WPS6798

POLICY RESEARCH WORKING PAPER 6798

What Factors Predict How Public Sector Projects Perform?

A Review of the World Bank's Public Sector Management Portfolio

Jürgen René Blum

Abstract

This paper uses regression analysis to identify which country context, reform content, process, and project management variables predict the performance of public sector management projects, as measured by the Independent Evaluation Group's project outcome ratings. The paper draws on data from a large sample of World Bank public sector management projects that were approved between 1990 and 2013. It contributes to an emerging literature that uses cross-country regressions to analyze public sector management reform patterns. The findings suggest that political context factors have a greater impact on the performance of public sector

management projects than on other projects. Specifically, public sector management projects perform better in countries with democratic regimes than autocratic ones. They fare better in the presence of programmatic political parties and in more aid-dependent countries. Project managers' subjective risk assessments predict performance in public sector management operations better than objective risk indicators. These findings suggest that the performance of public sector management projects would benefit from a better alignment of project design with political context and from a more open dialogue about risk between task team leaders and management.

What Factors Predict How Public Sector Projects Perform?

A Review of the World Bank's Public Sector Management Portfolio

Jürgen René Blum

Table of Contents

List of Acronyms

ATET	average treatment effect on the treated
CAE	Country Assistance Evaluation
CPIA	Country Policy and Institutional Assessments
CS	civil service
CSR	civil service reform
DAC	Development Assistance Committee
DEC	decentralization
DO	development objectives
DOD	debt outstanding and disbursed
DPI	database of political institutions
ED	education (sector board)
ES	evaluation summaries
EM	evaluation memoranda
DPL	development policy loan
ENV	environment (sector board)
EP	economic policy (sector board)
FH	Freedom House
FPD	financial and private sector development (sector board)
FY	fiscal year
GDP	gross domestic product
GIC	gender and inclusive (sector board)
GNI	gross national income
HE	health (sector board)
IBRD	International Bank for Reconstruction and Development
ICR	implementation completion report
ICRG	International Country Risk Guide
IDA	International Development Association
IEG	Independent Evaluation Group
IL	investment lending
IP	implementation progress
ISR	implementation supervision report
LAC	Latin America and the Caribbean
MDA	ministry, department, agency
MS+	"moderately satisfactory" or higher IEG outcome rating
MTEF	medium-term expenditure framework
ODA	official development assistance
OECD	Organisation For Economic Co-operation and Development
OED	Operations Evaluation Department
OLS	ordinary least squares
ORAF	Operational Risk Assessment Framework
OVB	omitted variable bias
P4R	Program-for-Results
PCN	project concept note
PCR	project completion report
PDIA	problem-driven iterative adaptation
PDO	project development objective
PEFA	public expenditure and financial accountability
PFM	public financial management
PPAR	project performance audit report

PPP	purchasing power parity
PS	public sector (sector board)
PSGB	Public Sector Governance Board
PSM	public sector management
RDV	rural development (sector board)
SDV	social development (sector board)
SP	social policy (sector board)
TR	trade (sector board)
TTL	task team leader
UD	urban development (sector board)
WAT	water (sector board)
WDI	World Development Indicators
WDR	World Development Report
WGI	World Governance Indicators

1 INTRODUCTION

What factors predict the performance of public sector management (PSM) projects? This paper explores this question through an econometric analysis of the World Bank's investment lending project portfolios approved between 1990 and 2013. It identifies observable data on country contexts, reform content and processes, and project management variables that predict the performance of PSM projects, as rated by the World Bank's Independent Evaluation Group (IEG).

The paper has two main objectives: first, to inform the broader debate on what works and why in PSM reforms; second, to manage uncertainty in PSM projects. For example, do tough PSM reforms fare better under autocratic rule—as in China, Rwanda, and Singapore—or under the incentives provided by inclusive political institutions? These and other questions are investigated. The resulting analysis may help manage uncertainty in PSM projects, by guiding the allocation of managerial attention and Bank budgets. It may also help to adjust risk in the Bank's PSM lending portfolio to its risk appetite.

The paper uses ordered probit regression estimates to identify performance predictors, employing a project sample of 1,097 PSM projects. It then compares these against a sample of 2,105 non-PSM projects, to ask if country characteristics affect PSM project performance to a greater or lesser extent than other World Bank projects. This difference is identified, among other ways, based on nearest-neighbor and subclassification matching estimators that match PSM projects with similar non-PSM projects in the same country.

The paper contributes to a small but growing set of studies that employ cross-country regression analysis to better understand PSM reform patterns. Such analysis has been constrained by lack of reliable comparative data on public management system properties. Among the few cross-country regression studies on PSM reform, that of Evans and Rauch (1999) on Weberianism and growth is a famous early example. More recently, a number of researchers have used the Public Expenditure and Financial Accountability (PEFA) data set (for example, Andrews 2009) or have coded particular finance management system properties, such as medium-term expenditure frameworks (MTEFs) (World Bank 2012a). This paper focuses on the corporate performance of World Bank PSM projects, and thus cannot be used to generalize about PSM reforms overall.

The remainder of this paper is structured as follows. Section 2 situates it within the existing literature on PSM reform. Section 3 sets out the theoretical questions and hypotheses that underlie the choice of variables included in the review. Section 4 presents underlying data; section 5 sets out identification strategies; and section 6 provides descriptive statistics, comparing the nature and targeting of PSM and non-PSM projects. Section 7 summarizes the major estimation results regarding the predicators of PSM

project performance and interprets them in the light of theoretical expectations. Section 8 concludes. The annex contains descriptive statistics and estimation results.

2 RELATED LITERATURE

This paper relates to three strands of research. In terms of substance, it contributes to (i) literature on PSM reform in developing countries and draws on (ii) broader political economy literature. By way of its methodology and empirical basis, it is part of (iii) a small but growing number of econometric studies that employ IEG outcome ratings of World Bank projects as their dependent variable.

The extensive literature on PSM reform in developing countries is dominated by qualitative work on public administration reform trajectories.[1] By contrast, few studies conduct cross-country regression analysis or evaluate the impact[2] of specific PSM reforms. Such analysis has been constrained both by a lack of consensus[3] on how to measure the performance of public administrations and by a lack of comparative data. Over the past decade, international organizations have made significant progress in generating such data: for example, the "Government at a Glance" of the Organisation for Economic Co-operation and Development (OECD, 2013), and the PEFA data,[4] which measure public financial management (PFM) arrangements in relation to an agreed normative framework in over 100 countries. These data sets have enabled pioneering research, such as that of Andrews (2010), who tests whether empirical PFM reform patterns are consistent with isomorphism theory. Despite these advances, comparative data on PSM systems remain patchy in terms of substance and geographic coverage.[5]

This paper contributes toward filling the gap in cross-country regression research on PSM reforms by relying on IEG outcome ratings of Bank PSM projects as the dependent variable. IEG ratings are available for a large number of PSM projects and countries, thanks to the World Bank's role as a leading financier of PSM reforms. Further, they do not imply a universal definition of a "good" PSM reform result. They measure how PSM projects perform *relative* to their pre-set objectives, not against an

---

[1] Examples of comparative qualitative studies include Pollitt and Bouckaert (2004) and Levy and Fukuyama (2010).

[2] Early examples of impact evaluations of public administration include Bandiera and others (2009) on a natural procurement experiment in Italy, Faguet (2004) on decentralization in Bolivia, as well as Dal Bo and others (2013) and Rasul and Rogger (2013) on civil service management questions.

[3] Achieving consensus on how to measure the performance of public administrations is challenging for three reasons. First, public bureaucracies serve a range of functions, and there is no single "good" model for the scope and prioritization of these functions. For example, the size and responsibilities of the public sector vary enormously across both the OECD and developing countries. As Isham, Kaufmann, and Pritchett (1995) point out, "deep conceptual differences about what governments ought to do [. . .] imply that efficacy cannot be inferred from the success and failure of achieving measured aggregate outcomes." Second, changes in broad, less controversial government performance measures—such as child mortality (Andrews, Hay, and Myers, 2010)—are typically hard to attribute to specific reforms and thus of limited use. Third, broad governance measures, such as the World Governance Indicators (WGI), have frequently been criticized for being loaded with (unfounded) assumptions about the institutional forms governments should take (see, for example, Andrews, Hay, and Myers 2010). The World Bank's own Country Policy and Institutional Assessment (CPIA) ratings (available since 1999) are among the few data sets that reflect some degree of consensus (at least among donors) and include ratings of public administration issues (CPIA scores 12 to 16).

[4] The OECD's International Budget Practices and Procedures Database is a second example of such a data set (http://www.oecd.org/gov/budget/database).

[5] The World Bank is currently launching a multidonor effort to expand the scope of comparative "Indicators of the Strength of Public Sector Management Systems"; http://go.worldbank.org/SGO4LFRSS0.

absolute performance standard. But they also have a major downside: a narrow focus on the World Bank's PSM projects and their performance from a corporate perspective. The contribution of this review to the literature on PSM reform is thus within a niche—it tests whether theoretical claims emerging from the broader literature on PSM reform can explain performance patterns in PSM projects—but its findings need not have external validity for government PSM reforms.

Second, this paper applies basic concepts from the political economy literature to predict PSM reform patters. It looks at how "inclusive" or "extractive" political (and economic) institutions (Acemoglu and Robinson 2012) or, similarly, "open-access" and "limited-access orders" (North, Wallis, and Weingast 2009) affect PSM projects. More specifically, it looks at the role of aid and natural resource rents in undermining the accountability relationship between the state and its citizens as taxpayers (see, for example, Bates 1992 and Knack 2002). The paper also considers political market imperfections (see Cruz and Keefer 2010), in particular the absence or existence of programmatic political parties as a potential determinant of PSM project performance.

Third, the paper adds to previous studies that employ IEG outcome measures by (i) exploring the performance determinants of public sector projects as a specific subset of the Bank's larger project portfolio; and (ii) by comparing performance determinants for PSM projects to performance determinants for the broader universe of Bank investment lending projects.

It closely relates to recent work by Denizer, Kaufmann, and Kraay (2011), who employ a large sample of World Bank–supported projects across sectors to identify "macro- [that is, country context] and micro- [that is, project management] correlates of World Bank project performance." Earlier work in this direction includes Dollar and Levin's (2005) review of "institutional quality and project outcomes in developing countries," which explores World Bank–supported projects across sectors and countries. Two studies stand out for their similarity to this review. Cruz and Keefer (2010) employ performance data on World Bank–supported public sector projects to test the theoretical prediction that the existence of programmatic political parties positively influences politicians' incentives for developing a well-performing public administration. An early paper by Isham, Kaufmann, and Pritchett (1995) closely relates to this paper in that it explores the impact of civil liberties and democracy (as measured by Freedom House indicators) on the performance of World Bank–supported projects (but takes no specific interest in the performance of public sector projects).

In sum, this paper contributes to the mostly qualitative literature on PSM reform by testing whether theory can help explain the performance ratings of World Bank PSM projects. Within this niche, it contributes to filling a gap of cross-country empirical research on the determinants of PSM reform

outcomes. It complements related studies that draw on similar data sets but pursue different sets of questions.

3   RESEARCH QUESTIONS

This review pursues three guiding research questions:

- *Question 1 (Q₁).* Which key country context, reform content, process, and project management characteristics predict the performance of PSM projects?

- *Question 2 (Q₂).* Do PSM projects perform differently from other World Bank projects in similar countries?

- *Question 3 (Q₃).* Do certain country characteristics affect PSM project performance to a greater or lesser extent than they affect the performance of other World Bank projects?

This section identifies observable factors expected to predict IEG ratings (Y): country context ($\vec{C}$), reform content ($\vec{R}$), reform process ($\overrightarrow{P}$)[6] and project management ($\vec{M}$) (see figure 1). Whether a particular factor is conducive or detrimental to PSM project performance is often debated, without consensus; this section summarizes the key causal arguments in both directions. Figure 1 provides an overview of observed factors included in the analysis and indicates which of three main research questions (Q1, Q2, and Q3) this review tests for each factor.

It is important to note up front that many factors that may influence PSM project performance cannot be observed in this study and may cause bias. This is also reflected in the fact that observed factors only predict a small share of the variation in PSM project performance (see section 5.5 for validity threats). Figure 1 points to examples of such unobserved factors, such as the specific MDA context, the engagement process with reform stakeholders, specific reform content, and project implementation arrangements. The correlates of PSM project performance identified should therefore be interpreted causally only with great caution.

---

[6]   The use of content, context, and process as useful categories for considering any type of change process is based on Armenakis and Bedeian (1999).

Figure 1. Overview of Observed and Unobserved PSM Project Performance Predictors

Country Context Factors ($\vec{C}$)	Q_1	Q_2	Q_3
Observed			
• Political and Civil Liberties	√	√	√
• Dependency on revenues from aid and natural resources	√	√	√
• Presence of Programmatic Political Parties	√	√	√
• Income, Growth and Human Development Indicators	√	√	-
• Administrative Capacity	√	√	-
Unobserved (examples)			
• "Meso" institutional context of the implementing MDA	-	-	-

Reform Process Factors ($\vec{P}$)	Q_1	Q_2	Q_3
Observed (during the first half of project implementation)			
• Recorded changes in Project Design («restructuring» or «upgrading»)	√	-	-
• Risk-Indicators (gathered during first half of project implementation)	√	-	some
Unobserved (examples)			
• Engagement Process with Government Counterparts in Project Preparation and Implementation	-	-	-
• Reform Winners and Losers and their Strategic Moves	-	-	-
• Project Adaptation during Implementation (short of «restructuring» or «upgrading»)	-	-	-

$$E\left[Y \mid \vec{C}, \vec{R}, \vec{P}, \vec{M}\right] = f\left(\vec{C}, \vec{R}, \vec{P}, \vec{M}\right)$$

Reform Contents Factors ($\vec{R}$)	Q_1	Q_2	Q_3
Observed			
• Broad type of PSM reform supported (PFM, CSR, Decentralisation, Tax)	√	-	-
Unobserved (examples)			
• Specific reform content	-	-	-
• Sequencing with prior reforms	-	-	-

Project Management Factors ($\vec{M}$)	Q_1	Q_2	Q_3
Observed (at approval / appraisal stage)			
• Lending Volume & preparation costs	√	-	-
• Time required until Project Approval and effectiveness	√	-	-
Unobserved (examples)			
• Preparing Bank Team	-	-	-
• Project Implementation Arrangements	-	-	-

Source: Author's own compilation.
Note: Check-marks indicate which of three research questions is considered in relation to a given factor.

3.1   Country Context: Political and Civil Liberties

There is a compelling political economy argument that government accountability is conducive to the performance of PSM reform projects (research question 1).[7] Where citizens and firms are able to hold the government to account, they can shift political elites' incentives from "taking" (rent-seeking) to "making" (provision of broad public goods). In Hirschman's evocative statement, "while markets create managerial discipline and induce efficacy through the exercise of choice, governments are principally disciplined through the exercise of voice."[8]

World Bank PSM projects typically seek to strengthen PSM institutions that facilitate "making" (concentrated costs, dispersed benefits)[9] and hinder "taking" (concentrated benefits, dispersed costs). Civil service projects often seek to strengthen meritocracy and reduce patronage. PFM projects seek to build systems that ensure that public money is used transparently and accountably and seek to limit

---

[7] In the *World Development Report 2004* (World Bank 2004) on public service delivery, for example, this argument is reflected in the concept of "long route accountability."
[8] Cited after Isham, Kaufmann, and Pritchett (1995).
[9] For an argument why the cost-benefit incidence of PSM reforms tends to be entrepreneurial, see Blum and Manning (2011).

discretion. One would thus expect more accountable governments to be more supportive of World Bank PSM projects, and these projects to perform better in the contexts of such governments.

But there are also powerful arguments that nonaccountable, authoritarian rule enables governments to push through tough PSM reforms. Empirically, it is possible to point to a number of "developmental states" with authoritarian governments that have been able to build well-performing administrations—such as China, Rwanda, and Singapore. One theoretical argument supporting this claim is that PSM reforms have a cost-benefit incidence that is particularly misaligned with electoral cycles—they tend to produce relatively certain short-term costs and uncertain long-term benefits (Schneider and Heredia 2003). Based on data from U.S. state governments, Moynihan (2008), for example, argues that democratically accountable state governments like to announce PSM reforms but tend to shy away from implementing them in full, which is costly. In brief, tough PSM reforms might be harder to do in systems with stronger checks and balances.

Levy and Fukuyama (2010) relate both arguments by arguing that the transformation of political institutions can but need not precede state capacity building. They contrast "transformational governance" with a "developmental state" trajectory. In the former, "political transformation has the potential to radically improve both the incentives and the means for state capacity building." The latter begins with state capacity building, while the route to the "transformation of political institutions [. . .] is a long-term and indirect one." Overall, the influence of political regimes on PSM project performance is controversial.

Arguably, PSM reform projects might be more difficult than non-PSM projects in contexts where voice mechanisms do not (or only weakly) check rent-seeking behavior (research question 3). One possible reason for this is that rent opportunities abound in the public sector (in the form of public money, contracts, jobs, and so on) and that PSM reforms are precisely about changing the rules (or the "government systems") that determine the allocation of these rents. Whereas public sector jobs and money are under the direct control of political elites, non-PSM projects may be easier to insulate from the influence of adverse political "taking" incentives.

Compare for example a (non-PSM) road construction project and a (PSM) procurement reform project in a country marked by the rent-seeking behavior of political elites. The procurement of Bank-funded roads would be tightly monitored, limiting rent-seeking opportunities. But this would not interfere with how the government awards contracts for roads financed from the national budget. Indeed, the resulting coexistence of (at least on the surface) very different parallel procurement systems—a discretionary one for government-financed projects and a competitive one for donor-financed projects—is typical in such contexts. By contrast, the ambition of a typical PSM procurement project is very different: to change the

client government's own procurement processes. Such a project explicitly seeks to introduce transparency and competition beyond "islands" of donor influence—and thus runs directly counter to rent-seeking interests that have a stake in preserving opacity and discretion in how contracts are awarded.

Arguments for why PSM projects should perform better than non-PSM projects in countries with weaker political and civic liberties are less obvious. One might simply be that PSM projects typically have much lower financing volumes (see annex table A.6) than non-PSM projects and thus are less attractive targets for rent-seeking.

That voice and accountability affect politicians' incentives to support PSM reform is a very "broad brush" claim. Both autocracies and democracies are heterogeneous, as are politicians' incentives to undertake PSM reform.[10] This paper covers two sets of more specific political economy factors—rentier-state-type hypotheses and the role of programmatic political parties.

3.2   Country Context: Dependence on Resource Revenues and Aid

The rentier-state argument suggests that resource revenues tend to reduce political pressures for administrative reform (research question 1). For example, politicians may be less accountable for how they spend oil revenues than they are for tax revenues. In addition, if resource revenues compose a large portion of total gross domestic product (GDP), the government may be able to afford inefficient financial management systems and an overstaffed or overpaid public service.

Whether aid dependence strengthens or weakens political incentives for PSM reform is contested (research question 1). On the one hand, donors may have more bargaining power with aid-dependent client governments, enabling them to push harder for PSM reform demands. Indeed, PSM reforms have historically been core to donor conditions. For example, development policy loans (DPLs) often involve PSM-related "prior actions"—a trend that has strengthened over the past two decades (see figure 2 for 1990–2011). Besides providing such an incentive for reform, aid may simply facilitate more useful information flows between donors and governments. High aid flows may entail more donor-government interaction and increase government exposure to exogenously "imported" reform ideas.

---

[10] See Blum and Manning (2011) for a more detailed discussion.

Figure 2. Public Sector Reforms as a Prerequisite for Development Policy Loans, 1990–2011
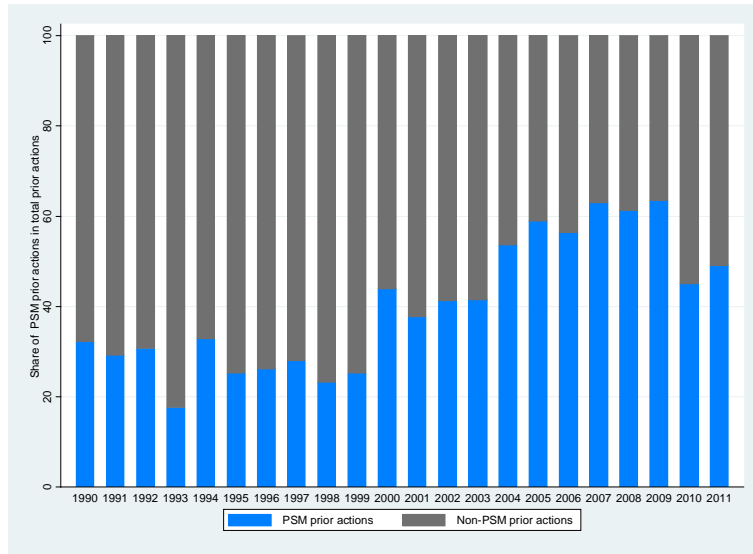
Note: "Prior actions" as reported here comprise both "prior actions" and "prior actions for future tranches," drawing on the DPL Prior Actions database as available on the World Bank's Web site. Consistent with the definition of PSM projects employed in this review, prior actions were identified as "PSM prior actions" if they were coded with theme codes 25 to 30 or sector codes "BC," "BH," or "BZ."

But large aid flows can also undermine PSM project performance. First, it may be wrong to assume that aid dependence necessarily strengthens donor bargaining power. If politicians in aid-dependent countries can expect that aid will continue flowing regardless of what they do, the threat of cutting aid loses credibility. Despite poor reform progress, donors may hesitate to cut aid exactly because the client country is needy. Or disparate, often competing donors may fail to coordinate in cutting aid. Second, even where implemented, reforms may not have lasting value; as Andrews (2009) argues, donor pressures may motivate governments to "mimic" the institutions donors would like to see—rather than actually improving government performance.[11] Meanwhile, aid rents, similar to resource rents, can weaken the incentive to reform because they undermine government dependence on citizens' tax payments—and, thus, government accountability.

3.3   Country Context: Programmatic Political Parties

There are strong arguments that Bank PSM projects are more likely to succeed in countries with programmatic political parties (research question 1). First, programmatic parties provide a vehicle for voters to hold politicians to account when an administration fails to deliver. Second, they can help solve politicians' collective action problems (see Aldrich 1995), providing the discipline that encourages them

---

[11] Andrews predicts that PSM reforms driven by "isomorphic mimicry" are likely to affect the visible, central, and de jure aspects of PSM institutions that are easy to observe by donors—even as they fail to affect the invisible, decentralized, and de facto aspects that may matter most to citizens.

to agree on providing broad public goods rather than targeted benefits to their constituents ("pork-barrel politics"; see, for example, Hasnain 2011).

This logic applies directly to public sector reforms. In a paper entitled "Programmatic Political Parties and Public Sector Reform," Cruz and Keefer (2010) find robust support for the claim that the presence of such parties improves the performance of World Bank PSM projects. They argue that PSM reforms typically aim to strengthen rule-based systems—for example, those governing the use of public money and jobs—and to limit ad hoc decisions on these issues by public administrators. Without party discipline, politicians may prefer an ad hoc over a rule-based process, enabling them to interfere in administrative decisions in favor of particular clients.

One might expect the presence of parties to affect PSM and non-PSM projects alike, as both typically aim to provide public goods (research question 3). But this is contestable. Overall, PSM reforms center on issues that are typically not urgent citizen concerns and are thus not part of parties' electoral platforms (Schneider and Heredia 2003). At the same time, some PSM reforms—such as public employment or pay reforms—are often unpopular with public employees and may require particularly strong party discipline.

This paper seeks to corroborate the findings of Cruz and Keefer. It tests them for a larger array of public sector projects, using a more extensive set of control variables and different estimation techniques. In addition, it tests whether the existence of programmatic parties distinctively predicts the performance of PSM projects more or less than non-PSM projects.

3.4   Country Context: Prior Administrative Capacity

Over the past decade, a strong consensus has emerged that PSM reforms need to be carefully tailored to preexisting administrative capacities—and that transplanting "best practices" from the OECD to developing countries (as in the style of New Public Management reforms) is risky. This consensus is, for example, reflected in Allan Schick's (1998) dictum of "basics first"[12] in "platform approaches"; in Andrews' (2010) classification of African countries into five distinct "PFM performance leagues";[13] and in the argument that the use of performance-related pay requires that an administration have an established culture of meritocracy.[14] It is also reflected in calls for setting modest PSM reform objectives.[15]

---

[12] The "basics first" dictum has been contested, however. Andrews (2006), for example, argues that moving toward a performance orientation in budgeting may not necessarily require or even benefit from an established, traditional input-based budgeting system.

[13] One related argument is that colonial heritage matters—in particular the difference between francophone and anglophone administrative traditions (for this argument, see, for example, Andrews 2009).

[14] See Pierskalla, Hasnain, and Manning (2012) for an in-depth review of the literature on pay flexibility.

[15] For example, Andrews, Pritchett, and Woolcock (2010) empirically estimate the speed of state modernization processes and conclude that governments take far longer to develop administrative capacity than is often assumed in development projects.

Despite this consensus, there are good reasons why donor PSM projects may remain overambitious and untailored to their contexts. As detailed in the World Bank's "Approach to Public Sector Management for 2011–2020" (World Bank 2012b), project managers may have incentives to set overly ambitious reform goals to sell their projects to the client and to meet expectations within the Bank. TLLs may find it less risky to adopt well-tested reform designs than to experiment with tailored approaches—while client governments may ask for "cutting-edge" approaches because they are politically attractive.

It is not obvious whether these dynamics imply that PSM projects should perform better or worse in countries with higher initial administrative capacity. If TTLs were able to tailor PSM projects perfectly to initial administrative capacity, then they should set more modest objectives in low-capacity countries and more ambitious objectives in high-capacity countries. In this case, initial administrative capability would not decide project performance. If TTLs, however, tend to set overly ambitious project targets, this might make projects perform worse in low-capacity countries, where the level of ambition risks being particularly out of tune with existing capacity. But one could also argue that decreasing marginal returns on reform and reform satiation in high-capacity countries might reduce the opportunities for "satisfactory" project implementation.

3.5 Country Context: Economic and Human Development

Economic and human development may positively influence the success of PSM reform projects through multiple channels. For example, countries with higher income levels may have a better-qualified and more-specialized workforce—and a better-financed public sector—and thus higher capacity for successful PSM reform. A better-educated public, meanwhile, may dispose of more means to hold government to account, even in authoritarian settings.

3.6 Reform Content

PSM reform projects have diverse aims and varying levels of difficulty (research question 1). For example, de jure reforms may be easier to achieve than de facto change. Introducing MTEFs has little in common with internal audit reform. Efforts to downsize the public administration will involve different challenges than pay reforms aimed at attracting and retaining qualified employees. While this calls for a granular distinction between different PSM reform content areas, because of data limitations this paper delineates only four broad groups of PSM reform: civil service, financial management, decentralization, and tax administration.

How does the performance of Bank projects in these four content areas compare? The debate on this question has been shaped by an IEG report on the World Bank's PSM lending portfolio, entitled "Public Sector Reform. What Works and Why" (World Bank 2008). The report ranks project performance across

different content areas. Employing an improved CPIA score (CPIA 12–16) as the dependent variable, the IEG report finds that "for all countries (with CPIA information), improvement was most likely (60–70 percent likely) in PFM (CPIA 13) and revenue administration (CPIA 14)." By contrast, the "quality of public administration (CPIA 15), which we take as civil service reform (CSR), had the lowest success rate, with fewer than 45 percent of borrowers in this area showing improvement" (World Bank 2008).

The present paper seeks to inform this debate, employing the IEG's project outcome ratings (rather than the CPIA scores)[16] as the dependent variable. It is well known within the Bank that some PSM reform projects perform above the Bank average (for example "tax policy and administration"), and others below (for example "administrative and civil service reform"—see section 7.2 and figure 6). But it is less clear whether the inherent risk of particular reform areas—or of country contexts—explains these differences. This report explores this question (research question 2).

3.7   Reform Process Factors

Reform process factors often predict project performance. They can help to flag risks early on in the implementation process,[17] direct managerial attention and resources toward risks, and encourage early course corrections. This review analyzes which risk indicators raised during the first half of the project implementation process are significant predictors of the IEG outcome ratings.

This investigation is bolstered by growing recognition that reform processes affect reform success. Employing Linsky and Heifetz's (1994) terminology, PSM reforms are fundamentally about solving "adaptive" problems, that is, about changing public servants' "values, attitudes, or habits and behavior," rather than mere "technical" problems. As the well-developed literature (for example, Schein 1999 and 2002) on reform management highlights, it is crucial that leadership be exercised to build buy-in.

Similarly, Andrews, Pritchett, and Woolcock's (2012) call for a "Problem-driven Iterative Adaptation" (PDIA) approach to building state capabilities emphasizes that performance problems should be locally defined, solved through a process of experimentation (rather than linear planning), and engage a broad set of actors. Within the Bank the PSM approach for 2011 to 2020 has put process issues center stage, by calling for a "diagnostic approach" toward project preparation (World Bank 2012b; Blum, Manning, and Srivastava 2012). A recently adopted results-based lending instrument (Program-for-Results, P4R) is seen

---

[16] Unlike CPIA, IEG outcome ratings do not provide a comparable measure of public administration quality. But they have the advantage of being unambiguously attributable to the respective project, whereas CPIA ratings capture broad improvements in PSM arrangements and change slowly, making it questionable to what extent they reflect the impact of Bank projects.

[17] It is noteworthy that these indicators are of course not the only—or major—risk management instruments that the Bank employs. Detailed peer review mechanisms during the project preparation phase and regular qualitative supervision reports by TTLs during implementation provide a much richer set of information. By comparison, the indicators included in this review are a reductive and mechanical way of measuring risks—but they have the advantage of being collected systematically across projects and are thus suited for a review of this nature.

as a potentially powerful way to ensure more flexibility and experimentation in the implementation process.

While data are unavailable on many important reform process characteristics—such as the breadth or intensity of stakeholder engagement—it is useful to ask which of the currently collected risk indicators actually predict IEG outcome ratings.

3.8   Project Management Factors

Project management factors include basic project characteristics, such as the committed lending amount, preparation and supervision costs, project preparation times, and so on.[18] Such factors are distinct from process or risk indicators, in that they are not explicitly designed to indicate risk, and they may be observed before project implementation starts.

Several of these factors are potentially useful predictors of project performance. The time required for project preparation[19] could indicate particularly careful preparation—but could also indicate a particularly controversial, challenging project. A long delay before the project actually becomes "effective" may indicate weak client commitment to the project.[20] Regarding loan size, smaller projects may get less high-level attention than large projects, both on the client and on the Bank side.

4   THE DATA

This paper analyzes project-level data from the World Bank's project management information system as well as country-level data from various sources. This section describes the project sample underlying this paper, the data employed, and associated measurement issues.

4.1   The Project Sample

The sample used for this paper includes all World Bank–supported investment lending projects[21] approved between FY 1990 and FY 2013 that had been closed and evaluated by the IEG at the time of the

---

[18] Identifiers of the Bank teams or task team leaders (TTLs) responsible for project preparation (and their performance track record) are not included in this paper. While they are observable in principle, data on preparing TTL identities have only been captured systematically since 2005. Such data are therefore not available for most of the project universe underlying this paper. It is important to note that Denizer, Kaufmann, and Kraay (2011) find that TTL identity and performance are significant predictors of project performance.

[19] The time required for project preparation is a reductive, but possibly telling, sign of the nature of this process. A first measure of preparation time is the "time to approval," that is, the time it took from the project concept note review (PCN) meeting to project approval by the Bank's Board of Directors. Above average times to approval could indicate (i) a particularly carefully prepared project or (ii) a particularly complex project that may be challenging to implement, or (iii) a project that is particularly controversial and therefore took a long time to agree on.

[20] One key step required between project approval and effectiveness is that the client government sign the loan agreement. A delay in this step may reflect a lack of urgency or unresolved points of contention. "Time to effectiveness" may thus be a signal of low client government commitment to the envisaged project, an important but hard to observe driver of project performance.

[21] Projects were included in the sample only if (i) they were approved between FY 1990 and 2013, (ii) their "lending instrument type" was classified as "investment" lending, and (iii) their committed amount was greater than zero.

preparation of the data set, that is, by June 30, 2013. Regional projects[22] that do not focus on a single country and projects in several selected countries were excluded from the sample for data availability reasons. The resulting sample comprises 6,149 investment lending (IL) projects approved between FY 1990 and FY 2013. The IEG outcome ratings are available for 3,202 of these projects, that is, for about half. As shown in figure 3, the majority of these 3,202 evaluated projects (light gray) date from the 1990s and early 2000s; fewer data are available for more recently approved projects, many of which have not yet been evaluated or are still active. Given this time lag, this paper primarily covers projects approved between 1990 and 2004. It therefore has a historical focus and does not necessarily reflect the performance patterns of projects approved after 2005 (which represent only about 7.5 percent of the projects reviewed).

Figure 3. Availability of IEG Outcome Ratings for Investment Lending Projects, by Fiscal Year Approved (1990–2013)
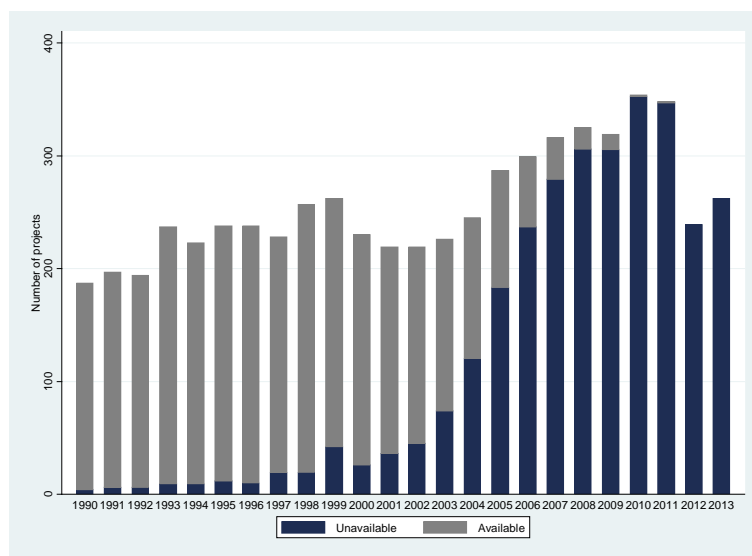
Source: Author's own compilation, based on World Bank Project Data.

4.2 Project Content Measures

Within this overall sample of 3,202 investment lending projects, PSM reform projects and other, "non-PSM" projects are distinguished. This paper employs a broad definition of PSM projects[23] that includes both (i) "upstream" projects primarily conducted at the center of government in core ministries and central agencies (Ministry of Finance, central HRM bodies, and so on), and (ii) "downstream," sector-specific projects (education, health, infrastructure) that have significant PSM reform components (at the

---

[22] Regional projects concern more than one client country.
[23] This broad definition will be used for all statistics reported hereafter, unless explicitly indicated otherwise. For a more detailed analysis of the "upstream" subset of projects focusing on the center of government and mapped to the Bank's "Public Sector Governance Board" (PSGB) please refer to Blum (2014).

MDA [ministry, department, agency] level). Based on this definition, 1,097 out of the 3,202 projects are PSM projects.

This definition of PSM draws on the Bank's classification of project themes and sectors.[24] It comprises all projects that have at least a 25 percent PSM component—whether by sector or theme. The delineation between PSM and non-PSM projects is thus based on a—somewhat arbitrary—25 percent cutoff value. In other words, more than half of the projects classified as PSM actually contain more than 50 percent non-PSM components (550 out of 1,097 PSM projects). These projects' performance ratings may therefore not necessarily reflect the performance of their PSM components, and the resulting estimates may be subject to significant noise.[25]

Within PSM projects, four PSM reform content areas can be distinguished based on theme codes: (i) civil service and administrative reform; (ii) public expenditure, financial management, and procurement; (iii) decentralization, and (iv) tax policy and administration.[26] Assigning PSM projects to a single (largest) PSM content category is problematic, because such projects are often designed as packages comprising several components relating to different PSM content categories (for example, both CSR and PFM). Rather than attempt to identify projects with a single major theme, this paper relies primarily on the percentages assigned to each public sector theme as a more accurate measure of a project's actual PSM reform content (see annex table A.5 for the average share of the six themes across projects both in and outside the public sector).

4.3   Project Performance Measures

This review uses the IEG's project outcome rating—available across Bank projects—as the dependent variable to measure project performance. This is the main tool used to analyze the performance of the Bank's lending portfolio, and has been employed in several related studies, including Isham, Kaufmann, and Pritchett (1995), Cruz and Keefer (2010), and Denizer, Kaufmann, and Kraay (2011).[27]

---

[24] When preparing new projects, Bank project managers assign percentage shares for up to five of these theme and sector classifications, each totaling 100 percent. "Theme" codes are meant to reflect "the goals/objectives of Bank activities," whereas sector codes are meant to reflect the "high-level grouping of economic activities based on the types of goods or services produced" and are "used to indicate which part of the economy is supported by the Bank intervention" (http://go.worldbank.org/CVGJVWWDF0).

[25] This holds particularly true as the thematic classifications of Bank projects only approximate the actual project content. Theme and sector classifications are entered during project preparation—but are rarely amended if later in the project cycle they no longer reflect the actual project focus (for example, after restructuring). Also, different sector units may have different ways of classifying projects, such that projects with actual public sector content may or may not be classified as such.

[26] The theme codes used for identifying PSM projects include codes 25 to 30. The PSM-related sector classifications include "Central Government Administration," "General Public Administration," and "Sub-National Government Administration." For detailed definitions of these theme and sector codes, please refer to http://go.worldbank.org/CVGJVWWDF0.

[27] While the IEG provides a number of other project ratings, this paper focuses on IEG outcome ratings. For details on the definitions of other project ratings, please refer to World Bank (2005). In particular, "IEG Institutional Development Impact Ratings" could be employed, which reflect the extent to "which a project improves the ability of a country or region to make more efficient, equitable and sustainable use of its human, financial, and natural resources" (Cruz and Keefer 2010). As Cruz and Keefer (2010) argue, for public sector reforms, "institutional development is precisely the point of the reforms" and these ratings may thus better reflect the performance of a project's PSM components than the IEG's overall outcome rating. But these ratings have several shortcomings. First, they are likely much less accurate measures than

IEG outcome ratings are meant to assess the extent to which "there were [. . .] shortcomings in the operation's achievement of its *objectives*, in its *efficiency* or in its *relevance*" (World Bank 2005). In other words, the outcome rating reflects the extent to which a project reached its (predefined) objectives, and with what degree of efficiency and relevance to broader development goals.

IEG outcome ratings are based on an ordinal six-point scale, ranging from "highly satisfactory" to "highly unsatisfactory."[28] Figure 4 shows the distribution of PSM and non-PSM project ratings along this scale for the 3,202 projects reviewed.[29] It shows that the majority of projects (~ 68.3 percent) are rated either "satisfactory" or "moderately satisfactory," whereas only about 25.3 percent are rated either "moderately unsatisfactory" or "unsatisfactory." The ratings at both extremes of the scale are used rarely—only about 1.9 percent of projects have "highly unsatisfactory" ratings and only about 4.5 percent are rated "highly satisfactory."

Figure 4. Distribution of IEG Outcome Ratings for Public Sector and Non-Public-Sector Projects
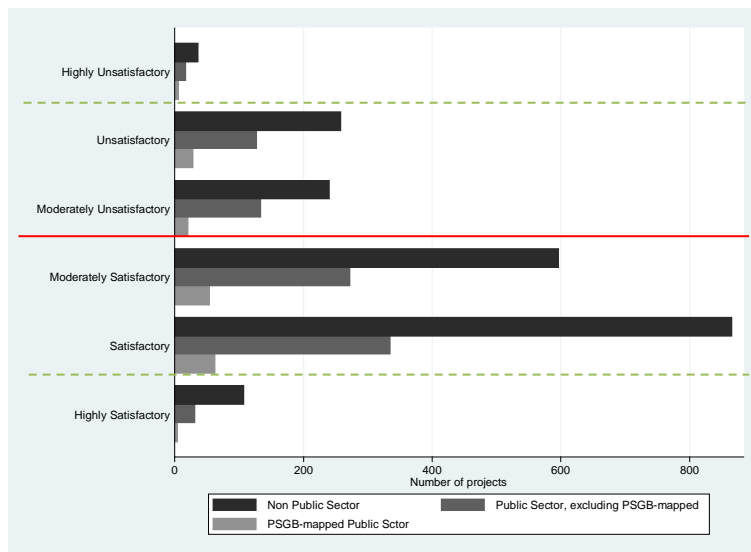
Source: Author's own compilation, based on World Bank Project Data.

Note: PSM projects mapped to the Public Sector Governance Board (PSGB) are typically "upstream" projects; others are typically "downstream" projects.

---

outcome ratings, because they bear less weight within the Bank and are more vaguely defined than the latter. In addition, these ratings are not available for about one-third of the projects in the sample underlying this review.

[28] The exact ratings, in descending order, are "highly satisfactory," "satisfactory," "moderately satisfactory," "moderately unsatisfactory," "unsatisfactory," and "highly unsatisfactory."

[29] As evaluation methods used by the IEG for establishing these ratings vary, this review systematically employs the "highest-quality" IEG evaluation available for each project. Document-based desk reviews by the IEG (conducted since 1995) are called "Evaluation Summaries" (ES) or "Evaluation Memoranda" (EVM). In addition, "a sample of about 25 percent of projects completed each year are selected by IEG for a more detailed ex-post evaluation" (Denzier, Kaufmann, and Kraay 2011), called "Project Performance Audit Reports" (PPARs). PPAR evaluations are typically conducted several years after project completion and based on field visits. Out of the total project universe of 3,202 projects underlying this review, 597 ratings are based on PPARs, 2,592 on ES or EVMs, and 13 on "Project Completion Reports/Notes" (PCRs/ PCNs), which were used for a few projects until the mid-1990s.

This review employs both (i) the original, ordinal six-point coding of the IEG outcome ratings and (ii) a simplified binary scale.[30] The binary scale collapses the six ratings into two groups of the three highest (at least "moderately satisfactory") and three lowest ratings (at most "moderately unsatisfactory"), as indicated by the red line in figure 4. Where possible, the ordinal ratings are used to accurately reflect the information contained in the data. The simplified binary data are employed because they are consistent with the rating terminology,[31] associated Bank staff incentives,[32] and prior research.

But the paper's findings suggest that the loss of information in the simplified data may be problematic. In ordered probit estimations with a nonsimplified dependent variable, most performance predictors change their "sign" (that is, whether they increase or reduce the likelihood of receiving one of the six IEG outcome ratings) between "moderately satisfactory" and "satisfactory," not between "moderately unsatisfactory" and "moderately satisfactory" as one would expect (see annex table A.10 and figures A.3 and A.4). Thus there seems to be a very meaningful threshold between "satisfactory" and "moderately unsatisfactory" projects. One explanation could be that "satisfactory" project ratings are mostly accurate, whereas "moderately satisfactory" ratings may often be used "mercifully" for projects that in fact deserve a "moderately unsatisfactory" rating. Sticking to the standard simplification despite this finding is not problematic for estimation results, because it will under- and not overestimate the respective predictor in question, by introducing noise in the "moderately satisfactory" or higher (MS+) category.

It is important to note two major limitations of the IEG outcome ratings. First, they do not adequately measure reform success or development outcomes. Rather, they measure to what extent a project performed "satisfactorily" from the Bank's corporate perspective. That is, they measure "the extent to which the [project's] major relevant objectives were achieved,"—which may or may not include differences in the client government's performance, or ultimately, development outcomes.[33]

Second, their comparability across projects is limited and they inherently suffer from endogeneity bias (see section 5.4 for details). Project objectives, which serve as the yardstick for the IEG outcome ratings, are negotiated between Bank teams and client governments and therefore clearly endogenous to the context. If TTLs, for example, set more modest PSM reform targets in countries with limited political and

---

[30] When the six-point ordinal rating is used in the data annex, it is referred to as "Ordinal IEG Outcome Rating." When the simplified binary rating is used, it is referred to as "Binary IEG Outcome Rating."

[31] The binary terminology is already explicit in the original scale, which distinguishes three "satisfactory" and three "unsatisfactory" ratings.

[32] Meeting the threshold of obtaining an at least "moderately satisfactory" rating is an important incentive for Bank staff. This incentive structure may also explain noticeable "threshold effects" in the distribution of the IEG outcome ratings (as shown in figure 4). Three thresholds are marked by a noticeable delta in the number or percentage of projects that get rated above or below the threshold. The first of these thresholds (in red) is the one between "moderately satisfactory" and "moderately unsatisfactory" projects that drives the simplified coding. The other two thresholds (in green) indicate that it is hard to get extreme ratings, and that they should not be considered "similar" to the midrange ratings. But as the number of projects with extreme ratings is very small, the error introduced by grouping them with midrange ratings should be limited.

[33] While relevance to overarching development goals is among the declared evaluation criteria, de facto evaluation practice suggests that the prime evaluation criterion for a project is whether it achieved its immediate objectives. For this reason, it seems a stretch to interpret IEG outcome ratings as a measure of the (much broader) concept of aid effectiveness (as, for example, Dollar and Levin suggest). The IEG project ratings are at best a rough correlate for aid efficacy. In addition, as, for example, Andrews (2011) argues, the IEG ratings tend to not sufficiently reflect important dimensions of PSM reform success, such as the development of "space" for future reforms.

civil liberties in anticipation of a challenging reform environment, the estimates in this review would underestimate the detrimental effect of limited liberties on the success of PSM reform.[34]

Despite these caveats (and as argued in section 3.2), IEG outcome ratings have the major advantage of not depending on normative claims about what PSM reforms should achieve across countries. Rather, they provide a measure of efficacy[35] in implementing a PSM reform plan that was agreed upon by and deemed acceptable by both the client country government and the Bank's Board of Directors.

4.4   Country Context Measures

The country-level data employed in this review come from multiple sources that include the Freedom House measures of Political Rights and Civil Liberties; the Polity IV institutionalized autocracy, anocracy, democracy, and combined polity scores; the Share of Programmatic Political Parties from the Database of Political Institutions; the CPIA 12-16 Ratings; the International Country Risk Guide (ICRG) Bureaucracy Quality Rating; as well as several variables from the World Development Indicators (WDI) data sets.[36]

4.5   Process and Risk Measures

Process or risk indicators routinely collected in implementation supervision reports (ISRs) include a set of 13 (binary, on/off) risk "flags," for areas such as financial management, disbursement delays, environmental safeguards, and so on. To limit the number of covariates included in estimations, only the flags relating to the country environment, country record, project management, safeguards and counterpart funds—as well as the total number of risk flags raised—were included in the final estimation.[37]

In addition, ISRs include two markers that TTLs may use to indicate how well a project is going: (i) "implementation progress," that is, whether project activities are being implemented as planned; and (ii) "development objectives," or how well the project is expected to achieve its ultimate development objectives. It should be noted that of these, the first is not entirely at the discretion of the TTL—if three or more risk flags are raised, it is automatically set to "moderately unsatisfactory" or less.[38]

---

[34] A third potential limitation is that the IEG ratings are ultimately subjective assessments conducted by IEG staff. But it can reasonably be assumed that these ratings do not systematically suffer from bias, given the IEG's institutional setup as an independent agency within the Bank. If this is true, subjectivity is merely a source of noise and not a concern for the purposes of this review.

[35] Isham, Kaufmann, and Pritchett (1995) introduce this term.

[36] Many of these measures have limited availability or may have questionable validity or consistency over time. Please refer to Blum (2014) for a more detailed discussion of these issues.

[37] This selection is based on theoretical expectations of the relevance of the respective risk flags, the frequency of their usage, and their suitability for improving the fit of the estimation model.

[38] Based on these risk ratings, projects are marked as "at risk" or as "proactivity" projects that require attention/action. Actions taken in response to these risks are also recorded, and required to move a project out of "at risk" status. "At risk" and "proactivity" statuses are also included as covariates.

4.6   Project Management Measures

A few key project attributes, including the loan type and size, the year of approval, project duration, and time elapses between different project milestones are directly available in the World Bank's management information system and included in the estimations.

5   IDENTIFICATION STRATEGY

This review employs three distinct sets of estimation models.

5.1   Ordered Probit Regression Models

The primary model for estimating which country context, reform content, process, and project management characteristics predict the performance of PSM projects (research question 1) is an ordered probit model. This choice seeks to do justice to the ordered nature of the IEG outcome ratings (see section 4.3). A probit model using the simplified binary dependent variable yields more intuitive results and serves as a robustness check. (The estimation results for these models are reported in annex tables A.7 to A.10.)

5.2   Heterogeneous Marginal Effect Estimations for PSM and Non-PSM Projects

The second set of regression estimators seeks to identify whether PSM projects are distinctly sensitive to certain country context factors, relative to non-PSM projects (research question 3). The estimated ordered probit regression models are the same as above, but include interaction terms between the PSM project marker variable and selected country context factors, to estimate differences in the marginal effects of country context characteristics on PSM and non-PSM projects (see annex figures A.2 to A.4).

5.3   Nearest Neighbor Matching

The third set of estimators seeks to identify whether PSM projects are distinctly sensitive to country context factors, by matching PSM projects with similar non-PSM projects in the same country. The review employs nonparametric matching estimators (nearest neighbor matching and matching based on subclassification) to assess whether PSM projects perform distinctively better or worse than projects in other policy fields/sectors in the same country (research question 2). Matching PSM projects with non-PSM projects within the same country serves to reduce bias in the estimated difference due to time-invariant, unobservable country characteristics.[39] Matching estimations also serve to corroborate the

---

[39] To the extent that it can be assured that both PSM projects and non-PSM comparators are approved and implemented around the same time, this matching also allows for reducing bias due to time-variant country characteristics.

findings on research question 3, by classifying countries based on similar context characteristics and matching PSM and non-PSM projects in each class.

Nearest neighbor matching estimators construct a counterfactual by matching "treated" PSM project observations to "nontreated," non-PSM project observations with similar covariates as the counterfactual, to estimate performance differences between both groups of projects. Equation 1 sets out the nearest neighbor estimation model for estimating the average treatment effect on the treated (ATET), that is, for projects with the country and project characteristics of an average PSM project (not an average project).

Equation 1. Nearest Neighbor Matching with Perfect Matching by Country

$$E[Y_1 - Y_0 \mid D = 1] = \frac{1}{N_1} \sum_{D_i=1} \left\{ Y_{iC} - \left(\frac{1}{M} \sum_{m=1}^{M} Y_{jCm(i)} \right) \right\}$$

where $Y_{jCm(i)}$ is the outcome of a non-PSM project implemented in the same country C as the PSM project I, such that $X_{jCm(i)}$ is the m-th closest value to X_i of a non-PSM project.

The matching criteria employed are two: exact matching[40] by country, and approximate matching by the fiscal year of approval. This ensures that the matched projects are approved in the same country (for about 80 to 95 percent of matches) and in a similar year, thereby reducing bias due to unobserved country context characteristics.[41] Annex table A.12 lists the estimation results for nearest neighbor matching and for matching based on subclassification.

5.4   Matching Based on Subclassification

The second nonparametric matching approach used is subclassification or coarsened exact matching: PSM projects are matched with non-PSM projects from exactly the same country, and that have been approved in a similar time period. More precisely, all projects are "sorted" into cells and are matched if they end up in the same cell. The cells are defined by country, and four-year approval intervals (that is, 1990–94, 1995–98, 1999–2002, 2003–06, and 2007–10). Formally,

Equation 2. Matching Based on Subclassification

$$E[Y_{PSM} - Y_{NON-PSM} | D = 1] = \sum_{k=1}^{K} (\bar{Y}_{PSM} - \bar{Y}_{NON-PSM}) \cdot \left(\frac{N_{PSM}^k}{N_{PSM}} \right),$$

---

[40] Within the specific algorithm used (see Abadie and others 2004), "exact" matching is implemented by recalculating the weight matrix for the different covariates, such that the weights of the variables for exact matching is multiplied by 1,000. This algorithm requires ordering of countries on an ordinal scale, so that a non-PSM project from a "neighboring" country can be picked if a match from the exact country is unavailable or too distant. For example, if Angola, Morocco, and Botswana were "neighboring countries" on the ordinal scale, the algorithm would look for suitable matches in Angola and Botswana if no suitable match for a PSM project in Morocco were found. The share of projects where matches from neighboring countries were used is small (~10 percent) and thus unlikely to introduce major bias. But within these 10 percent, the quality of matches from different countries matters. To ensure at least some comparability, countries are ordered based on their Freedom House ratings and, within the same Freedom House rating, by region.

[41] For each PSM project, three non-PSM project matches are identified, without using the same non-PSM project multiple times as a match (no "replacement"). The rationale for using multiple non-PSM matches for each PSM project is to ensure a sufficient sample size, since the number of PSM projects is relatively small.

with K denoting the total number of cells $\{X^1, \ldots, X^k, X^K\}$ that X is classified in. N^k_{PSM} is the number of PSM projects in cell k, and N_{PSM} is the total number of PSM projects in all cells. Y_{PSM} is the average success rate of PSM projects in cell k and $Y_{NON\text{-}PSM}$ is the average success rate for non-PSM projects in the same cell. $N^k_{PSM/NPSM}$ is a weighting term for each cell k, weighting it by the share of PSM projects it contains out of the total sample of PSM projects included in the estimation.[42] $E[Y_{PSM} - Y_{NON-PSM}|D = 1]$ denotes the "average treatment effect on the treated," rather than the "average treatment effect" $(E[Y_{PSM} - Y_{NON-PSM}])$. This seems appropriate since the population of interest is PSM projects (the "treated"), which are compared to "similar" non-PSM projects.

The advantage of matching based on subclassification is that it allows exact matching by country and approval period (though not by year), reducing bias due to time-invariant unobserved country characteristics.

5.5 Validity Threats

As is generally characteristic of regression research based on cross-country data, the (statistically significant) correlates of PSM project performance identified in this review should only be interpreted as *causes* of PSM project performance with great caution. Two major validity threats merit highlighting—one internal and one external.

The major internal validity threat is *endogeneity (omitted variable) bias.* As has been noted at the outset (see section 3), because of lacking data, this paper omits major unobserved factors that are likely to influence PSM project performance. Omitted factors include measures of—among other things—the specific MDA context (leadership, capacity, and so on), the engagement process with reform stakeholders, specific reform content, and project implementation arrangements.

Accordingly, observed factors explain only a small share of the variations in IEG outcome ratings. The best fit model predicts only about 19 percent of the variation in the IEG outcome ratings (based on the McFadden pseudo-R^2, see table 7), and reduces the proportional error of predictions by about 21 percent. This implies that about 80 percent of the variation in the IEG outcome ratings is explained by unobserved factors. Even if all relevant covariates could be controlled for, endogeneity bias would remain, simply because Bank teams design and set the objectives of PSM projects (which influence the IEG ratings) in response to (observed and unobserved) country context factors.

Within the limits of these caveats, this paper seeks to reduce obvious and evitable internal validity threats. It controls for observed potential performance correlates to the extent feasible and it controls for unobserved, time-invariant country context characteristics. PSM projects might affect context measures,

---

[42] The exact subclassification matching algorithm used is Coarsened Exact Matching, as suggested by Blackwell, Iacus, and King (2009).

such as the CPIA ratings. The paper addresses such concerns of reverse causality by controlling for context factors at baseline, that is, in the year of project approval.[43]

The major external validity threat is *sample selection bias*. One potential purpose of this review is to provide Bank teams with risk predictors for *potential* PSM reform projects. Risk predictors matter at the project design stage, when Bank teams face the questions: Should a PSM project be designed in a given country?—and how? Relative to the set of potential projects that Bank teams consider, those that are completed and evaluated is much smaller. Sample selection bias would occur if the "actual" project portfolio were nonrepresentative of "potential" projects in ways that affect project performance.[44] In this case, sample selection bias seems very likely because projects that have been approved and implemented likely had better chances of success than projects considered but not chosen. The broadest consequence of such selection bias is that project success rates are biased upwards.[45] They draw an overly optimistic picture of the odds of success and failure when considered at the design stage.

6 DESCRIPTIVE STATISTICS: COMPARING PSM AND NON-PSM PROJECTS

This section compares the performance, country targeting, and process and management characteristics (part of research question 2) of PSM and non-PSM projects.[46]

6.1 Comparing Project Performance

PSM projects on average perform worse than non-PSM projects in achieving at least "moderately satisfactory" (MS+) IEG outcome ratings. As figure 5 illustrates, the share of PSM projects (broadly defined) with MS+ IEG outcome ratings (at 69.4 percent) is 5.2 percentage points lower than for non-PSM projects (74.6 percent). This difference is statistically significant at the 1 percent significance level (see annex table A.1). For "upstream" PSM projects (narrowly defined), the success rate is very similar to broadly defined PSM projects, at 68.54 percent.[47]

---

[43] Bias due to reverse causality could still occur if a government eager to obtain a World Bank investment loan for PSM reforms undertakes PSM reform efforts prior to receiving the loan to demonstrate reform commitment. If these prior reforms led to improvements in the CPIA scores at baseline, these improvements would (in part) be caused by the prospect of the Bank's investment project. But this variant of reverse causality due to expectations is unlikely to play a major role.

[44] In other words, the conditional independence assumption—that is, that "PSM reforms" are selected in the sample independent of their potential outcome (and are thus exogenous)—is unlikely to hold.

[45] Bias would in particular occur if the selection were correlated with other project covariates, such as country context. If, for example, Bank teams filtered out more potential projects with poor potential outcomes in nondemocratic environments than in democratic environments, the regression estimates would underestimate the negative effect of nondemocratic environments on project success rates.

[46] Not all statistics underlying this section are reported in the data annex. In particular, a discussion of time trends in the IEG ratings of PSM and non-PSM projects between 1990 and 2013 are omitted, which could be driven by factors such as changes in the worldwide political context or the Bank's lending strategy. For details on this and other descriptive statistics, please refer to Blum (2014).

[47] Interestingly, when considering the IEG institutional impact ratings, no equivalent performance difference is observable. With an average of 48.05 percent of PSM projects (broadly defined) performing at least "moderately satisfactorily" and 46.74 percent of non-PSM projects, the performance difference is negligible and not statistically significant. This also holds for "upstream" PSM projects (see Blum 2014).

Figure 5. Average Share of Projects Rated "Moderately Satisfactory" or Better: PSM Projects versus Non-PSM Projects
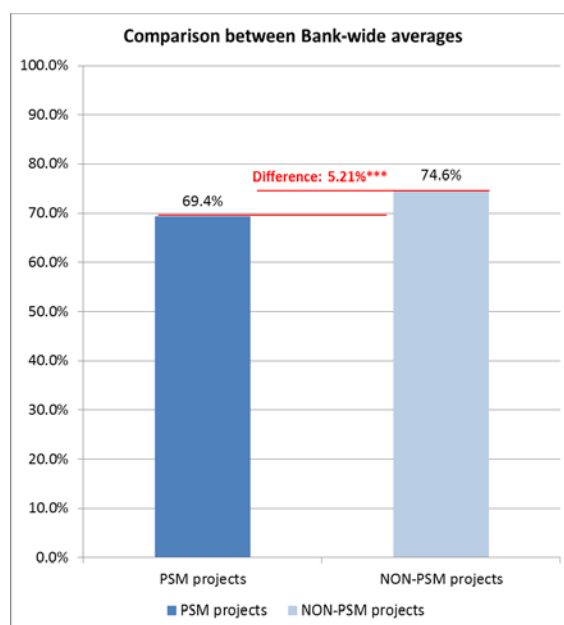
Source: Author's own compilation.

The inferior performance of PSM projects, relative to non-PSM projects, is in large part due to administrative and civil service reform (CSR) components. CSR components are the most widespread in PSM projects. Increasing the share of CSR components in a project by 10 percentage points makes it 2−3 percentage points less likely that the project will be rated "moderately satisfactory" or better (see annex table A.2).[48] The shares of PFM and decentralization components, meanwhile, do not predict performance ratings at statistically significant levels. In contrast to CSR, tax components appear to be associated with above-average performance ratings.[49] This finding is consistent with the findings of the IEG review of the World Bank's PSM portfolio conducted in 2008 (World Bank 2008).

6.2   Comparing Country Targeting

Overall, there are notable differences between the countries targeted by PSM projects and those targeted by non-PSM projects. While there is no significant difference in terms of per capita income, on average, PSM projects target less autocratic but more aid-dependent, slower-growing, and more unequal countries with lower levels of human development and lower levels of administrative capacity. For most context measures, these average targeting differences are not very large—but nevertheless statistically significant

---

[48]   This finding holds at the 95 percent significance level.
[49]   But this finding is only statistically significant in some specifications (see annex table A.2).

at high levels (typically 1 percent).[50] In regional terms, PSM projects are more concentrated in Africa and Latin America and the Caribbean (LAC) than non-PSM projects.[51]

PSM projects are on average targeted toward countries with a slightly higher degree of political and civil liberties than non-PSM projects. This finding holds based on all major indicators of political and civil liberties employed in this paper. For example, PSM projects were on average targeted toward countries with an average "polity" score of 2.6, compared to an average score of 1.8 for non-PSM projects (on a scale from -10 [most authoritarian] to 10 [most democratic]; see annex table A.3).[52] By contrast, there is no statistically significant difference in the number of political parties in countries with PSM projects and those with non-PSM projects (see annex table A.3).

PSM projects are targeted toward countries with significantly higher official development aid (ODA) flows (as a share of gross national income, GNI) than are non-PSM projects. On average, PSM projects target countries with net ODA flows of 9.5 percent of GNI, compared to 7 percent for non-PSM projects (see annex table A.4).[53]

There is some evidence, but not strong, that PSM projects are targeted based on needs—that is, toward countries with slightly lower initial administrative capability (see annex table A.3). On average, all CPIA scores relating to public administration (CPIA 12 to 16) are lower for the average country targeted by a PSM project than for the average country targeted by a non-PSM project. But these differences are small and only statistically significant for the quality of budgetary and financial management (CPIA 13, at the 10 percent significance level) and for transparency, accountability, and corruption in the public sector (CPIA 16, at the 5 percent significance level).[54]

On average, PSM projects are not targeted toward countries that are significantly richer or poorer (in terms of GDP per capita) than countries with non-PSM projects (see annex table A.4). But, interestingly, PSM projects appear to be targeted toward countries with a significantly lower growth rate. At the time of project approval, the average GDP per capita growth rate in countries with PSM projects was 2.2 percent (with a large standard deviation of 4.8 percentage points) compared to 2.9 percent for the average non-PSM project (see annex table A.4).[55]

---

[50] As a standardized measure that can help to compare the order of magnitude of these differences across covariates, annex tables A.3 and A.4 report the difference in means between PSM and non-PSM projects in terms of the number of standard deviations of the distribution of non-PSM projects.

[51] Please refer to Blum (2014) for details.

[52] This difference is statistically significant at the 10 percent significance level.

[53] For countries with "upstream" PSM projects, this difference is particularly striking. They receive an average of 12.4 percent of their GNI in form of (net) ODA flows—nearly twice as much as countries with non-PSM projects. This difference is statistically significant at the 1 percent significance level. One possible explanation for this could be that donors emphasize PFM reforms—the most important PSM reform theme in upstream projects—in countries that receive large amounts of ODA, with the view of enhancing governments' capacity to manage aid money through their PFM systems. For details, please refer to Blum (2014).

[54] Similarly small differences are manifest in terms of the WGI government effectiveness scores (statistically significant at the 10 percent significance level) and the ICRG Bureaucratic Quality Rating (statistically significant at the 5 percent significance level).

[55] This difference is statistically significant at the 1 percent significance level.

The average country targeted by PSM projects also performs worse on key human development indicators than the average country targeted by non-PSM projects. A life expectancy at birth of 62.2 years in PSM-target countries compares to 63.8 years in non-PSM countries and an under-5-year-old mortality rate of 88.7 per 1,000 compares to 77.8 per 1,000 in the average non-PSM country (see annex table A.4). While gross primary school enrollment rates are slightly higher in the average PSM country, secondary enrollment rates are slightly lower.

As previously noted, PSM projects are particularly targeted toward Africa and LAC.[56] About 35.5 percent of PSM projects are targeted toward the Africa region. By contrast, only 21.2 percent of non-PSM projects are targeted toward Africa. Similarly, about 25 percent of PSM projects are targeted toward the LAC region, compared with only 18.5 percent of non-PSM projects. In turn, PSM projects are underrepresented in the Middle East and North Africa, East Asia and Pacific, and South Asia regions. The percentage of PSM projects targeted to these regions is only about half as high as the percentage of non-PSM projects.

6.3   Comparing Reform Content Factors

PSM projects, by definition, contain a larger share of PSM components than do non-PSM projects (see section 3.2). As shown in annex table A.5, the PSM share in broadly defined PSM projects is 28 percent on average, compared to 7.1 percent in non-PSM projects.[57]

The dominant PSM component in PSM projects is CSR, representing 9.1 percent of total components on average. PFM and decentralization—next on the list of PSM components—are less important, with averages of 4.2 and 4 percent, respectively. Put differently, 22.1 percent of PSM projects have CSR components that represent 25 percent or more of the total project.[58] By contrast, only about 9.8 percent of these PSM projects have a share of PFM components greater than 25 percent.

6.4   Comparing Process and Risk Factors

During the first half of project implementation, TTLs flag most process or risk indicators with similar frequency across PSM and non-PSM projects—with three interesting exceptions. Project management is flagged in 25 percent of PSM projects, compared to 19 percent of non-PSM. Monitoring and evaluation, too, are flagged more often in PSM projects (13 percent) than in non-PSM (10 percent). Finally, the safeguards flag is raised less often for PSM projects (4 percent) than non-PSM (6 percent).[59] There is no

---

[56] For detailed descriptive statistics on regional targeting, please refer to Blum (2014).
[57] The PSM share in (narrowly defined) "upstream" PSM projects is on average 85 percent. All average component share differences between PSM and non-PSM projects are statistically significant at the 1 percent significance level.
[58] See Blum (2014) for detailed statistics.
[59] All differences reported here are statistically significant at the 1 percent significance level. See Blum (2014) for detailed descriptive statistics. That the safeguards flag is raised less frequently in PSM projects has an intuitive explanation. Compared to infrastructure projects or human or

statistically significant difference in how TTLs score the ISR ratings for progress (i) in implementation and (ii) toward development objectives. But, interestingly, 10 percent of PSM projects are flagged for "proactive" support during the first half of their implementation, compared to 7 percent of non-PSM projects.

6.5   Comparing Project Management Factors

PSM projects are on average much smaller than non-PSM projects, and slightly faster and less costly to prepare. They are, on average, only about half as big as non-PSM projects in terms of committed dollar amounts ($41.3 million compared to $82 million; see annex table A.6).[60] The average time required from PCN review to approval is 632 days for PSM projects, compared to 671 days for non-PSM projects. By contrast, the time required from approval to effectiveness does not differ at statistically significant levels between PSM and non-PSM projects. PSM projects are also cheaper to prepare—with an average lending preparation cost of $321,000 compared to $363,000 for non-PSM projects. PSM projects' lower preparation costs and shorter preparation times are roughly proportionate to their lower average project size.

7   ESTIMATION RESULTS

This section summarizes the estimation results and interprets them in view of the guiding research questions set out in section 3. Causal interpretations need to be subject to caution, primarily because omitted variable bias is likely (see section 5.2). This paper's findings should not be used to inform the targeting and design of PSM reform projects. Policy makers and Bank TTLs are concerned with the question of which reform approach will produce the desired results in a given project and country. But as the findings from this paper are about the average PSM project, they need not hold in any specific country.[61] As Hausmann, Klinger, and Wagner (2008) argue regarding the limitations of growth regressions, "there is no certainty that any given country is an average country in this particular respect."

---

social development projects, PSM projects primarily target the government administration and affect citizens and the environment indirectly. The risk that these projects have adverse or undesired side effects on citizens and the environment is thus lower than for other projects. Explanations for the more frequent occurrence of the "project management" and the "monitoring and evaluation" flags are less obvious.

[60] This difference is statistically significant at the 1 percent significance level.

[61] Testing for interaction effects of reforms with country-level contextual variables allows for accommodating some heterogeneity in a regression design, but only within tight limits. Pushing this argument further, an analysis based on country-level data also necessarily neglects intracountry variation of public sector capacities and incentives—for example, across regions or public agencies. Such variation is often crucial to take into account reform design, as region- or agency-specific problems may require specific reform approaches and reforms might work in some agencies or regions but not in others. Recanatini, Prati, and Tabellini (2005), for example, show that the levels of corruption and associated institutional arrangements (such as internal audits) vary enormously across agencies and regions in many countries.

7.1 Country Context Correlates of PSM Project Performance

Political and Civil Liberties

This paper finds that World Bank PSM reform projects are, on average, less likely to perform "satisfactorily" in countries where citizens have strongly limited political rights and civil liberties. Ordered probit regression estimates employing the Freedom House country classification suggest that PSM projects in nonfree countries are about 10 percentage points less likely to be rated either "satisfactory" or "highly satisfactory" than projects in "free" countries (see annex table A.8). Similarly, ordered probit regression estimates employing the Polity IV classification suggest that PSM projects in "autocracies" are about 13 percentage points less likely to be rated either "satisfactory" or "highly satisfactory" than projects in "anocracies" (see annex table A.9).[62] Somewhat surprisingly, the estimates using Polity IV measures also suggest that projects in democracies perform satisfactorily less often than project in anocracies. These performance differences are statistically significant at the 10 or 5 percent significance level in a number of specifications, but are not robust to all specifications.

These findings are consistent with the argument that political systems that favor "taking" over "making" are also less likely to invest in building state capacity. Bank PSM projects aimed at building this capacity are thus less likely to deliver the expected results in such contexts. The finding is inconsistent with the argument that a "developmental state" logic where authoritarian rule can facilitate tough PSM reforms and thus successful Bank PSM projects applies in the average country. But this finding about the average does not rule out the possibility that countries' developmental paths vary—as Levy and Fukuyama (2010) suggest—and that some might well follow a "developmental state" logic.

Estimates consistently suggest that PSM projects also have a distinctly lower success rate than non-PSM projects in nonfree countries with limited political and civil liberties. By contrast, there are no such differences in "partially free" and "free" countries. As shown in annex figure A.2, in "nonfree" countries the marginal probability of receiving "satisfactory" ratings for PSM projects is about 7−8 percentage points lower than for non-PSM projects. The probabilities for receiving below MS+ ratings are significantly higher (based on ordered probit regression estimates). This finding is confirmed in an equivalent estimation that uses the Polity scores instead of the Freedom House status (see annex figure A.3).

Matching estimates yield consistent results. As reported in annex table A.12, nearest neighbor matching estimates suggest that the share of PSM projects (broadly defined) that are rated MS+ in nonfree contexts is about 8.4 percentage points lower than for matched non-PSM projects in the same countries, at the 5

---

[62] Following the Polity IV terminology, "anocracies" are polities in the middle of the "democratic" continuum between "autocracies" and "democracies." On the Polity scale from -10 (full autocracy) to +10 (full democracy), anocracies are defined by scores ranging from -5 to + 5.

percent significance level. Subclassification matching estimates yield very similar findings (also annex table A.12), though only at the 10 percent significance level. There is no significant difference between PSM and non-PSM projects in partially free and free environments.

These findings suggest that PSM projects are distinctly more vulnerable to a lack of civil liberties and political rights than World Bank–supported projects in other reform areas. They are consistent with the hypothesis that PSM reforms are distinctly vulnerable to political economy logic of rent-seeking, possibly because they are harder to insulate from the influence of adverse political incentives than other reform areas.

Aid and Natural Resource Dependency

Natural resource dependency (as measured by the share of fuel, ores, and metals exports in GDP) is not a statistically significant predictor of PSM project performance.[63]

A key finding of this review is that PSM projects are more likely to succeed in countries with high aid flows, controlling for other contextual factors. Across different regression specifications (see annex tables A.7 and A.8), ODA received (as a percentage of GNI) is consistently positively correlated with PSM project performance, at high levels of statistical significance (mostly 1 percent significance level). But ODA flows only make a relatively small difference for PSM project performance. (Ordered) probit regression estimates suggest that a 10 percentage point increase in ODA (as a percentage of GNI) is on average associated with a 3−5 percentage point increase in the likelihood of receiving an IEG outcome rating of MS+. This finding seems particularly relevant in the light of the fact that PSM projects have been targeted toward countries that, on average, had higher aid flows than non-PSM projects. Interestingly, World Bank aid flows alone are not a significant predictor of project performance (see annex table A.7).

This positive correlation between aid flows and PSM project success rates is consistent with arguments that emphasize that donor bargaining power and/or government exposure to imported ideas may make governments more responsive to donor PSM reform pressures. But more donor pressure also heightens the risk that governments respond to such pressures with "isomorphic mimicry" and that, while project targets are met, real reform results are not achieved. This finding thus raises the question for subsequent research whether the objectives and results of PSM projects in more aid-dependent countries are indeed more prone to patterns of "isomorphic mimicry" than in non-aid-dependent countries.

Interestingly, whereas aid dependency is a statistically significant predictor of PSM project outcome ratings, it is not for non-PSM projects.[64] PSM projects are about 4−5 percentage points less likely to

---

[63] See Blum (2014) for details.

receive "satisfactory" IEG outcome ratings in countries with a low share of ODA in GNI than non-PSM projects and are about 3 percentage points more likely to receive "unsatisfactory ratings," at the 5 percent significance level. As ODA increases, this difference becomes statistically insignificant.[65]

One interpretation of this finding could be that donor bargaining power is particularly important for negotiating reforms that seek to establish "rule-based" institutions governing the public sector with client governments, because these reforms are harder to "insulate" from core government systems and thus are likely to meet with stronger resistance. If such bargaining power is lacking, PSM projects may therefore face lower odds of success than non-PSM projects. Why this difference does not persist in countries with very high aid dependency, as the matching estimates suggest, remains unclear.

Programmatic Political Parties

Consistent with Cruz and Keefer (2010), this paper finds that the share of political parties is a key country-level predictor of PSM project performance.[66] PSM projects in countries with 100 percent programmatic political parties are on average about 20 percentage points more likely to receive at least "moderately satisfactory" IEG outcome ratings than PSM projects in countries with no programmatic parties (see annex tables A.7 and A.9). This finding is very robust across specifications and holds at the 1 or 5 percent significance level in most specifications.

Adding to Cruz and Keefer's analysis, the paper finds that the existence of programmatic political parties is a distinctive predictor of PSM project performance, compared to non-PSM projects. Indeed, the share of programmatic political parties is not a statistically significant predictor for the performance of non-PSM projects.[67] As shown in annex figure A.4, based on ordered probit marginal effect estimations, PSM projects are about 5 percentage points less likely than non-PSM projects to receive "satisfactory" IEG outcome ratings in countries with a low share of programmatic political parties and are about 3–4 percentage points more likely to receive "unsatisfactory" ratings. This finding holds at the 5 percent significance level. It is confirmed by matching estimates, as reported in annex table A.15. In countries with a share of programmatic political parties smaller than 50 percent, PSM projects are on average 7–8

---

[64] This finding holds when excluding countries with extremely high aid dependency from the estimation (ODA as a share of GNI greater than 50 percent). See Blum (2014) for details.

[65] This finding is confirmed and nuanced by nearest neighbor and coarsened exact matching estimates that compare the average ratings of PSM and non-PSM projects in the same country, grouped by the level of aid dependency (see annex table A.17). These matching estimates suggest that there is no statistically significant difference between PSM and non-PSM projects in the share of projects rated MS+ in countries with a share of ODA in GNI that is smaller than 5 percent and in countries with a share greater than 20 percent. But for projects in countries with a share of ODA in GNI in the 5–20 percent bracket, the share of MS+ ratings for PSM projects received is 7–8 percentage points lower than for non-PSM projects. These projects represent a large share of the sample, with about one-third of the total number of projects.

[66] Regression estimates suggest that the success rate of PSM projects in countries with 100 percent programmatic parties is, on average, about 11–15 percentage points higher than in countries with no political parties (see annex table A.7), holding constant the degree of political and civil liberties and other context, content, and project management factors. These estimates, however, do not meet standard significance levels in all specifications, when controlling for other contextual, reform content, and project management factors.

[67] See Blum (2014) for details.

percentage points less likely to receive MS+ ratings than non-PSM projects. This difference disappears in countries whose share of political parties is greater than 50 percent.

This finding is consistent with the argument that it might be easier for donors to insulate non-PSM projects from adverse political incentives arising from a lack of political parties. A broader reading is that the emergence of political parties is a proxy indicator for unobserved characteristics of the politico-administrative system that influence government commitment to provide public goods. This finding points to the question of how programmatic parties influence public administration reform as a relevant area for further research.

Prior Administrative Ability

This review finds no clear evidence that PSM reforms are more likely to succeed in countries with higher initial administrative capacity. None of the CPIA scores relating to public sector management (CPIA 12 to 16)—as measured at baseline—predicts PSM project performance at statistically significant levels when controlling for other context, content, and project management factors.[68] Similarly, an alternative measure of administrative capacity, the ICRG Bureaucratic Quality Ratings, does not predict PSM project performance at statistically significant levels for any specification, although it is available for a much larger sample of PSM projects (674 out of a total of 1,097 PSM projects are included in the estimations in annex table A.7).

This finding can plausibly be interpreted in the sense that TTLs do a reasonably good job in adjusting the ambitiousness of their project objectives to the level of administrative capacity they find, so that the odds of reaching these objectives remain largely similar in low- and high-capacity countries.

Links between Economic and Human Development and PSM Reform

PSM projects are rated equally in poor and rich countries.[69] Basic human development indicators, such as life expectancy at birth and secondary school enrollment rates (gross), also do not predict project performance (see annex tables A.7 and A.8). PSM projects do perform slightly better in faster-growing countries, but the estimated coefficients are small. A 1 percentage point increase in annual growth rates at baseline is associated with a small, about 0.3 percentage point, increase in the likelihood of the project being rated MS+ (see annex table A.7). This estimate is statistically significant in some estimations.

---

[68] It is, however, possible that CPIA scores at baseline remain statistically insignificant because they are only available for a small sample of about 313 projects effectively included in the relevant estimation. This also holds true when only controlling for single CPIA indicators 13 (Quality of Budgetary and Financial Management) or 15 (Quality of Public Administration) without other CPIA controls. See Blum (2014) for details.

[69] See Blum (2014) for details. This finding may seem surprising, as one would expect countries with higher income levels to have more able administrations and higher success rates of PSM reforms. But "income level" is only a rough proxy for more specific, associated characteristics of countries that matter for PSM reform—such as administrative capacity and political incentives for reform (higher income countries are more democratic). At least once these more specific predictors of PSM reform success are controlled for (as in many specifications), it seems plausible that a measure of the level of economic development does not independently predict PSM project success.

It is interesting that economic growth rates are the sole nonpolitical economy country characteristic that predicts PSM (and non-PSM) project performance at statistically significant levels. One plausible explanation is that periods of growth open both fiscal and political space for more successful project implementation.

7.2   Reform Content Correlates of PSM Project Performance and PSM Project Targeting

This paper finds that none of the PSM project content measures systematically predicts PSM project performance at statistically significant levels, when controlling for other context and project management characteristics. The distinctively low performance of CSR projects and distinctively high performance of tax projects observed without such controls (see section 6.1 and annex table A.2) do not completely disappear but become statistically insignificant when such controls are introduced (see annex table A.7). CSR reform content measures are negative predictors of project performance in some specifications, but only at the 10 percent significance level.

The distinctive targeting of PSM projects explains a large share of the performance gap between PSM and non-PSM projects. Figure 6 compares how PSM and non-PSM projects perform with and without controlling for country context. It shows that the performance difference becomes small and statistically insignificant, when comparing PSM and non-PSM projects in the same country contexts.[70]

---

[70] The left-hand panel provides mere descriptive statistics, showing the 5.2 percentage point difference in the likelihood of being rated MS+ between PSM and non-PSM projects (see section 6). The right-hand panel shows the performance difference between PSM projects and (matched) non-PSM projects that were approved in the same countries at about the same time (see annex table A.11 for detailed matching estimation results). Regression estimates that control only for observable country context characteristics also point in this direction. When controlling for observable country context characteristics (including context measures, regional controls, and the "country record" flag), the estimated difference in the share of MS+ ratings between PSM and non-PSM projects shrinks by half, from 5.2 to 2.7 percentage points, and becomes statistically insignificant. See Blum (2014) for details.

Figure 6. Average Share of PSM and non-PSM Projects Rated MS+, With and Without (Nearest Neighbor) Matching by Country
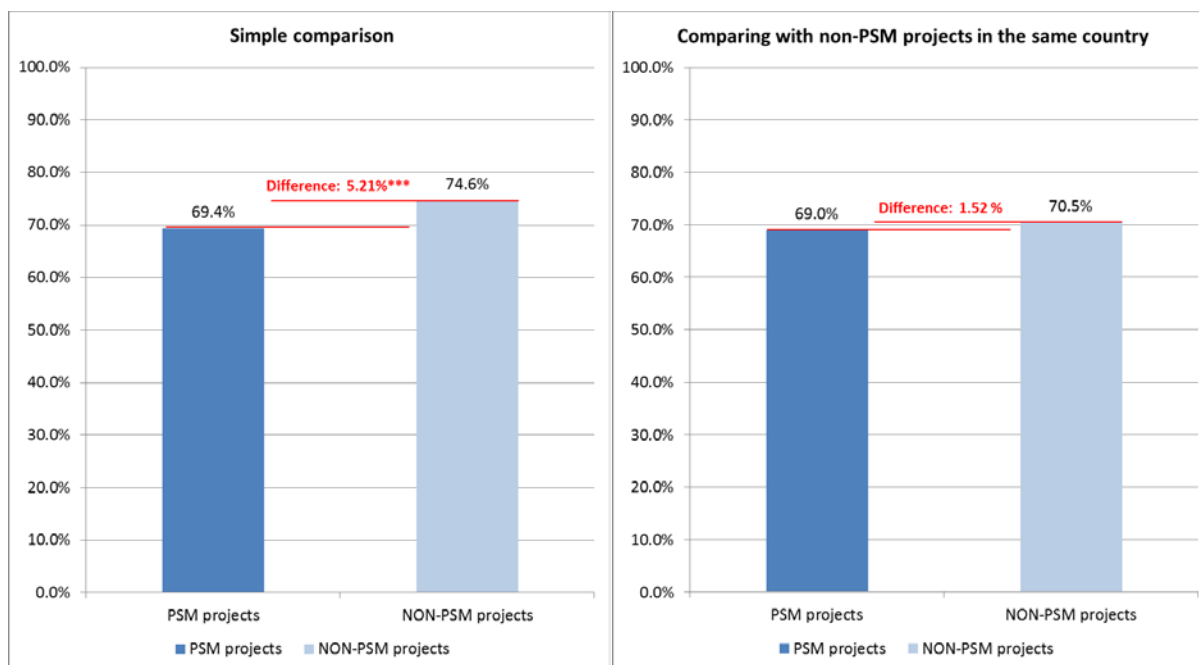
Source: Author's own compilation.

Note: The left-hand panel provides descriptive statistics; the right-hand panel is based on the nearest neighbor matching estimation results reported in annex table A.11.

Interestingly, this is the case despite the fact that PSM projects are particularly targeted to countries with observable characteristics that should favor their performance—in particular to more aid-dependent and less-autocratic contexts. Controlling for these characteristics should thus widen and not narrow the performance gap. The likely explanation is that other unobservable context characteristics that are controlled for in a matching estimation approach explain the performance gap.

Figure 7 presents findings, equivalent to those illustrated in figure 6, when disaggregating PSM projects based on their major content areas. It shows that the performance differences observed without controlling for country context (left-hand panel) are significantly reduced when comparing with non-PSM projects in the same country and become statistically insignificant—but do not fully disappear.[71]

Overall, these findings do not debunk the idea that particular types of PSM reform are challenging. But they do suggest that the performance difference observed between PSM and non-PSM projects is to a significant extent due to their different targets. Once these context characteristics are controlled for, performance differences shrink to become statistically insignificant.

---

[71] See Blum (2014) for details.

Figure 7. Average share of PSM and Non-PSM Projects Rated MS+, With and Without Matching by Country, by Subtheme
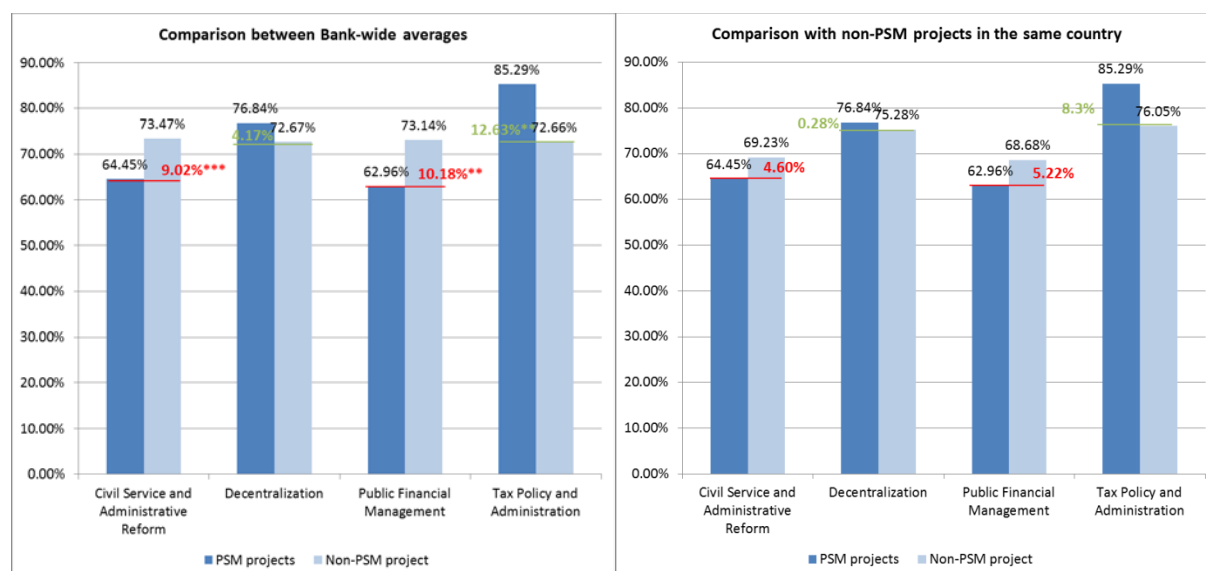
Source: Author's own compilation.

The finding that none of the PSM subthemes predicts project performance at statistically significant levels—when controlling for county contexts—challenges the idea of a performance "hierarchy" among PSM reform areas. For example, the view that PFM reforms are generally more successful than CSR and anti-corruption reforms has been reinforced by the IEG review "What Works and Why in PSM Reform" (World Bank 2008). The IEG review suggests that Bank lending for administrative and civil service reforms did less to improve the relevant CPIA score (CPIA 15) than Bank lending for PFM reform (CPIA 13). This paper's findings, which are based on project outcome measures, do not support such a view.[72]

7.3   Process and Risk Indicators

Three risk indicators tracked by TTLs in ISRs during the first half of project implementation provide very useful predictors of PSM project performance. The "country record" flag, the lowest ISR "development objective" rating provided by the TTL, and the "project at risk" flag consistently predict PSM project performance at high levels of statistical significance (1 or 5 percent significance level). None of the other risk indicators does so, when controlling for country context and basic project management factors. Jointly, these process indicators significantly improve the fit of the predictive model—from a McFadden pseudo-R^2 of 0.106 to 0.19. They also improve the proportional reduction of error compared to the modal category from 5.3 percent to 20.83 percent (see annex table A.7).

---

[72] The caveats of both measures of "project performance"—IEG project outcome ratings versus CPIA scores—however, need to be highlighted. The CPIA scores suffer from a major attribution problem (did the scores improve because of the World Bank intervention or for other reasons?) but arguably reflect properties of the public administration that are relevant to overall development outcomes. By contrast, the IEG project outcome ratings are clearly attributable to the project—but need not reflect any progress relevant to broader development objectives.

First, if a "country record" flag is raised during the first half of project implementation, the odds of receiving an MS+ IEG outcome rating on average decline by about 15 percentage points (see annex table A.7) for PSM projects. This finding is particularly relevant since the country record flag does not reflect the performance of the PSM project in question[73] but the performance of the Bank's overall lending portfolio in the respective country.[74] The country record flag was raised in about 17 percent of PSM projects[75] and thus should be considered as a telling, and relatively frequent, risk indicator. Second, TTL ratings of a PSM project's progress toward reaching development objectives (PDO ratings) during the first half of project implementation predict project outcome ratings well. An increase of the *minimal rating* used by TTLs for this measure by 1 point on the ordinal 6-point rating scale is on average associated with a 3.6 percentage point increase in the likelihood of receiving MS+ IEG outcome ratings (see annex table A.7). Finally, if a project receives an "at risk" status at least once during the first half of its implementation, its outcomes are on average about 6.2 percentage points less likely to be rated "moderately satisfactory" by the IEG (see annex table A.7).[76]

One relevant conclusion from these findings for the Bank's approach to monitoring risk in PSM projects is that the TTLs' subjective assessment of PSM project performance and riskiness is a much more telling risk predictor than the more standardized risk flags used to monitor project implementation progress.[77] This is unsurprising in the sense that TTLs possess rich information about their projects and are best positioned to assess risk. But they may not have the opportunity or incentives to fully reveal these risks to management. This finding suggests that making sure that TTLs' concerns are heard and that open conversations about risk are encouraged need to be core elements of the Bank's approach to handling risk. These steps can be complemented but not substituted for by a system of systematically measuring specific risks (as is done by the risk flags). The "Operational Risk Assessment Framework" (ORAF) introduced by the Bank in 2011 points in this direction, offering task teams a way to systematize subjective risk assessments.

---

[73] Three risk flags—"country record," "country environment," and "effectiveness delays"—do not capture project implementation performance. The other nine risk flags are linked to implementation performance.

[74] More precisely, the "country record" flag gets triggered when the Operations Evaluation Department's (OED's) evaluations find at least one of three conditions to be true: (i) the net disconnect—a measure of realism of regional staff's portfolio performance assessments[74] is 20 percent or higher; (ii) disbursements of projects associated with an unsatisfactory rating (by OED) are 40 percent or more of completed projects; (iii) the Country Assistance Evaluation (CAE) has been less than satisfactory in the previous five fiscal years. The country record flag thus should be considered as a Bank-specific country-context measure that signals tough lending environments, rather than as a project-specific measure.

[75] See Blum (2014) for detailed statistics.

[76] As a project's "at risk" status is triggered when either its "implementation progress" (IP) or "development objective" (DO) rating is below moderately satisfactory, it is closely correlated with these two measures—and it is surprising that it has distinctive predictive power for a project's final IEG outcome rating.

[77] According to the Bank's ISR guidelines, "the PDO rating is forward-looking, in that it assesses the likelihood that the PDO will be achieved." More precisely, "for projects in the early stages of implementation, for which intermediate outcomes are not yet observable, the PDO rating is based on (i) implementation performance ratings and achievement of scheduled outputs and (ii) judgments about the likelihood that major risks—factors outside the control of the project—can jeopardize the achievement of the project's outputs and/or outcomes." In practice, the PDO rating is seen as the one risk indicator in the ISR that the TTL can use with great discretion to signal to management a subjective assessment of a project's performance. Most of the other risk indicators—flags and the IP rating—are triggered based on more mechanical/standardized criteria with less discretion.

In comparing risk indicators that are predictive of the IEG outcome ratings for PSM and non-PSM projects, the development objectives rating is a highly significant predictor for both types of projects. By contrast, it is striking that the "country record" flag is associated with PSM projects having a 15 percentage point lower average likelihood of being rated MS+, but does not have predictive power for non-PSM projects.[78]

This difference raises the question of why PSM projects should be particularly vulnerable to contexts where the Bank's overall lending portfolio has a poor track record. While the answer is not obvious, one possible interpretation could be that the Bank's lending portfolio performs particularly poorly if World Bank–client government relations are difficult (for example, because objectives are misaligned). Projects with significant PSM components may be particularly affected by such difficulties, as they seek to affect the—possibly resistant—government administration directly. Other projects may operate in stronger isolation from government systems.

7.4   Project Management Correlates of Project Performance

None of the project management indicators observable at the time of a PSM project's approval—the committed amount, lending preparation costs, time to approval, or time to effectiveness—are statistically significant predictors of project performance, when controlling for country context and project content characteristics (see annex table A.7). But when not controlling for context and content characteristics, longer gaps between project approval and project effectiveness ("time to effectiveness") predict slightly lower chances of a project receiving MS+ IEG outcome ratings (see model 3 in annex table A.7). An increase in the time to effectiveness of PSM projects by 100 days is on average associated with a decrease in the likelihood of being rated MS+ by about 2.5 percentage points.

One interpretation of this finding is that long times to effectiveness can signal a lack of government commitment to a given PSM project. It seems plausible that times to effectiveness do not improve the model when country context is controlled for, because one would expect them to be caused, at least in part, by country context factors. Overall, including the four above potential predictors of project performance does not improve the fit of the model (see annex table A.7), suggesting that these characteristics are not useful risk indicators for PSM projects.

---

[78] See Blum (2014) for detailed statistics.

8   CONCLUSION

This paper's most noteworthy findings relate to the function of political economy factors as key predictors of how the World Bank's PSM projects perform. PSM projects perform better in countries with democratic than autocratic regimes; they fare better in more aid-dependent countries than in less; and they benefit from the presence of programmatic political parties. Importantly, these factors distinctly predict how PSM projects perform compared to non-PSM projects, suggesting that PSM projects are particularly sensitive to or harder to insulate from political contexts than non-PSM projects. This implies that the Bank might carefully consider how to align PSM project designs with political incentives.

While some case studies indicate that a few autocracies have been able to push tough PSM reforms, this paper suggests that this is the exception and not the rule, at least insofar as it is reflected in World Bank PSM project performance. Rather, this paper lends support to arguments that inclusive rather than exclusive political institutions create incentives that are conducive to better World Bank PSM project performance.

Above-average PSM project performance in more aid-dependent countries suggests that the bargaining power of donors compared with that of client governments may positively influence project outcomes. But it also points toward the risks that PSM project success will be on the surface only. Arguably, where client governments have less bargaining power, they also have less ownership and may pursue reforms for the sake of legitimacy in donor eyes, rather than performance (Andrews 2009). If this holds true on average, better project performance might simply reflect better compliance with donor demands but not better results on the ground—or benefits to citizens. As this paper exclusively relies on IEG ratings, it leaves room for more comprehensive research in this area.

Consistent with Cruz and Keefer's (2010) prior research, this paper finds that the existence of programmatic political parties is highly conducive to better PSM project performance. Adding to their research, it finds that the absence of programmatic political parties harms PSM project performance more than non-PSM. This finding is surprising, as many PSM reforms (with the exception of civil service reforms) seem less suited as electoral platforms than other broad public goods supported by World Bank projects, such as health or education service delivery or infrastructure. One possible explanation may be that while the absence of programmatic political parties would in principle be equally detrimental to PSM and non-PSM projects, it might be easier to insulate non-PSM projects from prevailing adverse incentives than PSM projects, because the latter directly seek to change the government systems that have been shaped by these incentives. This finding points to the relationship between the legislature and public administration reform in developing countries as a promising and underexplored field of research.

This paper does not find evidence that a country's administrative capacity at the time of a project's approval predicts PSM project performance. While this may seem surprising, it is likely explained by two factors. First, it may simply mean that TTLs realistically adjust the ambitiousness of their projects' objectives to the level of administrative ability they find at baseline. If this is the case, PSM projects in low-capacity environments would be equally likely to perform well as PSM projects in high-capacity environments. Second, the most reliable comparative data source on administrative capacity—the CPIA scores—is only available for a small share of the projects in the sample, making it more difficult to find statistically significant correlations.

A second key finding relates to the Bank's approach to managing risk in PSM operations. The review finds that subjective risk ratings provided by TTLs and the World Bank's overall portfolio performance in a given country are the most telling early risk indicators for PSM projects. By contrast, the set of 13 standardized risk flags utilized in ISRs does little to predict PSM project performance. This suggests that encouraging an open dialogue about risk between TTLs and management should be core to the Bank's risk management strategy going forward.

Finally, the paper finds that performance differences between different PSM reform content areas and the average Bank project are to a significant extent driven by the particular country contexts to which PSM projects are targeted. When comparing PSM projects with non-PSM projects in the same country and approved at a similar time, performance differences between PSM projects and non-PSM projects largely disappear. Performance differences between "low-risk" PSM reform content areas (such as taxes) and "high-risk" areas (PFM and civil service reform) shrink and become statistically insignificant, when controlling for country context. This suggests that a significant part of the "risk" of PSM projects is driven by location, rather than being inherent to what they attempt to achieve.

It is important to note that key factors that theoretically matter to PSM project performance remain omitted in this paper, due to data constraints—such as the immediate institutional context of the implementing client MDA. Such omitted variables pose a significant risk of bias, calling for caution in causally interpreting this review's findings. Overall, controlling for context and project covariates that are observable during the first half of project implementation improves the percentage of correctly predicted IEG outcome ratings for PSM projects from 71 to 77 percent, or reduces proportional prediction error by 21 percent. Adjusted and pseudo-R^2 suggest that these covariates can explain about 19 percent of IEG outcome ratings variation. That is not a negligible share, but clearly highlights that key predictors remain unobserved—predictors that account for the remaining 81 percent of variation.

Clearly, this finding to some extent reflects the inherent limits of quantitative cross-country regression research on public administration reform; many factors relevant to PSM reform success will always

remain hard to measure comparatively and can be captured more adequately in qualitative research. But it also points to the importance of generating better comparative data on public administration systems in developing countries—data that may not only serve to track reform progress but also fuel future comparative research on administrative reform.

REFERENCES

Abadie, Alberto, David Drukker, Jane Leber Herr, and Guido W. Imbens. 2004. "Implementing Matching Estimators for Average Treatment Effects in Stata." *The Stata Journal* 4 (3): 290–311.

Acemoglu, Daron, and James Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. New York: Crown Publishing.

Aldrich, John. 1995. *Why Parties? The Origin and Transformation of Party Politics in America*. Chicago: University of Chicago Press.

Andrews, Matthew. 2006. "Beyond 'Best Practice' and 'Basics First' in Adopting Performance Budgeting Reform." *Public Administration and Development* 26: 147–61.

Andrews, Matthew. 2009. "Isomorphism and the Limits to African Public Financial Management Reform." HKS Faculty Research Working Paper RWP09-012, Harvard Kennedy School, Cambridge, Massachusetts.

———. 2010. "How Far Have Public Financial Management Reforms Come in Africa?" HKS Faculty Research Working Paper Series RWP10-018, Harvard Kennedy School, Cambridge, Massachusetts.

———. 2011. "Does Change Space Influence Public Sector Management Project Success?" Note prepared for PRMPS as an input to the Public Sector Management Approach 2011–20, unpublished.

Andrews, Matthew, Lant Pritchett, and Michael Woolcock. 2010. "Capability Traps? The Mechanisms of Persistent Implementation Failure." Center for Global Development Working Paper No. 234, Center for Global Development, Washington, DC.

———. 2012. "Escaping Capability Traps through Problem-Driven Iterative Adaptation." Center for Global Development Working Paper No. 299, Center for Global Development, Washington, DC.

Andrews, Matthew, Roger Hay, and Jerrett Myers. 2010. "Governance Indicators Can Make Sense: Under-five Mortality Rates are an Example." HKS Faculty Research Working Paper RWP10-015, Harvard Kennedy School, Cambridge, Massachusetts.

Armenakis, Achilles A., and Arthur G. Bedeian. 1999. "Organizational Change: A Review of Theory and Research in the 1990s." *Journal of Management* 25: 293–315.

Bandiera, Oriana, Andrea Prat, and Tommaso Valletti. 2009. "Active and Passive Waste in Government Spending: Evidence from a Policy Experiment." *American Economic Review* 99 (4): 1278–308.

Bates, Robert. 1992. *Prosperity and Violence: The Political Economy of Development*. New York: W. W. Norton & Company.

Blackwell, Matthew, Stefano Iacus, and Gary King. 2009. "CEM: Coarsened Exact Matching in Stata." *The Stata Journal* 9 (4): 524–46.

Blum, Jürgen. 2014. "What Factors Predict How Public Sector Projects Perform? A Review of the World Bank's Public Sector Management Portfolio. Complete Version." Draft Working Paper, Governance and Public Sector Management Practice, World Bank, Washington, DC. http://go.worldbank.org/7GW473ZD10

Blum, Jürgen, and Nick Manning. 2011. "How Will We Move Towards a 'Working Theory of Public Sector Management Change' in Developing Countries?" Unpublished Working Paper, World Bank, Washington, DC.

Blum, Jürgen, Nick Manning, and Vivek Srivastava. 2012. "Public Sector Management Reform: Toward a Problem-Solving Approach." Economic Premise No. 100, World Bank, Washington, DC.

Cruz, Cesi, and Philip Keefer. 2010. "Programmatic Political Parties and Public Sector Reform." APSA 2010 Annual Meeting Paper. http://ssrn.com/abstract=1642962.

Dal Bó, Ernesto, Frederico Finan, and Martín A. Rossi. 2013. "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service," *Quarterly Journal of Economics* 128 (3): 1169–1218.

Denzier, Cevdet, Daniel Kaufmann, and Aart Kraay. 2011. "Good Countries or Good Projects? Macro and Micro Correlates of World Bank Project Performance." Policy Research Working Paper No. 5646, World Bank, Washington, DC.

Dollar, David, and Victoria Levin. 2005. "Sowing and Reaping: Institutional Quality and Project Outcomes in Developing Countries." Policy Research Working Paper No. 3524, World Bank, Washington, DC.

Evans, Peter, and James E. Rauch. 1999. "Bureaucracy and Growth: A Cross-National Analysis of the Effects of 'Weberian' State Structures on Economic Growth." *American Sociological Review* 64: 748−65.

Faguet, Jean-Paul. 2004. "Does Decentralization Increase Government Responsiveness to Local Needs? Evidence from Bolivia." *Journal of Public Economics* 88: 867−93.

Hasnain, Zahid. 2011. "Incentive Compatible Reforms: The Political Economy of Public Investments in Mongolia." Policy Research Working Paper Series No. 5667, World Bank, Washington, DC.

Hausmann, Ricardo, Bailey Klinger, and Rodrigo Wagner. 2008. "Doing Growth Diagnostics in Practice: A 'Mindbook'." Center for International Development Working Paper No. 177, Harvard Kennedy School, Cambridge, Massachusetts.

Isham, Jonathan, Daniel Kaufmann, and Lant Pritchett. 1995. "Governance and Returns on Investment: An Empirical Investigation." Policy Research Working Paper Series 1550, World Bank, Washington, DC.

Knack, Stephen. 2002. "Governance and Growth: Measurement and Evidence." IRIS Discussion Paper No. 02/15, Center for Institutional Reform and the Informal Sector, University of Maryland.

Levy, Brian, and Francis Fukuyama. 2010. "Development Strategies: Integrating Governance and Growth." Policy Research Working Paper No. 5196, World Bank, Washington, DC.

Linsky, Martin, and Ronald A. Heifetz. 1994. "Leadership on the Line: Staying Alive through the Dangers of Leading." Harvard Business School Press, Boston, Massachusetts.

Moynihan, Donald P. 2008. *The Dynamics of Performance Management: Constructing Information and Reform*. Washington, DC: Georgetown University Press.

North, Douglass, John J. Wallis, and Barry A. Weingast. 2009. *Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History*. Cambridge: Cambridge University Press.

Organisation for Economic Co-operation and Development (OECD). 2013. *Government at a Glance 2013*. OECD Publishing; 10.1787/gov_glance-2013-en.

Pierskalla, J. Henryk, Zahid Hasnain, and Nick Manning. 2012. "Pay Flexibility in the Public Sector: A Review of Theory and Evidence." Policy Working Paper No. 6043, World Bank, Washington, DC.

Pollitt, Christopher, and Geert Bouckaert. 2004. *Public Management Reform: A Comparative Analysis*. Oxford: Oxford University Press.

Rasul, Imran, and Daniel Rogger. 2013. "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service." Working paper, International Growth Centre, London School of Economics, June.

Recanatini, Francesca, Alessandro Prati, and Guido Tabellini. 2005. "Why Are Some Public Agencies Less Corrupt Than Others? Lessons for Institutional Reform from Survey Data." Paper Prepared for the Sixth IMF Jacques Polak Annual Research Conference on Reforms, International Monetary Fund, Washington, DC, November 3–4.

Schein, E. H. 1999. "Kurt Lewin's Change Theory in the Field and in the Classroom: Notes toward a Model of Managed Learning." Reflections 1, No. 1, Society for Organizational Learning, and the Massachusetts Institute for Technology.

———. 2002. "Models and Tools for Stability and Change in Human Systems." *Reflections* 4 (2). Society for Organizational Learning and the Massachusetts Institute for Technology.

Schick, Allan. 1998. "Why Most Developing Countries Should Not Try New Zealand Reforms." *World Bank Research Observer* 13 (1): 123–31.

Schneider, B. Ross, and Blanca Heredia, eds. 2003. *Reinventing Leviathan: The Politics of Administrative Reform in Developing Countries*. Miami: North-South Center Press at the University of Miami.

World Bank. 2004. *World Development Report 2004: Making Services Work for the Poor*. Washington, DC: World Bank.

———. 2005. *Harmonized Evaluation Criteria for ICR and OED Evaluations.* Washington, DC: World Bank.

———. 2008. *Public Sector Reform: What Works and Why? An IEG Evaluation of World Bank Support.* Washington, DC: World Bank.

———. 2012a. *Beyond the Annual Budget: Global Experience with Medium Term Expenditure Frameworks.* Washington, DC: World Bank.

———. 2012b. *World Bank's Approach to Public Sector Management 2011–2020: "Better Results from Public Sector Institutions."* Washington, DC: World Bank.

DATA ANNEX

A1. Descriptive Statistics

Table A.1 Difference-in-Means Test between PSM and Non-PSM Projects:
Mean Binary IEG Outcome Ratings ("success rate")

	Non-PSM Projects (complement of broad definition)			PSM Projects (broad definition)			Difference	
Variable	N	Mean	SD	N	Mean	SD	Difference in Means	robust SE
Binary IEG Outcome Rating ("Success Rate")	2,105	0.75	0.44	1,097	0.69	0.46	-.0521***	(0.017)
Binary IEG Institutional Impact Ratings ("Success Rate")	1,365	0.47	0.5	666	0.48	0.5	.0131	(0.023)

Figure A.1 Distribution of IEG Outcome Ratings by Public Sector Reform Content (categorical classification)

Bank average of 75 percent

Legend:
- highly satisfactory
- satisfactory
- moderately satisfactory
- moderately unsatisfactory
- unsatisfactory
- highly unsatisfactory
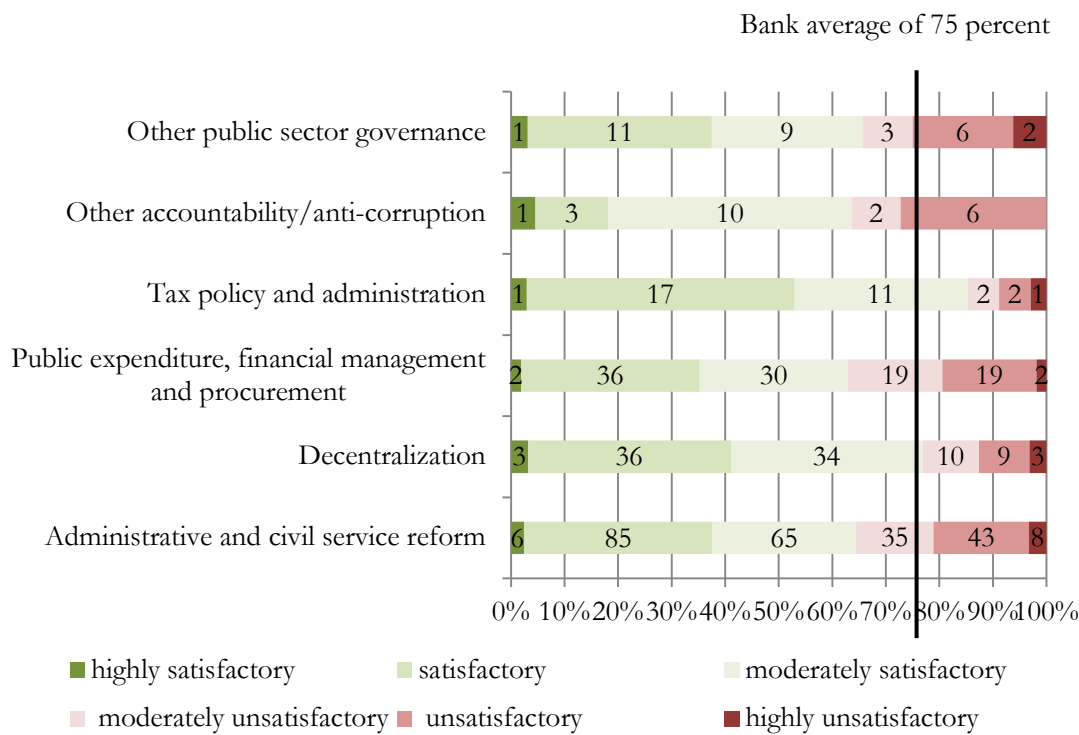
Note: A project is included in the respective theme if at least 25 percent of its components are classified with the respective theme code. The six public sector themes are not mutually exclusive, as many projects comprise > 25 percent of the components of different PSM themes.

Table A.2 Linear and Probit Regression Estimates: Performance Differences between PSM Themes
(measured as % of total project components)

Variable	Linear (OLS) Regression			Probit Regression (Marginal Effects Reported)		
	All Projects	PSM Projects (broad definition)	PSM Projects (narrow definition)	All Projects	PSM Projects (broad definition)	PSM Projects (narrow definition)
	(1)	(2)	(3)	(4)	(5)	(6)
	IEG Outcome Rating [binary]	IEG Outcome Rating [binary]	IEG Outcome Rating [binary]	IEG Outcome Rating [binary]	IEG Outcome Rating [binary]	IEG Outcome Rating [binary]
Administrative and Civil Service Reform	-0.00288***	-0.00213**	-0.00373**	-0.00268***	-0.00204**	-0.00365**
	(0.000916)	(0.000927)	(0.00166)	(0.000805)	(0.000863)	(0.00161)
Decentralization	0.000776	0.00163	-0.00195	0.000802	0.00175	-0.00205
	(0.00127)	(0.00133)	(0.00478)	(0.00132)	(0.00144)	(0.00447)
Public Expenditure, Financial Management, and Procurement	-0.00184	-0.00148	-0.00175	-0.00173	-0.00144	-0.00176
	(0.00124)	(0.00129)	(0.00160)	(0.00110)	(0.00120)	(0.00151)
Tax Policy and Administration	0.00280**	0.00308**	0.00214	0.00338	0.00384	0.00309
	(0.00135)	(0.00139)	(0.00159)	(0.00230)	(0.00243)	(0.00295)
Other Accountability/Anti-Corruption	-0.00164	-0.00174	-0.00355	-0.00149	-0.00165	-0.00338
	(0.00313)	(0.00316)	(0.00350)	(0.00285)	(0.00297)	(0.00313)
Other Public Sector Governance	0.000223	4.16e-05	-0.00421	0.000278	8.99e-05	-0.00400
	(0.00135)	(0.00150)	(0.00281)	(0.00140)	(0.00155)	(0.00259)
Constant	0.738***	0.709***	0.804***			
	(0.0185)	(0.0264)	(0.0659)			
Observations	3,202	1,097	178	3,202	1,097	178
R-squared	0.008	0.013	0.068			

Robust Standard Errors in Parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: IEG = International Evaluation Group.

Table A.3 Difference-in-Means Test between PSM and Non-PSM Projects:
Continuous Country Context Factors (non-WDI)

Variable	Non-PSM Projects			PSM Projects (broadly defined)			Difference in Means (absolute value)	Difference in Means in SD of Non-PSM projects	T-score	Robust SE	Missing Observations
	N	Mean	SD	N	Mean	SD					
Programmatic Political Parties											
Share of Political Parties	1,960	0.7	0.4	1004	0.6	0.4	-0.021	-0.053	-1.385	-0.015	238
WGI [all units SDN]											
WGI Control of Corruption	842	-0.5	0.5	490	-0.5	0.5	-0.043	-0.090	-1.517	-0.029	1,870
WGI Government Effectiveness	842	-0.4	0.5	490	-0.4	0.5	-0.048*	-0.099*	-1.724	-0.028	1,870
WGI Political Stability and Absence of Violence	848	-0.6	0.8	489	-0.6	0.8	0.014	0.017	0.302	-0.045	1,865
WGI Rule of Law	850	-0.5	0.6	490	-0.6	0.6	-0.065**	-0.116**	-2.028	-0.032	1,862
WGI Regulatory Quality	843	-0.3	0.6	490	-0.3	0.6	0.014	0.024	0.425	-0.033	1,869
WGI Voice and Accountability	850	-0.4	0.7	490	-0.3	0.6	0.110***	0.155***	2.909	-0.038	1,862
CPIA											
CPIA 12 Property Rights and Rule-base Governance	614	3.1	0.6	400	3	0.7	-0.034	-0.053	-0.819	-0.042	2,188
CPIA 13 Quality of Budgetary and Financial Management	614	3.6	0.6	400	3.5	0.6	-0.067*	-0.114*	-1.748	-0.038	2,188
CPIA 14 Efficiency of Revenue Mobilization	614	3.6	0.6	400	3.5	0.6	-0.055	-0.099	-1.516	-0.036	2,188
CPIA 15 Quality of Public Administration	614	3.3	0.5	400	3.2	0.5	-0.047	-0.087	-1.351	-0.035	2,188
CPIA 16 Transparency Accountability and Corruption	614	3.1	0.6	400	3	0.7	-0.082**	-0.134**	-1.993	-0.041	2,188
FH, POLITY, and ICRG											
FH Civil Liberties Rating [1–7]	2,057	4.2	1.4	1,088	4	1.3	-0.265***	-0.186***	-5.339	-0.05	57
FH Political Rights Rating [1–7]	2,057	4.1	1.9	1,088	3.9	1.8	-0.255***	-0.132***	-3.688	-0.069	57
POLITY Institutionalized Autocracy Score [0–10]	2,011	-0.3	15.3	1,064	-1.4	16.4	-1.054*	-0.069*	-1.733	-0.608	127
POLITY Institutionalized Democracy Score [0–10]	2,011	1.5	15.7	1,064	1.2	17.1	-0.337	-0.021	-0.533	-0.631	127
POLITY Combined Polity Score [–10–10]	1,986	1.8	6.4	1,047	2.6	5.9	0.770***	0.120***	3.310	-0.233	169
ICRG Democratic Accountability Rating [0–6]	1,754	3.4	1.4	884	3.6	1.3	0.166***	0.122***	3.073	-0.054	564
ICRG Bureaucracy Quality Rating [0–6]	1,754	1.9	0.8	884	1.8	0.8	-0.124***	-0.149***	-3.579	-0.035	564

Note: WGI = World Governance Indicators; CPIA = Country Policy and Institutional Assessments; FH = Freedom House; ICRG = International Country Risk Guide.

46

Table A.4 Difference-in-Means Test for PSM and Non-PSM Projects: Country Context Factors (WDI)

Variable	Non-PSM Projects			PSM Projects			Difference in Means (absolute value)	Difference in Means (in SD of non-PSM projects)	Robust SE	T-score	Observations	Missing Observations
	N	Mean	SD	N	Mean	SD						
GDP per Capita, PPP (constant 2005 international $)	2,059	3,855.1	3,413.9	1,090	3,964.6	3609	109.516	0.032	(132.692)	0.825	3,149	53
GDP per Capita Growth (annual %)	2,071	2.9	5.3	1,090	2.2	4.8	-0.693***	-0.130***	(0.188)	-3.693	3,161	41
Net ODA Received (% of GNI)	1,913	7	11.9	1,009	9.5	13.4	2.477***	0.209***	(0.501)	4.946	2,922	280
Net Bilateral Aid Flows from DAC Donors, Total (% of GDP)	2,073	0	0.1	1,090	0.1	0.1	0.013***	0.145***	(0.003)	4.025	3,163	39
IBRD Loans and IDA Credits (DOD, % of GDP)	1,945	0.1	0.1	1,040	0.1	0.2	0.031***	0.223***	(0.006)	5.262	2,985	217
External Debt Stocks, Public and Publicly Guaranteed (PPG) (DOD, % of GDP)	1,945	0.5	0.4	1,040	0.6	0.5	0.115***	0.273***	(0.019)	5.924	2,985	217
Fuel, Ores, and Metals Exports as a Share of GDP	1,875	4.00E+08	7.20E+08	972	3.40E+08	8.20E+08	-5.598e+07*	-0.078*	(31220550.613)	-1.793	2,847	355
Inflation, Consumer Prices (annual %)	1,996	51	289.9	1,035	55.6	341.9	4.626	0.016	(12.451)	0.372	3,031	171
School Enrollment, Secondary (% gross)	1,962	54.9	26.7	1,031	51.8	29.6	-3.148***	-0.118***	(1.101)	-2.860	2,993	209
Life Expectancy at Birth, Total (years)	2,076	63.8	8.7	1,090	62.2	9.9	-1.671***	-0.193***	(0.355)	-4.704	3,166	36
Population, Total	2,076	1.90E+08	3.80E+08	1,091	79117569.7	2.10E+08	-1.116e+08***	-0.298***	(10423965.768)	-10.709	3,167	35

Note: GDP = gross domestic product; PPP = purchasing power parity; ODA = official development assistance; DAC = Development Assistance Committee; IBRD = International Bank for Reconstruction and Development; WDI = World Development Indicators; IDA = International Development Association; DOD = debt outstanding and disbursed.

Table A.5 Difference-in-Means Test between PSM and Non-PSM Projects: Reform Content Variables

Variable	Non-PSM Projects			PSM Projects (broadly defined)			PSM Projects (narrowly defined)			Difference in Means (broad PSM and non-PSM)		Difference in Means (narrow PSM and non-PSM)	
	N	Mean	SD	N	Mean	SD	N	Mean	SD	Difference in Means	robust SE	Difference in Means	robust SE
Administrative and Civil Service Reform	2,105	0.5	3.1	1,097	9.1	16	178	17	20.5	8.591***	(0.487)	16.513***	(1.538)
Decentralization	2,105	0.9	3.9	1,097	4	9.7	178	3.5	9.1	3.020***	(0.305)	2.557***	(0.689)
Public Expenditure, Financial Management, and Procurement	2,105	0.2	1.7	1,097	4.2	12.1	178	18.2	21.8	4.029***	(0.368)	18.036***	(1.630)
Tax Policy and Administration	2,105	0	0.6	1,097	1.9	9.6	178	8.5	20.9	1.852***	(0.289)	8.465***	(1.559)
Total Public Sector (both themes and codes)	2,105	7.1	7.4	1,097	57.4	28.3	178	85.4	23.2	50.360***	(0.870)	78.385***	(1.742)

Table A.6 Difference-in-Means Test between PSM and Non-PSM Projects for Project Management Characteristics Observable at Approval/Effectiveness Stage

Variable	Non-Public-Sector Projects			Public Sector Projects			Difference in Means	Robust SE	Missing Observations
	N	Mean	SD	N	Mean	SD			
Committed Amount	2,105	82	96.6	1,097	41.3	64.4	-40.650***	(2.866)	0
Lending (~preparation) Costs Total [$ thousand]	1,957	362.7	220.4	1,022	321	219.8	-41.755***	(8.491)	223
Days from PCN Review to Approval	2,105	671.4	523.2	1,095	631.6	492.5	-39.743**	(18.747)	2
Days from Approval to Effectiveness	2,080	205.6	151.6	1,088	205.2	126.8	-0.369	(5.082)	34

Note: PCN = project concept note.

A2. Performance Predictors for Public Sector Management Projects

Table A.7 Probit Regression Estimates with Binary IEG Outcome Ratings: PSM Projects

		(1) IEG Outcome Rating [binary]	(2) IEG Outcome Rating [binary]	(3) IEG Outcome Rating [binary]	(4) IEG Outcome Rating [binary]	(5) IEG Outcome Rating [binary]
	Variable					
Country Context Covariates	DPI Share of Programmatic Parties	0.194***	0.199***	0.177**		0.132*
		(0.0676)	(0.0684)	(0.0699)		(0.0732)
	Nonfree Countries [dummy]	-0.0604	-0.0560	-0.0778		-0.169*
		(0.0734)	(0.0739)	(0.0796)		(0.0865)
	Free Countries [dummy]	0.0267	0.0230	-0.00857		-0.0224
		(0.0504)	(0.0498)	(0.0544)		(0.0537)
	ICRG Bureaucratic Quality Rating [0-6]	0.00780	0.00685	0.00864		0.0218
		(0.0298)	(0.0299)	(0.0291)		(0.0294)
	Net ODA Received (% of GNI)	0.00501**	0.00541***	0.00650***		0.00420
		(0.00195)	(0.00205)	(0.00219)		(0.00264)
	IBRD Loans and IDA Credits (DOD, % of GDP)	-0.135	-0.136	-0.0325		0.00236
		(0.226)	(0.221)	(0.266)		(0.236)
	External Debt Stocks, Public and Publicly Guaranteed (PPG) (DOD, % of GDP)	-0.0548	-0.0516	-0.0857		-0.0458
		(0.0647)	(0.0675)	(0.0764)		(0.0732)
	GDP per Capita, PPP (constant 2005 international $)	5.57e-07	6.32e-07	6.64e-07		2.94e-06
		(1.03e-05)	(1.04e-05)	(1.14e-05)		(1.18e-05)
	GDP per Capita Growth (annual %)	0.00687	0.00665	0.00615		0.00660
		(0.00447)	(0.00463)	(0.00504)		(0.00569)
	Inflation, Consumer Prices (annual %)	-0.000237	-0.000240	-0.000196		-0.000354
		(0.000189)	(0.000191)	(0.000158)		(0.000498)
	School Enrollment, Secondary (% gross)	0.000815	0.000649	-0.000179		-0.000981
		(0.00127)	(0.00122)	(0.00141)		(0.00129)
	Life Expectancy at Birth, Total (years)	-0.0120*	-0.0120*	-0.0101		-0.00518
		(0.00691)	(0.00706)	(0.00739)		(0.00690)
Reform Content Covariates	Administrative and Civil Service Reform		-0.00188	-0.00232*		-0.00238*
			(0.00129)	(0.00125)		(0.00125)
	Decentralization		0.000832	6.61e-05		0.000346
			(0.00192)	(0.00173)		(0.00181)
	Public Expenditure, Financial Management, and Procurement		-0.00187	-0.00205		-0.00190
			(0.00146)	(0.00147)		(0.00159)
	Tax Policy and Administration		0.000579	0.000822		0.00593
			(0.00270)	(0.00253)		(0.00388)
Project Management Covariates Observable at Effectiveness Stage	Committed Amount			1.77e-05	0.000168	6.97e-05
				(0.000170)	(0.000198)	(0.000186)
	Lending (~preparation) Costs Total ($ thousand)			7.25e-05	7.66e-06	5.74e-05
				(9.61e-05)	(7.63e-05)	(0.000103)
	Days from PCN Review to Approval			-4.88e-05	-3.40e-05	-5.26e-05
				(3.76e-05)	(3.17e-05)	(3.70e-05)
	Days from Approval to Effectiveness			-0.000227	-0.000276**	0.000241
				(0.000169)	(0.000126)	(0.000197)

Process and Risk Indicators Observable During the First Half of Project Implementation					
Country Environment Flag [dummy]					0.0335
					(0.0673)
Country Record Flag [dummy]					-0.155***
					(0.0599)
Project Management Flag [dummy]					-0.0647
					(0.0598)
Safeguards Flag [dummy]					-0.157
					(0.124)
Counterpart Funding Flag [dummy]					0.0131
					(0.0655)
Number of Risk Flags					-0.0128
					(0.0175)
Minimal ISR "DO" Rating					0.0364**
					(0.0164)
Minimal ISR "IP" Rating					0.0266
					(0.0192)
Projects at Risk					-0.0621**
					(0.0293)
Region Controls	Yes	Yes	Yes	Yes	Yes
Observations	674	674	619	1,012	579
McFadden's R^2	0.093	0.099	0.106	0.055	0.190
McKelvey and Zinova's R^2	0.200	0.213	0.223	0.113	0.379
Percent Correctly Predicted	72.55%	73.15%	71.57%	69.27%	77.03%
Proportional Reduction in Error	8.42%	10.40%	5.38%	0.64%	20.83%

t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: GDP = gross domestic product; PPP = purchasing power parity; ODA = official development assistance; DAC = Development Assistance Committee; IBRD = International Bank for Reconstruction and Development; IDA = International Development Association; DOD = debt outstanding and disbursed; PCN = project concept note; DO = development objectives; IP = implementation progress; ISR = implementation status and results report.

Table A.8 Ordered Probit Regression Estimates with Ordinal IEG Outcome Ratings, Employing Freedom House Measures: Marginal Effects Reported for Public Sector Projects

	Variable	Highly Unsatisfactory Rating	Unsatisfactory Rating	Moderately Unsatisfactory Rating	Moderately Satisfactory Rating	Satisfactory Rating	Highly Satisfactory Rating
		Marginal Probability of…					
Country Context Covariates	DPI Share of Programmatic Parties	-0.0221**	-0.0943***	-0.0562***	-0.0179**	0.162***	0.0286**
		(-2.22)	(-2.79)	(-2.80)	(-1.99)	(2.92)	(2.43)
	FH Freedom Status Rating: Not Free [dummy]	0.0142	0.0579*	0.0312*	0.00315	-0.0926*	-0.0139*
		(1.61)	(1.66)	(1.85)	(0.43)	(-1.82)	(-1.82)
	FH Freedom Status Rating: Free [dummy]	0.00128	0.00625	0.00401	0.00169	-0.0111	-0.00211
		(0.33)	(0.32)	(0.32)	(0.31)	(-0.32)	(-0.32)
	ICRG Bureaucratic Quality Rating [0–6]	-0.000504	-0.00215	-0.00128	-0.000408	0.00369	0.000650
		(-0.18)	(-0.18)	(-0.18)	(-0.18)	(0.18)	(0.18)
	Net ODA Received (% of GNI)	-0.000520*	-0.00222**	-0.00132**	-0.000421*	0.00380**	0.000671*
		(-1.89)	(-2.09)	(-2.10)	(-1.80)	(2.14)	(1.80)
	IBRD Loans and IDA Credits (DOD, % of GDP)	0.00702	0.0299	0.0178	0.00569	-0.0514	-0.00907
		(0.27)	(0.27)	(0.27)	(0.28)	(-0.27)	(-0.27)
	External Debt Stocks, Public and Publicly Guaranteed (PPG) (DOD, % of GDP)	0.00499	0.0213	0.0127	0.00405	-0.0365	-0.00645
		(0.68)	(0.67)	(0.69)	(0.65)	(-0.68)	(-0.69)
	GDP per Capita, PPP (constant 2005 international $)	-5.58e-08	-0.000000238	-0.000000142	-4.52e-08	0.000000408	7.20e-08
		(-0.05)	(-0.05)	(-0.05)	(-0.05)	(0.05)	(0.05)
	GDP per Capita Growth (annual %)	-0.000968**	-0.00413**	-0.00246**	-0.000785	0.00709**	0.00125**
		(-2.04)	(-2.17)	(-2.27)	(-1.54)	(2.24)	(2.22)
	Inflation, Consumer Prices (annual %)	0.00000343	0.0000146	0.00000871	0.00000278	-0.0000251	-0.00000443
		(1.03)	(0.93)	(0.95)	(0.82)	(-0.96)	(-0.92)
	School Enrollment, Secondary (% gross)	0.0000629	0.000268	0.000160	0.0000510	-0.000460	-0.0000812
		(0.51)	(0.54)	(0.52)	(0.51)	(-0.53)	(-0.50)
	Life Expectancy at Birth, Total (years)	0.000590	0.00251	0.00150	0.000478	-0.00432	-0.000762
		(0.96)	(0.85)	(0.92)	(0.82)	(-0.92)	(-1.01)
Reform Content Covariates	Administrative and Civil Service Reform	0.000181	0.000772	0.000460	0.000147	-0.00133	-0.000234
		(1.29)	(1.43)	(1.49)	(1.43)	(-1.48)	(-1.41)
	Decentralization	-0.000150	-0.000639	-0.000381	-0.000122	0.00110	0.000194
		(-0.81)	(-0.84)	(-0.81)	(-0.79)	(0.83)	(0.77)
	Public Expenditure, Financial Management, and Procurement	0.000172	0.000733	0.000437	0.000139	-0.00126	-0.000222
		(1.15)	(1.25)	(1.22)	(1.11)	(-1.23)	(-1.18)
	Tax Policy and Administration	0.000161	0.000686	0.000409	0.000130	-0.00118	-0.000208
		(0.57)	(0.60)	(0.61)	(0.57)	(-0.60)	(-0.58)
Other Covariate	Region Controls	Yes	Yes	Yes	Yes	Yes	Yes
	Project Management Controls at Effectiveness	No	No	No	No	No	No
	Process and Risk Indicators	No	No	No	No	No	No
	Project Management Controls ex post	No	No	No	No	No	No
	Observations	674	674	674	674	674	674
	McFadden's R^2	0.047					
	McKelvey and Zinova's R^2	0.141					
	Percent Correctly Predicted	41.54%					
	Proportional Reduction in Error	9.01%					

Marginal effects; *t* statistics in parentheses
(d) for discrete change of dummy variable from 0 to 1
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Estimates employ the same set of covariates as model 3 in annex table A.7. DPI = database of political institutions; GDP = gross domestic product; PPP = purchasing power parity; ODA = official development assistance; DAC = Development Assistance Committee; IBRD = International Bank for Reconstruction and Development; IDA = International Development Association; DOD = debt outstanding and disbursed.

Table A.9 Ordered Probit Regression Estimates with Ordinal IEG Outcome Ratings, Employing Polity IV Measures: Marginal Effects Reported for PSM Projects

	Variable	Highly Unsatisfactory Rating	Unsatisfactory Rating	Moderately Unsatisfactory Rating	Moderately Satisfactory Rating	Satisfactory Rating	Highly Satisfactory Rating
Country Context Covariates	DPI Share of Programmatic Parties	-0.0235**	-0.0995***	-0.0592***	-0.0183**	0.170***	0.0300***
		(-2.42)	(-3.15)	(-3.14)	(-2.07)	(3.32)	(2.76)
	Polity 2 Category= "autocracy" [dummy]	0.0169*	0.0663**	0.0380**	0.0113	-0.112**	-0.0208**
		(1.84)	(2.19)	(2.46)	(1.46)	(-2.46)	(-2.26)
	Polity 2 Category= "democracy" [dummy]	0.00991*	0.0423**	0.0262**	0.0115*	-0.0746**	-0.0154*
		(1.78)	(2.22)	(2.14)	(1.69)	(-2.18)	(-1.88)
	Other Country Context Controls	Yes	Yes	Yes	Yes	Yes	Yes
	Region Controls	Yes	Yes	Yes	Yes	Yes	Yes
	Project Content Controls	Yes	Yes	Yes	Yes	Yes	Yes
	Project Management Controls at Effectiveness	No	No	No	No	No	No
	Process and Risk Controls	No	No	No	No	No	No
	Observations	674	674	674	674	674	674
	McFadden's R^2	0.049					
	McKelvey and Zinova's R^2	0.147					
	Percent Correctly Predicted	40.65%					
	Proportional Reduction in Error	7.62%					

Marginal effects; *t* statistics in parentheses; (d) for discrete change of dummy variable from 0 to 1
* *p*< 0.10, ** *p*< 0.05, *** *p*< 0.01

Note: Estimates employ the same set of covariates as model 3 in annex table A.7. DPI = database of political institutions.

Table A.10 Ordered Probit Regression Estimates with Ordinal IEG Outcome Ratings for Project Management Variables Observable during the First Half of Project Implementation: Marginal Effects Reported for PSM Projects

	Variable	Highly Unsatisfactory Rating	Unsatisfactory Rating	Moderately Unsatisfactory Rating	Moderately Satisfactory Rating	Satisfactory Rating	Highly Satisfactory Rating
Project Management Covariates Observable at Effectiveness Stage	Committed Amount	-0.0000123	-0.0000646	-0.0000362	-0.0000146	0.000108	0.0000197
		(-0.61)	(-0.60)	(-0.61)	(-0.61)	(0.61)	(0.60)
	Lending (~preparation) Costs Total [$ thousand]	-0.00000358	-0.0000188	-0.0000106	-0.00000425	0.0000315	0.00000573
		(-0.45)	(-0.45)	(-0.45)	(-0.46)	(0.45)	(0.46)
	Days from PCN Review to Approval	0.00000394	0.0000207	0.0000116	0.00000468	-0.0000346	-0.00000630
		(1.10)	(1.19)	(1.16)	(1.12)	(-1.17)	(-1.20)
	Days from Approval to Effectiveness	-0.0000161	-0.0000846	-0.0000475	-0.0000191	0.000142	0.0000258
		(-1.11)	(-1.23)	(-1.21)	(-1.19)	(1.25)	(1.11)
Process and Risk Covariates Observable during the First Half of Project Implementation	Country Environment Flag [dummy]	-0.00515	-0.0296	-0.0181	-0.0107	0.0523	0.0112
		(-0.85)	(-0.82)	(-0.76)	(-0.57)	(0.78)	(0.65)
	Country Record Flag [dummy]	0.0131**	0.0625***	0.0317***	0.00482	-0.0981***	-0.0140***
		(2.08)	(2.92)	(3.07)	(0.97)	(-3.14)	(-2.70)
	Project Management Flag [dummy]	0.00397	0.0210	0.0116	0.00386	-0.0347	-0.00576
		(0.94)	(1.00)	(0.99)	(1.03)	(-0.99)	(-1.08)
	Safeguards Flag [dummy]	0.00823	0.0387	0.0194	0.00326	-0.0604	-0.00920
		(0.79)	(0.93)	(1.07)	(0.92)	(-1.00)	(-1.19)
	Counterpart Funds Flag [dummy]	-0.0000232	-0.000122	-0.0000685	-0.0000276	0.000204	0.0000372
		(-0.00)	(-0.00)	(-0.00)	(-0.00)	(0.00)	(0.00)
	Number of Risk Flags	0.000833	0.00438	0.00246	0.000989	-0.00733	-0.00133
		(0.56)	(0.57)	(0.57)	(0.54)	(-0.57)	(-0.52)
	Minimal ISR "DO" Rating	-0.00333*	-0.0175**	-0.00981**	-0.00395	0.0292**	0.00532**
		(-1.89)	(-2.25)	(-2.31)	(-1.60)	(2.28)	(1.99)
	Minimal ISR "IP" Rating	-0.00228	-0.0120	-0.00671	-0.00270	0.0200	0.00364*
		(-1.44)	(-1.56)	(-1.56)	(-1.57)	(1.58)	(1.78)
	Projects at Risk	0.00730**	0.0384***	0.0215***	0.00867**	-0.0642***	-0.0117**
		(2.02)	(3.10)	(3.25)	(2.04)	(-3.17)	(-2.18)
	Country Context Controls	Yes	Yes	Yes	Yes	Yes	Yes
	Reform Content Controls	Yes	Yes	Yes	Yes	Yes	Yes
	Region Controls	Yes	Yes	Yes	Yes	Yes	Yes
	Observations	579	579	579	579	579	579
	McFadden's R^2	0.095					
	McKelvey and Zinova's R^2	0.265					
	Percent correctly predicted	45.42%					
	Proportional reduction in error	15.05%					

Marginal effects; *t* statistics in parentheses; (d) for discrete change of dummy variable from 0 to 1.
* *p*< 0.10, ** *p*< 0.05, *** *p*< 0.01

Note: PCN = project concept note; DO = development objectives; IP = implementation progress; ISR = implementation status and results report.

A3. Matching Estimates: PSM Projects versus non-PSM Projects

Table A.11 Matching Estimates for the Share of Projects Rated MS+: PSM versus Non-PSM Projects

	PSM Projects (broadly defined)		PSM Projects (narrowly defined)	
	Nearest Neighbor Matching Estimates	Coarsened Exact Matching Estimates	Nearest Neighbor Matching Estimates	Coarsened Exact Matching Estimates
Average Treatment Effect on the Treated (ATET)	-0.0153	-0.017	-0.00905	-0.018
	(0.0199)	(0.018)	(0.0402)	(0.040)
Constant		0.706***		0.699***
		(0.011)		(0.016)
Observations	3,008	2,717	2,131	1,019
Percent Exact Country Matches	93.8%	100%	93.74%	100%

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: For this and the following tables, reported ATET estimates are based on nearest neighbor matching with exact matching by country (share of exact matches reported) and nearest neighbor matching by approval FY. For details on the matching algorithm, see Abadie and others (2004). Coarsened exact matching estimates are with strata defined by country and four-year brackets (1990 to 1993, 1994 to 1997, 1998 to 2001, 2002 following). For details on the matching algorithm, see Blackwell and others (2009).

Table A.12 Matching Estimates for the Share of Projects Rated MS+ in Nonfree, Partially Free, and Free Countries (based on Freedom House scores): PSM Projects versus Non-PSM Projects

	Nearest Neighbor Matching Estimates			Coarsened Exact Matching Estimates		
	(1)	(2)	(3)	(4)	(5)	(6)
	Nonfree Countries IEG Outcome Rating [binary]	Partially Free Countries IEG Outcome Rating [binary]	Free Countries IEG Outcome Rating [binary]	Nonfree Countries IEG Outcome Rating [binary]	Partially Free Countries IEG Outcome Rating [binary]	Free Countries IEG Outcome Rating [binary]
Average Treatment Effect on the Treated	-0.0838**	0.000811	0.00236	-0.071*	0.017	0.017
	(0.0414)	(0.0266)	(0.0433)	(0.039)	(0.025)	(0.040)
Constant				0.650***	0.710***	0.716***
				(0.021)	(0.016)	(0.025)
Observations	847	1,514	644	730	1,349	539
Percent Exact Country Matches	92.71%	93.16%	92.4%	100%	100%	100%

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: IEG = International Evaluation Group.

Table A.13 Matching Estimates for the Share of Projects rated MS+ in Autocratic, Anocratic, and Democratic Regimes (based on Polity Scores): PSM versus Non-PSM Projects

	Nearest Neighbor Matching Estimates			Coarsened Exact Matching Estimates		
	(1)	(2)	(3)	(4)	(5)	(6)
	Autocratic Regimes IEG Outcome Rating [binary]	Anocratic Regimes IEG Outcome Rating [binary]	Democratic Regimes IEG Outcome Rating [binary]	Autocratic Regimes IEG Outcome Rating [binary]	Anocratic Regimes IEG Outcome Rating [binary]	Democratic Regimes IEG Outcome Rating [binary]
Average Treatment Effect on the Treated	-0.0348	-0.0306	-0.00975	0.015	-0.044	0.008
	(0.0523)	(0.0325)	(0.0292)	(0.047)	(0.032)	(0.026)
Constant				0.596***	0.708***	0.747***
				(0.024)	(0.021)	(0.016)
Observations	668	1,012	1,283	576	872	1,168
Percent Exact Country Matches	89.66%	92.87%	95.41%	100%	100%	100%

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: IEG = International Evaluation Group.

Table A.14 Matching Estimates for the Share of Projects Rated MS+ in Countries with Low and High Shares of Programmatic Political Parties: PSM versus Non-PSM Projects

	Nearest Neighbor Matching Estimates		Coarsened Exact Matching Estimates	
	(1) Share of Programmatic Political Parties <=0.5	(2) Share of Programmatic Political Parties >0.5	(3) Share of Programmatic Political Parties <=0.5	(4) Share of Programmatic Political Parties >0.5
Average Treatment Effect on the Treated	-0.0837***	0.0283	-0.070**	0.028
	(0.0304)	(0.0266)	(0.029)	(0.024)
Constant			0.683***	0.724***
			(0.018)	(0.014)
Observations	1,261	1,744	1,093	1,524
Percent Exact Country Matches	92.40%	88.36%	100%	100%

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table A.15 Matching Estimates for the Share of Projects Rated MS+ in Countries with Low and High Shares of Programmatic Political Parties: PSM Projects versus Non-PSM Projects

	Nearest Neighbor Matching Estimates		Coarsened Exact Matching Estimates	
	(1) Share of Programmatic Political Parties <=0.5	(2) Share of Programmatic Political Parties >0.5	(3) Share of Programmatic Political Parties <=0.5	(4) Share of Programmatic Political Parties >0.5
Average Treatment Effect on the Treated	-0.120*	0.0813	-0.135**	0.053
	(0.0644)	(0.0506)	(0.065)	(0.051)
Constant			0.688***	0.707***
			(0.027)	(0.020)
Observations	876	1,255	359	626
Percent Exact Country Matches	89.00%	88.02%	100%	100%

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table A.16 Matching Estimates for the Share of Projects Rated MS+ in Countries by Aid Dependency (ODA as a Share of GNI): PSM versus Non-PSM Projects

	Nearest Neighbor Matching Estimates			Coarsened Exact Matching Estimates		
	(1) ODA as a Share of GNI <5% IEG Outcome Rating [binary]	(2) ODA as a Share of GNI >=5% &< 20% IEG Outcome Rating [binary]	(3) ODA as a Share of GNI >=20% IEG Outcome Rating [binary]	(4) ODA as a Share of GNI <5% IEG Outcome Rating [binary]	(5) ODA as a Share of GNI >=5% and < 20% IEG Outcome Rating [binary]	(6) ODA as a Share of GNI >=20% IEG Outcome Rating [binary]
Average Treatment Effect on the Treated	-0.00632	-0.0762**	-0.00353	0.000	-0.084**	-0.022
	(0.0284)	(0.0349)	(0.0639)	(0.025)	(0.033)	(0.067)
Constant				0.722***	0.694***	0.707***
				(0.014)	(0.021)	(0.049)
Observations	1,614	923	251	1,470	831	193
Percent Exact Country Matches	95.69%	92.03%	74.57%	100%	100%	100%

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: ODA = official development assistance; GNI = gross national income; IEG = International Evaluation Group.

Table A.17 Matching Estimates for the Share of Projects Rated MS+ in Countries by Aid Dependency (ODA as a Share of GNI): PSM versus Non-PSM Projects

|	Nearest Neighbor Matching Estimates			Coarsened Exact Matching Estimates		
	(1)	(2)	(3)	(4)	(5)	(6)
	ODA as a Share of GNI <5% IEG Outcome Rating [binary]	ODA as a Share of GNI >=5% &< 20% IEG Outcome Rating [binary]	ODA as a Share of GNI >=20% IEG Outcome Rating [binary]	ODA as a Share of GNI <5% IEG Outcome Rating [binary]	ODA as a Share of GNI >=5% and < 20% IEG Outcome Rating [binary]	ODA as a Share of GNI >=20% IEG Outcome Rating [binary]
Average Treatment Effect on the Treated	-0.00160	-0.146*	-0.0149	0.032	-0.199***	-0.006
	(0.0616)	(0.0743)	(0.105)	(0.061)	(0.072)	(0.095)
Constant				0.663***	0.740***	0.792***
				(0.021)	(0.031)	(0.054)
Observations	1,207	615	153	552	254	87
Percent Exact Country Matches	93.68%	97.03%	81.73%	100%	100%	100%

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: ODA = official development assistance; GNI = gross national income; IEG = International Evaluation Group.

A4. Marginal Effect Estimates of Country Context Variables on PSM and Non-PSM Projects

Figure A.2 Marginal Probability of Receiving Ordinal IEG Outcome Ratings in Nonfree, Partially Free, and Free Countries (based on Freedom House scores): PSM versus non-PSM Projects
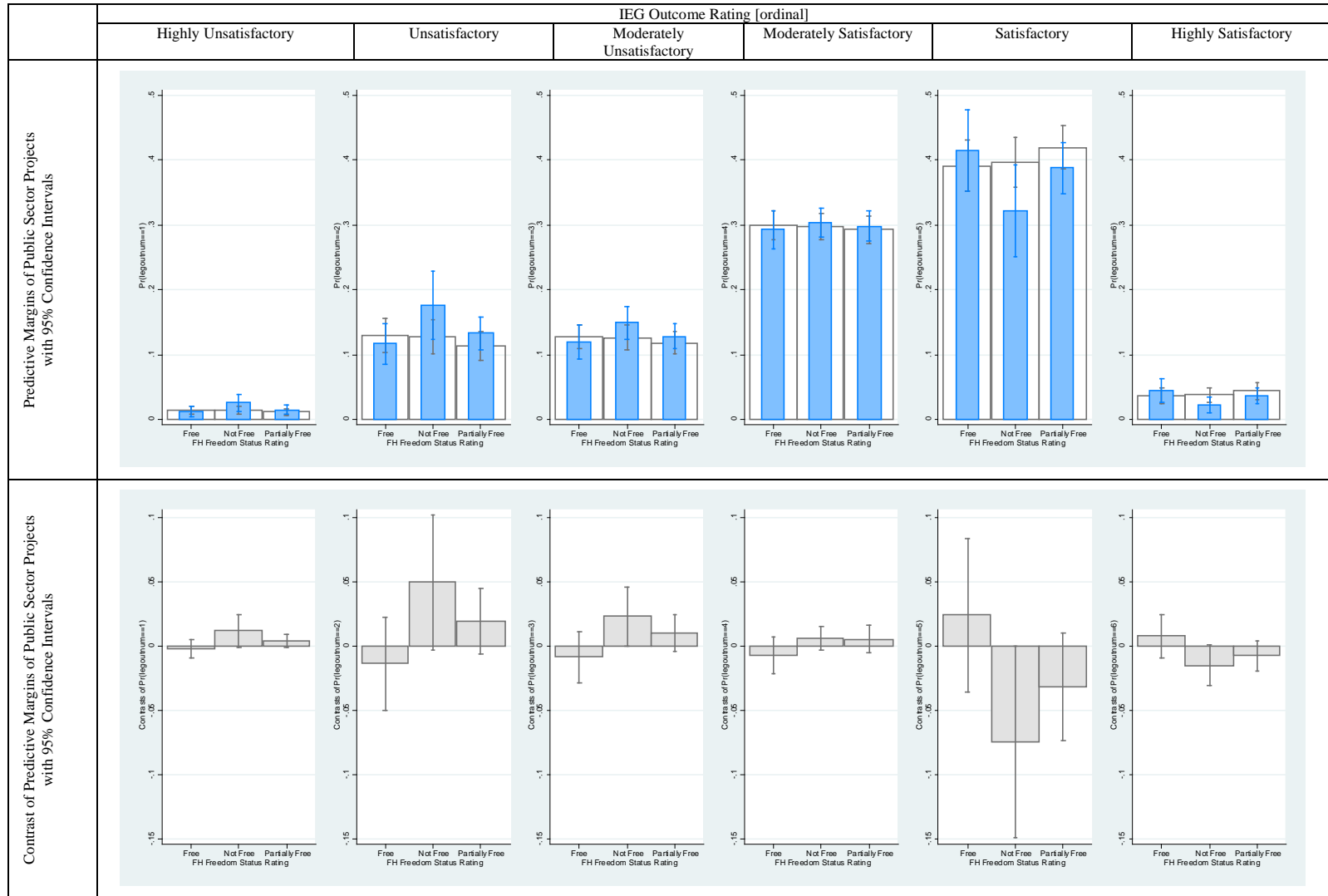
Note: Based on ordered probit estimations, this and the following figures contrast the marginal probability of receiving a particular IEG outcome rating (see column headings) for PSM projects and non-PSM projects, depending on a particular context variable (such as the Freedom House status in this figure). Note that for each context category, the marginal probabilities across all IEG outcome ratings sum to 100 percent. In this estimation, other country context and project management covariates observable at baseline are held constant at mean. The first row in each figure reports the marginal probability of receiving a particular IEG outcome rating for PSM projects in blue and for non-PSM projects in gray. The second row reports estimates of the difference between the marginal probabilities for PSM and non-PSM projects, corresponding to the height difference of the bars reported in the first row.

Figure A.3 Marginal Probability of Receiving Ordinal IEG Outcome Ratings in Autocratic, Anocratic, and Democratic Regimes (based on Polity Scores): PSM versus Non-PSM Projects
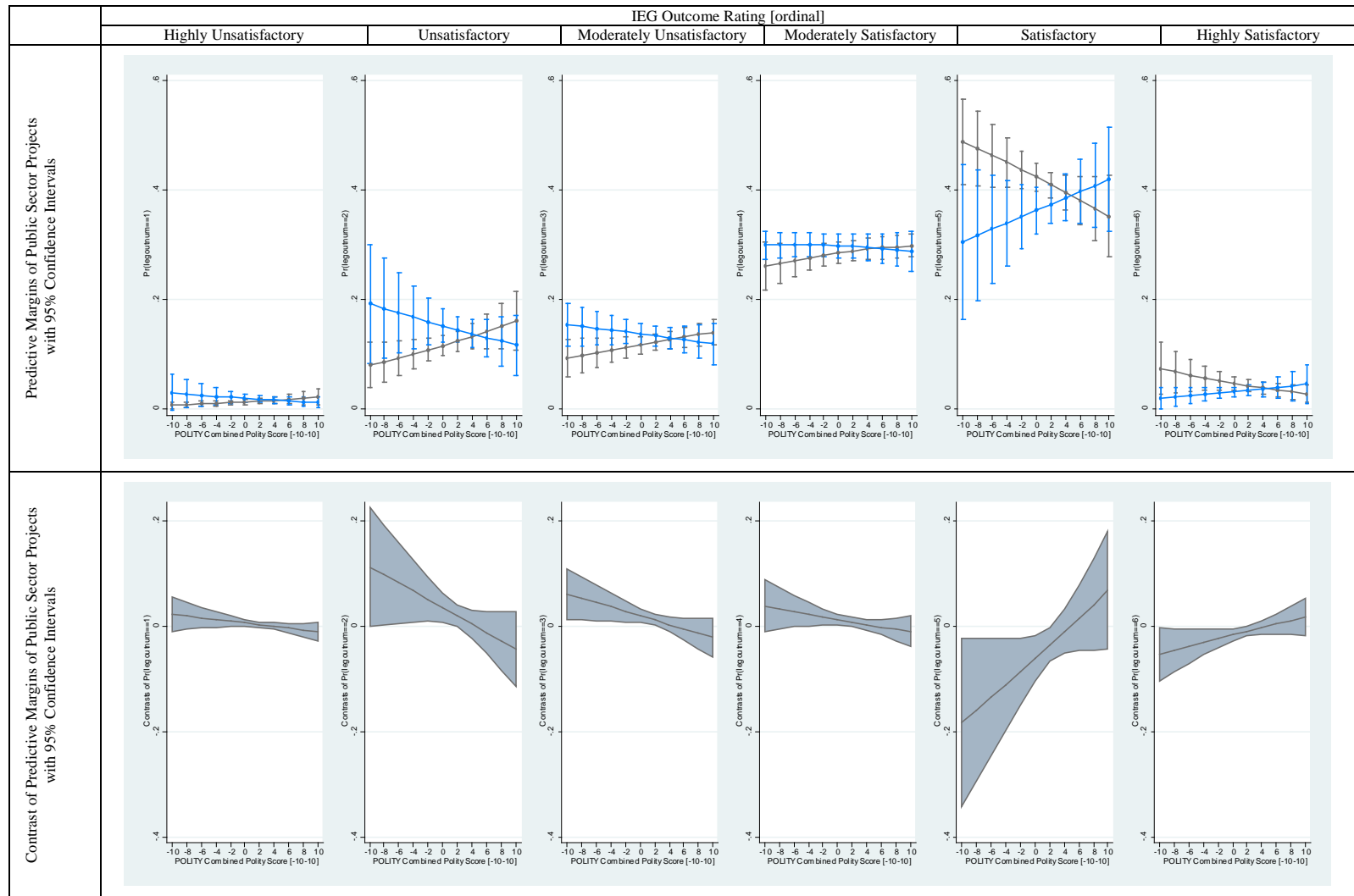
Figure A.4 Marginal Probability of Receiving Ordinal IEG Outcome Ratings in Countries, by Share of Programmatic Political Parties: PSM versus Non-PSM Projects