

## **Governance Indicators and the Level of Analysis Problem: Empirical Findings from South America**

DANIEL W. GINGERICH\*

Studies of the link between state capacity and development often utilize national-level governance indicators to explain fine-grained development outcomes. As capacity in some bureaucratic agencies matters more for these outcomes than capacity in others, this work proxies for capacity within the set of relevant agencies by using a measure of ‘mean’ capacity across all agencies in a polity. This practice is problematic for two reasons: (1) within-country, cross-agency diversity in capacity often overwhelms the variation encountered across public sectors considered in their entirety; (2) national-level reputations for capacity are not particularly informative about differences in capacity in functionally equivalent agencies in different countries. The article draws on the author’s survey of public employees in Bolivia, Brazil and Chile to establish these points.

Social scientists have long sought to understand the relationship between bureaucratic capacity and the prospects for development. Since about the year 2000, this research agenda has experienced exponential growth, a state of affairs largely attributable to the decision of the World Bank (in 1996) to begin explicitly considering the role of corruption in the success or failure of its lending programmes and the concomitant development of cross-national, perception-based indicators of corruption and bureaucratic professionalism (henceforth referred to as perception-based governance indicators, or PB-GIs). The dissemination of the PB-GIs immediately resulted in a wave of studies examining the ramifications of national-level bureaucratic capacity for aggregate economic growth, a wave which has since given way to a second generation of studies examining the impact of national-level bureaucratic capacity on a host of much more variegated development outcomes.<sup>1</sup>

\* Woodrow Wilson Department of Politics, University of Virginia, Charlottesville (email: [dwg4c@virginia.edu](mailto:dwg4c@virginia.edu)). The author would like to thank Jorge I. Dominguez, Peter Hall and Steven Levitsky for advice and support in carrying out this project. The helpful comments of John Echeverri-Gent, Eddy Malesky, David Waldner and three anonymous referees are greatly appreciated. In Bolivia, thanks go to Andrés Zaratti, Giovanna Mendoza, Walter Guevara Anaya, George Gray Molina, Juan Antonio Morales and René Antonio Mayorga; in Brazil, to David Fleischer, Paulo Calmon, Yves Zamboni Filho, Anali Cristino Figueirido and Gustavo Freitas Amora; in Chile, to Claudio Fuentes, Gonzalo Biggs, Alejandro Ferreiro, Steven Reifenberg and the late Luciano Tomassini. The data utilized in this study are available from the author upon request. Additional material is to be found in an appendix online at <http://dx.doi.org/10.1017/S0007123412000403>

<sup>1</sup> Paulo Mauro, ‘Corruption and Growth’, *Quarterly Journal of Economics*, 110 (1995), 681–712; Paulo Mauro, ‘The Effects of Corruption on Growth, Investment and Government Expenditure: A Cross-Country Analysis’, in Kimberly A Elliott, ed. *Corruption in the Global Economy* (Washington: Institute for International Economics, 1997), pp. 83–108; Daniel Kaufmann, Art Kraay, and Pablo Zoido-Lobaton, ‘Governance Matters’, World Bank Policy Research Working Paper No. 2196 (1999); Peter Evans and James E Rauch, ‘Bureaucracy and Growth: A Cross-National Analysis of the Effects of “Weberian” State Structures on Economic Growth’, *American Sociological Review* 64 (1999), 748–65.

This article argues that the use of the national-level PB-GIs is not well suited to the aims of this second generation of scholarship. To the degree that bureaucratic capacity matters for the development outcomes in question, it is capacity within a relatively small subset of the constituent units of the bureaucratic apparatus that tends to drive the outcomes of interest. Moreover, capacity as measured by a national-level summary statistic for the entire public sector is an unreliable proxy for capacity within the set of these relevant agencies. The reason that the national-level indicators of capacity, which ostensibly measure the central tendency within the bureaucratic apparatus, do not proxy well for capacity in particular state agencies is twofold: (1) within-country bureaucratic diversity often overwhelms variation encountered across public sectors considered in their entirety; (2) national-level reputations for bureaucratic capacity are not particularly informative about differences in capacity in functionally equivalent agencies in different countries.

The study evaluates the validity of these claims by drawing upon data collected by the author as part of a large-scale survey of public employees conducted in Bolivia, Brazil and Chile from June 2003 to November 2005. The choice of countries included in the study presents a particularly hard test for these hypotheses. This is because the three countries have highly distinctive and well-established national reputations on phenomena such as patronage and corruption, with Bolivia being characterized in the literature as an ideal-typical patrimonial state, Chile characterized as nearly an ideal-typical Weberian state, and Brazil characterized as occupying a middling position between these two extremes. If national-level differences in capacity ever dominate intra-state differences in capacity or are generally informative about the relative performance of specific types of agencies, one would expect this to hold true for this set of cases.

This article's empirical analysis unfolds in four steps. First, the article draws upon the public employees survey data to examine three key components of bureaucratic capacity – the level of politicization of personnel management and routine decision making, the stability of bureaucratic career paths, and formal and informal mechanisms of corruption control – showing that the evidence generally points to greater agency-level differences than country-level differences. Secondly, the article shows that the relative position of any two countries' national-level reputations on bureaucratic capacity does not, in general, correspond to relative differences in employee perceptions about bureaucratic capacity in comparable agencies found in the two countries. More specifically, the data show that if one compares 'sister' agencies (i.e. agencies with similar administrative functions) across two countries, the agency belonging to the country with the superior national-level reputation for bureaucratic capacity is quite often *not* the agency whose employees are most inclined to perceive their work environment as having Weberian features. Thirdly, as a check on the quality of employee perceptions, the article shows that employees' perceptions of bureaucratic capacity correspond to patterns of actual behaviour within bureaucratic agencies that are consistent with the perceptions. In particular, the article uses the survey data from Brazil to show that the relationship between having obtained one's post via a public examination (*concurso público*) and having authority in one's job over the allocation of public contracts, licences and loans varies systematically across agencies in a manner aligned with employee perceptions of bureaucratic capacity. Finally, the article utilizes cross-national data on the onset of banking crises during the 1990s to illustrate the pitfalls associated with using measured associations between a given development outcome and national-level indicators of bureaucratic capacity as a basis for inferring the impact of bureaucratic capacity within the substantively most relevant state agencies on that outcome.

## GOVERNANCE INDICATORS AND THE LEVEL OF ANALYSIS PROBLEM

In providing an empirically informed methodological critique of the manner in which PB-GIs are used in an important strand of development research, this article contributes to a growing literature which calls into question the inordinate reliance on those indicators in contemporary investigations of the economic and social consequences of state capacity. Existing critiques of the indicators have focused primarily on either their accuracy or conceptual merit.

With respect to the former, a number of studies have argued that perceptions of capacity and actual levels of capacity may differ substantially, especially to the degree that capacity is (inversely) manifested in corruption. In this respect, various works have noted that perceptions of corruption have only a tenuous relationship with the personal experience of corruption, that citizens do not do a very good job of detecting corruption even in projects taking place in their immediate vicinity, that expert opinions about the frequency of citizens' experiences with corruption do not match up with the actual frequencies, and that the determinants of citizens' perceptions of corruption are generally different from the determinants of experienced corruption.<sup>2</sup> Other authors have called into question the accuracy and usefulness of the PB-GIs by suggesting that they exhibit 'halo effects', i.e. that they are driven in part by knowledge of past successful development outcomes such as growth.<sup>3</sup> With respect to conceptual issues, several authors have suggested that some PB-GIs tend to conflate evaluations of particular policies with evaluations of capacity, leading to a distinct (pro-free market) ideological bias in the indicators.<sup>4</sup> On this same point, others have claimed that certain existing PB-GIs suffer fundamental definitional confusion and fail to tap into distinct aspects of state capacity.<sup>5</sup>

The nature of my critique is quite distinct from those listed above. I do not wish to make the case that there is a fundamental flaw intrinsic to the indicators themselves. Rather, I seek to draw attention to an important level of analysis problem evident in the manner in which they have been employed in the literature. With increasing frequency, students of development utilize the PB-GIs to study the impact of state capacity on

<sup>2</sup> Claudio Weber Abramo, 'How Much Do Perceptions of Corruption Really Tell Us?' *Economics: The Open-Access, Open-Assessment E-Journal*, 2 (2008), 1–56; Richard Rose and William Mishler, 'Experience versus Perception of Corruption: Russia as a Test Case', *Global Crime*, 11 (2010), 145–63; Benjamin A. Olken, 'Corruption Perceptions vs. Corruption Reality', *Journal of Public Economics*, 93 (2009), 950–64; Mireille Razafindrakoto and François Roubaud, 'Are International Databases on Corruption Reliable? A Comparison of Expert Opinion Surveys and Household Surveys in Sub-Saharan Africa', *World Development*, 38 (2010), 1057–69; Daniel Treisman, 'What Have We Learned about the Causes of Corruption from Ten Years of Cross-National Empirical Research?' *Annual Review of Political Science*, 10 (2007), 211–44.

<sup>3</sup> Edward Glaeser, Rafael La Porta, Florencio Lopez-de-Silanes and Andrei Shleifer, 'Do Institutions Cause Growth?' *Journal of Economic Growth*, 9 (2004), 271–303.

<sup>4</sup> Christiane Arndt and Charles Oman, *Uses and Abuses of Governance Indicators* (Paris: OECD Development Centre, 2006); Marcus J. Kurtz and Andrew Shrank, 'Growth and Governance: Models, Measures, and Mechanisms', *Journal of Politics*, 69 (2007), 538–54.

<sup>5</sup> Melissa Thomas, 'What Do the Worldwide Governance Indicators Measure?' *European Journal of Development Research*, 22 (2010), 31–54; Laura Langbein and Stephen Knack, 'The Worldwide Governance Indicators: Six, One, or None?' *Journal of Development Studies*, 46 (2010), 350–70. For a response to these criticisms, see Daniel Kaufmann, Art Kraay and Massimo Mastruzzi, 'The Worldwide Governance Indicators Project: Answering the Critics', World Bank Policy Research Working Paper No. 4149 (2007).

outcomes which are primarily shaped by specific subsets of institutions within the public sector and only secondarily, if at all, by others. To give a sense of the frequency of this practice and its scholarly influence, Table 1 lists nineteen studies that draw upon the PB-GIs in this way along with the citation count for each study.<sup>6</sup> These studies combine for a total of 4,208 citations, a clear testament to the important role they have played in shaping scholarly debate on the relationship between state capacity and development.

The outcomes examined in these works include the presence of banking crises, the inflation rate, indicators of pollution and environmental protection, indicators of citizen health, indicators of citizen education, measures of infrastructure quality, and the size of the unofficial economy. For all these outcomes, it is easy to identify the specific institutions for which capacity should be most important: bank regulators for banking crises, environmental protection agencies for pollution, ministries of health and public hospitals for health outcomes, tax services for the size of the unofficial economy, etc. Yet rather than measuring capacity in the most relevant institutions, these works tap into bureaucratic capacity using only the national-level PB-GIs. Supposing these measures adequately capture the overall central tendency of bureaucratic capacity in a polity, this data collection strategy is equivalent to using the mean level of bureaucratic capacity across agencies in the public sector to proxy for capacity within the relevant agencies.

Is this a legitimate way to draw inferences about whether or not bureaucratic capacity is responsible for variation in the outcomes in question? A brief foray into linear regression theory helps shed light on this question. Suppose that the true relationship between a particular development outcome and bureaucratic capacity is captured by the linear function  $Y_i = \alpha + \beta X_{j,i} + \varepsilon_i$ , where  $i$  indexes countries,  $X_{j,i}$  denotes capacity within the relevant institution(s)  $j$  in country  $i$ , and  $\mathbb{E}[\varepsilon_i] = 0$ . Now imagine that the analyst eschews measuring  $X_{j,i}$  and instead uses a measure of average state capacity (across institutions) equal to:

$$A_i = (1/K)(X_{j,i} + \sum_{k \neq j}^{K-1} X_{k,i}),$$

where  $X_{k,i}$  is capacity in institution  $k \neq j$  contained within country  $i$ , and  $K$  is the total number of institutions. The analyst then performs a linear regression of  $Y_i$  onto  $A_i$ , with the resulting coefficient on  $A_i$  equal to  $\tilde{\beta}$ . Treating  $X_{j,i}$  and  $A_i$  as fixed, the expression for the bias of  $\tilde{\beta}$  in this setting is  $\text{Bias}(\tilde{\beta}) = \mathbb{E}[\tilde{\beta}] - \beta = \beta(\psi - 1)$ , where  $\psi$  is the coefficient on  $X_{j,i}$  in a regression of  $A_i$  onto  $X_{j,i}$ .

This simple linear theory derivation tells us that bias will be small in absolute value (implying that the use of  $A_i$  in place of  $X_i$  is more or less legitimate) under one of two scenarios. The first, rather obvious, scenario would be that each of the  $X_{k,i}$  are close in value to  $X_{j,i}$  for all  $i$ , meaning that the central tendency of bureaucratic capacity is a good guide to capacity within the constituent units of the state. In essence, this scenario requires that interstate variation in capacity swamp intrastate variation in capacity. The second scenario is that the  $X_{k,i}$  differ from  $X_{j,i}$ , perhaps substantially, but that there is a near perfect one-to-one correlation between  $A_i$  and  $X_{j,i}$  across countries. Substantively,

<sup>6</sup> The PB-GIs utilized include the World Bank Institute's Worldwide Governance Indicators (WGI), Transparency International's Corruption Perceptions Index (TI-CPI), and indices generated by Political Risk Services Group's *International Country Risk Guide*, the World Economic Forum's Global Competitiveness Survey, and Business International.

TABLE 1 *Partial List of Studies Using Perception-based Governance Indicators to Examine the Relationship between Bureaucratic Capacity and Narrowly Defined Development Outcomes\**

Study	Development outcome(s) analysed	Governance indicator(s) used	Findings	Times cited*
Asli Demirgüç-Kunt and Enrica Detragiache, 'The Determinants of Banking Crises in Developing and Developed Countries', <i>IMF Staff Papers</i> , 45 (1998), 81–109.	Banking crises	PRS-ICRG-law & order	Greater administrative capacity (as measured by law & order index) is associated w/ a smaller probability of a banking crisis	733
Simon Johnson, Daniel Kaufmann and Pablo Zoido-Lobaton, 'Regulatory Discretion and the Unofficial Economy', <i>American Economic Review</i> , 88 (1998), 387–92.	Unofficial economy as % GDP	TI-CPI, WEF-bribery	Greater corruption is associated with a larger unofficial economy	455
Daniel Kaufmann, Art Kraay and Pablo Zoido-Lobaton, 'Governance Matters', World Bank Policy Research Working Paper No. 2196 (1999).	Infant mortality, adult literacy rate	WGIs (all)	Each WGI found to be strongly related to mortality & literacy	1232
Fahim A. Al-Marhubi, 'Corruption and Inflation', <i>Economics Letters</i> , 66 (2000), 199–202.	Inflation rate	TI-CPI, BI-corruption, BI-bureaucratic efficiency	Greater corruption & less bureaucratic efficiency are associated with higher inflation	78
Eric Friedman, Simon Johnson, Daniel Kaufmann, and Pablo Zoido-Lobaton, 'Dodging the Grabbing Hand: The Determinants of Unofficial Activity in 69 Countries', <i>Journal of Public Economics</i> 76 (2000), 459–93.	Unofficial economy as % GDP, tax revenues as % of GDP	PRS-ICRG corruption index, WEF-bribery, TI-CPI	Greater corruption is associated with a larger unofficial economy (except in transition countries) and smaller tax revenue	552

TABLE 1 (Continued)

Study	Development outcome(s) analysed	Governance indicator(s) used	Findings	Times cited*
Gil Mehrez and Daniel Kaufmann, 'Transparency, Liberalization and Banking Crises', World Bank Policy Research Working Paper No. 2286 (2000).	Banking crises	PRS-ICRG corruption index	No statistically significant relationship between corruption and the probability of a banking crisis	30
Sanjeev Gupta, Hamid Davoodi, and Erwin Tiongson, 'Corruption and the Provision of Health Care and Education Services', IMF Working Paper No. 00/116 (2001).	Child mortality, infant mortality, low birthweight, primary school dropout rate	PRS-ICRG corruption index, WBI-graft (CCI)	In various specifications, corruption was found to have a negative impact on outcomes in health & education	136
George Abed and Hamid Davoodi, 'Corruption, Structural Reforms, and Economic Performance in Transition Economies', in George Abed and Sanjeev Gupta, eds, <i>Governance, Corruption, and Economic Performance</i> (Washington, D.C.: IMF, 2002), pp. 489–537.	Inflation rate	TI-CPI, BI-corruption, PRS-ICRG	Differences in corruption levels are not found to be consistently related to differences in inflation	110
Vito Tanzi and Hamid Davoodi, 'Corruption, Public Investment, and Growth', in George Abed and Sanjeev Gupta, eds., <i>Governance, Corruption, and Economic Performance</i> (Washington, D.C.: IMF, 2002), pp. 280–99.	Paved roads in good condition, power outages, telecommunications faults, water losses, % of railway diesels in use	BI, PRS-ICRG	Corruption reduces paved roads in good condition and increases power outages; no significant relationship w/ other outcomes	622

TABLE 1 (Continued)

Richard Damania, Per G. Fredriksson, and Muthukumara Mani, 'The Persistence of Corruption and Regulatory Compliance Failures: Theory and Evidence', <i>Public Choice</i> , 121 (2004), 363–90.	Perceived compliance w/ international environmental agreements (WEF)	TI-CPI, WBI-CCI	Corruption found to undermine compliance with environmental agreements	46
Heinz Welsch, 'Corruption, Growth, and the Environment: A Cross-Country Analysis', <i>Environment and Development Economics</i> , 9 (2004), 663–93.	Various indicators of air & water pollution	WGI-graft (CCI)	Corruption found to have a substantial direct and indirect (via income) on pollution	57
Axel Dreher and Thomas Herzfeld, 'The Economic Costs of Corruption: A Survey and New Evidence', Working Paper, University of Konstanz (2005).	Enrolment in secondary school, life expectancy, inflation	PRS-ICRG corruption index	Corruption reduces school enrolment in one specification but not the other; corruption reduces life expectancy; corruption reduces inflation in one specification but not the other	25
Daniel C Esty and Michael E Porter, 'National Environmental Performance: An Empirical Analysis of Policy Results and Determinants', <i>Environment and Development Economics</i> , 10 (2005), 391–434.	Urban particulates, urban SO <sub>2</sub> concentration	WEF (public sector competence, irregular payments, administrative corruption)	Public sector competence not significant predictor of particulates or SO <sub>2</sub> ; corruption measures significant predictors of both particulates & SO <sub>2</sub>	25
Maureen Lewis, 'Governance and Corruption in Public Health Care Systems', Center for Global Development Working Paper No. 78 (2006).	Measles immunization coverage, under-5 mortality rate	WGIs (GE, CCI, VA)	GE increases immunization rates; exerts no impact on mortality rate	57

TABLE 1 (Continued)

Study	Development outcome(s) analysed	Governance indicator(s) used	Findings	Times cited*
Matthew A. Cole, 'Corruption, Income and the Environment: An Empirical Analysis', <i>Environmental Economics</i> , 62 (2007), 637–47.	CO <sub>2</sub> and SO <sub>2</sub> emissions	PRS-ICRG corruption index	Corruption found to have a substantial direct and indirect effect (via income) on pollution	11
J. Eslava-Schmalbach, H. Alonso, H. Oliveros, H. Gaitan and C. Agudelo, 'A New Inequity-in-Health Index based on Millenium Development Goals: Methodology and Validation', <i>Journal of Clinical Epidemiology</i> , 61 (2008), 142–50.	Inequity-in-Health Index (IHI): underweight children, child mortality, child death from malaria, all deaths from malaria, births attended by skilled health personnel, measles immunization	TI-CPI	Study reports strong correlation between corruption & IHI; no causal interpretation is given	9
Sören Holmberg, Bo Rothstein, and Naghmeh Nasiritousi, 'Quality of Government: What You Get', <i>Annual Review of Political Science</i> , 12 (2009), 135–61.	Subjective health, life expectancy, infant mortality, child mortality, HIV prevalence, environmental sustainability, air & water quality, access to drinking water, CO <sub>2</sub> emissions, deforestation	WBI-GE	Using univariate regressions, government effectiveness is found to be associated with improvements in listed outcomes in health & the environment	9



TABLE 1 (Continued)

Cuneyt Koyuncu and Rasim Yilmaz, 'The Impact of Corruption on Deforestation', <i>Journal of Developing Areas</i> , 42 (2009), 213–22.	Rate of deforestation	TI-CPI, PRS-ICRG, BI	Positive – albeit statistically weak – relationship found between corruption & deforestation	21
Mogens K. Justesen, 'Democracy, Dictatorship, and Disease: Political Regimes and the Fight against HIV/AIDS', Working Paper, University of Southern Denmark (2010).	Proportion of HIV patients treated w/ anti-retroviral therapy	WBI-CCI	Corruption has an inconsistent and weak relationship to anti-retroviral therapy for HIV patients	0

\*According to Google Scholar as of 14 May 2010.

this scenario permits capacity to differ across agencies contained within a given country but requires that the relative placement of agencies be uniform across countries.

In the pages that follow, I use the public employees survey data from Bolivia, Brazil and Chile to underline the implausibility of the two requirements elaborated above: greater diversity in capacity across states than within them and uniformity in the relative placement of (functionally similar) agencies. To anticipate the study's findings, even in a set of cases which are highly likely to encounter evidence in favour of the existence of these conditions, the data point unambiguously in the opposite direction. Thus, it seems highly unlikely that these conditions would hold more generally, implying, in turn, that inferences built around an examination of the link between sector-specific development outcomes and national-level PB-GIs are on fairly shaky ground. Moreover, this article's examination of the onset of banking crises provides concrete evidence of the potentially nefarious consequences of using the aforementioned strategy. In this regard, I have found that the association between crisis and a national-level bureaucratic capacity indicator is actually signed opposite to the association between crisis and an indicator of capacity within bank regulators.

#### REPUTATIONS FOR BUREAUCRATIC CAPACITY IN BOLIVIA, BRAZIL AND CHILE

Why are Bolivia, Brazil and Chile appropriate cases to study in order to gauge the importance of cross-agency variation in bureaucratic capacity? The stylized facts which emerge from the standard references on bureaucratic performance in these countries point to one important reason: the cases are depicted in the literature as inhabiting sharply distinct positions on a continuum of bureaucratic capacity, with the Bolivian public administration purportedly approaching the neo-patrimonial ideal type, the Chilean public administration purportedly approaching the Weberian ideal type, and the Brazilian public administration typically depicted as occupying a middling position between the two poles. To the extent that we find that intra-state, agency-level variation overwhelms interstate variation in this set of cases, then we will have a strong basis for making the case that analysis of variation in capacity at the agency level deserves a greater place at the academic table than it has heretofore received. Let us consider the standard descriptions of the bureaucracy in the cases.

In order to describe the basic contours of the Bolivian state, observers from the World Bank have made extensive use of the term 'informality'.<sup>7</sup> In this regard, the Bolivian public sector is purportedly characterized by the primacy of informal norms and practices over formal rules and standard operating procedures. Entry into the public administration and advancement is supposedly based largely on partisan affiliation, not merit, and the assiduous cultivation of political contacts is characterized as the principal means of rising up the ranks. In this context, the public sector has come to be characterized by high rates of turnover, low morale and low institutional memory.<sup>8</sup> Most accounts of the Bolivian public administration emphasize that clientelistic relations within the state are hegemonic. As one high-level Bolivian functionary put it to the author in 2003: 'take 100, subtract the percentage of employees who say their institution belongs to

<sup>7</sup> World Bank, *Bolivia: From Patronage to a Professional State*, 2 vols., Report No. 20115-BO (2000).

<sup>8</sup> Alberto Leyton, 'Administrative Reform Program: The Case of Bolivia', in Shahid A. Chaudhry, Gary J. Reid and Waleed H. Malik, eds, *Civil Service Reform in Latin America and the Caribbean: Proceedings of a Conference* (Washington, D.C.: World Bank, 1994), pp. 158–65.

the *cuoteo político* [the spoils system then orchestrated by Bolivia's traditional parties] and there you have the percentage that are lying.<sup>9</sup> In short, the stylized depiction of the Bolivian bureaucracy is one of a homogeneous, inert and inefficient apparatus, long treated as a teat from which patronage-minded parties have suckled.

Academic accounts of the nature and evolution of the Brazilian public administration share certain commonalities with the Bolivian case as well as important differences. The logic of patrimonialism is acknowledged to have installed itself in wide swathes of the Brazilian state and to have had a historically important role in establishing bureaucratic careers.<sup>10</sup> Not surprisingly, this feature of the Brazilian state has contributed to recurrent bouts of corruption.<sup>11</sup> However, the Brazilian bureaucracy also exhibits a number of features that would tend to set it apart from the highly patrimonial reputation of the Bolivian bureaucracy. The country exhibits a competitive public examination system for entry into many low-level and mid-level posts and it boasts the existence of a national school of public administration (ENAP). Moreover, in contrast to the Bolivian case, where a high level of informality in the execution of bureaucratic procedures is often pointed to as a symptom of low bureaucratic capacity, students of Brazilian public administration have long argued that, if anything, the country suffers from the opposite type of ailment: excessive and unwieldy legal formalism in personnel management.<sup>12</sup> Given this apparent coexistence of rigid legalism and clientelistic practices within the same state apparatus, the Brazilian public administration is typically described as one which falls roughly in the midpoint between the extremes of the patrimonial and Weberian ideal types.<sup>13</sup>

Of the three cases in the present study, the Chilean public sector is perceived as most closely approximating the conditions of a highly professional, Weberian bureaucracy. According to observers at the World Bank, the Chilean state 'is effective in delivering public services, its public servants are professionalized – competent and honest – and its operations are predictable'.<sup>14</sup> Although journalistic works have occasionally described instances of misproprieties within the Chilean bureaucracy, most discussions of Chilean public administration emphasize that such misbehaviour is more the exception than the rule, and that as a whole the Chilean public administration is well described by the constituent features of the ideal-typical Weberian bureaucracy.<sup>15</sup> One of the origins of the relatively high level of probity encountered in the Chilean public sector supposedly rests

<sup>9</sup> Confidential interview with an official belonging to the Unidad de Análisis de Políticas Sociales y Económicas (UDAPE), La Paz, Bolivia, July 2003.

<sup>10</sup> Richard Graham, *Patronage and Politics in 19th Century Brazil* (Palo Alto, Calif.: Stanford University Press, 1990); Ben Ross Schneider, *Politics within the State: Elite Bureaucrats and Industrial Policy in Authoritarian Brazil* (Pittsburgh: University of Pittsburgh Press, 1991).

<sup>11</sup> David V. Fleischer, *Corruption in Brazil: Defining, Measuring, and Reducing* (Washington, D.C.: CSIS, 2002).

<sup>12</sup> Luis C. Bresser-Pereira, *Reforma do Estado para a Cidadania: A Reforma Gerencial Brasileira na Perspectiva Internacional* (Brasília: ENAP, 1998); Robert T. Daland, *Exploring Brazilian Bureaucracy: Performance and Pathology* (Washington, D.C.: University Press of America, 1981).

<sup>13</sup> Daland, *Exploring Brazilian Bureaucracy*; Edson Nunes, *A Gramática Política do Brasil: Clientelismo e Insulamento Burocrático* (Rio de Janeiro: Jorge Zahar Editor, 1997); Marcelo D. de Figueiredo Torres, *Estado, Democracia e Administração Pública no Brasil* (Rio de Janeiro: Editora FGV, 2004).

<sup>14</sup> Shahid J. Burki and Guillermo E. Perry, *Beyond the Washington Consensus: Institutions Matter* (Washington, D.C.: World Bank, 1998), p. 132.

<sup>15</sup> For a compilation of administrative improprieties, see Maura Brescia, *Manos Limpas: La Corrupción en las Empresas Públicas (de Todos los Chilenos), 1980–2001* (Santiago: Editorial Mare Nostrum, 2001).

with the considerable power and independence of its supreme auditing institution (SAI), the Comptroller General of the Republic, or CGR.<sup>16</sup> Historically, a formal examination system has not been utilized as it is in Brazil, although candidates have long been required to satisfy a rather stringent set of educational standards for entry to middle-level and technical posts.<sup>17</sup>

These accounts of bureaucratic quality in the three cases find resonance in various cross-national measures of bureaucratic effectiveness and corruption control. Table 2 presents rankings and percentile values of indicators for the three countries produced by three different organizations: the World Economic Forum, the World Bank Institute and Transparency International. On all of the indicators, Chile scores highest among the three cases, followed by Brazil, which is in turn followed by Bolivia. On the World Economic Forum's indicator of civil service independence (a measure of the independence of the civil service from political pressure), the difference between the Chilean and Brazilian cases is slight. However, for all other indicators – one designed to capture the frequency with which public funds are diverted from their official purposes (WEF), one designed to measure overall government effectiveness (WBI) and two designed to capture overall levels of corruption (WBI and TI) – the Chilean bureaucracy is ranked well ahead of the Brazilian bureaucracy, which in turn, is ranked well ahead of the Bolivian bureaucracy. In fact, Chile typically ranks among the top 30 per cent of all countries on such metrics, whereas Bolivia typically ranks among the bottom 30 per cent and Brazil typically resides close to the median.

Given the very different characterizations of the overall level of bureaucratic quality in the three cases, they provide the study with a unique opportunity to evaluate the claim that differences encountered across institutions in the same polity may be equally or more important than those encountered across entire public sectors. In essence, the use of the three cases here follows in broad strokes the 'crucial case' strategy elaborated by Harry Eckstein.<sup>18</sup> If one finds evidence that the level of institutional diversity within these cases exceeds the diversity encountered between the public sectors considered in their entirety – a surprising finding given the well-established national level reputations of the cases – then it is likely that this state of affairs obtains more generally.<sup>19</sup>

#### THE PUBLIC EMPLOYEES SURVEY

The public employees survey conducted in Bolivia, Brazil and Chile included 2,859 respondents from thirty different public agencies in the three countries. A concerted

<sup>16</sup> Mario Marcel, 'Effectiveness of the State and Development Lessons from the Chilean Experience', in Guillermo E. Perry and D. M. Leipziger, eds, *Chile: Recent Policy Lessons and Emerging Challenges* (Washington, D.C.: World Bank, 1999), pp. 265–325.

<sup>17</sup> James Petras, *Politics and Social Forces in Chilean Development* (Berkeley: University of California Press, 1969); Arturo Valenzuela, 'Parties, Politics and the State in Chile: The Higher Civil Service', in Ezra N. Suleiman, ed., *Bureaucrats and Policy Making: A Comparative Overview* (New York: Holmes and Meier, 1984), pp. 242–79.

<sup>18</sup> Harry Eckstein, 'Case Study and Theory in Political Science', in Fred I. Greenstein and Nelson W. Polsby, eds, *Handbook of Political Science: Strategies of Inquiry* (Reading, Mass.: Addison-Wesley, 1975), pp. 79–137.

<sup>19</sup> While in theory it might have been possible to find cases even further apart on overall reputations for bureaucratic capacity than the countries considered here, almost any such alternative set of cases would be likely to exhibit such marked differences in the internal organization of the public bureaucracy, language and cultural background that performing the type of survey-based comparisons presented in these pages would be infeasible.

TABLE 2 *Ranking and Percentile of the Three Cases on Cross-national Measures of Bureaucratic Effectiveness and Corruption*

Indicator	Bolivia		Brazil		Chile	
	Rank	Percentile	Rank	Percentile	Rank	Percentile
World Economic Forum						
Civil service independence (2000)	54th of 56	4%	35th of 56	30%	28th of 56	46%
Diversion of public funds (2003)	73rd of 80	8%	52nd of 80	35%	24th of 80	70%
World Bank Institute						
Government effectiveness (2004)	134th of 204	34%	84th of 204	59%	28th of 204	87%
Corruption control (2004)	152nd of 205	26%	89th of 205	57%	18th of 205	92%
Transparency International						
Corruption perceptions index (2003)	106th of 133	16%	54th of 133	58%	20th of 133	85%

attempt was made to select agencies with comparable functions across the three countries. To this end, the survey included agencies responsible for generating government revenue (tax services, customs agencies), agencies responsible for monetary policy and/or supervision of the banking sector (central banks, banking superintendencies), agencies responsible for overseeing infrastructure projects (road services), agencies responsible for managing development projects and agrarian reform (development ministries, social investment funds, land reform agencies), agencies responsible for managing and allocating social security contributions, as well as agencies responsible for regulating the non-bank financial sector (securities and insurance regulators).<sup>20</sup> All of the institutions selected for inclusion were national-level institutions (municipal or other sub-national level institutions were excluded). Moreover, only employees that worked in the central office of a given institution were included in the survey (regional offices were excluded). Fifty per cent of eligible employees in each institution were chosen at random to participate in the survey (up to a maximum of 250).

The surveys were self-administered and anonymous. They were delivered to and collected from individuals selected for participation in the study by trained undergraduate assistants as well as the author. Response rate by country varied from 69 per cent (Brazil) to 88 per cent (Chile). Response rate by institution varied from 42 per cent to 100 per cent, with the median equal to 81 per cent.

#### ANALYSIS

The central claims of this article emerge from the application of a distinct statistical procedure (or set of procedures) to the survey data. In what follows, I outline the survey questions utilized in the data analysis and describe the statistical procedures employed.

CLAIM 1: *Variation in bureaucratic capacity across agencies is generally greater than variation in bureaucratic capacity across nations.*

In order to assess the validity of this first claim, I examine the relative importance of agency-level differences versus national-level differences on individual survey responses to questions concerning the degree of politicization, the stability of career paths, and the strength of corruption control mechanisms in the environments in which public employees work. A variety of questions for each dimension of bureaucratic capacity

<sup>20</sup> The institutions included in the survey were as follows: *Bolivia*: Central Bank ( $n = 183$ ), Customs Service ( $n = 107$ ), Ministry of Peasant Affairs ( $n = 73$ ), Ministry of Economic Development ( $n = 67$ ), Ministry of Sustainable Development ( $n = 89$ ), National Institute of Agrarian Reform ( $n = 53$ ), National Tax Service ( $n = 138$ ), National Road Service ( $n = 109$ ), Productive Social Investment Fund ( $n = 43$ ), Agrarian Superintendency ( $n = 16$ ), Superintendency of Banks ( $n = 60$ ), General Superintendency ( $n = 23$ ), Superintendency of Hydrocarbons ( $n = 27$ ), Superintendency of Pensions, Securities and Insurance ( $n = 50$ ); *Brazil*: Central Bank ( $n = 200$ ), Federal Revenue Service, RF ( $n = 127$ ), Development Corporation for the São Francisco and Parnaíba Valleys, CODEVASF ( $n = 75$ ), Ministry of Agrarian Development, MDA ( $n = 46$ ), Ministry of Social Development, MDS ( $n = 140$ ), National Agency of Land Transport, ANTT ( $n = 76$ ), National Department of Transportation Infrastructure, DNIT ( $n = 198$ ), National Institute of Colonization and Agrarian Reform, INCRA ( $n = 129$ ), National Social Security Institute, INSS ( $n = 166$ ), Superintendency of Private Insurance, SUSEP ( $n = 69$ ); *Chile*: Customs Service ( $n = 152$ ), Institute of Agricultural Development ( $n = 84$ ), National Development Corporation, CORFO ( $n = 98$ ), Superintendency of Banks ( $n = 69$ ), Superintendency of Pension Fund Managers ( $n = 74$ ), Superintendency of Securities and Insurance ( $n = 117$ ).

were included in the analysis: six questions on politicization were examined, along with five different questions on the stability of bureaucratic career paths and three on the strength of formal and informal corruption control measures. The hope was that by analysing responses to diverse questions on each topic, the study would have a basis for conclusions about the relative weight of agency-level versus country-level variation in capacity that would not be overly dependent upon the specific details of question wording.

A statistical procedure developed for drawing conclusions about the relative importance of two or more categorical variables on an outcome of interest is the Bayesian hierarchical ANOVA (BHA) model.<sup>21</sup> For each question corresponding to each dimension of bureaucratic capacity explored in this article, this is the procedure that I use to assess whether or not the variation in responses attributable to agency-level differences is greater than that attributable to national-level differences. The model is roughly analogous to multivariate regression performed using only batches of binary predictor variables, where each batch corresponds to a particular source of variation in the outcome. However, unlike classical (unconstrained) regression, the BHA model presumes that the coefficients on the indicator variables corresponding to a particular source of variation (a given ‘batch’) are drawn from a common distribution, a normalization which permits one to calculate a separate coefficient for each value of each categorical variable included in the model. As a consequence, the standard deviation of the coefficients on the indicator variables corresponding to a particular source of variation can be used as a measure of the importance of that source of variation for the outcome. With two sources of variation – agencies and countries as studied here – the standard deviations of the coefficients for each source of variation can be compared in order to assess the relative importance of each. Gelman and Hill show that this approach can be fruitfully applied in situations in which one’s response data is both normally and non-normally distributed.<sup>22</sup>

In what follows, my statistical inferences concern the relative magnitude of  $s_\gamma$ , the finite population standard deviation of the coefficients on the agency indicators, relative to that of  $s_\delta$ , the finite population standard deviation of the coefficients on the country indicators. One benefit of adopting a Bayesian approach in the setting considered here is that it allows the analyst to inspect the simulated posterior density of these two quantities. If the mean value of the marginal posterior density of  $s_\gamma$  is greater than the mean value of that of  $s_\delta$  for a given survey question, this implies that the variation in individual survey responses attributable to agency-level differences is greater than that attributable to country-level differences for the question. If the inequality is reversed, this would imply that the opposite is the case. For each question analysed, I use the BHA model to calculate the mean values of  $s_\gamma$  and  $s_\delta$ . Moreover, by explicitly comparing values of these two quantities across draws from the simulated posterior, I am able to straightforwardly calculate the probability of the null hypothesis that  $s_\delta > s_\gamma$  (a task which would be substantially more complicated in a non-Bayesian framework). In this way, the Bayesian approach facilitates the crucial task of hypothesis testing.

In the specification of each Bayesian hierarchical ANOVA model, agencies were nested within countries. Moreover, as all of the survey responses analysed were binary or ordinal in form, the ANOVA models were specified using logit or ordered logit link functions to model the relationship between the sources of variation and the responses.

<sup>21</sup> Andrew Gelman and Jennifer Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge: Cambridge University Press, 2007), chap. 22.

<sup>22</sup> Gelman and Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*, pp. 490–4.

The Appendix presents a formal description of the BHA models used in this article, including all of the requisite specification and estimation details.

Before appraising the results, let us consider the question format. The set of questions chosen for analysis has the virtue that it contains a mix of formats which, alternatively, solicit judgements about importance (e.g., ‘To what extent do you agree or disagree with the following statement: Political connections are important in determining who obtains a post in my agency?’), solicit descriptions of behaviour (e.g., ‘Do you have an active role in a political party?’), or solicit explicit comparisons (e.g., ‘Please rank the following four factors in terms of the relative importance of each for dismissals in your agency.’) Relative to a scenario where question format solely prompts generic judgements about importance, the mix of formats utilized here reduces the risk of the results being coloured by national anchoring, i.e. survey responses being determined in part by what respondents believe to be the norm in the country as a whole. Such anchoring is potentially a concern because, were it to have a strong presence in the responses to the questions analysed here, it could lead agency-level variation in responses to swamp national-level variation even if the true amount of variation at the two levels were similar. Fortunately, we know that responses to the behavioural questions are not susceptible to such anchoring, since what is being solicited in these circumstances is a specific aspect of personal experience, not a judgement of any kind which could be clouded by an implicit national baseline. Moreover, the explicit comparison questions are also not likely to be susceptible to such anchoring dynamics, since by prompting only relative evaluations of importance of specific factors for a given bureaucratic task, a common baseline is already built in to the question format. In any case, as the reader will see, the general pattern of greater attributable variation encountered at the agency-level than at the national-level is found across all three question formats, which should mollify concerns that national anchoring or any other pathology tied to question format is driving the results.

### *Bureaucratic Politicization*

I begin by decomposing the sources of variation in responses to questions on bureaucratic politicization. Politicization is understood to be low when the following conditions obtain. First, hiring decisions depend primarily on the education, training and work experience of candidates – and only secondarily or not at all on partisan affiliation or political contacts. Secondly, the criteria for promotions are clearly explained to public employees, and consist primarily of indicators of job performance and/or experience, not on political or personal affinities with superiors. Thirdly, dismissals occur as a result of poor performance or unethical behaviour, and perhaps occasionally as a result of budgetary shortfalls or the shifting priorities of the government. As a rule, dismissals of significant numbers of employees are generally not tied to cabinet reorganizations or the electoral cycle. Finally, the role played by elected officials in relation to the bureaucracy is primarily one of establishing the direction of broad policy initiatives. Political intrusion into the day-to-day implementation of policy is kept to a minimum.

Table 3 presents the results of the BHA analysis. Of the six questions about politicization, four queried respondents about the importance of political criteria in hiring, promotions or dismissals in the agency in which they work. For each of these questions, the variation in responses attributable to agency-level differences was quite large ( $s_y$  parameters estimated to be in the 0.8–1.0 range). Moreover, for all such questions the standard deviation of the coefficients on the agency indicators was greater



TABLE 3 Two-way Analysis of Variance for Questions on Politicization, the Institutionalization of Bureaucratic Career Paths, and Corruption Control

Survey question	Est. sd of coefficients		Pr( $s_\gamma < s_\delta$ )
	Agency ( $s_\gamma$ )	Country ( $s_\delta$ )	
<i>Politicization</i>			
Political connections are important in determining who obtains a post in my institution*	0.83 (0.05)	0.39 (0.17)	0.00
Promotions in my institution depend more on good performance in the workplace than on political affiliation or influence*	1.01 (0.07)	0.37 (0.23)	0.01
Indicate with a 1 below the factor that has the greatest influence on dismissals in your institution, 2 the factor that has the second greatest influence, and so on until you have ranked the four factors (outcome = 1 if rank of political factors was 1 or 2, outcome = 2 if rank of political factors was 3, outcome = 3 if rank of political factors was 4)	0.91 (0.26)	0.46 (0.26)	0.06
Indicate with a 1 below the factor that has the greatest influence on promotions in your institution, 2 the factor that has the second greatest influence, and so on until you have ranked the four factors (outcome = 1 if rank of political connections was 1 or 2, outcome = 2 if rank of political connections was 3, outcome = 3 if rank of political connections was 4)	0.79 (0.07)	0.28 (0.18)	0.01
Do you have an active role in a political party?	1.17 (0.20)	0.57 (0.44)	0.11
Party leaders can easily punish employees who do not obey the orders of the governing parties*	0.33 (0.05)	0.82 (0.09)	1
<i>Institutionalization of Bureaucratic Career Paths</i>			
Number of years served in institution (1 = less than 3 yrs, 2 = 3 to 5 yrs, 3 = >5 yrs)*	1.00 (0.09)	0.55 (0.28)	0.05
Number of years served in current post (1 = less than 3 yrs, 2 = 3 to 5 yrs, 3 = >5 yrs)*	0.77 (0.08)	0.60 (0.20)	0.21
How would you characterize the likelihood that you lose your job in the near future?†	0.60 (0.06)	1.17 (0.14)	1
Prior to beginning current post, the respondent had another post in the same institution	0.66 (0.10)	0.32 (0.16)	0.04
Indicate with a 1 below the factor that has the greatest influence on promotions in your institution, 2 the factor that has the second greatest influence, and so on until you have ranked the four factors (outcome is rank given to workplace performance)	0.95 (0.07)	0.24 (0.18)	0

TABLE 3 (Continued)

Survey question	Est. sd of coefficients		Pr( $s_\gamma < s_\delta$ )
	Agency ( $s_\gamma$ )	Country ( $s_\delta$ )	
<i>Corruption Control</i>			
If someone were to misappropriate the resources of your institution, how probable do you think it would be that an internal audit, an external audit, or some other type of investigation would occur as a result? ‡	0.55 (0.06)	0.47 (0.16)	0.36
If an audit or another type of investigation did occur, what do you think would happen to the person who misappropriated the resources? ¶	0.48 (0.06)	0.62 (0.15)	0.83
Would you say that the majority of employees in your institution would file a complaint ( <i>denuncia</i> ) if they were aware that someone had misappropriated the resources of the institution? 1 = yes, 2 = no.	0.34 (0.05)	0.15 (0.09)	0.03

*Notes:* Standard deviations in parentheses. Due to the absence of responses in certain categories, the Brazilian agency ANTT was removed from the analysis. \*1 = strongly agree, 2 = agree, 3 = disagree, 4 = strongly disagree. †1 = 'highly probable' or 'probable', 0 = 'improbable' or 'highly improbable'. ‡ 1 = highly probable, 2 = probable, 3 = improbable, 4 = highly improbable. ¶ 1 = the person would probably be judged in a court of law and sanctioned, 2 = the person would probably be judged in a court of law but not sanctioned, 3 = the person would probably neither be judged in a court of law nor sanctioned.

than that of the country indicators ( $s_\delta$  parameters estimated to be in the 0.3–0.5 range), thus indicating that agency-level differences were a more important source of variation in responses than country-level differences. By and large, the differences in the quantities of interest were statistically significant by conventional standards: the probability of the null hypothesis of greater variation in responses attributable to country-level differences than agency-level differences was estimated as being less than 0.01 for three of these questions and 0.06 for the fourth.

One question in this series asked respondents if they would classify themselves as playing an active role in a political party. The variation in responses to this question attributable to both agency-level and country-level differences was quite large (with Bolivian public employees more likely to be partisan activists than their Brazilian or Chilean counterparts). Yet the agency-level component of variation was nevertheless twice as large as the country-level component: the estimated (mean) value of  $s_\gamma$  for this question was equal to 1.2, whereas the estimated value of  $s_\delta$  was 0.6. It would thus seem that, while differences in the intensity of patronage politics across countries are important, the differences across agencies may be even more so.

The only question on politicization for which country-level differences clearly swamped agency-level differences as a source of variation was one that prompted respondents to denote the extent to which they agreed or disagreed with the proposition that 'party leaders can easily punish functionaries who do not obey the orders of the governing parties'. The Bolivian respondents clearly stood out in this regard, with 88 per cent of respondents responding that they 'strongly agreed' or 'agreed' with the proposition, as

opposed to 62 per cent and 66 per cent of the Brazilian and Chilean public employees, respectively, who responded in this way.<sup>23</sup> All told, perceptions about the sanctioning capacity of party leaders aside, most aspects of politicization appear to vary more across the agencies included in this study's sample than across the countries.

### *Stability of Bureaucratic Career Paths*

An important component of Weberian bureaucracy is the existence of stable career paths which permit employees to cultivate knowledge and skills conducive to the efficient functioning of their institutions as well as promotion criteria tied to performance in the workplace. The questions included in this series queried respondents about the number of years they had served in their agency, the number of years they had served in their current post, the likelihood they would assign to losing their job in the near future, whether they had been recruited for their current job internally, and the importance they believe workplace performance holds for promotions in their agency. Agency-level differences were highly relevant as a source of variation in responses to these questions ( $s_\gamma$  parameters in the 0.6–1.0 range).

In four of the five questions contained in this set, the variation attributable to agency-level differences was found to be greater than that attributable to country-level differences. For three out of these four questions, in turn, the null hypothesis of greater variation in responses attributable to country-level differences than agency-level differences could be rejected using standard thresholds of statistical significance. For only a single question in the series (the one which prompted employees to assess the likelihood of job loss in the near future) was the variation attributable to country-level differences found to be greater than that attributable to agency-level differences. In this instance, one again finds a large 'Bolivia effect': 68 per cent of respondents in Bolivia indicated that it was 'highly probable' or 'probable' that they would lose their jobs in the near future, compared to 18 per cent and 40 per cent of Brazilian and Chilean respondents, respectively.

### *Strength of Corruption Control Measures*

A final series of questions explored respondents' perceptions about the strength of formal and informal corruption control mechanisms. These questions asked respondents to assess the likelihood of an internal or external investigation being opened in response to a hypothetical misappropriation of resources, the likelihood of a legal sanction being imposed upon the guilty party if an investigation did occur, and whether the respondents' colleagues would be likely to serve as whistleblowers if they were privy to an act of malfeasance.

In general, agency-level differences were less relevant as a source of variation in responses in this category of questions than those on politicization and the stability of bureaucratic career paths ( $s_\gamma$  parameters in the 0.3–0.6 range). For two of the three questions, the variation in responses attributable to agency-level differences outweighed that attributable to country-level differences (although only in one case – the question on

<sup>23</sup> It is likely that the distinctiveness of the answers of the Bolivian respondents to this question derives in part from features of electoral rules in that country which place substantial political power in the hands of party leaders. See Brian Crisp, 'The Nature of Representation in Andean Legislatures and Attempts at Institutional Reengineering', in Scott P. Mainwaring, Ana M. Bejarano and Eduardo Pizarro Leongómez, eds, *The Crisis of Democratic Representation in the Andes* (Palo Alto, Calif.: Stanford University Press, 2006), pp. 204–25.

colleagues' willingness to be whistleblowers – were the differences in attributable variation statistically significant by conventional standards). For the question on the likelihood of legal sanction given an investigation into malfeasance, the importance of country-level differences marginally trumped that of agency-level differences. Given that the question taps largely into perceptions of the competence of the public prosecutor's office and/or judiciary, both national-level features of the legal environment, this finding is not particularly surprising. Rather, what would seem to be most noteworthy about the responses to this question is the similarity in the levels of variation attributable to country-level and agency-level differences.

All told, public sector bureaucrats' responses to questions about politicization, the stability of bureaucratic career paths, and corruption control mechanisms in their work environment in Bolivia, Brazil and Chile appear to be driven at least as much by the particular characteristics of the agency to which they belong as by the contours of the national-level bureaucratic apparatus in which the agency is located. This is not to suggest that there are not exceptions to this pattern – both in terms of their perceptions of (a lack of) autonomy vis-à-vis party leaders and in terms of their subjective assessments of the likelihood of job loss in the near term, Bolivian public employees plainly stand out from their counterparts in Brazil and Chile. Nevertheless, the weight of evidence provided in Table 3 clearly points to a pattern of greater attributable variation in responses due to agency-level differences than national-level differences.

### *Potential Objections*

At least two potential objections could be raised to my strategy of using data from a survey of public employees to gauge the relative magnitude of cross-national versus cross-agency differences in bureaucratic capacity. First, it could be argued that there may exist a generalized tendency among public employees to describe their agencies in a manner which conforms to the stereotype of a high performance institution, irrespective of their agencies' true performance levels. This could potentially be due to concerns about making one's agency look bad (perhaps leading to a reduced budget in the future if negative survey results are publicized) or it could be due to a psychological drive to feel that one is among the best in the public sector. Secondly, and relatedly, it could be argued that public employees working in agencies that are clearly suffering from various manifestations of low bureaucratic capacity may choose not to participate in a survey which would lead them to divulge information about such a sorry state of affairs, presumably for the same reasons articulated above.

While I believe issues of misrepresentation and non-response are legitimate and important concerns in public employee surveys as a general matter, it is fairly straightforward to show that they do not constitute a serious threat to the results presented above. Consider the first objection. If the stated tendency was operative, cross-agency variation in responses would be fairly muted (creating a bias against a finding in favour of Claim 1) and even the worst-performing agencies on a given survey question would not perform particularly poorly in an absolute sense. In fact, the statistical analyses above have already shown that cross-agency variation is generally quite large. Moreover, the survey results can be inspected directly to assess whether there is any reason to believe that respondents are reticent to cast their agencies in an unfavourable light.

Figure 1 is instructive in this regard. Presenting the responses to two questions on politicization of personnel matters, the figure provides a clear sense of both the large

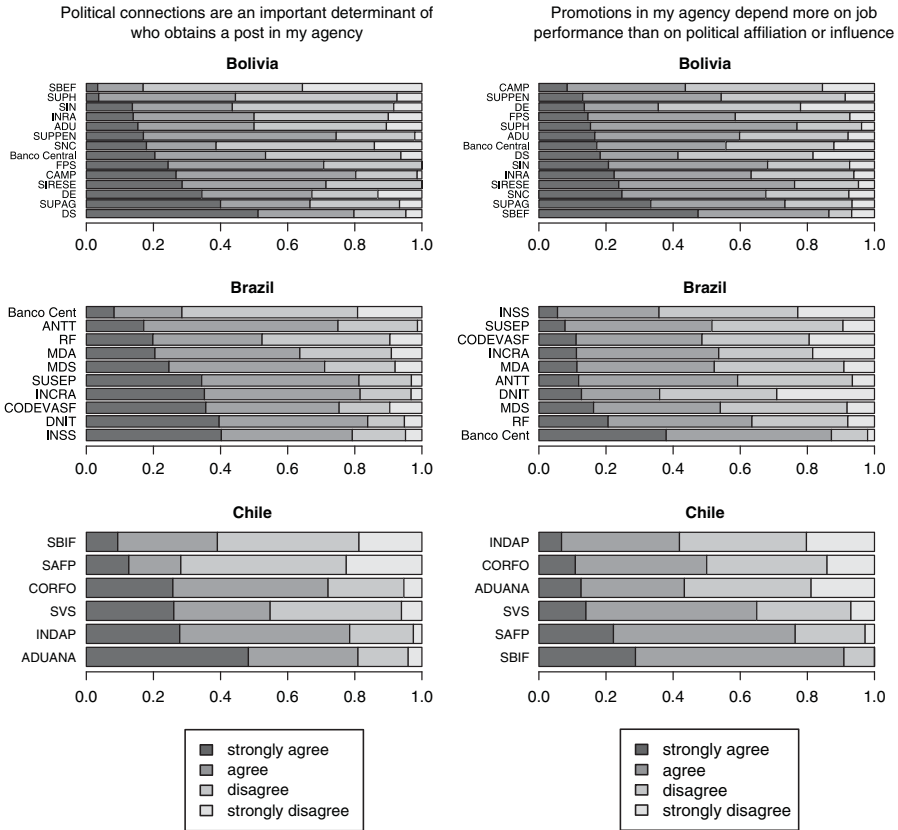


Fig. 1. Agency-level diversity in each polity in perceptions of politicization of personnel matters

agency-level diversity that exists within countries and the general willingness of respondents to describe their agencies in often starkly negative terms. In fact, many employees were more than willing to provide responses that cast their agency in the worst possible light. For instance, 51 per cent of respondents in Bolivia’s Ministry of Sustainable Development, 48 per cent of respondents in Chile’s Customs Service, and 40 per cent of respondents in both Brazil’s National Department of Infrastructure and Transport and its Social Security Institute indicated that they strongly agreed that political connections were important for hiring decisions within their institutions. With respect to factors that influence promotions, 59 per cent, 57 per cent, 64 per cent and 64 per cent, respectively, of the respondents in these same institutions stated that they disagreed or strongly disagreed that workforce performance was more important than political affiliation or influence. Responses to other questions (not shown in the figure) illustrate the same point. Asked to rank the importance of political factors in dismissals (relative to poor job performance, unethical behaviour and insubordination), more than 70 per cent of respondents in Bolivia’s Ministry of Sustainable Development, Ministry of Economic Development and Ministry of Peasant Affairs all ranked this factor as most important, as did 56 per cent of respondents in Brazil’s National Department of Infrastructure and Transport. These types of results are clearly inconsistent with the

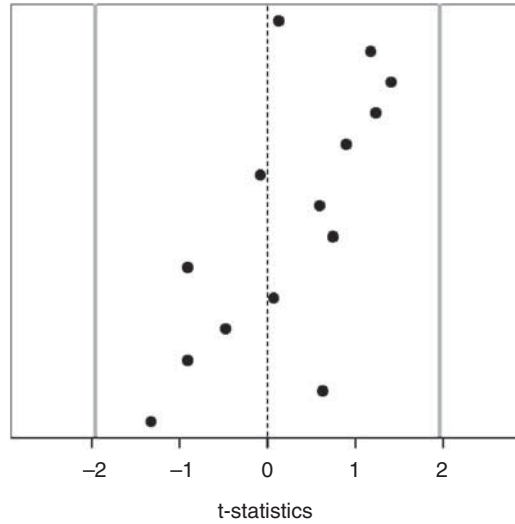


Fig. 2. *T-statistics from pairwise regressions of agency-level response rates on agency-level average responses to questions about bureaucratic capacity*

Note: Black circles denote *t*-statistics corresponding to the coefficient on (the average value of) responses for a given question about bureaucratic capacity. Vertical gray lines denote the 95% threshold for statistical significance.

notion the bureaucrats are inherently predisposed to protect the reputation of their agencies through their survey responses.

Now consider the second objection. As before, if the behavioural tendency identified by this objection was indeed operative, then one consequence would again be muted cross-agency variation in responses, because those individuals with negative opinions and experiences would not be inclined to participate in the survey; clearly, this is not what the study finds. Another consequence would be that agency-level response rates should be strongly associated with bureaucratic capacity measures across the three dimensions discussed above, since employees in poorly performing agencies would be less inclined to participate. This second implication can be evaluated directly. To conduct such an evaluation, I performed a series of pairwise regressions which regressed the agency-level response rate on the agency-level average value of the responses to each question listed in Table 3. I then examined the *t*-statistics corresponding to the coefficient on the average responses at the agency-level, which are presented in Figure 2. If the employees' inclination to participate in the survey was an underlying function of the level of bureaucratic capacity within their agencies, one would expect a number of these *t*-statistics to be greater than 1.96 in absolute value (the standard threshold for statistical significance). As the figure illustrates, there was not a single question in Table 3 for which this was the case. Thus, concerns about non-participation being driven by bureaucratic capacity appear to be unfounded in the dataset studied here.

CLAIM 2: *The relative position of countries' national-level reputations for bureaucratic capacity by and large does not correspond to differences in perceptions about capacity for employees who belong to different countries but serve in agencies that perform the same function.*

Simply because I find evidence of greater agency-level variation in responses to questions about bureaucratic capacity than country-level variation, it does not necessarily follow that studies which use national-level indicators of capacity to explain development outcomes shaped largely by a small subset of bureaucratic agencies are misguided in their approach. One potentially reasonable response to my findings would be to admit that although agency-level variation in capacity is large – perhaps generally greater than country-level variation – yet claim that the variation in capacity that one finds across agencies within a given country tends to follow the same pattern across countries. More specifically, one could argue that the position of particular types of agencies in the overall distribution of capacity is relatively consistent across countries, where ‘type’ refers to administrative function (e.g. road services are always politically penetrated and thus have low capacity relative to other agencies in a country; central banks are usually autonomous and thus have high capacity relative to other agencies in a country). If this is indeed the case, and if the distribution of capacity has the same basic shape across countries, then it would follow that the ranking of countries on mean values of bureaucratic capacity would be equivalent to the ranking of countries on the capacity of agencies of a particular type. In other words, one could claim that a road service in a high capacity state always has greater capacity than a road service in a low capacity state, a central bank in a high capacity state always has greater capacity than a central bank in a low capacity state, and so forth. Were this true, then using national-level measures of capacity as proxies for capacity in specific parts of the public sector would not be as problematic as this article would have it seem.

The analysis conducted in this section is designed to assess the reasonableness of this response to the findings presented above. My evaluation proceeds in two steps. First, in the interest of economy of presentation, I estimate Weberianness Perception Scores (WPSs) for each respondent in the survey using a Bayesian factor analytic technique appropriate for observed variables that are ordinal in form. The WPSs provide a one-dimensional summary measure of the degree to which each respondent perceives his work environment as conforming to the characteristics of an ideal-typical Weberian bureaucracy. Secondly, I examine differences in the distributions of the WPSs across pairs of ‘sister agencies’: agencies in different countries that execute similar tasks. In so doing, this article is able to gauge whether differences in the distributions of perceptions of bureaucratic capacity held by employees in different countries but in agencies with the same administrative function correspond to the countries’ national-level reputations for capacity embodied in the PB-GIs.

### *Generating the Weberianness Perception Scores*

The WPSs are constructed using responses for each of the subjective (perception-oriented) questions presented in Table 3. The responses to all of these questions have a very important feature in common. They can all be thought of as observable manifestations of a single latent (unobservable) factor: the survey respondent’s belief about the degree to which the environment in which he works approximates the Weberian ideal-type of stable, meritocratic bureaucracy. Given the conceptualization of the survey responses as emanating from this factor (and given the fact that the observed responses are ordinal in form), I use the tools of item response theory to estimate the perception scores. In particular, I use the two-parameter ordinal item response model with a probit link function to model the relationship between the ‘manifest’ variables (the survey responses)

and the latent variable (perceptions of 'Weberianness').<sup>24</sup> My approach to estimation of the model was Bayesian, and, in particular, followed a strategy developed by Kevin Quinn.<sup>25</sup> The Appendix presents the posterior density summary and estimation details for the model.

### *Comparing 'Sister' Agencies*

Table 4 specifies the institutional pairings for each comparison of sister agencies across countries. The scores for more than a single agency were pooled in cases in which multiple agencies perform the task(s) denoted under the category heading. In Brazil, for instance, both monetary policy and banking regulation fall under the purview of the central bank, whereas in Bolivia these two activities are the separate domains of the central bank and superintendency of banks, respectively. Thus, in comparing employee perceptions of bureaucratic capacity in agencies dealing with monetary policy and banking regulation, the scores for the two Bolivian institutions are pooled. Instances in which such pooling was necessary should be clear from the table.

Figure 3 compares the distributions of employee perceptions in the paired 'sister agencies' using quantile–quantile (Q–Q) plots. For any two countries, the QQ plots present the quantiles of the distribution of the WPSs for individuals who belong to the agency (or agencies) of the indicated type in one country plotted against the quantiles of the distribution of the WPSs for individuals who belong to the agency (or agencies) of the indicated type in the other country. To the degree that the observations in the plot are arrayed on or close to the 45° line, this constitutes evidence that the distributions of perception scores for individuals in agencies of the indicated type in the two countries are similar to one another. If the observations arrayed in the plot lie above (below) the 45° line for certain ranges of the WPS, this indicates that there is a greater concentration of scores in those ranges for the country indicated on the *y* axis (*x* axis) than the country on the *x* axis (*y* axis). In each plot contained in the figure, the country with the reputation for higher bureaucratic capacity is located on the *x* axis. Therefore, any plot in which observations are concentrated below the 45° line constitutes support for the hypothesis of greater perceptions of capacity in the agency or agencies belonging to the country with the better national-level reputation for capacity, whereas any plot in which observations are concentrated above the 45° line constitutes support for the hypothesis of greater perceptions of capacity in the agency or agencies belonging to the country with the worse national-level reputation for capacity.

Out of the fourteen paired comparisons of sister agencies presented in the figure, only for eight could one definitely conclude that the distribution of perceptions differed across the two groups (based on the two-sample Kolmogorov–Smirnov test for the equality of distributions). Of these eight cases, perceptions of bureaucratic professionalism were evenly split: in four instances they were found to be higher in the agencies belonging to the country with the better national-level reputation for capacity and in four instances they were found to be lower in the agencies belonging to the country with the better national-level reputation for capacity. For the four categories of agency type which permitted comparisons of sister

<sup>24</sup> Valen E. Johnson and James H. Albert, *Ordinal Data Modeling* (New York: Springer, 1999).

<sup>25</sup> Kevin M. Quinn, 'Bayesian Factor Analysis for Mixed Ordinal and Continuous Responses', *Political Analysis*, 12 (2004), 338–53. The package MCMCpack written for the R statistical computing environment was used to calculate the posterior density of the model parameters of interest as well as the perception scores. See Andrew Martin and Kevin M. Quinn, 'MCMCpack 0.4–8' (available at <http://mcmcpack.wustl.edu>).



TABLE 4 *Sister Agencies (as Determined by Administrative Function)*

	Bolivia/Brazil		Bolivia/Chile		Brazil/Chile	
	Bolivia	Brazil	Bolivia	Chile	Brazil	Chile
Monetary Policy and/or Banking Regulation	Central Bank, Superintendency of Banks	Central Bank	Superintendency of Banks	Superintendency of Banks	Central Bank	Superintendency of Banks
Taxation and/or Customs Regulation of Insurance, Pensions, and Securities	Customs Service, Tax Service Superintendency of Pensions, Securities and Insurance	Federal Revenue Service (RF) Superintendency of Private Insurance	Customs Service Superintendency of Pensions, Securities and Insurance	Customs Service Superintendency of Pension Fund Managers, Superintendency of Securities and Insurance	Federal Revenue Service (RF) Superintendency of Private Insurance	Customs Service Superintendency of Securities and Insurance
Agrarian Reform	National Institute of Agrarian Reform	National Institute of Colonization and Agrarian Reform	National Institute of Agrarian Reform	National Institute of Agrarian Development	National Institute of Colonization and Agrarian Reform	National Institute of Agrarian Development
Road Services	National Road Service	National Department of Transportation Infrastructure				
Ministries	Ministry of Sustainable Development, Ministry of Economic Development, Ministry of Peasant Affairs	Ministry of Social Development, Ministry of Agrarian Development				

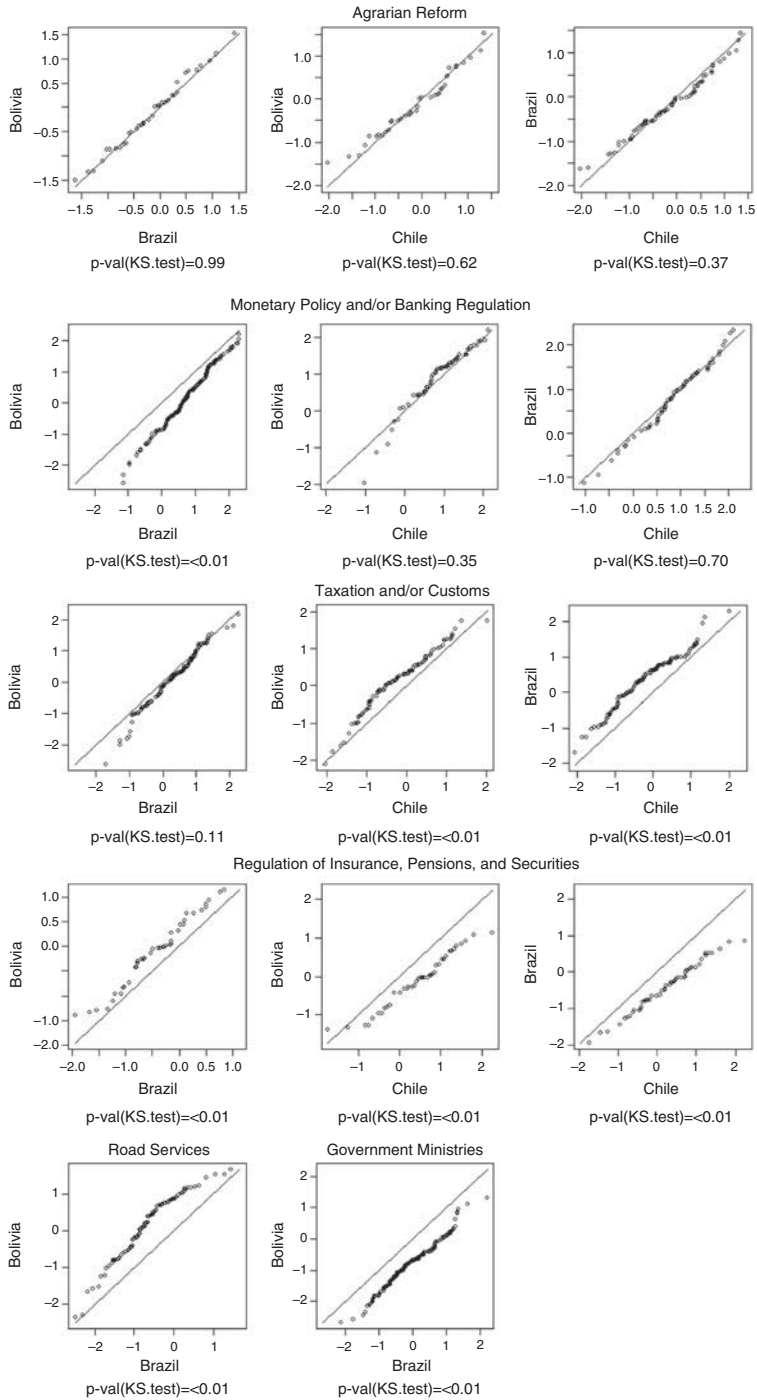


Fig. 3. Q-Q plots of perceptions of Weberianism in sister institutions (as determined by administrative functions). Note: See Table 4 for the list of agencies being compared in each plot: p-val K-S test refers to the probability of the null hypothesis that the distributions of the factor scores being compared are equal (according to the two-sample Kolmogorov-Smirnov test).

agencies across multiple pairs of countries, in no instance do the national-level reputations for bureaucratic capacity correctly predict the differences in perceptions of Weberianness for all country pairs. All told, this analysis should lead one to doubt the claim that relative differences in national-level measures of bureaucratic capacity adequately capture differences across nations in capacity within particular types of state agencies.

CLAIM 3: *Employee perceptions of bureaucratic capacity correspond to patterns of actual behaviour within bureaucratic agencies that are consistent with the perceptions.*

A potentially reasonable objection to the findings of the previous section could be that employee perceptions of professionalism and actual professionalism may be two different things. In particular, one could argue that perceptions of capacity swing widely across agencies within a given country but actual levels of professionalism do not. In order to address such concerns, in this section I examine the extent to which the one-dimensional Weberian Perception Scores are informative about actual patterns of behaviour within state agencies.

One crucial aspect of bureaucratic behaviour consists of staffing decisions for posts whose duties entail the selection of firms that will receive public contracts, government licences or government-backed loans. Because the choices made by individuals in such posts may greatly affect the profitability of private firms, opportunities for corruption commonly arise for their holders. Whether or not these opportunities for corruption translate into actual corruption depends in part on how the slots for such posts are filled. In this regard, a wealth of evidence drawing upon myriad forms of data (cross-national, agency-level narrative case studies) has shown that meritocratic recruitment procedures are associated with a lower incidence of corruption among bureaucrats in the developing world.<sup>26</sup> Given this state of affairs, one would expect that public agencies which have high levels of professionalism would have very different types of recruitment processes for the aforementioned posts than public agencies with low levels of professionalism. Specifically, one would expect that highly Weberian agencies would be characterized by recruitment procedures for these posts that emphasize the technical merit of candidates and eliminate any sense of indebtedness to political patrons, whereas more patrimonial agencies would be characterized by recruitment procedures that give technical merit short shrift and intensify candidates' sense of indebtedness to patrons.

The logic undergirding this expectation is easy to articulate. In the more professional agencies, administrators will seek to organize hiring practices so as to minimize the incidence of corruption, especially in those posts where the financial consequences of corrupt behaviour are the greatest, since – as administrators in a Weberian environment – their

<sup>26</sup> James Rauch and Peter Evans, 'Bureaucratic Structure and Bureaucratic Performance in Less Developed Countries', *Journal of Public Economics*, 75 (2000), 49–72; Daniel Kaufmann, Massimo Mastruzzi and Diego Zavaleta, 'Sustained Macroeconomic Reforms, Tepid Growth: A Governance Puzzle in Bolivia?' in Dani Rodrik, ed., *In Search of Prosperity: Analytic Narratives on Economic Growth* (Princeton, N.J.: Princeton University Press, 2003), pp. 334–98; World Bank, *Understanding Public Sector Performance in Transition Countries: An Empirical Contribution*, Report No. 30357 (2003); Yemane Desta, 'Designing Anti-Corruption Strategies for Developing Countries: A Country Study of Eritrea', *Journal of Developing Societies*, 22 (2006), 421–49; Odd-Helge Fjeldstad, 'Corruption in Tax Administration: Lessons for Institutional Reforms in Uganda', in Susan Rose-Ackerman, ed., *International Handbook on the Economics of Corruption* (Cheltenham, Glos.: Edward Elgar, 2006), pp. 484–511.

prestige and advancement depends upon the efficient use of resources within their agency. In the more patrimonial agencies, often effectively cash cows for political parties, the incentives of administrators will be exactly the opposite: they will seek to organize hiring processes in order to encourage acts of corruption, since they and/or their party will anticipate sharing in the proceeds of this activity. In operational terms, what this discussion implies is that if the perceptions of bureaucratic capacity represented by the WPSs are informative about the true Weberian–patrimonial nature of public agencies, then agencies should differ in their recruitment procedures for these posts in a manner in keeping with the perception scores: agencies with high scores (Weberian) filling these posts through meritocratic procedures and agencies with low scores (patrimonial) filling these posts through non-meritocratic procedures.

Fortunately, the Brazilian component of the public-employee survey allows the study to test this proposition in a direct way. Each respondent was asked if his job ‘involved decisions regarding the content of contracts between private firms and public institutions’ and, subsequently, whether his job ‘involved decisions regarding licences or loans conceded to private firms or citizens’. One third of employees responded in the affirmative to the first question and 12 per cent did so for the second. The survey also asked respondents whether they had obtained their posts through a public examination. This was clearly the modal form of entry in Brazil: approximately three quarters (72 per cent) of Brazilian respondents declared that they had obtained their post in this way.

Using the responses to the former set of questions as dependent variables, the study specified two logistic regression models, one specifying the relationship between having taken a public examination and the likelihood of having a post which involves control over contracts, and the other specifying the relationship between having taken a public examination and the likelihood of having a post which involves control over licences and loans. The models have two important features. First, they include a full set of agency dummy variables, thus accommodating the fact that employee duties in some agencies are much more likely than in others to include the allocation of contracts, licences and loans. Secondly, they permit the impact of having taken a public examination to vary according to the type of agency in which a respondent works.

By agency type, I refer to the level of bureaucratic professionalism as represented by the Weberianness perception scores. Because the number of observations per agency was limited, agencies were divided into two types: agencies for which the WPS for the median employee in the agency is low (‘low WPS’ agencies; these correspond to a low level of bureaucratic professionalism), agencies for which the WPS for the median employee in the agency is at a moderate or high level (‘high WPS’ agencies; these correspond to a relatively high level of bureaucratic professionalism). Based on the median WPS in each agency, three agencies were placed in the first category (DNIT, INSS, SUSEP; median WPS equal to  $-0.70$ ,  $-0.42$  and  $-0.36$ , respectively) and seven in the second (MDS, INCRA, MDA, CODEVASF, ANTT, RF, BCdoB; range of median WPSs  $-0.20$  to  $0.86$ ).

Let  $i$  subscript respondents and  $j \in \{1, \dots, J\}$  subscript agencies. Each model was specified as  $\Pr(y_i = 1) = \text{logit}^{-1}(v_i)$ , where

$$v_i = \mu + \sum_j^{J-1} \gamma_j a_i(j) + \phi \text{exam}_i + \lambda (\text{exam}_i \times \text{low.WPS}_i) + \mathbf{X}_i \beta,$$

$y_i$  is the response of respondent  $i$ ,  $\mu$  is an intercept term,  $a_i(j)$  is an indicator variable equal to 1 if individual  $i$  belongs to agency  $j$ , 0 otherwise,  $\gamma_j$  captures the effect on the

TABLE 5 *Logit Models of Relationship between Public Examinations and Authority over Public Contracts, Licences and Loans*

	Contracts		Licences & loans		
	Estimate	se	Estimate	se	
$\mu$	0.68	0.41	-0.90	0.51	
$\phi$	0.51	0.21	0.43	0.31	
$\lambda$	-1.12	0.34	-1.18	0.48	
agency	BC	-2.75	0.39	-1.82	0.44
	CODEVASF	-1.27	0.39	-1.26	0.46
	DNIT	-0.89	0.39	-1.15	0.46
	INCRA	-2.33	0.39	-1.68	0.47
	INSS	-1.66	0.41	-0.51	0.47
	MDA	-1.96	0.44	-2.41	0.77
	MDS	-1.72	0.34	-2.94	0.57
	RF	-3.17	0.42	-2.30	0.52
SUSEP	-1.56	0.48	-0.13	0.54	
Years in public sector	0.13	0.08	-0.06	0.10	
Years in agency	-0.05	0.07	0.02	0.10	
Internal hire	0.02	0.17	0.34	0.22	
Formal education	university	-0.04	0.18	0.39	0.28
	postgraduate	-0.06	0.20	0.31	0.31
<i>N</i>	1,136		1,132		

response of membership in agency  $j$  relative to the omitted agency (ANTT),  $exam_i$  is an indicator variable equal to 1 if individual  $i$  had taken a public examination, 0 otherwise,  $low.WPS_i$  is an indicator variable equal to 1 if individual  $i$  belonged to one of the three more weakly professionalized agencies, 0 otherwise,  $\phi$  captures the effect on the response of having taken an examination for members of the more professionalized agencies,  $\phi + \lambda$  captures the effect on the response of having taken an examination for members of the relatively weakly professionalized agencies, and  $X_i$  represents confounding variables.

If the WPSs truly tap into an actual Weberian–patrimonial dimension, one would expect to find that  $\phi > 0$  and that  $\phi + \lambda < 0$ . That is to say, the effect of having taken a public examination on the likelihood of having a post with substantial control over the allocation of valuable benefits to firms should be positive in the more Weberian agencies and negative in the more patrimonial agencies.

Table 5 presents the results of the analysis. As expected, in the low WPS agencies having taken a public examination is negatively and statistically significantly related to authority over contracts and authority over licences and loans. Also as expected, in the high WPS agencies one finds that having taken a public examination is positively related to authority over contracts and authority over licences and loans (although the impact is statistically significant only in the case of contracts). The analysis revealed quite a bit of variation across agencies in terms of the likelihood of functionaries controlling access to state benefits; not surprisingly, Brazil’s National Agency of Land Transport (ANTT) and the National Department of Infrastructure and Transport (DNIT) stood out as agencies particularly likely to have employees with some control over public contracts. The only confounder with a statistically significant impact on the outcomes of interest was the

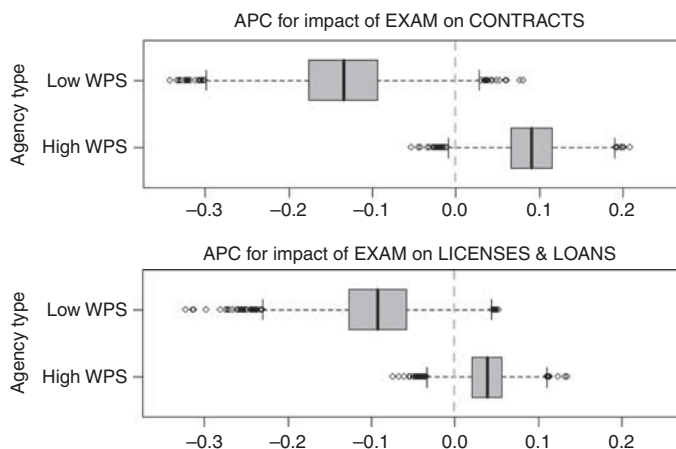


Fig. 4. Estimated average predictive comparisons for the probability of having authority over public contracts (upper panel) and the allocation of licences and loans (lower panel) as a function of having one's job through a public examination (Brazilian agencies only).

Note: Impact of having obtained one's post via a public examination on authority over contracts, licences and loans is allowed to vary in the logistic regression models across two agency types: agencies with low Weberian perception scores (DNIT, INSS) and agencies with high perception scores (BCdoB, RF, ANTT, CODEVASF, INCRA, MDA, MDS). Displayed in the figure are the estimated average predictive comparisons associated with having taken a public examination, broken down according to agency type.

number of years served in the public sector, which was positively related to having authority over contracts.

In order to provide the reader with a sense of the magnitude of the effect of having taken a public examination on authority over the allocation of state benefits, Figure 4 presents boxplots of the average predictive comparisons (APCs) associated with having taken an examination for low WPS agencies and high WPS agencies.<sup>27</sup> The APC is a measure of the average effect of having taken a public examination for individuals within the indicated type of agency. In the case of authority over public contracts, the differences in the APCs for the low WPS agencies and the high WPS agencies were quite large. Within the former, the APC was estimated as  $-0.13$  whereas in the latter, the APC was estimated as  $0.09$ . In layman's terms, this means that, on average, having taken a public examination was associated with a  $0.13$  reduction in the probability of having authority over contracts in the agencies perceived of as faring relatively poorly on professionalism by their employees, whereas having taken a public examination was associated with a  $0.09$  increase in the probability of having authority over contracts in the agencies perceived of as being at a moderate or high level of professionalism by their employees. For the case of authority over licences and loans, the APC associated with having taken an examination in the low WPS agencies was  $-0.10$ , whereas in the high WPS agencies the value of this statistic was estimated as  $0.04$ .

In sum, bureaucrats' perceptions about the level of professionalism in the agencies in which they work correspond to important behavioural manifestations of bureaucratic capacity.

<sup>27</sup> Andrew Gelman and Ian Pardoe, 'Average Predictive Comparisons for Models with Nonlinearity, Interactions, and Variance Components', *Sociological Methodology*, 37 (2007), 23–51.

Agencies which were perceived to lack many of the key components of Weberian bureaucracy were relatively unlikely to place authority over potentially exploitable state benefits in the hands of employees who had gone through a competitive examination process, whereas agencies with a greater correspondence to the Weberian ideal type were relatively likely to place this authority in the hands of employees who had gone through such a process.

COMPARING AGENCY-LEVEL AND OMNIBUS MEASURES OF BUREAUCRATIC QUALITY IN THE STUDY OF BANKING CRISES

Utilizing the public employees survey data from Bolivia, Brazil and Chile, I have thus far shown that the assumptions which would justify the use of national-level measures of bureaucratic quality to study sector-specific development outcomes are likely to be generally invalid. In this final empirical section of the article, I take the next logical step and empirically demonstrate the pitfalls which may be associated with substituting an omnibus measure of bureaucratic quality in place of the appropriate agency-level one. I do so by examining an issue of considerable importance for the developed and developing world alike: the onset of systemic banking crises.

An examination of banking crises is well suited to my aims in this article for two reasons. First, to the degree that bureaucratic quality plays a role in the incidence of such crises, it is fairly clear which component of the bureaucratic apparatus should be most important, namely, the bank supervision authority. This is so practically by definition – one of the explicit mandates of such institutions is to perform the oversight activities necessary to ensure that banking crises do not occur. Secondly, high quality datasets have been assembled on the incidence of major banking crises as well as the legal and behavioural characteristics of bank supervision authorities around the world. This makes it possible to compare explicitly the associations between bank crises and measures of bureaucratic quality in supervision authorities versus associations between crises and measures of the quality of the bureaucracy as a whole. Recall that my central admonition in this article is that the impact of bureaucratic quality at the national level may be unreflective of the impact of bureaucratic quality within the particular institutions that matter the most for a particular fine-grained development outcome. An examination of banking crises allows the paper to assess – in a focused way – if this is an admonition whose violation has substantively important consequences.

In operational terms, the outcome examined here consists of an indicator variable denoting the presence of a systematic banking crisis during the 1990–99 period. The variable is drawn from the World Bank's Database of Banking Crises, which defines a systemic banking crisis as a scenario in which any of the following occur: (1) emergency measures are taken to assist the banking system; (2) large-scale nationalizations of banks take place; (3) non-performing bank assets reach at least 10 per cent of total assets; (4) banking sector rescue operations take place at a cost of at least 2 per cent of gross domestic product (GDP).<sup>28</sup>

I seek to compare the association between this outcome and an omnibus (national-level) measure of bureaucratic quality to the association encountered between this outcome and a suitably chosen measure of bureaucratic quality within bank supervision authorities, in both cases conditioning on a small set of covariates likely to exert an

<sup>28</sup> Gerard Caprio, Daniela Klingebiel, Luc Laeven and Guillermo Noguera, 'Banking Crises Database,' World Bank (2003). Available at: [http://www1.worldbank.org/finance/html/database\\_sfd.html](http://www1.worldbank.org/finance/html/database_sfd.html).

independent impact on the emergence of banking crises. The omnibus measure of bureaucratic quality I utilize is drawn from the Political Risk Service's *International Country Risk Guide* (ICRG). It is equal to the average value of the variable 'bureaucratic quality' measured in December 1990 and December 1991.<sup>29</sup> ICRG's measure is the only omnibus indicator of bureaucratic quality available prior to the onset of banking crises in the 1990s (and thus has the virtue of being uncontaminated by negative halo effects). In order to measure bureaucratic quality within the bank supervision authority, I created an indicator of bank regulator autonomy designed to reflect institutional and behavioural conditions within such agencies that make them less susceptible to capture by politicians and private sector actors. The indicator is based upon entries contained in Barth, Caprio and Levine's Bank Regulation and Supervision Database.<sup>30</sup> It is an additive index, ranging in value from 0 to 3, which is made up of three component factors. The first is a binary variable equal to 1 if the head of the supervision agency is appointed by another member of the bureaucracy (such as the head of the Central Bank), and not by political authorities such as the president, the parliament or a king (0 otherwise). The second is a binary variable equal to 1 if the head of the supervision agency cannot be directly dismissed by the aforementioned political authorities (0 if she can). The final component is a binary variable equal to 1 if, based on survey responses of national bank supervisory authorities, the supervision agency is characterized as one in which bank supervisors do not frequently work in the banking industry after leaving their jobs as regulators (0 if they frequently do work in the banking industry after leaving their jobs). Given these coding rules, higher levels of the index indicate greater levels of functional autonomy from the demands of political elites and banks than do lower levels. The covariates I utilize consist of the logarithm of GDP per capita in 1990, the amount of domestic credit provided by the banking sector as a percentage of GDP in 1990, and an indicator variable denoting the existence of a deposit insurance law in force during the 1990s.<sup>31</sup>

With this data in hand, I proceeded to conduct two logistic regression analyses, one which regressed the banking crisis indicator onto the ICRG measure of bureaucratic quality and covariates and another which regressed the banking crisis indicator onto the bank regulator autonomy measure and covariates. Table 6 presents the results of these regressions. The most important aspect of the table concerns the value of the coefficients on the two bureaucratic quality variables. In the model in which ICRG measure was utilized as an explanatory variable, the coefficient on this variable was statistically insignificant and of the 'wrong' sign (i.e. it was positive, indicating that higher bureaucratic quality increases the probability of a banking crisis). In the model in which the indicator of bank regulator autonomy was instead utilized as an explanatory variable, the coefficient on this variable was negative and statistically significant at conventional levels, indicating that low bureaucratic quality within bank supervision authorities was an important predictor of the outbreak of a systematic banking crisis in the 1990s.

<sup>29</sup> Post-communist states are given the bureaucratic quality score which corresponds to the first year they enter the ICRG dataset.

<sup>30</sup> James R. Barth, Gerard Caprio Jr and Ross Levine, 'The Regulation and Supervision of Banks around the World: A New Database', World Bank Policy Research Working Paper No. 2588 (2001). Data available at: <http://econ.worldbank.org>.

<sup>31</sup> The first two indicators come from World Bank, 'World Development Indicators' (2011). Available at: <http://data.worldbank.org/data-catalog/world-development-indicators>. The latter indicator comes from the aforementioned Barth, Caprio and Levine, Bank Regulation and Supervision Database.



TABLE 6 *Logit Models of Bureaucratic Quality and Systematic Banking Crises in the 1990s (Bank Regulator-specific Measure Compared to Omnibus Measure)*

Outcome: existence of systematic banking crisis during 1990–99 period				
	Model 1		Model 2	
	Coefficient	se	Coefficient	se
ICRG bureaucratic quality	0.27	0.25	–	–
Index of bank regulator autonomy	–	–	–0.53	0.24
ln(GDP per capita)	–0.84	0.30	–0.55	0.20
Deposit insurance	0.79	0.59	0.59	0.57
Domestic credit/GDP	0.00	0.01	0.00	0.01
	<i>N</i> = 76		<i>N</i> = 76	
Expected impact of bureaucracy measure on probability of banking crisis for a randomly selected polity in the sample (comparison is sample minimum vs. sample maximum)				
	Estimate	se	5%	95%
Index of bank regulator autonomy	–0.32	0.14	–0.54	–0.07
ICRG bureaucratic quality	0.21	0.20	–0.15	0.49

Thus, in this example, the use of a focused, agency-specific measure of bureaucratic quality versus an omnibus measure of the same makes all the difference in the world in terms of drawing conclusions about the relevance of state capacity for the health and stability of the financial sector. Use of the former would erroneously lead an analyst to conclude that bureaucratic quality was irrelevant for the onset of banking crises in the 1990s, or worse, that it actually somehow had a stimulative effect on such crises. In contrast, when a reasonable agency-level measure is utilized, it becomes immediately clear that bureaucratic quality – operationalized here as institutionalized autonomy from political and private sector capture – did in fact appear to dampen the likelihood of crisis during this period.

The difference in conclusions that one would reach based upon the use of one measure or the other is not just a matter of sign and statistical significance but also one of deep substantive importance. Drawing upon the estimated regression models, the bottom section of Table 6 presents the expected impact of each bureaucracy measure on probability of banking crisis for a randomly selected polity in the sample. In order to calculate these estimates, the bureaucracy measures were varied from their sample minimum values to their sample maximum values. The table reports that such a change in the index of bank regulator autonomy would be expected to produce a reduction in the probability of a banking crisis of 0.32 for a randomly selected polity in the sample. A similar change in the ICRG bureaucratic quality index would be expected to produce an increase in the probability of a banking crisis of 0.21 for a randomly selected polity in the sample, although in this case the large confidence intervals around the estimate do not permit one to exclude the possibility of zero impact.

The take-home point of this exercise is this: if an analyst or policy maker were to rely naïvely upon the national-level measure of bureaucratic quality, she would be wrongly led to believe that improving the professionalism and capacity of bank regulators is not likely

to reduce the incidence of banking crises. Were she to use the more appropriate agency-level measure, she would discover that the professionalism and capacity of bank regulators is likely to have significant impact in reducing the incidence of such crises. Empirical analyses based upon the presumption that the behaviour of the mean bureaucratic agency reflects that of those agencies most relevant to the problem at hand can lead analysts profoundly astray.

## CONCLUSION

The existence of significant within-country variation in bureaucratic performance has long been reflected in important work produced by political scientists and the larger development community.<sup>32</sup> Nevertheless, students of development have become increasingly inclined to rely upon national-level PB-GIs in producing assessments of the impact of bureaucratic capacity on a whole host of fine-grained development outcomes. In cases where the outcomes in question are likely to be driven substantially more by certain state agencies than others, I have argued that reliance upon the national-level PB-GIs may ultimately lead to faulty inferences about the impact of bureaucratic capacity. Drawing primarily upon data on bureaucratic performance in Bolivia, Brazil and Chile, I have attempted to evaluate the reasonableness of approaches that rely upon national-level measures of bureaucratic capacity by carefully examining the empirical veracity of the implicit assumptions that national-level differences in capacity outweigh those found at the agency level and that such differences correspond to differences in the performance of specific types of agencies.

The verdict of the article is straightforward on these points. Agency-level differences in capacity generally appear to outweigh national-level differences. For the vast majority of survey questions tapping into aspects of bureaucratic capacity such as politicization, the stability of bureaucratic career paths, and corruption control, an employee's response depended more on the particular agency in which he worked than on the country in which he resided. Moreover, I have shown that the capacity of agencies performing similar functions in different countries often did not reflect the national-level reputation for capacity of the countries to which they belonged. In other words, agency-level differences in capacity across countries were not predictable based on the national-level reputations for capacity. Although there can be no guarantee that these findings will hold for a larger sample of countries than those considered here, the fact that Bolivia, Brazil and Chile have such radically different reputations for bureaucratic effectiveness at the national level suggests that this is likely to be the case.

The substantive importance of these findings was emphasized by an examination of data on banking crises in the 1990s. Explicitly assessing the validity of utilizing a national-level omnibus measure of bureaucratic capacity in place of a measure designed to tap into bureaucratic capacity within the most substantively relevant agency, I have found that the association between bureaucratic capacity and crisis using the omnibus measure was completely unrepresentative of the association between crisis and the more appropriate agency-level measure. Thus, the findings of the article suggest that not only are the assumptions which would justify the use of mean-level measures of capacity in place of

<sup>32</sup> J. Edgardo Campos and Sanjay Pradhan, ed., *The Many Faces of Corruption: Tracking Vulnerabilities at the Sector Level* (Washington, D.C.: World Bank, 2007); Robert Putnam, with Robert Leonardi and Raffaella Y. Nanetti, *Making Democracy Work: Civic Traditions in Modern Italy* (Princeton, N.J.: Princeton University Press, 1993).

agency-level ones not likely to be met in practice, but that the inferences gleaned from doing so may be highly misleading.

Although the tenor of this article is certainly critical of the often indiscriminant use of national-level PB-GIs in the literature on the political economy of development, it is important to emphasize that their use is not necessarily inappropriate for studying *all* the types of development outcomes that may be of interest to social scientists. Rather, the appropriateness of their use ultimately rests on the scope of the dependent variable being examined. On one hand, finely grained development outcomes which are logically tied much more strongly to bureaucratic capacity within a specific subset of the agencies contained within the bureaucratic apparatus than to others are clearly not good candidates for the use of these omnibus indicators. In these cases, political scientists would do well to identify the state agencies most relevant for the outcome of interest and attempt to measure capacity within these directly, rather than assuming that national-level measures of capacity adequately represent the conditions in these agencies. Researchers interested in how bureaucratic capacity affects banking crises should examine capacity within bank supervision agencies, researchers interested in how capacity affects measures of pollution should examine capacity within environmental protection agencies, researchers interested in how capacity affects academic performance should examine capacity within education ministries (and their dependencies), and so forth. On the other hand, outcomes whose logical relationship to the state apparatus is highly multi-dimensional in nature, such as economic growth, may be more appropriate candidates for the use of national-level PB-GIs. This is because many, if not all, the agencies of the state make a contribution to the genesis of such outcomes. An empirical strategy which concentrated on studying the role of capacity only within a narrow sliver of the state apparatus would seem difficult to justify in such circumstances.

What are the implications of this study for future research? The fact that national-level reputations for overall bureaucratic capacity may correspond only weakly, if it all, with the efficacy of particular state agencies points to a potentially fruitful research agenda for enterprising comparativists: cross-national comparative analysis of bureaucratic performance at the sectoral level. Such an agenda recommends a number of important and intriguing puzzles. To give one example: why, in spite of massive corruption directed towards party politics and electoral campaigns in Brazil, Mexico and Venezuela, does one find in Brazil a state-owned oil company (Petrobras) that is relatively insulated from partisan politics and is among the most competitive and efficient in the world (to the degree that it was recently able to help make Brazil self-sufficient in oil), whereas in Mexico and Venezuela one finds state-owned oil companies (Pemex and PDVSA) that have been routinely used as cash cows for the ruling parties and are characterized by enormous waste and inefficiency? With three quarters of the world's oil reserves in the hands of state-owned enterprises, an answer to the above question – carefully formulated using general principles – could have an appreciable effect on social welfare around the globe.

Another implication is that policy makers interested in bureaucratic reform may find it useful to learn from successful and unsuccessful forms of bureaucratic organization within their own states. Towards this end, some social scientists have used in-depth case studies to identify the sources of success in bureaucratic performance using agency-level or programme-level variation within particular countries.<sup>33</sup> Although of potentially great

<sup>33</sup> Judith Tendler, *Good Government in the Tropics* (Baltimore, Md.: Johns Hopkins University Press, 1997).

utility, such work remains a rarity in comparative politics. Of course, surveys of public employees such as the one presented here are also exceptionally well suited for this purpose. Employee surveys have the advantage of providing a means to map out in a fairly systematic way patterns of cross-agency variation in capacity within a given country, thus making the appropriate targets of reform and the path to it easier to discern. In sum, a healthy appreciation for the variation in capacity that exists within the state recommends a number of exciting paths for comparative research and this article indicates why these are paths worth taking.