sector. Only coefficient estimates – linear and quadratic term – and cluster-robust standard errors of conditions variable are reported. * significance at 10%; ** significance at 5%; *** significance at 1%.

Table 8: Falsification test

| Dependent variable | CPIA A | CPIA B | CPIA C |
|-------------------------------------|---------------------|-----------------------|----------------------|
| number of cumulative PSG conditions | -.00009 (.007) | .002 (.006) | .004 (.004) |
| PSG conditions, squared | .000005 (.00004) | -.000009 (.00003) | -.00002 (.00002) |
| economic management conditions | .022 (.013)* | . | . |
| econ. mgmt. cond., squared | -.0003 (.0001)** | . | . |
| structural policies conditions | . | .009 (.004)** | . |
| struct. pol. cond., squared | . | -.00005 (.00002)** | . |
| social policies conditions | . | . | .016 (.008)** |
| soc. pol. cond., squared | . | . | -.0002 (.00009)** |
| log of per capita GDP | .216 (.076)*** | .225 (.059)*** | .192 (.069)*** |
| aid over GDP | -1.426 (.604)** | -.779 (.421)* | -1.182 (.622)* |
| Political Rights | -.076 (.032)** | -.134 (.023)*** | -.109 (.029)*** |
| <i>country fixed effects</i> | <i>yes</i> | <i>yes</i> | <i>yes</i> |
| <i>year fixed effects</i> | <i>yes</i> | <i>yes</i> | <i>yes</i> |
| Observations | 1761 | 1761 | 1761 |
| Number of countries | 139 | 139 | 139 |
| Number of instruments | 125 | 125 | 125 |
| Wald statistic | 135.05 | 234.43 | 174.08 |
| <i>p-value</i> | 0.0001 | 0.0001 | 0.0001 |
| Hansen J-test | 92.23 | 101.21 | 99.12 |
| <i>p-value</i> | 0.789 | 0.559 | 0.617 |
| Diff-in-Hansen test | 23.63 | 29.31 | 30.46 |
| <i>p-value</i> | 0.701 | 0.397 | 0.342 |

Regression results from estimating GMM model with CPIA A, CPIA B and CPIA C as dependent variables. CPIA cluster A average measures the quality of economic management, CPIA cluster B average the quality of structural policies and CPIA cluster C average measures the quality of policies for social inclusion and equity. * significance at 10%; ** significance at 5%; *** significance at 1%.

Appendices

Appendix A. Country Policy and Institutional Assessment

The CPIA scores are designed to measure government policies and institutions, rather than outcomes. The set of criteria are revised periodically to reflect changes in the collective knowledge of practitioners and specialists - both inside and outside the World Bank – regarding policies and public sector management institutions that matter for these outcomes. The criteria are grouped into 4 “clusters” as follows:

- A. Economic Management
 1. Macroeconomic Management
 2. Fiscal Policy
 3. Debt Policy

- B. Structural Policies
 4. Trade
 5. Financial Sector
 6. Business Regulatory Environment

- C. Policies for Social Inclusion/Equity
 7. Gender Equality
 8. Equity of Public Resource Use
 9. Building Human Resources
 10. Social Protection and Labor
 11. Policies and Institutions for Environmental Sustainability

- D. Public Sector Management and Institutions
 - 12. Property Rights and Rule-based Governance
 - 13. Quality of Budgetary and Financial Management
 - 14. Efficiency of Revenue Mobilization
 - 15. Quality of Public Administration
 - 16. Transparency, Accountability, and Corruption in the Public Sector

For each criterion, countries are rated on a scale of 1 (low) to 6 (high). A 1 rating corresponds to a very weak performance, and a 6 rating to a very strong performance. Intermediate scores of 1.5, 2.5, 3.5, 4.5 and 5.5 may also be given. For the years 1995-1997, countries were rated on a scale of 1 to 5. Scores have been rescaled for this research to a scale of 1 to 6. See OPCS (2009) for a detailed elaboration of the scoring procedure.

Appendix B. Additional System GMM regression

Table B.1: Additional System GMM regression

| Dependent variable | CPIA cluster D |
|--|-----------------------|
| number of cumulative conditions | .006 (.003)* |
| number of cumulative conditions, squared | -.00003 (.00001)** |
| log GDP per capita (PPP) | .24 (.056)*** |
| aid over GDP | -.301 (.545) |
| Political Rights | -.135 (.028)*** |
| <i>country fixed effects</i> | <i>yes</i> |
| <i>year fixed effects</i> | <i>yes</i> |
| Observations | 1690 |
| Number of instruments | 30 |
| Wald statistic | 199.8 |
| <i>p-value</i> | 0.0001 |
| Hansen J-test | 11.13 |
| <i>p-value</i> | 0.517 |
| Diff-in-Hansen test | 0.10 |
| <i>p-value</i> | 0.951 |

Note: cluster-robust standard errors are reported. Coefficients estimated with forward orthogonal deviations and level equations for IV style instruments. Collapsed instrument matrix. Lags 5 to 10 used . * significance at 10%; ** significance at 5%; *** significance at 1%.