Maximizing the Reliability of Cross-National Measures of Presidential Power

David Doyle and Robert Elgie

British Journal of Political Science / FirstView Article / December 2014, pp 1 - 11

DOI: 10.1017/S0007123414000465, Published online: 23 December 2014

David Doyle
Associate Professor of Latin American Politics
Department of Politics and International Relations
University of Oxford
Fellow of St Hugh's College
St Margaret's Road
OX2 6LE
UK
david.doyle@politics.ox.ac.uk

Robert Elgie
Professor of Government and International Studies
School of Law and Government
Dublin City University
Dublin 9
Ireland
robert.elgie@dcu.ie

This article aims to maximize the reliability of presidential power scores for a larger number of countries and time periods than currently exists for any single measure and in a way that is replicable and easy to update. We begin by identifying all of the studies that have estimated the effect of a presidential power variable, clarifying what scholars have attempted to capture when they have operationalized the concept of presidential power. We then identify all of the measures of presidential power that have been proposed over the years, noting the problems that are associated with them. To generate our new set of presidential power scores, we draw upon the comparative and local knowledge embedded in existing measures of presidential power. Employing principal component analysis together with the expectation maximization algorithm and maximum likelihood estimation, we generate a set of presidential power scores for a larger set of countries and country time periods than currently exists, reporting 95 per cent confidence intervals and standard errors for the scores. Finally, we discuss the implications of the new set of scores for future studies of presidential power.

Estimating the effect of presidential power

There is now a large body of work that has estimated the outcome of variation in presidential power. To identify studies of presidential power systematically, we searched a selection of leading comparative politics journals. Using the term 'presidential power variable', we searched the *American Journal of Political Science*, *American Political Science Review*, *British Journal of Political Science*, *Comparative Political Studies*, *Comparative Politics*, *European Journal of Political Research*, *International Studies Quarterly*, *Journal of Politics*, *Political Research Quarterly*, and

World Politics. All articles published to the end of 2011 that included an estimation with a presidential power variable were recorded. These articles were also consulted to identify whether or not they referred to other books or journal articles that might include such an estimation. In total, 49 studies that included an estimation of presidential power were identified.[1] The distribution of this work confirms that scholars are increasingly choosing to estimate the effect of presidential power generally. Four were published from 1995-1999 inclusive and 10 from 2000-2004 inclusive, whereas 25 articles were published from 2005-2009 inclusive with 10 in 2010 and 2011 alone. In all but four of these studies presidential power was operationalized explicitly or implicitly as an explanatory variable. In these 45 studies, the dependent variable ranged widely across topics such as economic reform; democratic consolidation; the level of protectionism; the effective number of parties; cabinet composition; voter turnout; and many others. In 30 of these 45 studies, variation in presidential power was confirmed to have a significant effect on the outcome under investigation.

What are scholars trying to capture when they estimate the effect of presidential power? In 11 of the 49 studies we identified, scholars focused on only a specific aspect of presidential power. For example, Cheibub wished to explain variation in budget balances in democratic systems.[2] Consistent with his focus, he operationalized a presidential power variable, but only in terms of the president's power over budgetary policy and the president's veto power rather than presidential power generally. Thus, when scholars wish to test a particular theory of

[1] A full list of these studies is available in the online material.

[2] Cheibub 2006

presidential power, there is evidence that they have estimated the effects of only the specific elements of presidential power relating to that theory.

In the remaining cases, though, scholars stated that they wished to estimate the effect of presidential power generally. A very small number of scholars were more precise about what they understood by this term. For example, Biglaiser and DeRouen stated that they were trying to capture "centralized executive authority".[3] Hicken and Stoll understood presidential power to be "the degree to which power is concentrated in the presidency within the national level of government".[4] Most scholars, though, stated only that they were interested in the effects of a general term such as presidential power or powers, presidential strength, presidential authority, executive power, executive authority, or an equivalent term. While there could be semantic differences between these terms, there is no discussion of such differences. Scholars have been using them synonymously. With regard to the terms presidential power and executive power, there were studies that used the terms presidential power and executive power as direct synonyms.[5] However, there were studies that estimated the effect of variation in the level of constraints on the executive in the system of checks and balances. They operationalized Polity's XCONST variable or Henisz's POLCON variable.[6] There were also studies where

[3] Biglaiser and DeRouen 2004, 567

[4] Hicken and Stoll 2008, 1110

[5] Roper 2002, 253; Clark and Wittrock 2005, 475 and 479

[6] Polity's XCONST variable is available at http://www.systemicpeace.org/inscr/inscr.htm with data through to 2011 inclusive. Henisz's POLCON variable is available at: http://www-management.wharton.upenn.edu/henisz/ with data now through to 2012 (polconiii variable). Both accessed 25 April 2013. We excluded studies that estimated the effect of constraints on the executive

scholars estimated both a presidential power variable and the XCONST executive constraints variable separately.[7] In short, scholars were able to distinguish presidential power from executive constraints more broadly. We excluded studies that estimated solely the effect of executive constraints.

Overall, we identified 38 studies where scholars tried to estimate the impact of presidential power generally. They used different terms to refer to this phenomenon, but we can be confident that whatever term they used they were trying to capture the extent to which the presidency was a powerful actor within the national government, rather than either some specific power of the institution or the position of the executive within the system of checks and balances more broadly.

Existing measures of presidential power

How have scholars tried to estimate the impact of presidential power generally? A number of the 38 studies we identified drew up a discrete measure of presidential power with cross-national country scores. Most studies, though, relied on a measure that had been drawn up by other scholars whose sole aim was to generate a set of presidential power scores rather than to estimate the empirical effect of variation in the scores. These measures were often available only either in specialist journals or in online datasets. Therefore, to identify the full set of presidential power measures that have been proposed over the years, it was necessary to move beyond a search of leading journals. To that end, a separate Google Scholar search was conducted

from the start. So, they are not included in the set of 49 studies of presidential power that we identified.

[7] For example, Doyle 2010

using terms such as 'presidential power measure' and 'index of presidential power'. We identified 19 separate and original measures of presidential power,[8] plus a further 16 studies that used one of these measures but both/either reported scores for a different set of countries and/or gave countries different scores from the original study.[9] Thus, we have a dataset of 35 measures of presidential power.

The methodology used across the 35 measures is relatively consistent. The measures are all based on a set of individual indicators of presidential power. Often, the indicators are binary. If a president enjoys a particular power, then a value of 1 is assigned for that indicator. Otherwise, a value of 0 is recorded. Sometimes the indicators are ordinal. For example, Shugart and Carey propose ten indicators of presidential power with each indicator having a range of 0-4.[10] Presidents are then awarded a score within this range for each indicator. Whether the indicator scores are binary or ordinal, the total score for presidential power is invariably the aggregate of the scores for each indicator. This generates a set of cross-national presidential power scores for particular time periods.

While there are now many different measures of presidential power, there are empirical and theoretical problems with them. First, while none of the measures aimed to capture the personal power of individual presidents, the measures did capture two different manifestations of presidential power. Some were derived solely from constitutional indicators of presidential power, whereas others were based on a mix of constitutional and behavioral power, meaning the power of the

[8] A full list of these studies is available in the online material.

[9] A full list of these studies is available in the online material.

[10] Shugart and Carey 1992

presidency in "actual political practice".[11] There are problems with measuring the constitutional powers of presidents because constitutions can be imperfect measures of actual political power. However, there are also problems with measuring the behavioral power of presidents because there is the risk of capturing the impact of factors such as party competition, rather than the power of the presidency itself.

Second, even if we confine ourselves to measures of one type of presidential power, the correlation between the different measures can be relatively low. For example, comparing only those measures that are based on indicators of constitutional powers, the pairwise correlation between the Shugart and Carey and Johannsen measures is -0.19.[12] The same figure for the Hellman and Frye measures is 0.50, even though both are measuring presidential power scores only in Central and Eastern European and the former Soviet Union.[13] Inevitably, this means that empirical results are likely to be sensitive to the particular measure that is used.[14]

Third, there is great variation in the country coverage of the different studies as well as the time periods that were covered. Only three of the 35 measures covered a large number of countries across political regimes generally.[15] Some focused on only one particular region, such as Latin America, Eastern Europe and the former Soviet Union, or Africa. Others selected on the basis of a different analytical criterion. For example, Tavits reports the scores for 23 countries but only those with

[11] Siaroff 2003, 303

[12] Shugart and Carey 1992; Johannsen 2003. These figures are based on the way in which the scores are calculated that is described in the next section.

[13] Hellman 1996; Frye 1997

[14] Tavits 2009, 48

[15] Shugart and Carey 1992; Siaroff 2003; Johannsen 2002

weak presidencies.[16] What is more, scholars have now been proposing presidential powers scores for nearly 20 years. However, these scores are not updated after publication. Given that constitutions are often amended, reported presidential power scores can soon go out of date. This means that countries sometimes cannot be reliably included in an estimation even if a presidential power score for that country exists.

Finally, there are problems of construct validity. Fortin has shown that the indicators of any given measure of presidential power are not necessarily capturing a single latent construct.[17] She performed factor analysis on a dataset that pooled Shugart and Carey's presidential power scores with Frye, Hellman and Tucker's scores.[18] These scores are based on ten indicators of presidential power, capturing two different dimensions, one relating to the president's executive powers with four indicators and another relating to the president's legislative powers with six indicators. However, Fortin found that seven of the ten indicators cluster into a single factor with eigenvalues greater than 1 and "with no evidence of separate latent constructs for legislative and non-legislative powers".[19] She also pointed out that the process of aggregating the scores for the individual indicators is problematic. She states: "[a]ggregation produces homogeneity claims, meaning that equal scores are substitutable or equivalent".[20] However, she noted that "each score can be obtained through broad combinations of different powers, and should thus

[16] Tavits 2008

[17] Fortin 2013

[18] Shugart and Carey 1992; Hellman and Tucker 2000

[19] Fortin 2013, 97

[20] Fortin 2013,107

not be considered homogenous in terms of causal analyses".[21] She goes on to argue that for any given measure "not all items hypothesized to capture the concept of presidential power seem to matter equally in accounting for composite scores" and that "not all potentially relevant items were tested".[22] She concludes that existing indices of presidential power have "limited validity".[23]

Generating a new set of presidential power scores

We wish to generate a time-series cross-sectional (TSCS) dataset of presidential power scores with country years as the units of observation. To do so, we resist the temptation to construct a new measure of presidential power from scratch. Fortin's study shows that any measure of presidential power is likely to suffer from a basic problem of construct validity.[24] From this, she concludes very skeptically, effectively questioning whether any measure of presidential power is likely to be valid. We agree with her analysis, but draw a different conclusion. Most social science concepts, such as voter turnout, social equality, corruption, and so on, suffer from equivalent problems of construct validity. For that reason, we prefer to place the emphasis on the reliability of the data that underpins the concept we are trying to capture. Specifically, we wish to avail of the expert information embedded in existing measures, but in a way that generates a more reliable set of cross-national presidential power scores.

[21] Fortin 2013,107

[22] Fortin 2013,108

[23] Fortin 2013,108

[24] Fortin 2013

To maximize the reliability of our new set of scores, three elements are emphasized. First, we focus solely on measures that record the constitutional power of presidents. To be sure, constitutions can be sometimes imperfect indicators of presidential power, but the overall reliability of our new set of measures is increased by referring solely to information in publicly available documents than by including essentially contestable judgments about presidential power in practice. Five of the 35 measures of presidential power that we identified provided scores for the behavioral power of presidents.[25] Excluding them leaves 30 measures. For the purposes of our methodology, two measures of constitutional presidential power that scored only a single country were also excluded, leaving a database of 28 measures from which to generate our new set of scores.

Second, we wish to draw upon all of the expert information in these 28 studies, but we wish to generate new scores in a way that indicates their general reliability. This allows researchers to make a decision as to whether to include particular countries in any estimation of presidential power. Therefore, standard errors and 95 per cent confidence intervals are reported for each of our presidential power scores.[26]

Third, we wish to maximize the reliability of our scores by accounting for systematic variation between the 28 measures of presidential power and so reducing

[25] See the online materials for a full list.

[26] We calculate the confidence intervals and standard errors on the basis of the original prespow 1 and 2 scores (i.e. before normalization). They provide a measure of the degree of certainty among the constitutive measures. Their primary purpose is to provide researchers with an easy way of establishing whether a given country score is capturing a shared understanding of presidential power among researchers for that country or whether there are divergent opinions.

the impact of any idiosyncratic measures. To do so, principal-component analysis (PCA) is employed. If certain measures are found to vary systematically from others, then it is possible to adjust for the relative importance of those measures when generating our new presidential power scores.

To begin, we identify the time period covered by the presidential power score for all of the different countries in each of the 28 original datasets. There can be more than one time period for a given country. For example, there are two time periods for Albania, 1991-1997 inclusive and 1998-2012 inclusive. These periods correspond to the first post-communist constitution that came into force in 1991 and the new constitution that was promulgated in 1998. Eight of our 28 datasets recorded a presidential power score for Albania for the 1991-1997 period and three for the later period. Overall, there are scores for a total of 116 countries and 181 country time periods. There was a maximum of four time periods for a number of countries, including Chile and Slovakia, and a maximum of 17 presidential power scores for one country time period, namely Romania 1991-2012. The mean number of scores per country time period was 2.7, the modal category was one score for 54 country time periods, and the median number of scores per country time period was two. Therefore, the data is in country time period format. Country scores do not change on a yearly basis. They change only when the constitution is amended in a way that alters that country's presidential power score. For example, there are two lines for Argentina in the dataset; one for the years 1984-1994 and another for the period from 1995 onwards, following the constitutional amendments in August 1994.

With information about the time period for each country, the first new measure can be calculated (prespow1). Given that presidential power scores are calculated differently across many of the different datasets, a set of mean

normalized scores is generated. For each of the 28 datasets, each country score was normalized on the basis of the following formula: (x minus minimum possible value)/(maximum possible value minus minimum possible value). For example, Shugart and Carey recorded a score of 17 for Panama 1972-2012 on their scale from 0-40.[27] Therefore, the Shugart and Carey normalized presidential power score for Panama was 0.43 in a range from 0-1. A score for Panama was recorded in four of the 28 datasets. The average of these four normalized scores was 0.47, generating a raw (prespow1) measure. The whole set of country scores was then normalized to generate a range from 0-1 to facilitate comparison with our second set of scores below. The final normalized (prespow1) score for Panama is 0.45. The full set of raw and normalized prespow1 scores with standard errors and 95 per cent confidence intervals are reported in Table 2 in the online materials.[28] A selection of scores is provided in Appendix 1a of this article.

To calculate our second new measure, principal component analysis (PCA) was employed. This method relies on a correlation or covariance matrix. However, there are large gaps in our sample. Any individual measure of presidential power covers only a specific subset of countries and country years. For example, Shugart and Carey may have good coverage of the Americas, but no African countries are

[27] Shugart and Carey 1992

[28] It is important to note that we calculate the confidence intervals on the basis of the original presidential power scores that generate our raw scores and not the final normalized 0-1 prespow 1 and 2 scores. This has the effect that the confidence intervals we report for a given country do not necessarily provide a measure of confidence relative to the prespow scores for any other country. Rather, they provide a measure of confidence in the level of concordance among the original set of scores for that country.

included. Moreover, Shugart and Carey's scores were reported as of 1992. As a result, their data will only partially overlap with Hicken and Stoll, who code presidential power not only for the Americas but also for countries in Asia, Africa and Eastern Europe and who also have the opportunity to record scores for more recent country years.[29] Therefore, before we can apply PCA, the issue of missing data needs to be addressed.

We do this by following the method of analyzing incomplete data suggested by Truxillo and performing PCA by using maximum likelihood estimation with the expectation-maximization (EM) algorithm.[30] This approach is an alternative to multiple imputation and is particularly suited to PCA, for while principal components can be explicitly computed, as Chen notes, we can also derive the principal components with an EM approach.[31] This allows us to use the EM as a means to estimate the data we are missing. In essence, this is an iterative procedure, which, without explicitly deriving the sample covariance, enables us to determine the subspace spanned by the dominant eigenvector.[32] The initial step in this approach involves computing the maximum likelihood estimates of the mean vector and covariance matrix for our set of 28 presidential power measures.[33] These estimates are derived from an iterative expectation-maximization algorithm.[34] The

[29] Hicken and Stoll 2008; Shugart and Carey 1992

[30] Truxillo 2005

[31] Chen 2002, 4.

[32] Chen 2002, 5.

[33] Given our data is in country constitution format, the starting point for each country, is the first constitution where we have a measure of presidential power.

[34] Truxillo 2005, 3; see the website for the full code.

EM algorithm provides estimates of the missing data based on the observed values within the dataset, that is, the existing measures of presidential power. In doing so, it estimates parameters that take into account any dependencies in the missingness among our measures of power.[35] So, the Expectation (E) step fills in the gaps in our data. The now complete data, including all observed and estimated data points, are processed with maximum-likelihood estimation, or the Maximization (M) step. This provides the updated mean vector and covariance matrix estimates. This process is repeated until the "maximum change in the estimates from one iteration to the next does not exceed a convergence criterion".[36] That is, with the new data from the M step, the E is repeated, followed again by the M step, and so on. This iterative process continues until we derive reliable estimates of the missing data matrix.

With complete data, we can then perform PCA. This method seeks a linear combination of potentially correlated variables and extracts the maximum variance from them. The resulting principal component (Y_1) is weighted by the degree to which each original variable explains the variance in the underlying orthogonal dimension.[37] That is,

$$Y_1 = a_{11}X_1 + a_{12}X_2 + \ldots + a_{1p}X_p$$

Each of the 28 measures of presidential power can be treated as a separate variable. Using PCA, a single presidential power score can be generated for each country

[35] Truxillo 2005, 3.

[36] Truxillo 2005, 3; for the EM algorithm more generally, see McLachlan and Krishnan 1997.

[37] See for example Flury 1988; Jolliffe 2002.

time period using the information from all 28 measures.[38] The resulting measure is a linear *weighted* construct of all existing power measures.[39] Using this technique, we can control for variation across the 28 measures of presidential power, reducing the impact of idiosyncratic measures on the final presidential power score that we report. This method allows us to weight the contribution of each existing measure of presidential power. Thus, the prespow2 scores are a linear construct of all existing presidential power variables, which are *weighted* by their rotated component scores.[40] These scores capture the underlying variance explained by each measure of power. The Kaiser-Meyer-Olkin measure of sampling adequacy is quite high, lending credence to our low-dimensional representation of presidential power. In a final step, the raw scores are normalized to generate a range from 0-1. The full set of raw and normalized prespow2 scores with standard errors and 95 per cent confidence

[38] We use the command pcamat in Stata 13. This allows the covariance matrix C to be specified as a $k \times k$ symmetric matrix. We also avoid biased estimates by specifying the column-wise minimum n as suggested by Truxillo 2005.

[39] If our data exhibits serial dependence, then this will undermine the assumptions needed to perform PCA on longitudinal data (for example, Wansbeek and Meijer 2000). However, our raw data is in country time period format, meaning that we do not perform PCA with a long time series, which is where we are most likely to find this type of auto-correlation. In fact, the results of a Wooldridge test for serial correlation indicates that this is not an issue for our data (see Drukker 2003). Nonetheless, as a further robustness test, we also performed a dynamic PCA with panel data. Please see the online material for more detail on this analysis. The result of the dynamic PCA model has a correlation with prespow2 of over 0.93.

[40] As part of this process, each measure of presidential power is standardized with mean 0 and variance 1.

intervals are reported in Table 3 in the online materials.[41] A selection of scores is provided in Appendix 1b of this article.

Discussion

We have generated a set of presidential power scores for a greater number of countries and country years than any existing dataset. By accounting for the idiosyncrasies of existing measures, we have maximized the reliability of our set of scores relative to any existing measure. By using publicly available measures, our method is replicable. Our scores also have the potential to be dynamic. Our method makes it easy to include new measures of presidential power and generate updated prespow1 and prespow2 scores. In fact, additional measures would be welcomed, as they will help to further increase the reliability of the scores. To be sure, if scholars wish to test a particular theoretical proposition about a certain aspect of presidential power, such as veto power or decree power, then they should construct their own measure and estimate its effect.[42] However, if scholars wish to examine the effect of presidential power generally, which has been the purpose of the vast majority of

[41] It is important to note that we calculate the confidence intervals on the basis of the original prespow 1 and 2 scores that generate our raw scores and not the final normalized 0-1 prespow 1 and 2 scores. This has the effect that the confidence intervals we report for a given country do not necessarily provide a measure of confidence relative to the prespow scores for any other country. Rather, they provide a measure of confidence in the level of concordance among the original set of scores for that country.

[42] Martínez-Gallardo 2012. We would like to thank Cecilia Martínez-Gallardo for drawing our attention to this important point.

studies to date, there is great benefit to be gained from the scores we have generated. With this aim in mind, two points should be emphasized.

First, for both of our measures standard errors and 95 per cent confidence intervals for each country year have been reported. This allows the basic reliability of any individual score to be identified, meaning that scholars can make an informed choice about whether or not to include a country in their estimation. For example, there are only two original scores for Cyprus (1960-) and both are very different. The normalized Hicken and Stoll score is 0.325, while for Shugart it is 1.[43] Cyprus is the only presidential system in Europe. Therefore, the relatively high scores for Cyprus in the Appendices might be considered to have good face validity (prespow1 = 0.64, prespow2 = 0.70). However, the Tables in the Appendix show that the confidence intervals for Cyprus are very large, reflecting the differences in the original measures. The way that the scores have been generated and reported gives scholars the opportunity to decide whether or not to include Cyprus in any estimation. Some may wish to include it because of what they might consider to be good face validity. Others may wish to exclude it because of the large confidence intervals. We make no recommendation, but we provide the grounds on which scholars can make an informed choice.

Second, we also provide the grounds on which scholars can decide which set of scores to use in comparative analysis. Figure 1 compares the range of standard errors for the prespow1 and prespow2 scores for the different regions. It suggests that the prespow2 scores increase the range of the standard errors for Latin America, but decrease it for both Africa and Asia. The effect on the scores for

[43] Hicken and Stoll 2008; Shugart 1996

presidents in European countries is minimal. This suggests that scholars wanting to estimate the effect of presidential power solely in Latin America might wish to use the prespow1 scores. However, scholars wanting to estimate the effect of presidential power solely in Africa might wish to use the prespow2 scores. Scholars who wanted to estimate the effect of presidential power across all regions might also wish to use the prespow2 scores, because on balance the reliability of the whole set of scores is probably slightly greater, even if the range of the standard errors in Latin America is increased. Again, we make no firm recommendation because the choice will be sensitive to the precise case selection with which the scholar is working. However, we provide information with which the scholar can make an informed decision.

Figure 1 about here

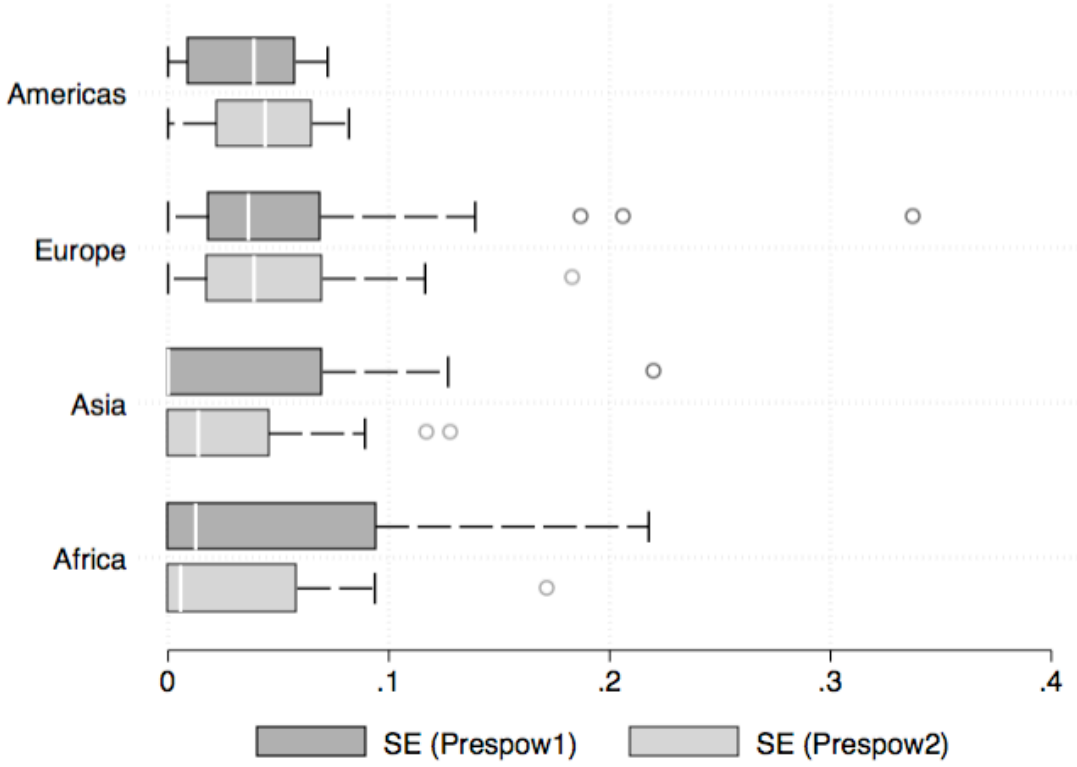
Conclusion

Studies have increasingly demonstrated that presidential power has an impact on a wide range of political outcomes. However, there are many separate measures of presidential power. By pooling the comparative and local knowledge present in 28 existing measures, we have generated a new set of presidential power scores for a larger number of countries and a longer time series than before. We have also maximized the reliability of these scores by deriving them solely from measures based on constitutional indicators of presidential power and by using a method that accounts for the idiosyncrasies of country scores in existing measures. In addition, by reporting the standard errors and the confidence intervals for all the country years in our measures, we have provided information with which scholars can make

an informed choice about whether or not a particular country should be included in an estimation and which of our measures should be used in comparative studies. Overall, we encourage people to keep developing new measures of presidential power and to update existing measures for as many countries and as long a time period as possible. The advantage of our approach is that new country scores can be easily incorporated into the method we have used, creating the potential for country coverage to be further extended, for existing country scores to be updated, and for cross-national measures to become even more reliable.[44]

[44] We will provide updated scores at our website www.presidential-power.com.

Figure 1: The Distribution of Standard Errors for prespow1 and prespow2



Note: The thick white lines within each box represent the median. The outside edges of the boxes represent the 25% quantile and the 75% quantile. The whiskers report outliers. The hollow circles represent data points 1.5 times outside of the interquartile range.

References

Biglaiser, Glen, and Karl DeRouen. 2004. The Expansion of Neoliberal Economic Reforms in Latin America. International Studies Quarterly 48: 561-578.

Cheibub, José Antonio. 2006. Presidentialism, Electoral Identifiability, and Budget Balances in Democratic Systems. American Political Science Review 100 (3): 353-368.

Chen, Haifeng. 2002. Principal Component Analysis with Missing Data and Outliers. Technical Report, Robust Image Understanding Laboratory.

Clark, Terry D., and Jill N. Wittrock. 2005. Presidentialism and the Effect of Electoral Law in Postcommunist Systems: Regime Type Matters. Comparative Political Studies 38 (2): 171-188.

Doyle, David. 2010. Politics and Privatization: Exogenous pressures, domestic incentives and state divestiture in Latin America. Journal of Public Policy 30 (3): 291-320.

Drukker, David M. 2003. Testing for Serial Correlation in linear panel-data models. The Stata Journal 3 (2): 168-177.

Fortin, Jessica. 2013. Measuring presidential powers: Some pitfalls of aggregate measurement. International Political Science Review 34 (1): 91-112.

Flury, B. 1988. Common Principal Components and Related Multivariate Methods, New York: John Wiley and Sons.

Frye, Timothy. 1997. A Politics of Institutional Choice: Post-Communist Presidencies. Comparative Political Studies 30 (5): 523-552.

Frye Timothy, Joshua S. Hellman, and J. Tucker 2000. Database on political institutions in the post-communist world. Unpublished manuscript, Columbia University, New York.

Hellman, Joel. 1996. Constitutions and Economic Reform in the Postcommunist Transitions. East European Constitutional Review 5: 46-56.

Hellman, Joel, and Joshua Tucker. 2000. Coding of Presidential Powers [unpublished database]. Available from http://www.wws.princeton.edu/~jtucker/pcelections.html. Accessed March 2005.

Hicken, Allen, and Heather Stoll. 2008. Electoral Rules and the Size of the Prize: How Political Institutions Shape Presidential Party Systems. Journal of Politics 70 (4): 1109-1127.

Johannsen, Lars 2003. IPA: The Index of Presidential Authority. Explorations into the measurement and impact of a political institution. Paper prepared for the ECPR Joint Sessions of Workshops Edinburgh, Mars 28–April 4, 2003.

Jolliffe, I. T. 2002. Principal Component Analysis, New York: Springer.

Martínez-Gallardo, Cecilia. 2012. Out of the Cabinet: What Drives Defections from the Government in Presidential Systems? Comparative Political Studies. 45 (1): 62-90.

McLachlan, Geoffrey and Thriyambakam Krishnan. 1997. The EM algorithm and extensions. New York: John Wiley and Sons.

Roper, Steven D. 2002. Are All Semipresidential Regimes the Same? A Comparison of Premier-Presidential Comparison Regimes. Comparative Politics 34 (3): 253-272.

Samuels, David J., and Matthew S. Shugart. 2003. Presidentialism, Elections and Representation. Journal of Theoretical Politics 15 (1): 33-60.

Shugart, Matthew S. 1996. Executive-Legislative Relations in Post-Communist Europe. Transition 13 December 1996: 6-11.

Shugart, Matthew S., and John M. Carey. 1992. Presidents and Assemblies: Constitutional Design and Electoral Dynamics. Cambridge: Cambridge University Press.

Siaroff, Alan. 2003. Comparative presidencies: The inadequacy of the presidential, semi-presidential and parliamentary distinction. European Journal of Political Research 42: 287-312.

Tavits, Margit. 2008. Presidents with Prime Ministers. Do Direct Elections Matter? Oxford: Oxford University Press.

Tavits, Margit. 2009. Direct Presidential Elections and Turnout in Parliamentary Contests. Political Research Quarterly 62 (1): 42-54.

Truxillo, C. 2005. Maximum likelihood parameter estimation with incomplete data. Proceedings of the Thirtieth Annual SAS Users Group International Conference, Philadelphia, PA.

Wansbeek, Tom, and Erik Meijer. 2000. Measurement Error and Latent Variables in Econometrics. Amsterdam, Elsevier.

Appendix 1a Sample of prespow1 scores

Country Year	Raw score	Standard errors	95% Confidence Intervals for raw scores		Normalized score
Albania 1991-1998	0.214	0.037	0.127	0.301	0.181
Albania 1998 - …	0.175	0.027	0.055	0.295	0.141
Argentina 1984-1994	0.317	0.032	-0.098	0.732	0.288
Argentina 1994 - …	0.430	0.049	0.294	0.567	0.407
Chile 1891-1925	0.400				0.375
Chile 1925-1969	0.500				0.479
Chile 1969-1973	0.542	0.058	-0.191	1.275	0.523
Chile 1989 - …	0.587	0.058	0.427	0.747	0.570
Cyprus 1960 - …	0.663	0.338	-3.626	4.951	0.648
Panama 1972 -	0.474	0.046	0.326	0.621	0.452
Romania 1992 - …	0.280	0.033	0.211	0.349	0.250
Zimbabwe 1991-2008	0.752	0.189	-0.061	1.565	0.742

Appendix 1b Sample of prespow2 scores

Country Year	Raw score	Standard errors	95% Confidence Intervals for raw scores		Normalized score
Albania 1991-1998	-0.181	0.038	-0.270	-0.091	0.215
Albania 1998 - …	-0.203	0.116	-0.704	0.298	0.185
Argentina 1984-1994	-0.069	0.078	-1.059	0.921	0.363
Argentina 1994 - …	0.052	0.047	-0.078	0.182	0.523
Chile 1891-1925	0.041				0.509
Chile 1925-1969	0.105				0.593
Chile 1969-1973	0.143	0.025	-0.169	0.456	0.644
Chile 1989 - …	0.293	0.069	0.102	0.485	0.843
Cyprus 1960 - …	0.183	0.184	-2.154	2.520	0.697
Panama 1972 -	0.137	0.033	0.031	0.243	0.636
Romania 1992 - …	0.006	0.034	-0.067	0.079	0.463
Zimbabwe 1991-2008	0.172	0.044	-0.019	0.362	0.682