WPS8645

Policy Research Working Paper 8645

Is It the Journey That Matters?

A Fresh Look at the Impact of World Bank Policy Lending

Peter Moll
Lodewijk Smets

WORLD BANK GROUP
Operations Policy and Quality Vice-Presidency
November 2018

Abstract

This paper investigates the impact of World Bank development policy operations on the quality of economic policy during the period 1998–2015. A new theoretical framework distinguishes among three effects that have been conflated hitherto: (a) marginal impacts of additional policy actions within the current year; (b) length of the policy engagement with client countries, and (c) changes over time in the marginal impact of policy actions. The analysis focuses on policy actions that are particularly relevant for the quality of economic management. Consistent with past research, robust panel estimations indicate that development policy financing has a positive effect on the quality of government economic policy. The econometric work suggests that the nature of the policy dialogue and quality of the engagement with government matter more than the sheer number of policy actions adopted. There is also tentative evidence that although the positive impact is sustained over time, the initial years of an engagement are the most productive for improvement in government economic policy. This may be linked to the fact that over time the reform program changes from 'first-generation' to more complex 'second-generation' policy actions.

This paper is a product of the Country Economics Unit, Operations Policy and Quality Vice-Presidency. 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://www.worldbank.org/research. The authors may be contacted at pmoll@worldbank.org and lodewijks@iadb.org (corresponding author).

Is It the Journey That Matters? A Fresh Look at the Impact of World Bank Policy Lending

Peter Moll and Lodewijk Smets[1]

Keywords: structural adjustment, adjustment lending, development policy lending, budget support, World Bank, policy-based lending, aid effectiveness

JEL Classification Codes: O10, O19, O20, O57

[1] Moll: Senior Economist, Operations Policy and Country Services, The World Bank, Washington, DC. Email address: pmoll@worldbank.org. Smets: Senior Economist, Inter-American Development Bank, Trinidad and Tobago and affiliated researcher, LICOS Centre for Institutions and Economic Performance, KU Leuven, Belgium. Email address: lodewijks@iadb.org (corresponding author). Manuela Francisco initiated this research task and participated in its design. The team has benefited from technical discussions with Christian Gonzalez, Kamer Karakurum-Özdemir, Stephen Knack, Eric Le Borgne, Dorsati Madani, Toru Nishiuchi, Lucy Pan, Mariela Paredes Alanes, Peter Siegenthaler, Vivek Suri, Ashley Taylor, and Theo Thomas.

1. Introduction

Sound economic and social policies are important if countries wish to prosper and achieve sustainable development. Public policies are needed to provide public goods, address externalities, aim for equitable redistribution and support the good functioning of the economy. As the financial and economic crisis of 2008 illustrated, it is far from guaranteed that good policies arise endogenously.

That is why many organizations are engaged in supporting policy reform processes in recipient countries, with international financial institutions such as the World Bank and the IMF playing a leading role. Take the World Bank. In 1980 the organization launched its first non-project lending instrument to support policy change in recipient countries, called structural adjustment lending (SAL). As a new lending instrument, SAL would support countries to improve policies to achieve development results. The policy-based instrument provided budget finance upon completion of a program of policy reforms.[2] In its early years, SAL emphasized economic stabilization and correction of balance of payments distortions. In the 1990s, the emphasis of SAL shifted towards protecting the poor from the adverse effects of the needed policy changes (Dreher, 2002).

As structural adjustment lending generated concerns from within the Bank and from borrowing countries (World Bank, 1989), several studies investigated its effectiveness. A broad consensus in the literature is that traditional policy-based lending, as it was conceived and practiced until the end of the 1990s, was not entirely successful (see, e.g., Devarajan et al., 2001; Svensson, 2003; Easterly, 2005).

Recognizing the limitations of its impact, the World Bank modified its approach towards policy-based lending, notably by sharply reducing the number of multi-tranche operations, in favor of single-tranche operations and programmatic series of operations.[3] Other changes included strengthened country 'ownership' of lending programs by using countries' own development strategies and a reduction in the average number of conditions in its loans (Koeberle, 2003; World Bank, 2004, 2006). A new operational policy, OP 8.60, was issued in

[2] The financing did not fund the reforms but was provided in the form of general budget support.

[3] In multi-tranche operations (consisting of typically two or three tranches), the authorities completed a set of prior actions, whereupon the operation was presented to the Board and the first tranche disbursed. The authorities then needed to complete another set of policy actions to receive the second tranche, and a third to receive the third tranche, and so on. The operation was inflexible because if circumstances in the country changed, the second and subsequent tranche release conditions could not be changed easily; this required a bureaucratically arduous restructuring. Under development policy lending from late 2004, it became more common to have single tranche operations or programmatic series. The latter refers to a series of operations linked by a common set of development objectives and anticipated results, in which each operation was presented to the Board so that the policy actions could readily be adjusted in the light of changed circumstances. This flexibility in turn permitted more predictability in disbursement.

September 2004, including a name change from *structural adjustment lending* to *development policy lending.*[4]

In contrast to the extensive research evaluating the first two decades of adjustment lending,[5] there is not much systematic work investigating more recent episodes of policy-based lending. One notable exception is Smets and Knack (2016), who investigate the impact of World Bank development policy lending on the quality of economic policy, covering the period 1995-2008. The authors find that the quality of policy increases, but at a diminishing rate, with the cumulative number of policy loans. When considering the number of reform actions, quadratic specifications indicate that additional conditions may even reduce the quality of policy beyond some point.

This study takes a new look at the data and adds value in five different ways. First, a new theoretical framework is developed, which distinguishes among three sources of non-linearity in the relation between policy actions and the quality of government economic policy, namely (i) contemporaneous changing marginal returns to policy actions, (ii) the impact of the length of the engagement, and (iii) changes in the marginal returns to policy actions over time. Second, this study extends the time frame and investigates the impact of World Bank policy-based lending over the period 1998-2015.[6] The third contribution of the paper is the creation of a new data set involving a complete reclassification of all policy actions from 1980 to 2015. Fourth, we test for non-stationarity and cointegration since the data are time series. And finally, we rigorously test different functional forms based on a defined data generating process. This facilitates the interpretation and permits sharper findings.

Following Smets and Knack (2016), one of our dependent variables is the quality of economic management, as measured by the World Bank's Country Policy and Institutional Assessments (CPIA). The other is the Economic Freedom Index of the Heritage Foundation. The main independent variable of interest is the number of reforms (policy actions) related to economic management. We test various linear and non-linear model specifications. The key finding of the paper – from robust panel estimations – is that the exact number of policy actions makes no difference to the quality of government economic policy, while *the fact that there was at least one policy action* does make a significant difference.[7] We interpret this finding to mean that the *process of generating economic reform* results in improved quality of government policy.

---

[4] In 2014 with the mainstreaming of policy-based guarantees into OP8.60 the instrument was renamed as Development Policy Financing.

[5] See Smets and Knack (2016) for a review of that literature.

[6] The term "policy-based operations" covers structural adjustment from 1980 to 2005, and development policy operations from 2004 to 2015. The two modes overlapped briefly in 2004 and 2005.

[7] This does not mean that development policy loans should only include one policy action. The policy matrix adopted in DPLs is an articulation of the government's reform program, signifying the authorities' commitment to achieving the development objectives therein, whether this requires a handful of prior actions or a long list of them.

We also find tentative evidence that the earlier years of policy engagements between the World Bank and recipient countries tend to have a larger impact on the quality of policy. In the later years of the engagement, the overall impact of World Bank operations continues to be positive, but appears to be not as great as before. We relate this finding to the evolution from first- to second-generation reforms.

The remainder of this report is structured as follows. In the next section we describe the theoretical approach and in section 3 the data sources used. Section 4 outlines the econometric strategy while section 5 discusses the main empirical results. Robustness tests are presented in section 6. Section 7 offers some explanations for the two main findings of this paper. Section 8 concludes.

2. Theoretical approach

We theorize that in the short term the execution of a policy action results in an improvement in the quality of government economic policy. Labeling the number of policy actions taken in year t as A_t, a straightforward way of depicting the relationship with the quality of government economic policy, Q_t, is:

$$\Delta Q_t = \alpha A_t + \varepsilon_t , \qquad (1)$$

with ε_t being other observable and unobservable factors influencing policy choice. Suppose the initial level of quality in year 0 is Q_0. Then building up the data generating process we get

$$Q_1 = Q_0 + \Delta Q_1$$

$$= Q_0 + \alpha A_1 + \varepsilon_1$$

$$Q_2 = Q_1 + \Delta Q_2 + \varepsilon_2$$

$$= Q_0 + \alpha A_1 + \alpha A_2 + \varepsilon_1 + \varepsilon_2,$$

and so on, so that

$$Q_t = Q_0 + \alpha \sum_{i=1}^{t} A_i + \sum_{i=1}^{t} \varepsilon_i \qquad (2)$$

However, the impact of policy actions on the change in quality in a particular year might not be linear. A simple and flexible way of allowing for contemporaneous decreasing or increasing returns is to raise the number of policy actions to a power γ, which can be estimated, thus:

$$\Delta Q_t = \alpha A_t^{\gamma} \qquad (3)$$

An estimated γ of 1 would indicate that policy actions continue to be effective, irrespective of their number. A γ less than one would give decreasing returns, and greater than one increasing returns. A γ of zero would mean that the number of policy actions does not really matter but the *fact that there was at least one policy action* matters.

Then building up the same data generating process as above, the equivalent relationship stated in levels is

$$Q_t = Q_0 + \alpha \sum_{i=1}^{t} A_i^{\gamma} + \sum_{i=1}^{t} \varepsilon_i \qquad (4)$$

We now turn to the impact of the length of the policy engagement. It could be that in the initial years of a policy engagement the impact of development policy lending is small, but that it tends to grow with time. This might be the case if the partners—Bank and client country—familiarize themselves over time with the most effective ways of working together. Similarly, if the Bank injects innovative ideas into the policy mix and these are gradually taken up by the authorities, the subsequent DPF might be more effective than the last.

On the other hand, it could be that the initial years of a policy engagement are the most productive, e.g. if there is only a given number of important reforms to be done, such as the shift from a state-dominated economy to a market-driven economy, and if these can be accomplished within, say, ten years. From this point onwards, the impact of further policy actions might continue to be positive – in the form of improved competitive rules for the private sector, improved efficiency of the licensing process by government (e.g. as revealed by the Doing Business indicators) and strengthened fiscal controls. But it is possible that the impact of these reforms is less. Another rationale for stronger impacts of policy-based lending in the early years of the engagement might be that government decision-makers and Bank staff choose to favor the most impactful reforms first, and move to refinements in later years.

A simple way of depicting the differential impacts of long engagements is the following. Define I_t as an indicator of whether there is a policy action in year t, thus $I_t = 0$ if $A_t=0$, and $I_t = 1$ if $A_t>0$. Then define the length of the engagement as the number of years in the past in which there has been a policy action, viz. $\sum_1^t I_t$. The simplest assumption is that the length of the engagement has a linear impact on the change in the quality of policy, as represented by the coefficient β, thus:

$$\triangle Q_t = \beta \sum_{i=1}^{t} I_i \qquad (5)$$

Long engagements whose productivity blossoms have β>0, while long engagements whose impact declines have β<0.

Then combining equations (3) and (5):

$$\triangle Q_t = \alpha A_t^{\gamma} + \beta \sum_{i=1}^{t} I_i \qquad (6)$$

Building up the data generating process we have

$$Q_t = Q_0 + \alpha \sum_{i=1}^{t} A_i^{\gamma} + \beta \sum_{i=1}^{t} \sum_{j=1}^{i} I_j + \sum_{i=1}^{t} \varepsilon_i \qquad (7)$$

Some observable factors that are expected to influence the quality of government policy include: GDP per capita, aid over GDP and political rights (see more detail below). We add these variables to equation (7) and assume a vector of coefficients δ for these. As time specific events

– such as the East Asian financial crisis or the 2008-2012 Great Recession – and time-invariant country characteristics may also influence policy choice, we add country (ϑ_k) and year dummies (τ_t) to equation (7). Indexing countries by k, we get:[8]

$$Q_t = Q_0 + \alpha\sum_{i=1}^{t} A_{i,k}^{\gamma} + \beta\sum_{i=1}^{t}\sum_{j=1}^{i} I_{j,k} + X_{t,k}\delta + \theta_k + \tau_t + \sum_{i=1}^{t}\epsilon_i \qquad (8)$$

3. Data

In this study we analyze the association of World Bank operations with the quality of economic policy. Following Smets and Knack (2016), one of our dependent variables measures the quality of economic management, as derived from the World Bank's CPIA ratings.[9] The CPIA assessments are subjective ratings of 16 policy indicators, grouped into four "clusters", updated annually by World Bank staff.[10] Possible scores on each indicator range from one to six, including half-point increments (e.g. 3.5). For this analysis, our main dependent variable is the simple average of CPIA clusters A and B, which reflect policies aimed at achieving economic growth and poverty reduction by means of encouraging the market mechanism, and directing government activity towards rule-making, the regulation of markets, and compensation for market failures such as natural monopoly. Cluster A covers macroeconomic and debt policy, and cluster B structural policies, including those about trade, the financial sector, and the regulation of private enterprise.[11] The mean score of this CPIA-based policy quality indicator in our sample is 3.6, with a standard deviation of 0.48.[12]

On the one hand, the CPIA is arguably an appropriate policy measure because its content reflects the views of World Bank management and staff regarding what policies are most conducive to poverty reduction and the effective use of aid resources. On the other hand, the CPIA indicators reflect the subjective judgments of World Bank staff. However, they are correlated with conceptually-related objective indicators, as well as with subjective indicators produced by other organizations. The CPIA cluster A and B average is correlated in the expected direction with macroeconomic indicators such as inflation (r = -0.12) or government debt (r = -0.43). It is also strongly correlated with the 'economic risk' composite of the International

[8] Since equation (8) indicates that residuals are serially correlated, we cluster standard errors at the country-level in our estimations.

[9] See the definitions of variables, and their sources, in Table 13, p. 29.

[10] See OPCS (2009) for a detailed description of the 16 indicators and the assessment procedure used to generate them. A brief description of the CPIA is given in Annex B.

[11] In addition to the macroeconomic and structural policies of clusters A and B, the CPIA includes cluster C on human development and social and environmental policies, and cluster D on public sector governance and institutions. See Bogetic and Smets (2017) and Smets and Knack (2018) for an analysis of the impact of World Bank lending on these clusters.

[12] For further detail on summary statistics see Table 14, p. 29.

Country Risk Guide (ICRG) – an index including GDP per capita, real GDP growth, annual inflation rate, budget balance and current account balance as components.

The CPIA is available for the period 1995-2015, but we discard the first three years and use only 1998-2015, on account of significant changes in the rating system that were introduced in 1998. Specifically, the ratings were from 1 to 5 up to 1997 but from 1 to 6 thereafter, and rescaling the earlier ratings would be arbitrary. There have been further changes in the CPIA since 1998 but most of these had to do with additional questions about the social sectors and fiduciary matters, which increased the total number of questions to 20 for a time, before returning to the current 16 questions. However, during this period of change the questions for Clusters A and B did not change and hence we can use these with some confidence.

It is desirable to have a second dependent variable drawn from a source unrelated to the World Bank. This would have the benefit of avoiding the accusation of bias which might be applied to the CPIA: for the personnel who develop the CPIA overlap with, or may have influence over, the personnel who run development policy operations. A task manager of a SAL/DPF may be tempted to upgrade the relevant parts of the CPIA. Another potential source of bias is that for the poorer countries, whose funding is from the International Development Association (IDA), the country allocation rises with the CPIA by formula, so that an ambitious country director may be tempted to push up the CPIA in order to garner more funds, which would then be likely to be applied to policy lending.

Our second dependent variable is the Economic Freedom Index of the Heritage Foundation (see https://www.heritage.org/index/about), which cannot be impugned for the above-cited biases. The Index measures economic freedom based on twelve quantitative and qualitative factors, grouped into four broad categories:[13]

1. Rule of Law (property rights, government integrity, judicial effectiveness)
2. Government Size (government spending, tax burden, fiscal health)
3. Regulatory Efficiency (business freedom, labor freedom, monetary freedom)
4. Open Markets (trade freedom, investment freedom, financial freedom)

Each of the twelve economic freedoms is graded on a scale of 0 to 100. A country's overall score is derived by averaging these twelve economic freedoms, with equal weight being given to each. We use a modified version of the Index excluding the parts on government integrity, judicial effectiveness and the tax burden, so as to ensure maximum correspondence in subject matter with that of Clusters A & B of the CPIA.

[13] Some of the components do not measure the quality of policies per se, but rather policy outcomes. For instance, Government Size considers actual government spending, rather than the extent to which fiscal policies achieve stabilization and allocative efficiency.

The key independent variable of interest is the number of policy actions. To this end, a new data set was created using the original program documents of all policy-based loans from the beginning of structural adjustment in 1980, until 2015. This entailed reclassification of all 21,509 policy actions – which term covers specifically listed "actions already taken", tranche release conditions, and prior actions. The objective with the reclassification was to identify all policy actions that could have had an impact on Clusters A and B of the CPIA, viz. questions 1 to 6. Note that the classification exercise did not attempt to assess the quality or strength of the policy actions, but only their subject matter. Portmanteau actions involving two or more sub-actions were unbundled. Loans and credits were retained in the sample if they had one or more policy actions that were germane to Clusters A and B.[14]

The reclassification delivers a more precise variable of interest and, arguably, less attenuation bias in our results vis-à-vis other studies such as Smets and Knack (2016). Of the 1,634 fast-disbursing operations from 1980 to 2015, 222 were dropped because they were exclusively sectoral (e.g. education, health, social protection), 62 were dropped as they were sub-national, and 209 were dropped because they were supplemental loans/credits and thus bore no explicit policy actions, and 32 were dropped because they lacked policy actions, as frequently happened in the 1980s when many adjustment operations were hybrids of policy-based operations and investment operations. This left a total of 1,109 economic reform operations.

[14] One exception was made. Budget allocations are generally germane to CPIA question 2 about fiscal matters. But many purely sectoral operations such as education or health include a lone budget allocation. It would not make sense to include these many operations which were not of any significance for economic reform, purely on account of a lone budget allocation which was placed there in order to ensure that the resources would be made available for the relevant sectoral ministry. So sectoral operations with lone budget allocations were not included in the sample.

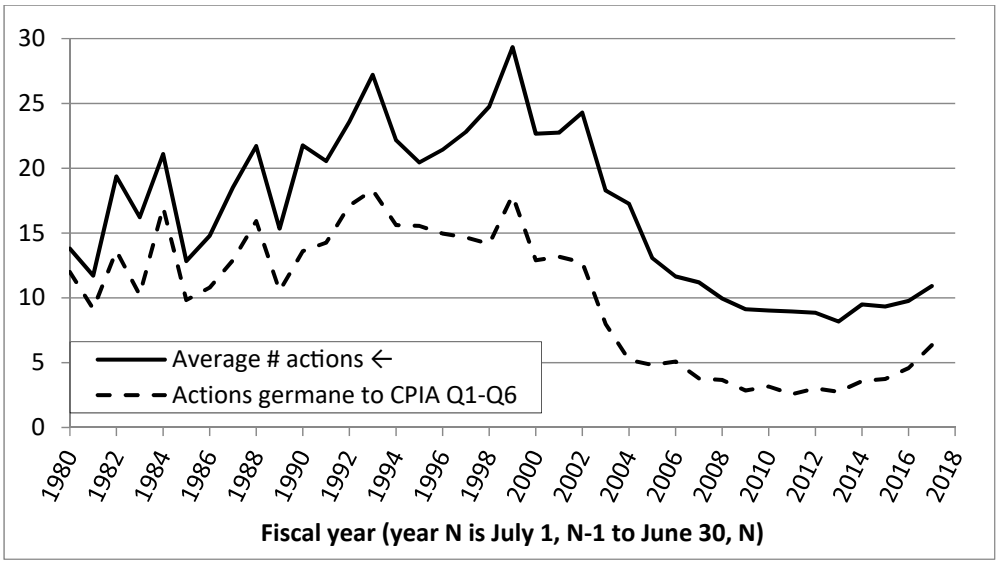

Figure 1. Average number of policy actions in SAL and DPF, and number of policy actions germane to Q1-Q6 of the CPIA, by fiscal year

Figure 1 shows that the number of policy actions germane to Clusters A and B of the CPIA – viz. economic reform actions – was about 10-12 in the early 1980s, and gradually rose to about 16 by 1999, after which it fell to about 5 by 2004. Whereas economic reform constituted the bulk of the conditionality in fast-disbursing operations in the 1980s, by the 2000s economic reform constituted between a third and half of all the policy actions recognized.

A second task of the reclassification exercise was to identify the timing of policy actions. In some previous work it had been assumed, *faute de mieux,* that all the policy actions had been taken by the time of Board presentation. This is not the case for multi-tranche operations, where the prior actions are taken by Board presentation but the policy actions for the second tranche are taken, on average, 1.5 years later than Board presentation, and the policy actions for the third and subsequent tranches are taken, on average, 2.5 years after Board presentation. Of all 1,634 operations, 503 had at least two tranches, and 161 had three or more. Hence this reclassification and correction of the timing of policy actions eliminated a source of measurement error in our key variables of interest.

A third task of the reclassification exercise was to exclude policy actions in tranches that were never disbursed owing to the authorities' non-fulfilment of the tranche release conditions. In the 1,634 operations from 1980 to 2015, there were 73 such instances. In previous work the policy actions associated with these tranches would have been counted as though they had actually been taken. We have excluded this form of error in measurement of the key independent variables.

A fourth task of the reclassification was to derive the length of the relationship between the Bank and the client country. This required classification of all the policy actions from the

earliest adjustment operations in 1980, so as to deliver an accurate count of the number of years in which at least one policy action was taken in the area of *economic* reform.

As mentioned above, we have a matrix of control variables X_{it}. Gross domestic product can be expected to have positive impacts on the quality of government policy, acting through the educational and skill levels of government employees and economic advisers. Furthermore, as many factors influence a country's welfare, GDP per capita also serves as a general control. Hence we include a variable for GDP per capita at purchasing power parity. Aid from other donors could have direct or indirect effects on policy reform. To capture these impacts, we include total aid over GDP as a control variable. Following Besley and Persson (2011) among other studies, we include a measure of democracy, specifically the Freedom House index of political freedoms.

4. Econometric strategy

The power term in γ in equation (8) poses a problem for estimation because it would require a complex non-linear estimation procedure in a panel estimation context and with a requirement for testing for cointegration. We are unaware of econometric software which might make such a procedure tractable. For practical reasons, therefore, we conducted the econometric work in two stages: (i) finding the value of γ and testing for functional form; and (ii) having established the value of γ, performing stationarity and cointegration tests.

(i) Finding the value of γ and testing for functional form

To settle on a value for γ we applied OLS to equation (8), using a grid for γ from 0 to 1.0. We applied three criteria: the within R-squared, the Akaike information criterion and the Bayesian information criterion. As can be seen in Table 1, by all three criteria, the best fit occurred when γ was 0.0 with the CPIA as the dependent variable. With the Economic Freedom Index as the dependent variable, there is a clear trend of increasing fit when γ decreases, but the test statistics are less discriminatory with gamma values close to 0.

Table 1. A grid to determine γ in equation (8)

	Value of γ	Within R-sq.	Akaike's information criterion	Bayesian information criterion
Dependent variable: Clusters A&B of the CPIA				
	1	0.109	828.6	954.7
	0.5	0.130	774.9	901.0
	0.1	0.147	729.2	855.3
	0.0	0.148	722.6	848.7
Dependent variable: Economic Freedom Index				
	1	0.179	14345.5	14491.3
	0.5	0.203	14270.6	14416.4
	0.1	0.217	14226.7	14372.5
	0.0	0.217	14227.0	14372.8

Note: The number of observations was 2,280 for the samples using the CPIA as the dependent variable, and 2518 for the samples using the Economic Freedom Index as the dependent variable.

We discuss the implications of this important finding in section 7. Then we compare this fit with that given by substituting A^{γ} with log A, thus:

$$\Delta Q_t = \alpha \log(A_t) \quad (9)$$

Building up the data generating process, we obtain

$$Q_t = Q_0 + \alpha \sum_{i=1}^{t} \log(A_i) + \text{other variables} \quad (10)$$

which is equivalent to

$$Q_t = Q_0 + \alpha \log \Pi_{i=1}^{t} A_i \quad (11)$$

Note that this is different from the customary

$$Q_t = Q_0 + \alpha \log(\sum_{i=1}^{t} A_i) \quad (12),$$

where the second term would be hard to interpret because it does not stem from a plausible data generating process.

A third possible functional form is to substitute A^{γ} with A and its square, so that equation (1) becomes

$$\Delta Q_t = \alpha_1 A_t + \alpha_2 (A_t)^2 + \text{other variables} \quad (13)$$

Building up the data generating process, this yields

$$Q_t = Q_0 + \alpha_1 \sum_{i=1}^{t} A_i + \alpha_2 \sum_{i=1}^{t} A_i^2 \quad (14)$$

Note that this is very different from the conventional process of testing for nonlinearity which would have been

$$Q_t = Q_0 + \alpha_1 \sum_{i=1}^{t} A_i + \alpha_2 (\sum_{i=1}^{t} A_i)^2 \quad (15)$$

Since the term $\alpha_2(\sum_1^t A_t)^2$ does not stem from a theoretical kernel in a data generating process, it would be hard to interpret.

The result of these tests of functional form – testing equation (8) against (10), and (8) against (14), was conclusive: the fit of equation (8) was better as gauged by the within R-squared (equation (8) had 0.15, while (10) had 0.13 and (14) had 0.12). J-test and Cox-Pesaran tests for non-nested models showed likewise (see Table 2).

Table 2. Non-nested tests of functional form

Non-nested equations compared	**Dep. var.**	**J-test**		**Cox-Pesaran test**	
		Statistic	P-val.	Statistic	P-val.
Cumulated policy actions (8) versus cumulated logged policy actions (10):					
H_0: (8), H_1: (10)	CPIA A&B	-0.8	0.43	1.60	0.055
H_0: (10), H_1: (8)	CPIA A&B	2.92	0.004	-8.35	0.00
H_0: (8), H_1: (10)	Ec. Fr. Index	0.27	0.78	-0.85	0.20
H_0: (10), H_1: (8)	Ec. Fr. Index	1.65	0.101	-5.94	0.0
Cumulated policy actions (8) versus their level and square (14):					
H_0: (8), H_1: (14)	CPIA A&B	-0.69	0.49	1.43	0.076
H_0: (14), H_1: (8)	CPIA A&B	3.64	0.00	-13.8	0.00
H_0: (8), H_1: (14)	Ec. Fr. Index	0.14	0.885	-0.48	0.32
H_0: (14), H_1: (8)	Ec. Fr. Index	2.51	0.013	-11.8	0.00

The statistics in the first two lines of the table (enclosed in the shaded box with heavy lines) may be interpreted thus: the null that (8) with cumulated policy actions is preferred over (10) with cumulated logged policy actions is not rejected because the P-value is 0.43, while the null that (10) is preferred over (8) is rejected because the P-level is 0.004, well below the standard threshold of significance of 0.01 or 0.05. All the tests—within R-squared, J, and Cox-Pesaran—point in the same direction, namely that the best-fitting model is (8).

(ii) Stationarity and cointegration tests

Since the levels equation includes variables that grow over time, such as GDP per capita and the sum of the number of all past policy actions, it is likely that at least some of the variables are integrated of order 1 or more. So panel unit root tests were conducted.

Table 3. Panel unit root tests

Variable	Breitung λ	P-value	# panels, # periods
Q_{it} = CPIA clusters A & B	-0.598	0.27	70, 18
Q_{it} = Economic Freedom Index	-1.98	0.024	63, 21
$\sum_{i=1}^{t} A_i^{0.0}$ = # policy actions raised to the power γ = 0.0, cumulated	13.0	1.00	64, 18
$\sum_{i=1}^{t}\sum_{j=1}^{i} I_j$ = cumulated length of engagement	29.54	1.00	63, 21
gdpcap_PPP = GDP per capita at purchasing power parity	21.0	1.00	64, 18
Aid/GDP	-4.24	0.0000	64, 18
Political rights	-1.23	0.110	64, 18

Note: the Breitung test has H_0: Panels contain unit roots, versus H_1: Panels are stationary. Lags and time trends are not included but in all cases their inclusion did not change the qualitative findings.

The central result of Table 3 is that five of the variables – CPIA clusters A & B, cumulated policy actions, cumulated length of engagement, GPD per capita and the index for Political Rights – contain at least some unit roots. The remaining variables, the Economic Freedom Index, and Aid/GDP, are stationary. Hence, we have to use a cointegration approach, by estimating a model in levels with equation (8) using panel OLS with fixed effects[15] and by testing for cointegration using the Kao test.

The Kao cointegration test has H_0: no cointegration versus H_1: all panels are cointegrated, which means that the variables move together. Since the sample is unbalanced some of the cointegration tests such as the Westerlund do not work; we use the Kao instead. This performs the regression of equation (8) and tests the residuals for stationarity using the Dickey-Fuller t-statistic. According to Table 4, the null of no cointegration is rejected.

Table 4. Kao test of cointegration of equation (8)

Variable	Statistic	P-value
Unadjusted modified Dickey-Fuller t	-2.71	0.0034
Unadjusted Dickey-Fuller t	-6.12	0.0000

Note: H_0: No cointegration, versus H_1: All panels are cointegrated. Number of panels: 145; average number of periods: 13.6. A full set of year dummies is included.

[15] OLS gives super-consistent estimates if the model cointegrates, but the usual OLS standard errors are not correct. Stock and Watson have suggested a procedure to get accurate standard errors, but this is unlikely to be available in panel form. So, for the time being we restrict our attention to the OLS standard errors.

5. Results of the main model

The estimates of the main model are presented in Table 5. As was expected, the coefficient α of 0.0741 is significant, indicating that adopting one or more policy actions results in an improvement in government economic policy. As anticipated, GDP per capita is associated with higher quality of government economic policy. The coefficient β on the length of the engagement is -0.0025. Aid/GDP carries a positive and significant coefficient. The coefficient on political rights is statistically insignificant.[16]

While the coefficient on the variable for policy actions is significant, it should be noted that the sheer number of policy actions does not matter materially to the corresponding level of policy quality. That is, the best fit was obtained when γ was zero. The crucial thing is that there be *at least one* germane policy action. It may be that the existence of such prior actions is proxying for the overall value of the policy engagement (see section 7). As to the size of the impact on policy: by equation (3), $\Delta Q_t = \alpha A_t^{\gamma} = 0.0741$ since γ is zero.

The coefficient on the length of the engagement, as noted, is negative and significant. The cumulated length of the engagement has an implied quadratic effect. If in equation (8) we substitute 0 for γ, 0.0741 for α, and -0.00250 for β (cumulative length of engagement), we have:

$$Q_t = Q_0 + 0.074 * t - 0.0025 * (t+1) * t/2$$

$$= Q_0 + 0.07285 * t - 0.00250 * t^2$$

Solving, Q is maximized when t = 29.1 years, and at that point the marginal impact on Clusters A&B of the CPIA of an additional year's engagement is zero. This does not imply that any particular country is in fact deriving zero marginal benefits from economic reforms in the context of Bank budget support. Instead, what it means is that on average, the marginal benefit of policy engagements is largest in the earlier years and tends to decline later.

[16] Two further variables were entered: (1) mean years of education of citizens of the country, which is intended to capture the educational level of the civil service, because better skills may facilitate higher quality of government economic policy; and (2) fuels and ores exports as a percentage of GDP, which is intended to capture the dependence of the country on extractives. With the "resource curse" in mind, one might theorize that a government that can rely on "easy" taxes in the form of royalties on extractives, such that it is free from accountability to its taxpayers, may engage in more rent seeking and be less concerned with improving the quality of economic policy. Both variables proved to be far from statistically significant in all the estimations. They were dropped so as not to clutter the presentation. Furthermore, we also added lagged values of $\sum_{i=1}^{t} A_i^{0.0}$, but these were statistically insignificant and left out.

Table 5. Estimates of model (8) using Clusters A & B of the CPIA as the dependent variable

Variable	**Main model equation (8)**	
Dependent var.: CPIA clusters A & B (Q_{it})	**Coeff.**	**Robust std. err.**
$\sum_{i=1}^{t} A_i^{0.0}$ = # policy actions raised to γ = 0.0, cumulated	0.0741	0.0149***
Cumulated length of engagement	-0.00250	0.000852***
GDP per capita at PPP	0.0000231	6.87e-06***
Aid/GDP	0.508	0.199**
Political Rights	-0.00214	0.0224
Country & year fixed effects	Yes	
N	2280	
Number of countries	145	
R-squared, within	0.150	
R-squared, overall	0.160	

Note: * = significance level of 10% or less; ** = 5% or less; *** = 1% or less. Standard errors are adjusted for 145 clusters.

The partial residual plot in Figure 2 conveys a strong notion that there is some non-linearity in the relationship between the number of policy actions and the CPIA outcome. First a regression was performed of equation (8), omitting the term $\alpha \sum_{i=1}^{t} A_{i,k}^{\gamma}$, but retaining all the remaining variables including country and time dummies. The residuals were then plotted, and a non-parametric smooth was fitted using an Epanechnikov kernel (red line).

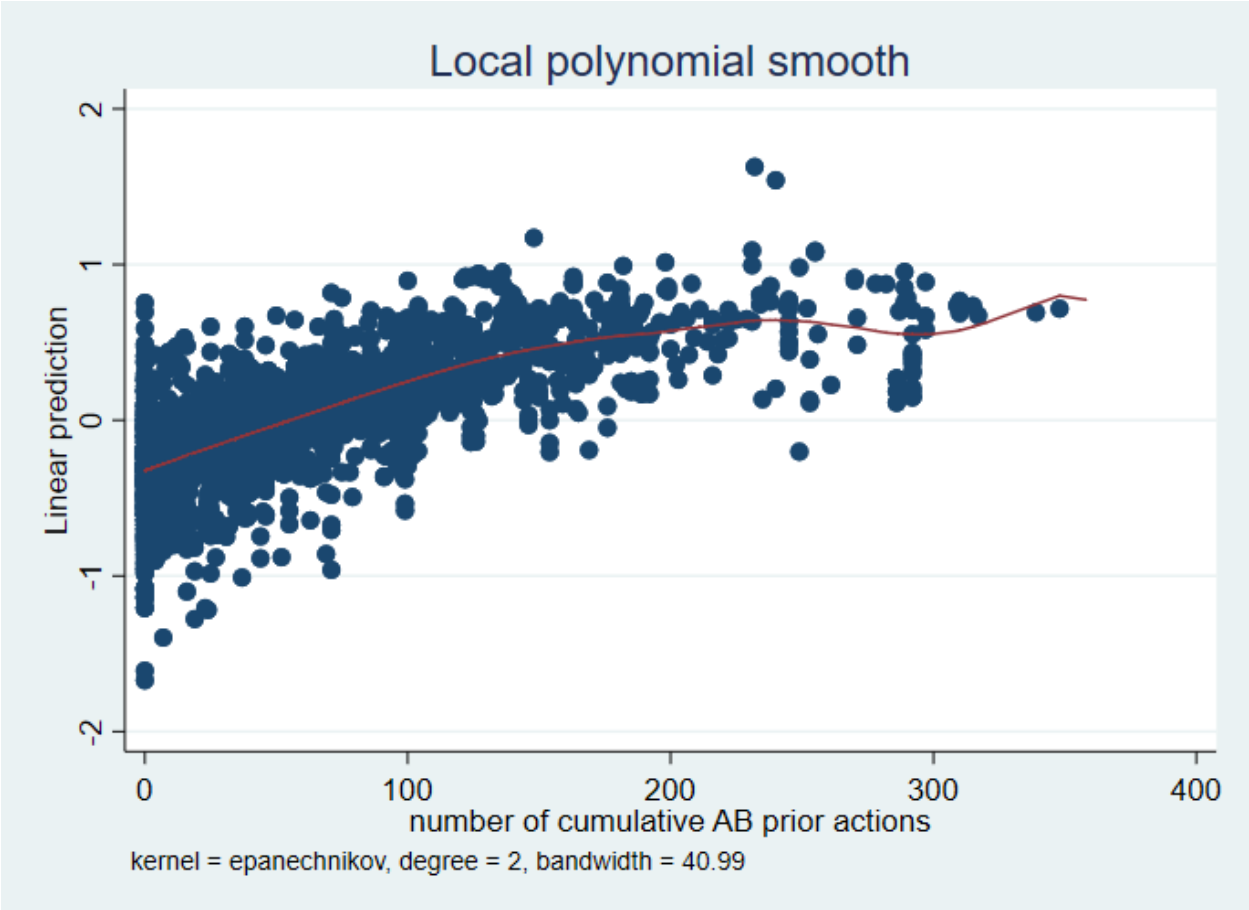

Figure 2. Partial residual plot of CPIA and cumulative policy actions

It seems that there is a positive relation between policy actions and Clusters A and B of the CPIA, and that the relationship has some curvature. The non-linearity could arise from two elements: within-year declining marginal impacts of policy actions (or in the extreme, as we have found, zero marginal impacts beyond the first such policy action), or a decline of the marginal impacts of policy actions over time. The former effect is captured by equation (8); the latter turns out to be unimportant in further estimations which are reported in the section on robustness tests and in Annex C.

For another view of the data we present, in Figure 3, a partial residual plot with, on the horizontal axis, $\sum_{i=1}^{t} A_i^{0.0}$, the number of policy actions raised to γ = 0.0, and cumulated. Now that we are counting only the first policy action in a particular year and in a particular country, the curvature noted in Figure 2 has vanished and the relationship is linear.

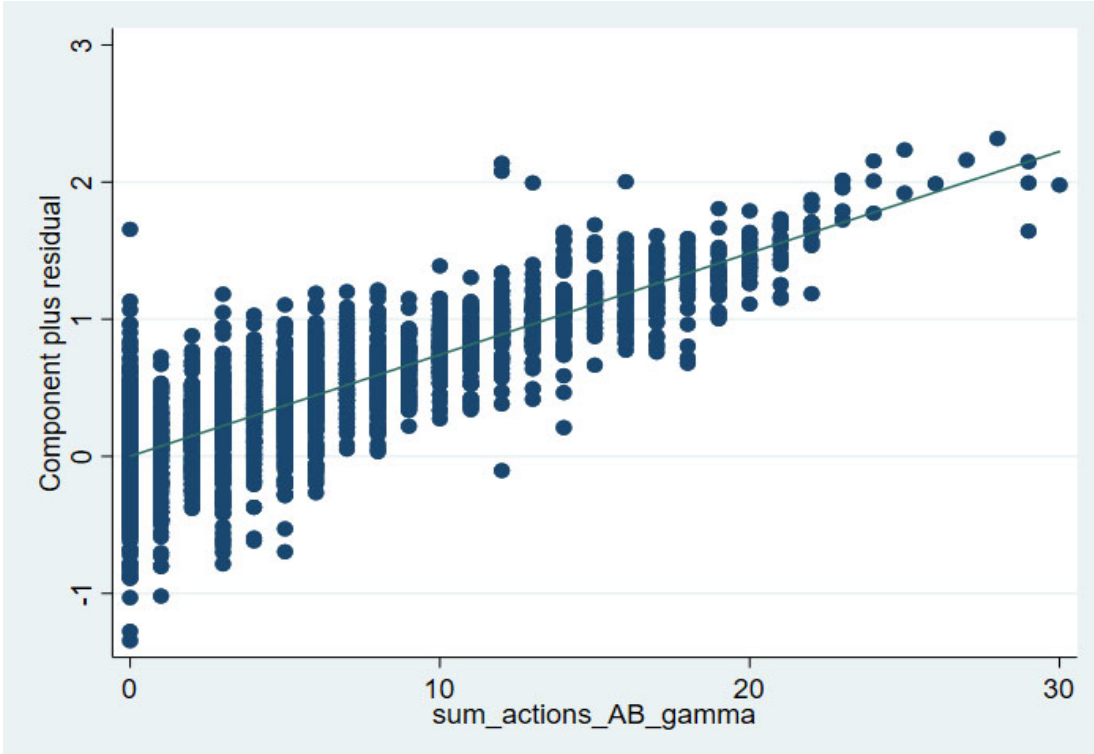

Figure 3. Partial residual plot of CPIA against $\sum_{i=1}^{t} A_i^{0.0}$, the number of policy actions raised to γ = 0.0, and cumulated

As a check on the validity of the cointegrating equation (8), we also present the error correction model, which is the differenced form of equation (8), plus an error correction term comprising the lagged residuals from the cointegrating relationship. Setting γ to zero, we get:

$$\Delta Q_t = \alpha A_t^0 + \beta \sum_{i=1}^{t} I_j + \Delta X_t \delta + \zeta \hat{\epsilon}_{t-1} + \tau_t + \epsilon_t \qquad (17)$$

Since all the variables in the error correction model are I(0), the coefficients can be consistently estimated with OLS.

Table 6. Estimates of the error correction model (17) with CPIA as the dependent variable

Variable	**Equation (8)**	
Dependent var.: change in CPIA clusters A & B (ΔQ_{it})	**Coeff.**	**Robust std. err.**
$A_t^{0.0}$ = # policy actions raised to γ = 0.0	0.0574	0.0103 ***
Length of engagement	-0.00201	0.0008067**
Change in GDP per capita at PPP	0.0000492	0.0000216**
Change in Aid/GDP	0.996	0.0925
Change in Political Rights	-0.00241	0.0101
Residual from eq. (8), lagged one period	-0.0494	0.00977***
Year fixed effects	Yes	
N	2115	
Number of countries	145	
R-squared	0.084	

Note: * = significance level of 10% or less; ** = 5% or less; *** = 1% or less. Standard errors are adjusted for 145 clusters.

The coefficient estimate on the number of policy actions is positive and statistically significant, as expected, and is within two standard errors of the estimate in levels of equation (8). Likewise, the coefficient on the length of the engagement is negative and statistically significant, as anticipated, and is less than a standard deviation from the estimate in levels of equation (8).

We do not have priors about the impact in the short term of GDP per capita, aid/GDP or political rights. Importantly, the coefficient on the residual from the cointegrating equation is negative and statistically significant, reinforcing the notion that the short-term movements in the system tend to converge. This can be seen in this way: suppose that in time t-1 there is an increase in Q_t; then in the current period t, ΔQ_t is smaller due to the negative coefficient on $\hat{\varepsilon}_{t-1}$, thereby tending to equilibrium. Thus, the variables in equation (8) tend to move together.

Next, we present the estimates deploying our second dependent variable, namely the Economic Freedom Index of the Heritage Foundation. The index has been described above in the section about the data set. The results of the main model (8), using the Economic Freedom Index instead of the CPIA as the dependent variable, are presented in Table 7.

Table 7. Estimates of model (8) with the Economic Freedom Index as the dependent variable

Variable	Equation (8)	
Dependent var.: Economic Freedom Index	**Coeff.**	**Robust s.e.**
$\sum_{i=1}^{t} A_i^{0.0}$ = # policy actions raised to γ = 0.0, cumulated	0.994	0.248***
Cumulated length of engagement	-0.0487	0.0132***
GDP per capita at PPP	0.000102	0.000134
Aid/GDP	-19.0	5.77***
Political Rights	-0.205	0.304
Country & year fixed effects	Yes	
N	2518	
R-squared, within	0.217	
Number of countries	154	

Note: * = significance level of 10% or less; ** = 5% or less; *** = 1% or less. Standard errors are adjusted for 145 clusters.

The key element in this estimation is that cumulated policy actions (raised to the power 0) have a positive and statistically significant impact on the quality of government policy as measured by the Economic Freedom Index. The coefficient on the cumulated length of the engagement is negative and statistically significant. These findings buttress the earlier findings using the CPIA as the dependent variable. GDP has the expected sign (positive) but is insignificant. Aid/GDP is negative this time, and significant.

6. Robustness tests

In this section we present the results from three robustness tests: quantile regression, a Tobit model and split samples.[17] The first deploys a 50% quantile regression (or Least Absolute Deviations regression) which allows for contamination, viz. the possibility that some elements of the data come from a different distribution than that of the bulk of the data. Under these circumstances, OLS does not deliver consistent estimates. Methods have been developed to cope with *single* outliers, e.g. Cook's statistic, but this would be inadequate in our case – a large data set with several RHS variables – since there may be clusters of observations in n-tuple space which might not be detected by Cook's statistic, but which could greatly distort the coefficient estimates. A more rigorous way of coping with such situations of potential contamination is to use robust regression approaches. LAD or 50% quantile regression eliminates the distorting effect of y-outliers. The results are presented in Table 8.[18]

[17] Attempts have been made to implement a more general correction for endogeneity by lagging the variable of interest. However, given the cumulated nature of the error term, this strategy is not feasible. Future work may aim to address potential endogeneity bias more explicitly.

[18] The STATA module QRPD was used, which is adapted for panel regression.

Table 8. Estimates of model (8), using 50% quantile regression

	Dep. var: CPIA clusters A & B		Dep. var.: Economic Freedom Index	
Independent variables	**Coeff.**	**Robust s.e.**	**Coeff.**	**Robust s.e.**
$\sum_{i=1}^{t} A_i^{0.0}$	0.0500	0.0137***	0.443	0.183**
Cumulated length of engagement	-0.00180	0.00082**	-0.0128	0.0094
GDP per capita at PPP	0.0000179	5e-06***	0.000062	0.000073
Aid/GDP	-0.00357	0.149	-13.3	4.74***
Political Rights	-0.0512	0.0330	-0.786	0.268***
Country & year fixed effects	Yes		Yes	
N	2280		2518	
Number of countries	145		154	

Note: * = significance level of 10% or less; ** = 5% or less; *** = 1% or less. Standard errors are adjusted for 145 clusters.

The coefficients retrieved by quantile regression with the CPIA as the dependent variable are similar in size and significance to those retrieved by least squares (as reported in Table 5, p. 15). The coefficient on policy actions is 0.0500, which is within two standard errors of the coefficient estimated by least squares. The coefficient on the cumulated length of the engagement is statistically significant, and at -0.00180 it is less than a standard deviation away from the estimate by least squares. An exception is the variable Aid/GDP which is now statistically insignificant. There does not seem to have been much if any contamination of the sample, to judge by these results.

Similar comments may be made of the coefficients with the Economic Freedom Index as the dependent variable. The coefficient on policy actions is 0.443 which is statistically different from zero but is much smaller than the coefficient from least squares of 0.994 (see Table 7) and more than two standard deviations distant from it. The coefficient on the length of the engagement is negative, as before, but now it is not significant at conventional levels (p=0.174), which may be due to some measurement error in the dependent variable.

Since our dependent variables are censored – CPIA between 0 and 6 and Economic Freedom Index between 0 and 100 – we have estimated a Tobit model as a second robustness test. Results from this estimation are presented in Table 9, for both CPIA and Economic Freedom as the dependent variable. Table 9 indicates that censoring does not affect the main results: coefficient estimates and significance levels for policy actions and length of engagement are similar to those from the base model.

Table 9. Estimates of Tobit regression

	Dep. var.: Q_{it} = CPIA clusters A & B		Dep. var.: Economic Freedom Index	
Independent variables	**Coeff.**	**Robust s.e.**	**Coeff.**	**Robust s.e.**
$\sum_{i=1}^{t} A_i^{0.0}$ = # policy actions raised to γ = 0.0, cumulated	0.0741	0.0148***	0.9936	0.183***
Cumulated length of engagement	- 0.0025	0.0008***	- 0.0486	0.0130***
GDP per capita at PPP	0.000023	6.84e-06***	0.0001	0.0001
Aid/GDP	0.5077	0.1978***	-18.98	5.73***
Political Rights	- 0.00214	0.0222	- 0.204	- 0.302
Country & year fixed effects	Yes		Yes	
N	2280		2518	
Number of countries	145		154	

The third test of robustness is to examine whether the relationship between policy actions and the quality of government economic policy has changed over time. To do this we split the sample at breakpoints from 2000 to 2011, and estimate the main equation (8) for both the earlier sub-period and the later. The results are reported in full in Annex C, and may be summarized as follows: using the CPIA as the dependent variable, there is a hint of a decline in the coefficient on policy actions (α) at some point 2002 and 2007, but the change is not statistically significant. Using the Economic Freedom Index as the dependent variable, there are fluctuations over time, but these are not statistically significant. Hence, overall, we are not able to conclude that there are significant changes over time in the relationship between development policy operations and the quality of economic policy.

7. Discussion

We have established two 'stylized facts' about economic policy lending:[19] policy-based operations have a positive impact on the quality of economic policy; and the number of policy actions in a given year makes no difference beyond the first. We have also found some tentative evidence that the impact of a policy engagement is strongest in the early years and, while remaining positive, tends to fall in later years.

First we consider why the number of policy actions—beyond the first—does not matter. We surmise that this is because the engagement process or "policy dialogue" is key, not so much

[19] It is important to note that these findings are limited to economic policy actions, and may or may not be generalizable to policy-based operations in other areas.

the conditions that come after it (see Smets, 2018).[20] Indeed a large theoretical and empirical literature shows that without recipient commitment, conditional financing is unlikely to induce policy reform (see, e.g, Kilby, 2005; Svensson, 2000; World Bank, 2001; World Bank, 2005).[21]

Several aspects of the process are likely to support improvements in government economic policy. First, policy-based lending requires much information. Typically, a loan or credit will be preceded by studies and reports such as a public expenditure review, a country economic memorandum, sectoral studies, an investigation of the public debt, and so on. Sound evidence about the costs and benefits of policy proposals may reduce uncertainty about the distributional impacts of reform and thus induce government buy-in (Fernandez and Rodrik, 1991; Majumdar and Mukand, 2004). Relatedly, when countries lack the technical capacity to identify high-quality public policies, policy makers will be inclined to stick to the status quo. Providing high-quality evidence may lead to an increased valuation of public policies and a willingness to engage in reform (Besley and Persson, 2011).

Second, there is regularly a process of policy dialogue which occurs before a formal discussion of an operation, and which always involves discussions about macroeconomic 'adequacy'. This process may include seminars or official visits to other countries with similar economic reform challenges. Policy debate has the power to influence the mindset of policy makers, which may result in an increased commitment to reform (see Smets, 2018). Haggard and Webb (1994) note that in the long run, the transmission of ideas may be more important for policy reform than conditional financing.

Third, between the formal request for an operation and the signing of a loan agreement, several months may pass. The government will typically use this time to consult with the populace about the reform process. Khemani (2017) argues that communication by leaders is a critical component for reforms to be successful.

Finally, the list of prior actions (and/or tranche release conditions) in the program document that is finally approved by the Board of the Bank is not the sum total of the government's reform effort. The preparation process typically identifies a broader swathe of reforms, which the list of prior actions adumbrates but does not comprehensively define. The requirement that the country's macroeconomic policy stance be 'adequate' involves the World Bank team in lengthy policy debates and often helps governments to identify policy reforms, some of which might not be reflected in the program matrix. Sometimes task team leaders will

[20] A potential econometric reason why we do not find an effect of the number of policy actions is measurement error. That is, policy actions in policy-based lending are heterogeneous in scope and may have a differing impact, so that their number measures with error the 'true' force of the reform. This gives rise to attenuation bias which may diminish the estimate of γ. However, we performed a rough econometric test using lagged values and failed to find a heterogeneous impact of policy actions.

[21] This is not to say that conditions and financing are of no use whatsoever. For instance, conditions are needed for fiduciary reasons while budget support funds may help in closing the financing gap in recipient country budgets (Koeberle et al., 2005).

prefer not to list the worthy reforms that have already been adopted but will formulate a shorter list of elements that are still outstanding, in order to focus and incentivize efforts on the remaining agenda. Some of the latter prior actions may be likened to the wedge-shaped keystone of a masonry arch: the keystone may be one of fifty or a hundred wedges but without it the structure would collapse.

With this metaphor in mind, the count of prior actions and/or tranche conditions is an imperfect way of gauging the total reform effort. To quote the paper's title, it could be that the journey (process) is what matters. The tally of reforms done, or papers produced, or policy actions agreed to, is a contributing factor but perhaps not a critical one.[22]

An example is given from the Economic Management and Private Sector Operation (EMPSO) in Mozambique in the period 2000 to 2003 (P049878). As of 2000 there were only five fixed telephone lines per 100 inhabitants, among the lowest in Africa. There were high mobile and international tariffs, reflecting *inter alia* the monopoly of the incumbent, Telecomunicações de Moçambique (TDM), over international services.

In 2000, a meeting was held with the Minister of Transport and Communications at which the World Bank telecommunications staff and the Country Director presented the advantages of opening up the system to competition. Although the Minister was persuaded, his senior staff were not convinced. Nevertheless, the exploratory discussions went ahead and eventually the continued debate paid off, shifting the mindset of government officials. Furthermore, the Bank offered technical assistance for the regulation of the sector. Later a Communications Sector Reform Project was initiated by the Bank so as to strengthen the telecoms regulator and provide the conditions for a competitive market.

The monopoly of TDM was ended by decree and the government agreed to an auction of a second mobile license (the first being held by the incumbent). A second license was granted to Vodacom in August 2002. Meanwhile the EMPSO recognized, as a second tranche condition, the submission of a new telecommunications law to parliament which would provide for competitive allocation of licenses. The EMPSO operation was approved by the Bank's Board in late August 2002 and the second tranche was released in 2004.

The effects in the real economy took but a matter of months. For the first time ever, TDM ran an aggressive advertising campaign, sending out trucks of young people in orange uniforms to hold banners promoting the TDM mobile operation MCel. Prices of mobile phones fell even

[22] One way of demonstrating directly that it is the process that induces policy improvement would be to obtain data on analytical work done by the World Bank, the government and other instances and construct appropriate variables which could be inserted into the main equation (8).

before the second license was awarded. By 2004 there were 760,600 mobile phones and by 2008, 4.4 million.

The econometrician's only evidence for all this is one second tranche condition, that relating to the submission of the telecoms law. This is rather like the tip of the iceberg; the count of policy actions in the operation's matrix reflects only a part of the overall reform process and does not record the effect of persuasion and the Bank's technical assistance. In this case, it was the policy engagement that delivered the measured improvement in policy, not the sheer number of policy actions agreed upon.

As noted, the coefficient on the cumulative length of the engagement was negative in the least squares estimates using both dependent variables, and the quantile regressions with CPIA, but not in the quantile regression with the Economic Freedom Index. Hence the finding is best interpreted with caution and referred to as 'tentative'. On the assumption that there is something real behind the estimated coefficient, it is worth examining why the impact of policy-based lending may be larger in the initial years than in later years.

The question may be raised: is this due to a decline in the quality of the Bank's loans? One measure of the quality of the lending instruments is the evaluation performed by the Independent Evaluation Group (IEG) of the Bank. Every loan is assessed and its outcome awarded a rating—Highly Satisfactory, Satisfactory, Moderately Satisfactory, Moderately Unsatisfactory, Unsatisfactory or Highly Unsatisfactory. For ease we draw the line after Moderately Satisfactory (MS). A "Satisfactory" outcome means that the development policy objective of the loan was achieved, which in practice means that the bulk of the anticipated results defined in advance in the program document have been attained.

Table 10. Percentage of policy-based operations rated MS+ by length of the engagement

	0-5 yrs	6-10	11-15	16-20	21-30	Total
% MS+	73.5	82.9	73.8	66.9	76.0	75.1

Note: MS+ is the union of Highly Satisfactory, Satisfactory and Moderately Satisfactory. The sample is identical to that used for the estimation of equation (8), viz. SAL or DPF operations approved between 1998 and 2015 and which included at least one policy action germane to questions 1 to 6 of the CPIA.

Table 10 shows that the outcome ratings of policy-based operations average around 75.1 percent moderately satisfactory or better, and there appears to be no link with the length of the engagement. So the quality of the Bank's loans does not appear to be a factor driving this result. We need to look elsewhere.

One reason why the impact of policy-based lending may be larger in the initial years and smaller later may be the shift from first-generation actions to second-generation policy actions. In the early stages of the Bank's engagement with a country client, policy-based operations more frequently address first generation problems such as the removal of price controls or reducing the number of licenses required to operate a business—'stroke of the pen' reforms that do not

require high capacity on the part of government officials, that can be completed quickly, and that deliver quick returns. The subsequent reforms, which might be termed second-generation reforms, frequently need more government capacity, take longer, and may deliver lower returns in the short term, even if they deliver high returns in the long term.

An example is given from the agricultural sector.[23] All 467 policy actions in agriculture between 1995 and 2014 were classified into nine broad categories, as in the first column in Table 11. The idea was to distinguish between first-generation and second-generation reforms, where, generally speaking, first-generation reforms are less complex and their impact is of brief gestation, while second-generation reforms are more complex and their impact is of longer gestation.

Table 11. Policy actions in agriculture, 1995 to 2014, classified by degree of complexity

	# policy actions		**Total**
	'95-'04	**'05-'14**	**'95-'14**
"First generation reforms"			
Privatization, divestiture	62	1	63
Elimination of price controls, liberalization, deregulation, elimination of subsidies, water tariffs, elimination or restrictions of the scope of licensing, abolition of rationing, demonopolization	84	6	90
Tax (on income, or value added tax; not export or import tariffs)	6	2	8
Import and export duties, tariffs, and customs; non-tariff barriers, quantitative restrictions, duty drawbacks, trade agreements, foreign investment restrictions	57	2	59
Sub-total	*209*	*11*	*220*
"Second generation reforms"			
Institutions (including creation of institutions for research, extension, watershed management, training, monitoring and evaluation, crop insurance), decentralization, civil service reform in agriculture, procurement, reorganization, restructuring, corporatization, commercialization, agricultural state owned enterprises, outgrower systems, transparency, marketing information systems, product standards and certification, agriculture and environment services, input (fertilizer and seed) support programs, service delivery charters, veterinary and food safety	30	59	89
Creation or restructuring of funds; establishment or modification of security stocks or grain stocks; government purchases of agricultural products; restructuring of agricultural banks	12	6	18
Policy statements; studies and evaluations, including environmental impact assessments; functional reviews; passage of laws and regulations	13	18	31
Budget preparation for agriculture, Medium Term Expenditure Frameworks for agriculture, management information systems for agriculture, Public Expenditure Review for agriculture, budget flexibility appropriations, budget allocations to agriculture, public expenditure tracking in agriculture, auditing in the agriculture ministry	6	30	36

[23] In Annex D more examples can be found regarding the shift from first-generation reforms to second-generation policy actions.

Sub-total	*61*	*113*	*174*
Sample: All prior actions in the agricultural sector, 1995 to 2014, irrespective of whether the operation was sectoral or multi-sectoral. Prior actions focusing on land, fisheries, climate and environment without referring specifically to agriculture were omitted.			

In the period from 1995 to 2004 there were 209 first and 61 second generation reforms. In the period from 2005 to 2014 there were only 11 first generation reforms, but 113 second generation reforms. Part of the reason for this sharp contrast is that in the period 1995-2004 several former Soviet Union countries took adjustment loans, which typically focused on the transition to market economies, thus incorporating many 'stroke-of-the-pen' reforms such as abolition of price controls. Having started in semi-socialist mode, their CPIA had plenty of scope to improve on the basis of reforms with quick returns. It is no surprise that these earlier operations produced larger impacts on the CPIA. The CPIA had been developed in the 1980s and early 1990s, at which time one of the pervasive economic reforms called for was the shift from a planned to a market economy; and so it is no surprise that the CPIA weighs this element more heavily than it does the second-generation reforms. In the period 2005-2014, most of those 'stroke-of-the-pen' reforms had been completed, leaving knotty institution-building and other long-term issues with extended gestation periods to be dealt with.

A second illustration of the shift to second-generation reform is given by quantile regressions of the main model (8), at the 25th, 50th (as in Table 8, p. 20) and 75th percentiles.[23] To save space we report not the full quantile regression in each of the six cases, but only the coefficient on policy actions, in Table 12.

Table 12. Coefficients on policy actions and on the length of the engagement in the main model (8), at different percentiles, using quantile regression[24]

	CPIA		**Economic Freedom Index**	
Regression quantile	**Policy actions (α)**	**Length of engagement (β)**	**Policy actions (α)**	**Length of engagement (β)**
25th percentile	0.0780***	-0.00288***	0.704***	-0.0171
50th percentile	0.0500***	-0.00180**	0.443**	-0.0128
75th percentile	0.0355**	-0.00127*	0.466**	-0.0138

Notes: (1) The coefficient on policy actions is α in equation (8), on the assumption that the coefficient γ is zero. Thus the independent variable for policy actions becomes $\sum_{i=1}^{t} A_i^{0.0}$ = # policy actions raised to γ = 0.0, cumulated. (2) * = significance level of 10% or less; ** = 5% or less; *** = 1% or less. Standard errors are adjusted for 145 clusters.

The impact of policy actions is much stronger for countries whose CPIA or Economic Freedom Index is near the 25th percentile than for those more successful countries whose ratings are near the 75th percentile. In fact the 'punch' from policy actions at the 25th percentile is

[24] The STATA module QRPD was used, which is adapted for panel regression.

approximately double that at the 75th. This corroborates the sense that many country economists have, that it is much easier to raise the CPIA from 2.0 to 3.0 than to raise it from 4.0 to 5.0.

8. Summary and concluding remarks

In this study we investigate the impact of World Bank development policy operations on the quality of economic policy, covering the period 1998-2015. (Note that purely sectoral operations, and the areas of public financial management, tax and governance are beyond the scope of this paper.) We have developed a new theoretical framework that provides clear criteria for interpretation of the estimated coefficients, and distinguishing clearly among three effects which have been conflated in previous work: (a) differing marginal impacts of additional policy actions within the current year; (b) the effect of the length of the policy engagement with client countries, and (c) changes over time in the marginal impact of policy actions. We have also created a new data set by reclassifying all policy actions ever taken in the context of Bank budget support operations, and distinguishing the tranche to which such policy actions belonged and hence their timing. Since the data set involves time series, we have deployed standard cointegration techniques.

The key finding of the paper is that policy actions have a positive effect on the quality of economic policy, whether the latter is measured by Clusters A and B of the Bank's CPIA or by the Economic Freedom Index of the Heritage Foundation. If there is a policy debate in a particular year which issues in at least one germane policy action in a development policy program, then the CPIA is likely to rise by, on average, about 0.074. The number of germane policy actions in a particular year does not seem to make any measurable difference. The key element is that there be at least one policy action, and this is likely proxying for the policy debate and the overall engagement.

There is also tentative evidence that the initial years of an engagement are the most productive of improvement in government economic policy. The policy engagement continues to have a positive impact in later years, but the impact may be less. We present evidence suggesting that this is because after a long engagement the opportunities for first-generation reforms such as agricultural liberalization or reduction of import tariffs may be exhausted. While many reforms are still needed, these second-generation reforms take longer and require higher levels of skill and commitment on the part of civil servants because they have to do with complex institutional reform.

Some areas for future research may be mentioned. First, it would be worth conducting a *direct* test of the supposition that it is the process of generating a policy-based loan that induces the subsequent improvement in government economic policy. For example, one way to do so would be to re-estimate the main equation (8) incorporating variables for the amount of analytical and advisory support provided . Second, further research may show up considerable heterogeneity among different sub-groups of the sample. For instance, there may be regional

factors that differ and that would deliver different insights about the impact of policy-based lending. Another aspect of heterogeneity could be explored by splitting up the dependent variables by the six economic questions in the CPIA. So, for instance, the impact of policy-based lending may differ as between interventions in trade and interventions in private sector regulation. Third, one could study the impact of purely sectoral operations (education, health, infrastructure etc.), if suitable dependent variables could be identified. Finally, given that macroeconomic crises often usher in periods of fundamental structural reform, it may be worth relaxing our assumption of a constant coefficient on the cumulative length of the engagement, so as to explore the possibility of cycles of reform effectiveness.

References

Besley, T., Persson, T., 2011. Pillars of Prosperity. Princeton University Press, Princeton and Oxford.

Bogetic, Z. & Smets, L., 2017. Association of World Bank policy lending with social development policies and institutions. Policy Research Working Paper Series 8263, The World Bank.

Devarajan, S., Dollar, D., Holmgren, T., 2001. Aid and reform in Africa: Lessons from ten case studies. World Bank, Washington, D.C.

Dollar, D., Svensson, J., 1998. *What explains the success or failure of structural adjustment programs?* Macroeconomic and growth group, The World Bank, April 1998. Also: *Economic Journal* 110 (466): 894-917, October.

Dreher, A., 2002. The development and implementation of IMF and World Bank conditionality. Hamburg: HWWA Discussion Paper 165.

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.

Kapur, D., Lewis, J. P., Webb, R., 1997. The World Bank: Its first half century. Volume 1. History. Brookings Institution Press, Washington, D.C.

Koeberle, S. G., 2003. Should policy-based lending still involve conditionality? The World Bank Research Observer 18 (2), 249–273.

Moll, P., Geli, P., Saavedra, P., 2015. *Correlates of success in World Bank development policy lending*. Policy Research working paper no. WPS 7181. Washington, DC: World Bank Group. January.

Paris Declaration on Aid Effectiveness, 2005. Issued jointly by 91 countries, most international aid organizations, and 14 civil society organizations. On the web at http://www.oecd.org/dac/effectiveness/34428351.pdf

Navia, P., Velasco, A., 2003. The politics of second-generation reforms. In: Kuczynski, P.-P., Williamson, J. (Eds.), After the Washington Consensus: Restarting Growth and Reform in Latin America. Institute for International Economics.

OPCS, 2009. Country policy and institutional assessments: 2009 assessment questionnaire. Operations Policy and Country Services, World Bank.

Rodrik, D., 2006. Goodbye Washington consensus, hello Washington confusion? Journal of Economic Literature 44 (4), 973–987.

Smets, L., and Knack, S., 2018. World Bank policy lending and the quality of public sector governance. Economic Development and Cultural Change, 67: 29-54.

Smets, L., and Knack, S., 2016. World Bank lending and the Quality of Economic Policy. The Journal of Development Studies 52: 72-91.

Smets, L., 2018. Supporting Policy Reform from the Outside. mimeo

Svensson, J., 2003. Why conditional aid does not work and what can be done about it? Journal of Development Economics 70 (2), 381–402.

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, 1989. Adjustment Lending: An Evaluation of Ten Years of Experience. World Bank, Washington, D.C.

World Bank, 2004. 2003 Annual Review of Development Effectiveness: The effectiveness of Bank support for policy reform. World Bank, Washington, D.C.

World Bank, 2006. Development Policy Retrospective 2006. World Bank, Washington, D.C

Annex A. Variables and summary statistics

Table 13. Variable definitions and sources

Variable	Definition	Source
CPIA Cluster A	Assessment of the quality of a country's economic management	World Bank
CPIA Cluster B	Assessment of the quality of a country's structural policies	World Bank
Number of prior actions	Count of prior actions germane to Clusters A & B of the CPIA	World Bank
GDP per capita	GDP per capita, PPP	World Development
Aid over GDP	Net Overseas Development Assistance and official aid over GDP (international $)	Based on WDI
Political Rights	Measure for political rights	Freedom House
Economic Freedom Index	An index of twelve elements of economic freedom	Heritage Foundation

Table 14. Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.
Average of CPIA Clusters A & B	3.57	0.48	1	5.85
Policy actions germane to CPIA Clusters A & B, cumulated	62.4	64.4	0	348
Policy actions germane to CPIA Clusters A & B, squared, and then cumulated	1147.4	1969.5	0	15132
Policy actions germane to CPIA Clusters A & B, raised to the power 0 [25]	0.29	0.45	0	1
Policy actions germane to CPIA Clusters A & B, raised to the power 0, and cumulated [26]	6.56	5.80	0	30
Length of engagement, viz. number of years in which there was at least one policy action germane to Clusters A & B	6.56	5.80	0	30
Length of engagement, cumulated	82.3	90.0	0	542
GDP per capita, PPP	6304.7	5792.5	399.9	48710.7
Aid/GDP	0.0712	0.1083	-0.0243	1.520
Political Rights	3.785	1.958	1	7
Economic Freedom Index of the Heritage Foundation	63.4	9.69	17.2	91.7

[25] This means setting γ=0 in equation (3) and finding $A_t^{0.0}$.

[26] This means setting γ=0 in equation (4), and finding $\sum_{i=1}^{t} A_i^{0.0}$.

Annex B. 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 four "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 rating of 1 corresponds to very weak performance, and a rating of 6 to 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. Rather than rescale the CPIA for these years, we decided to omit these years altogether and take the sample from 1998 to 2015 only.

To convey an idea of how the CPIA has changed over the years, it is instructive to compare the guidance given in 1999[27] and that in 2015.[28] For the sake of compactness, only the criteria for a rating of '2' and of '5' are given, and that only for Question 4 on trade.

In 1999, the guidance for question 4 on trade, for ratings of '2' and '5' was:

> "2. Average tariff (weighted by global trade flows) is high (over 30%). High and erratic import and/or export barriers, including quantity restrictions and duty exemptions. Customs authorities make discriminatory or ad hoc exemptions and valuations. Administered foreign exchange regime with multiple exchange rates.
>
> "5. Average tariffs (weighted by global trade flows) is low (10% or less), with low dispersion and insignificant or no quantitative restrictions or export taxes. No trading monopolies. Indirect taxes (e.g. sales, excise, surcharges) do not discriminate against imports. Efficient and rule-bound customs administration. IMF Article 8 status. Minimal or no foreign exchange restrictions on long-term investment capital inflows."

In 2015, the guidance for question 4 on trade, for ratings of '2' and '5' was:

> "2. a. Very discretionary and discriminatory trade regime with widespread and discretionary use of highly trade-restrictive NTBs (e.g. quantitative restrictions); no consultation for NTMs; high and variable taxation of imports or exports (e.g., average MFN tariffs plus all other taxes collected at the border above 15%), with discretionary and nontransparent exemptions; haphazard application of special import regimes, and refunds (e.g., duty drawbacks) rarely made; several changes in export regulations each year; severe limitations on service trade with high level of discretion (e.g., STRI between 60 and 80); very limited data transparency (e.g., through collection and publication of statistics).
>
> "b. Some trade facilitation strategy at national or agency level, some one-way dialogue with stakeholders. Some ad hoc data available on clearance times. Difficult to obtain latest and complete information on trade-clearing procedures and formalities. Appeal mechanism established but difficult to access. Widespread corruption in border management agencies, but visible attempts to address the issues. Heavy reliance on physical inspection of goods, but some risk-based selectivity (e.g., over 75% Customs physical inspection). Frequent delays and unreliability in the clearance of goods. Customs declarations submitted in paper form but processing supported by IT system in Customs. Limited access to foreign transport service providers (e.g., freight for both domestic and transit movement is allocated by quota system).
>
> "5. a. Low barriers and transparent trade regime with limited use of NTBs; formal review process for NTMs, low taxation of imports or exports (e.g., average MFN tariffs plus all other

[27] See International Development Association, 2000. *IDA Country Performance Rating Process: Annual Report 1999.* February.
[28] See The World Bank Group, 2015. *CPIA 2015 Criteria.* September 24, 2015.

taxes collected at the border between 5-10%) with little variability; very few and limited special import regimes; moderate use of anti-dumping regulations (between 2 and 5 new measures in the year); few restrictions on service trade (e.g., Services Trade Restrictiveness Index between 10 and 25); trade data beyond regular statistics is made available for policy analysis.

"b. Clear and consistent trade facilitation strategies at national and agency levels. Formal processes for consultation and cooperation with stakeholders. Performance standards established and monitored and data on clearance times publicly available. Information on trade-clearing procedures and formalities is published and accessible via the Internet. The hierarchy of regulatory framework and the links are clear. Appeal mechanisms are established and accessible. Few instances of corruption in border management agencies. Risk management used extensively (e.g., < 20 % Customs physical inspection). Speedy and predictable processing of goods (e.g., some facilitated procedures for traders with good compliance records). IT system used by Customs and other key border management agencies, with some interconnectivity. Largely open access to foreign transport service providers (e.g., transit transport services are fully liberalized within an economic grouping and transit routes are not specified on the permit)."

The 1999 and 2015 criteria differ in that (a) the 2015 criteria are more detailed and longer, (b) the 2015 criteria have been modernized by, for instance, frequent references to the internet which was not a factor in 1999, and (c) the 2015 criteria refer to indexes which were unavailable in 1999, such as the Services Trade Restrictiveness Index.

On the whole, however, it appears that the overall direction of the criteria was rather similar in 2015 to what it was in 1999. For instance, in 1999 to get a '2', the customs authorities "make discriminatory or ad hoc exemptions and valuations", while in 2015 to get a '2', the customs authorities run a "very discretionary and discriminatory trade regime with widespread and discretionary use of highly trade-restrictive NTBs … with discretionary and nontransparent exemptions [and] haphazard application of special import regimes".

Annex C. Robustness test: Has the relationship between development policy operations and the quality of economic policy changed over time?

We examine changes over time by splitting the full sample (1998 to 2015) into sub-periods. We define breakpoints from 2000 to 2011, and estimate (8) for both the earlier sub-period and the later. As before, we estimate γ by a grid search. We find that in almost all subsamples the optimal "between" R-squared occurs where γ=0. For brevity's sake we report only the coefficient estimates for α, their significance, the corresponding estimate for γ, and the size of the sub-sample. We start with the estimates using Clusters A & B of the CPIA as the dependent variable.

Table 15. Estimates of α and γ in model (8) (with CPIA as dependent variable), for different sub-periods

	Coefficients					Coefficients			
Sub-SMPL	**α**	**Γ**	**N**		**Sub-SMPL**	**α**	**γ**	**N**	**Change**
98-99	-		266		00-15	0.0645***	0.0	2014	
98-00	-		399		01-15	0.0616***	0.0	1881	
98-01	0.00150*	1.40	534		02-15	0.0552***	0.0	1746	
98-02	0.00485*	1.00	669		03-15	0.0526***	0.0	1611	
98-03	0.0891***	0.0	804		04-15	0.0494***	0.0	1476	-45%
98-04	0.109***	0.0	938		05-15	0.0467***	0.0	1342	-57%
98-05	0.113***	0.0	1061		06-15	0.0429**	0.0	1219	-62%
98-06	0.111***	0.0	1188		07-15	0.0394**	0.0	1092	-65%
98-07	0.105***	0.0	1315		08-15	0.0277*	0.0	965	
98-08	0.104***	0.0	1443		09-15	0.0177	0.05	837	
98-09	0.100***	0.0	1571		10-15	0.00978	0.40	709	
98-10	0.0947***	0.0	1699		11-15	0.00905	0.55	581	

Note: * = significance level of 10% or less; ** = 5% or less; *** = 1% or less. Standard errors are adjusted for 145 clusters. OLS used. Complete sets of country dummies and year dummies are included in all regressions. Before regressing, it was found by a grid search that γ=0. Cells are left empty in the few cases when there was no convergence; in these cases the estimated coefficient α was statistically insignificant.

In Table 15 we contrast the earlier with the later periods, e.g. in the second line, we compare 1998-2000 with 2001-2015. We are on the surest grounds when we compare periods *both* of whose coefficients are significant at 5% or better: this is the shaded (yellow) part of the table. (In the remainder of the table there are too few observations to get significant estimates, making it hazardous to draw conclusions.) The estimate of α is larger in each earlier period than in the equivalent later period: for instance, in the earlier period 1998-2003 the coefficient is 0.0891, and in the later period 2004-2015 the coefficient is 0.0494. The decline between the earlier and the later periods varies between 45% and 65%. The differences are bordering on

statistical significance. For instance, the 5% confidence interval for the estimate of α of 0.0891 for 1998-2003 is [0.040, 0.139] which includes the point estimate of 0.0494 for the period 2004-2015. On the basis of this comparison alone we could not conclude that there is a *statistically significant* decline. However, the changes in all four of the yellow-shaded periods are in the same direction, and the changes using breakpoints from 2007 onwards suggest – not prove – that there is a real effect in the earlier period but none in the later. Putting all of this together it appears that at some point between 2003 and 2006 there may have been change in the relationship between policy actions and the quality of government policy as measured by Clusters A & B of the CPIA. However, we will have to await more or better data sets before we can be sure.

To explore further the possibility of regime change, we deploy our second dependent variable, the Economic Freedom Index of the Heritage Foundation. As described above in the section about the data set, the content of the Economic Freedom Index largely overlaps with the content of the CPIA. We split the samples in the same way as we did for Table 15. The results are presented in Table 16.

Table 16. Estimates of α and γ in model (8) (with Economic Freedom Index as dependent variable), for different sub-periods

	Coefficients					Coefficients			
Sub-SMPL	**α**	**γ**	**N**		**Sub-SMPL**	**α**	**γ**	**N**	**Change**
98-99	0.786	0.0	253		00-15	0.849***	0.0	1953	
98-00	0.463**	0.46	385		01-15	0.774***	0.08	1821	-9.3%
98-01	0.442**	0.42	512		02-15	0.697***	0.16	1694	3.8%
98-02	0.624***	0.28	639		03-15	0.654***	0.26	1567	1.5%
98-03	0.598***	0.30	767		04-15	0.606***	0.34	1439	8.1%
98-04	0.676***	0.24	895		05-15	0.585***	0.34	1311	1.6%
98-05	0.376**	0.34	1002		06-15	0.583***	0.30	1204	45.4%
98-06	0.300*	0.36	1111		07-15	0.618***	0.20	1095	
98-07	0.336*	0.28	1220		08-15	0.717***	0.04	986	
98-08	0.444*	0.16	1329		09-15	0.757***	0.0	877	
98-09	0.636**	0.02	1458		10-15	0.712***	0.0	748	8.4%
98-10	0.767**	0.0	1586		11-15	0.653***	0.0	620	-14.9%
98-11	0.822***	0.02	1711		12-15	0.695***	0.0	495	-18.1%
98-12	0.824***	0.04	1833		13-15	0.639**	0.0	373	-27.3%
98-13	0.833***	0.04	1959		14-15	0.0956	0.62	247	

Note: The column "Change" is the percentage change in $\Delta Q_t = \alpha A_t^{\gamma}$ between the earlier and the later period, after substituting the estimated values for α and γ, and 5 policy actions for A_t.

Again, we contrast the earlier with the later periods, e.g. in the second line, we compare 1998-2000 with 2001-2015. Yellow shading indicates periods *both* of whose coefficients are significant at 5% or better, and where comparisons are best made. In order to make direct comparisons we cannot now compare the estimated α and γ directly, because they work together

nonlinearly. Instead we need to substitute the estimated α and γ into equation (3), namely $\Delta Q_t = \alpha A_t^{\gamma}$, for the earlier period and the later period, having settled on a figure for A_t. We propose that we use 5 policy actions for A_t inasmuch as from 2003 to 2015 the average number of policy actions per operation in the realm of "economic reform" was about 5 (see Figure 1, p. 9). We present the percentage change from the earlier to the later period in the final column entitled "Change".

It is hard to discern any clear change over time in Table 16. There is a suggestion of an increased effect of 45 percent as between period 1998-2005 and 2006-2015, but this increase is not supported by the adjoining periods; and in any case the estimate of α for the period 2006-2015 does not differ statistically from the estimate for 1998-2005 owing to the wide confidence intervals. Similarly the apparent decline of the impact, of 27.3 percent, between the period 1998-2012 and 2013-2015, is not statistically significant. We conclude that the impact of development policy operations on the Economic Freedom Index has fluctuated somewhat over time but there is no discernible trend.

Annex D: Additional illustrations of first- vs. second-generation reforms

Another illustration of the evolution from first- to second-generation reform is given by the Bank's budget support for Romania. In the 1990s Romania transformed its economy with a swift and ambitious creation of markets and divestments of most state-owned enterprises. In 1991, the government instituted a land reform program that resulted in 80 percent of agricultural land being privately owned. The housing stock was privatized. The majority of state-owned companies were transformed into commercial entities. Prices were liberalized from mid-1993, including elimination of price controls on industrial and agricultural products, energy and agricultural products, energy products and consumer goods. A VAT system was introduced. An open foreign trade regime has been instituted, and a more realistic and liberalized exchange rate established.

One of the key elements here was privatization. The Bank assisted in this with the Financial and Enterprise Sector SAL of 1996 (P008773). The operation covered privatization, the stimulation of private investment, financial sector supervision, and the opening of the Bucharest Stock Exchange. A Board presentation condition was that 1,500 firms be privatized; a second tranche condition was that a cumulative 2,750 firms be privatized, and a third tranche release conditions raised this to 3,600. In 2004-2006 the three-part Programmatic Adjustment Loan series[29] had among its prior actions to complete the privatization of the manufacturing sector, to privatize the two remaining commercial banks, and to close down the privatization agency APAPS.

The Bank's involvement with Romania continued with further DPF between 2010 and 2015, with a focus on the Medium Term Expenditure Framework (MTEF), social assistance, health and education reforms, and further reform of the financial sector.

Thus, the first period up to about 2006 dealt to a large extent with first-generation reforms; having completed the bulk of these reforms, the country proceeded in the latter period, from 2010 to 2015, to deal with second-generation reforms such as the MTEF. The impact on the economic reform elements of the CPIA, viz. Clusters A&B, was in accord: the average was 3.0 in 1998, and had risen to 4.0 by 2008, whereupon its progress slowed, rising to only 4.08 by 2015. Similarly, the Economic Freedom Index rose swiftly from 58.5 in 1998 to 69.5 by 2008, and then slowed, rising to 72.9 by 2015.

A final illustration of the shift to second-generation reform comes from the Bank's involvement with reducing tariffs on imports. The scope for tariff reduction was large in the 1990s when many less developed countries had average tariffs exceeding 20 percent.[30] Several

[29] The first PAL was P008791.

[30] Consider the following weighted average tariffs on manufactured products, in Sub-Saharan Africa up to the year 2000: Kenya 23% in 1994, Malawi 22% in 1997, Mauritius 27% in 1998, Nigeria 58% in 1995,

operations financed by the World Bank focused on reducing average tariffs. For instance, the Industry and Trade Policy SAL for Madagascar in 1987 (P001511, p. 37f) had second tranche release conditions eliminating all import prohibitions, simplifying the system of import tariffs, and reducing their maximum to 80 percent. Release of the third tranche was conditional upon adoption of a four-year tariff reduction program aimed at reducing average protection from its current level of 46 percent to 35 percent.

Similarly, in Kenya in 1990, by operation P001653 (paragraph 2.09), the average tariff was to fall from its current level of 20 percent by five percentage points. In Sierra Leone in 2005, the average tariff of 17.5 percent was to come down to 13.3 percent, supported by the Fourth Economic Rehabilitation and Recovery Grant in 2005 (P083477, paragraph 68).

Scope for further reductions in tariff levels soon fell, and with it the level of ambition of programs supported by policy-based reform. In Mauritius, the weighted mean of tariffs on manufactured products was 27% in 1998; this had fallen to 14% by 2004[31] and to 6.5 percent by 2006.[32] The Bank supported Mauritius with the First Trade and Competitiveness DPL in 2006, in virtue of which the average tariff rate was to fall by two percentage points (P101570, paragraph 85).

Rwanda 26% in 1993, Seychelles 25% in 2000, Tanzania 19% in 1988, Zimbabwe 39% in 1996. See https://data.worldbank.org/indicator/TM.TAX.MANF.WM.AR.ZS?contextual=region&end=2016&locations=ZG-ZA-MW-MG-MU-ZW-TZ-SZ&name_desc=true&start=1988&view=chart.

[31] Indicator TM.TAX.MANF.WM.AR.ZS in the World Bank's open data source; derived from the World Integrated Trade Solution system, in turn based on data from United Nations Conference on Trade and Developments Trade Analysis and Information System database, and the World Trade Organization's Integrated Data Base and Consolidated Tariff Schedules database.

[32] The 2006 number is from the program document of P101570, paragraph 17.