Leadership or Luck?
An Analysis of World Leaders, Governors, and Mayors

Christopher R. Berry and Anthony Fowler
crberry@uchicago.edu; anthony.fowler@uchicago.edu
Harris School of Public Policy
University of Chicago

Abstract

Anecdotal evidence suggests that some political leaders are more effective than others, causing better outcomes for their citizens. However, observed differences between leaders could be attributable to chance variation rather than differences in ability. We develop a quantitative test of leader effects and implement it for world leaders, U.S. governors, and U.S. mayors and for several outcomes including economic growth and crime. We find clear evidence that world leaders matter for GDP growth and some evidence that U.S. governors matter for crime, but we also obtain several surprising null results. Our test can be applied to virtually any setting with leaders and an objective outcome of interest, so its continued application could improve our understanding of where, when, and why leaders matter.

The U.S. economy grew at an annual rate of over 6 percent during Harry Truman's second term, faster than under any other postwar president. George H. W. Bush presided over an economy growing at less than 1 percent on average during his second term, the worst record over the same period (see Blinder and Watson 2016). Violent crime in New York City fell by more than 56 percent in New York City in 1990's, under the administration of Mayor Rudolph Giuliani, compared to a decline of 28 percent nationwide (Corman and Mocan 2005). By contrast, under the leadership of Rahm Emmanuel, homicides have increased 70 percent since 2014 in Chicago, which was the only major U.S. city to experience an increase in murders in both 2015 and 2016 (Kennedy and Abt 2016; Sanburn 2016).

How much credit do leaders such as Truman and Giuliani deserve for the outcomes that happened on their watch? How much blame should fall at the feet of those like Bush and Emmanuel? Do the decisions and actions of leaders change the course of events, or are some merely (un)lucky, holding office at a time when other factors would have generated largely the same outcomes regardless of who sat behind their desk? This question—whether leaders matter—has fascinated scholars for centuries. Today, colleges and universities offer advanced degrees in leadership, and airport bookstores feature the latest bestsellers on the topic, implying a settled conclusion that leaders matter. Yet there has been relatively little in the way of rigorous empirical analysis to demonstrate the effects of leaders on outcomes of interest to social scientists.

In this paper we aim to make two contributions to the study of leadership. First, we introduce a new method for statistically testing leader effects, which has several advantages relative to other methods that have been used in the literature. Second, whereas the extant literature has focused on assessing the effects of national leaders on the economy, we extend our analysis to include subnational leaders and non-economic outcomes.

Related Literature

The subject of leadership has fascinated scholars at least back to the ancient Greeks. And for just as long, scholars have debated whether leaders matter in shaping outcomes for better and worse. Thucydides chronicled the great leaders, such as Pericles, whose particular decisions and abilities, in his view, determined the outcome of the Peloponnesian War. Meanwhile, Plato extoled the virtues of the philosopher-king while advocating a system of education and selection that would make any individual leader replaceable by another who would make similar decisions.

In the nineteenth century, Thomas Carlyle (1859) advanced the still-influential "great man" view of leadership with his declaration that, "The history of the world is but the biography of great men." But his contemporary, Karl Marx (1852), argued that historically determined social and economic forces constrain the choices available to leaders, making any individual relatively unimportant in the course of events.

More recently, the role of leadership has been overshadowed in political science and economics by the role of institutions in determining the fate of nations. Building on the work of Douglas North (1990), a dominant theme in the contemporary literature on economic development is that good institutions are the fundamental cause of long-run economic growth (e.g., Acemoglu et al. 2005), although empirically identifying the effects of institutions is not unproblematic (e.g., Glaeser et al. 2004).

While institutions have taken center stage in contemporary scholarship on economic development, leaders have not been entirely forgotten. An emerging formal literature has begun to tackle fundamental questions of the nature and source of leadership (see Levi and Ahlquist 2011). Notable contributions include Dewan and Squntani (2015) on relations between a leader and followers, Canes-Wrone (2006) on the distinction between leadership and pandering, and Dewan and Myatt (2008) on leadership and communication.

From an empirical perspective, a seminal paper by Jones and Olken (2005) rekindled interest in the study of leadership by providing the most credible evidence to date that political leaders matter. They use unexpected deaths of world leaders while in office as a source of exogenous variation in leadership. They show that the rate of economic growth in a country changes when a leader dies in office. Their key empirical test is whether the change in economic growth between the last two years of one leader's tenure and the first two years under the successor is greater than would be expected by chance. They find strong evidence of abnormal variation in growth around random leader transitions, which implies that leaders matter. Their estimates imply that leader quality can explain at least 1.5 percentage points of the variation in economic growth. They show further evidence that unexpected leader turnover leads to changes in economic growth in autocracies, but not in democracies.

The Jones and Olken findings have been extended in several ways by subsequent authors. Besley et al. (2011) find that educated leaders exert a positive effect on economic growth. Within the Jones-Olken framework, they show that a transition from a more educated to a less educated leader results in a reduction in economic growth, while a transition from a less educated to a more educated leader leads to a boost in economic growth. Yao and Zhang (2015) analyze the effects of city leaders in China on local economic growth, taking advantage of the fact that leaders regularly move, so that the same person may be mayor of multiple cities over the course of her career. Yao and Zhang use all leader transitions in their analysis rather than identifying unexpected transitions as in Jones and Olken, and they find mixed results depending on the statistical test they consider.

Easterly and Pennings (2016) challenge Jones and Olken's conclusion that leaders matter more in autocracies than in democracies. Specifically, Easterly and Pennings estimate leader fixed effects in a growth model and show that the variance of the fixed effects is at least as large in democracies as in autocracies. While there is more total variability in growth rates in autocracies, they contend that the amount of variation attributable to leaders is higher in democracies.

A Quantitative Test of Leader Effects

Our goal is to test whether leaders matter for particular outcomes of interest. The basic idea involves regressing the outcome on leader fixed effects, recording a summary statistic of fit, and then simulating the distribution of summary statistics that we would expect under the null. We start by converting each outcome variable to proportionate growth by subtracting by the lagged value and then dividing by the lagged value. This accounts for variation in the level of each outcome across geographic units and provides additional statistical precision. Next, we de-mean the data by year to remove consistent time trends across units.[1] This step is not necessary for identification but it provides additional statistical precision. Then, we identify each geographic unit and time period for which we have a complete data set with no missing years. We drop any unit-periods for which there was only one leader, since they provide no leverage for estimating the effects of leaders. Then, we code indicator variables for each individual leader and regress growth on these leader fixed effects. In cases where the same leader served in multiple unit-periods, we create a new indicator for each leader-unit-period. After running this regression, we record the r-squared statistic, indicating the proportion of variation in growth explained by the leader fixed effects.

In and of itself, this r-squared statistic is not particularly informative. A high value could reflect leader effects, but it could also reflect within-unit variation in growth unrelated to leader effects, or it could suggest that the regression with many independent variables over-fit random variation in growth. Therefore, we need a strategy for simulating the distribution of r-squared statistics that we would expect under the null hypothesis of no leader effects. To do this, we

[1] We could also de-mean by unit but this would have no impact on our subsequent p-values. Our inferential strategy implicitly accounts for variation in growth across units.

randomly permute the ordering of leaders within each unit-period, keeping the tenure of each leader the same as in the real data set but varying the order in which each leader served. For each random permutation, we again regress growth on leader fixed effects and record the r-squared statistic. We repeat this procedure many times and plot the distribution of statistics to estimate the expected distribution under the null of no leader effects. The proportion of random permutations that produce an r-squared statistic greater than that from the real data is an estimated p-value testing the null hypothesis of no leader effects.[2]

<u>Summary of our Quantitative Test of Leader Effects</u>

1. Recode outcome of interest as proportionate growth.
2. De-mean by year to remove consistent time trends across units.
3. Regress growth on leader fixed effects and record the r-squared statistic.
4. Randomly permute leaders within each unit-period, sampling each leader stint as a block.
5. Repeat #4 many times, recording the proportion of permutations yielding a higher r-squared.

The logic of these random permutation tests is as follows. There are three ways we can get a high r-squared statistic when we regress growth on leader indicators. First, there could be leader effects, and this is what we'd like to identify. Second, there could be serial correlation or genuine trends in growth over time within units even in the absence of leader effects, and the leaders who happened to serve in good times will get credit for this in the regression. Third, the leader fixed effects could be over-fit to random, year-to-year fluctuations in growth or even measurement error, further inflating the r-squared statistic. Therefore, in order to test for leader effects, we'd like our random permutation tests to incorporate the last two factors but not the first.

In our random permutations, the number of fixed effects in each regression is held constant, and the distribution of tenure across leaders is also held constant. This means that the extent of overfitting is the same in the real data and the permuted data. Furthermore, if there is serial

[2] These hypothesis tests are one sided because there is no reason to expect the real r-squared statistic to be smaller than the expected r-squared statistic under the null. If some leaders are better than others, this will only increase the value of the real r-squared statistic.

correlation or unit-specific time trends, this will inflate the r-squared from the permuted regressions in the same way, in expectation, as it inflates the r-squared from the real regression. In either case, some leaders might wrongly receive credit for good times. If the r-squared from the real data is much larger than that from the random permutations, this is an indication that ebbs and flows in growth coincide with the intervals of time in which different leaders served, suggesting that some portion of that r-squared statistic can be attributed to leader effects rather than just serial correlation or chance.

Our practice of sampling each leader as a block and maintaining the same distribution of contiguous periods of service in our permutations is important. To our knowledge, the approach in the literature closest to our own is a robustness check from Yao and Zhang (2015, p. 420). However, instead of sampling each leader as a block, Yao and Zhang sample each year independently such that leaders' terms are no longer contiguous in the random permutations. This approach accounts for the possibility of overfitting discussed above, but it does not account for the possibility of serial correlation or unit-specific time trends, and as a result, this test could reject the null even if there are no leader effects.

To fix ideas, consider the simplest possible example where our test would allow us to say something about leader effects. Suppose there is 1 unit with 2 leaders and 3 periods. Without loss of generality, suppose Leader A served during the first two periods, and Leader B served during the last period. In this simple example, there are only two ways to permute the leaders. We can assign Leader A to the first two periods—as in the real world, or we can assign her to the last two periods. If Leader A is better than Leader B, or vice versa, we would expect the growth values from the first two periods to be more similar to each other than they are to the value from the third period, and the real data will give a higher r-squared statistic. If there are no leader effects but there is random noise or serial correlation, either permutation is equally likely to give a higher r-squared.

This simple example illustrates several features and limitations our approach. First, identification comes from leaders who serve different periods of time. If there were 4 periods and each leader served two periods, both permutations would yield the same r-squared. Next, our procedure is poorly behaved when there are few leaders per period. In the example above, the p-value can only take values of 1 or .5, but we'd like to have a uniform distribution of p-values if there are no leader effects. Asymptotic refinement improves with more units and/or periods, so long as there is variation in lengths of service. Furthermore, our procedure requires that we put some structure on the timing of leader effects. Suppose a leader's policy decisions primarily affect growth in the next year. In that case, we would have no opportunity to detect leader effects in the simple example above. As a first pass, we will assume that leaders can only affect outcomes in the years in which they serve, but we can also conduct separate tests where we assume that their effects are lagged by one year, two years, etc. Lastly, our approach does not require us to hypothesize that one particular leader is better than another. For the purposes of this study, we are agnostic about which leaders are better; we test whether some leaders are different from others in ways that matter for various outcomes of interest.

Comparison with Other Methods

Our method resembles those used in other papers that rely on leader fixed effects in one way or another (e.g., Bertrand and Schoar 2003; Easterly and Pennings 2016). Our approach to inference is different, however. We do not assume that the r-squared from a set of leader-specific fixed effects will be zero when leaders do not actually affect the outcome. Because of random noise in the data, serial correlation, and unit-specific trends, leader fixed effects will contribute to r-squared even when leaders do not matter. Using the adjusted r-squared, as Bertrand and Schoar (2003) do, accounts for the problem of overfitting to random noise, but it does not address the other issues. Our method accounts for all these factors, without requiring additional assumptions about the nature of the serial correlation or the unit-level trends. In this sense, our approach is more conservative and will be less prone to detecting leader effects in cases where there are none.

A preview of one of our results illustrates the idea that the r-squared or adjusted r-squared statistics alone are insufficient for assessing leader effects. In a subsequent analysis, we estimate whether U.S. mayors affect employment in their cities. When we randomly permute the terms of service for various mayors, there should be no leader effects because terms of service are random. Nonetheless, the average r-squared statistic from our random permutations in this setting is .606, meaning 61 percent of the variation in employment across cities appears to be explained by leader fixed effects, even though the leader variables correspond to random times. The adjusted r-squared, which is designed to mitigate overfitting is still .525. If we first de-mean by city, the r-squared and adjusted r-squared statistics are .343 and .209, respectively. Clearly, none of these statistics can separate leader effects from other forces because the numbers are large even when there are no leader effects. As we'll see, although the r-squared and adjusted r-squared statistics are large when we regress city employment on mayor fixed effects, they are no larger than we would expect by chance if mayors do not affect employment.

Our approach also offers some advantages relative to the method of Jones and Olken (2005). They use a similarly nonparametric inferential approach, comparing the changes in economic growth in periods with leadership transitions to the distribution of changes in periods when there are no transitions. However, their analysis only includes leader transitions arising from unexpected deaths—i.e., those due to accident of illness in office. While unexpected deaths provide plausibly exogenous changes in leadership, identification comes from a relatively small number of leader transitions. Specifically, they have only 57 such unexpected transitions from a panel of 130 countries since 1945. Our approach utilizes information from virtually all leaders and all countries.

A concern with focusing on unexpected leader deaths is that the resulting changes in economic growth may also reflect whatever disruptive effects happen as the result of an unanticipated transition of power. To the extent that unexpectedly changing leaders is disruptive to the government or economy, this disruption will be reflected in the Jones and Olken estimates. While they take measures to reduced the possibility that disruption contaminates their results—such as excluding the first year or two after the transition—some concerns remain. Besley et al. (2011) show that the average change in growth following an unexpected leader transition is negative. In particular, they show that there is a 0.2 percentage point reduction in annual growth during the 5 years after an unexpected transition in leadership. Absent some disruptive effects, it is hard to see why the average effect of random leader transitions should be negative and statistically significant.

Our method also offers a purely practical advantage in terms of its generalizability. The Jones and Olken method requires not only identifying leader transitions due to death in office, but also knowledge of the cause of the leader's death. While uncovering such information may be feasible for world leaders, it may be impractical or impossible in the case of more obscure leaders serving in less prominent offices. Even for the large U.S. cities that we study in this paper, we sometimes had difficulty finding the names of the mayors who served in the past. We suspect that finding detailed information about their cause of death would require heroic effort. Fortunately, our method does not require additional information beyond the leaders' terms of service, and so should be more applicable to studying a wide range of leaders.

We also note some limitations of our method, which are shared in common with others in the literature. Importantly, we cannot saying anything about the performance of any individual leader or make direct comparisons of two leaders. We are testing a particular notion of what it means for leaders to matter, that variation in outcomes across leaders is greater than would be expected by chance. Our method enables us to assess whether the variation due to leaders is statistically significant in this sense, but not whether any individual leader's effect is significant.

Alternatively, we can think of our approach as testing the sharp null hypothesis that all leaders are equal, with respect to a particular outcome of interest. We conclude that leaders matter when we can reject the hypothesis that all leaders are equal. One reason we might fail to reject the null is if selection reduces the variation among those who become leaders. For instance, electoral competition may result in only leaders above a certain level of quality being elected, or only leaders who share some common set of preferences and beliefs. If these leaders perform equally well, there will be no leader effects according to our definition. But such a null finding would not imply that replacing a sitting leader with a random person from the population would not matter.

Monte Carlos Assessing the Properties of the Test

To assess the properties of our quantitative test of leader effects, we have conducted a series of Monte Carlo simulations. First, we show that if there are no leader effects, our test will not detect them. Under the null hypothesis, p-values should be uniformly distributed between 0 and 1, and this is exactly what we find so long as there are enough time periods within each unit. We simulate data sets with a certain number of units and time periods. For our initial simulations, we assume that each leader's tenure is randomly drawn uniformly from integers between 1 and 5. We simulate an extra 5 periods for each unit and remove the first 5 periods. This ensures that the simulated data set starts in the middle of some leaders' tenures, making the simulations more similar to our subsequent analyses with real data. Growth in the first period is drawn independently from a standard normal distribution, and growth in each subsequent period is growth in the previous period plus random variation drawn from a standard normal distribution. This means there is random variation in growth from year to year and there is also serial correlation over time within each unit, but growth is unrelated to leaders.

In Figure 1, we vary the number of units and periods in the simulated data set in order to assess the performance of our test across these two factors. For every combination of units and periods, we generate 1,000 different simulated data sets. For each simulated data set, we implement our procedure and estimate a p-value from 19 random permutations. This means there are 20 possible values our p-value estimate can take, so if the procedure performs well, we should see each p-value 5 percent of the time. Figure 1 allows the number of units and number of periods to be 5, 10, or 20, showing all 9 possible combinations. Even with only 5 units and 5 periods, the distribution of p-values is uniform. These simulations confirm that random noise and serial correlation do not contaminate the results of our test, and our p-values are reliable even with a small number of units and periods. Also notice that our procedure never requires the researcher to specify the nature of serial correlation in the data. If growth is unrelated to leaders' tenures, our test will not wrongly detect leader effects.

More to come here . . .

Applications to World Leaders, Governors, and Mayors

We examine the effects of three types of leaders on economic growth: national leaders; U.S. governors; and mayors of the top 100 U.S. cities. For mayors and governors, we also estimate leader effects on crime rates. We utilize data on world leaders from Archigos (Goemans, Gleditch, and Chiozza 2009) version 4.1, and we collected data on U.S. governors and U.S. mayors from various sources. Our data sets indicate which leader held each leadership position in each year. In cases where multiple leaders served in the same position in the same year, we record which leader held the position for the longest period of time. We utilize data on GDP by country and year from the

Maddison Project (Bolt and van Zanden 2014). Data on U.S. state- and county-level income per capita and employment come from the Bureau of Economic Analysis. For U.S. cities, we focus on the 100 largest cities by 2015 population. Because income and employment estimates are not available by city, we match each city to its home county and drop the two cities that are not the largest in their county (Long Beach, CA and Mesa, AZ). Data on state- and city-level per capita property crime and violent crime come from the Federal Bureau of Investigation.

<u>World Leaders</u>

Following the methodology described above, we analyze the effects of world leaders by regressing economic growth, after de-meaning the data by year, on a set of leader dummies. As shown in Table 1, the r-squared from this regression is .286. It would be a mistake, however, to conclude that the r-squared is due to leader effects. Even in the absence of any leader effects, we would expect this regression to have some explanatory power because the leader dummies will pick up some random variation in the data as well as serial correlation within countries over time. Indeed, we see from Table 1 that the average r-squared from the placebo regressions is .251. In other words, even a randomly permuted set of leader dummies can "explain" about 25 percent of the variation in economic growth. Genuine leader effects, according to our method, are reflected in the *difference* between the placebo r-squared and the r-squared from the actual data. In this case, that difference is .035 percent, which is to say leader effects explain roughly 3.5 percent of the variation in economic growth. This difference is highly significant. In fact, the r-squared from the regression on the actual data is larger than the r-squared from all 1,000 regressions on permuted data sets, as shown in Figure 2.

We next analyze the difference in leader effects for democracies and autocracies, as classified according to Polity IV scores. Jones and Olken (2005) classify countries according to their status at the time of a leader transition. Easterly and Pennings (2016), on the other hand, classify counties according to their average policy scores over time. Relying on average scores, however, means that countries will have a constant classification in the analysis even if they in fact transition from autocracy to democracy, or vice versa, during the study period. To overcome this problem, we divide countries into three categories: those that were always democracies, those that were always autocracies, and those that changed their status at least once over time. 31 country-periods in our data were always autocracies, 42 were always democracies, and 90 transitioned at some point. We then conducted analyses separately for each set of countries using our method.

We find evidence of leader effects in each case, although the effects are only significant at conventional levels for transitional countries ($p < .01$). Effects for democracies ($p = .075$) and autocracies ($p = .12$) are marginally significant. We are hesitant to make too much of these differences in statistical significance given the accompanying differences in the sample size for each category. The magnitude of the leader effects—which we measure as the difference in the r-squared between the regression on the actual data and the average placebo regression—is greatest in democracies and smallest in autocracies, although the differences in magnitudes across categories are not statistically distinguishable. Like Easterly and Pennings (2016), then, we fail to find support for Jones and Olken's (2005) conclusion that leader effects are greater in autocracies. If anything, leader effects are as large or larger in democracies and transitional countries as in autocracies. Furthermore, the smaller r-sqaured for autocracies even in the placebo regressions is consistent with the finding from Easterly and Pennings that random variation in growth is greater in these countries.

U.S. Governors

We extend our analysis to governors following the same basic methodology. We have data on state level personal income from the Bureau of Economic Analysis starting in 1930, and for employment starting in 1970. While these panels are shorter than we had for world leaders, our Monte Carlo simulations make us confident that our method can still detect leader effects if they are present (formal power analyses will appear in a later draft). As Table 2 and Figure 3 show, however, there is not much evidence of leader effects on these outcomes. Indeed, in the case of employment, the difference in r-squared between the actual and placebo regressions is actually negative.

Perhaps we should not be surprised to find that governors do not affect economic outcomes. Jones and Olken (2005) find substantial effects of world leaders on monetary policy but only weak evidence for fiscal policy. Governors may not have access to important levers, like monetary policy, that drive economic outcomes. If so, then we have been looking in the wrong place for evidence that governors matter. We should look at outcomes that are potentially shaped more directly through state-level actions. Pursuing this idea, we next examine the effects of governors on crime rates. Using data from the FBI's Uniform Crime Reporting database from 1961 to 2012, we analyze the effects of governors on property and violent crime rates. We find significant governor effects on property crime rates ($p = .04$), and marginally significant effects for violent crime ($p = .12$).

<u>Mayors of the Top 100 U.S. Cities</u>

We next study the effects of mayors on the same outcomes. We focus our analysis on the 100 largest U.S. cities according to 2015 population. For each city, we collected data on the names and service dates of mayors dating back at least to 1970, when our data on local income and employment begin. At the time of this writing, we have complete mayoral data for 70 of the top 100 cities, and we will continue to compile data for the remainder. An important caveat is that annual economic data for U.S. cities are not available. Instead, we use county-level economic estimates from the BEA. As discussed above, we match each city to its county and drop the two cities in our data set that are not the largest city in their county. We can also weight the regressions according to the city's share of county population, which is perfectly allowable within our methodology and has no impact on our subsequent results.

Can mayors reasonably be expected to affect economic growth in their cities? The existing literature offers differing perspectives. One school of thought argues that Tiebout (1956) competition forces mayors to single-mindedly pursue economic development (Peterson 1981), with some going so far as to argue that cities are governed as "growth machines" that pursue development at the expense of all other priorities (Logan and Molotch 1987). Others contend that mayors are relatively weak executives that lack the power to control basic service delivery, much less to drive economic growth in their cities (Yates 1977). We find little evidence of mayoral effects on income and employment in their counties, as shown in Table 3 and Figure 4. In both cases, mayoral fixed effects appear to explain a substantial amount of the variation in the outcome, but we see that the placebo dummies "explain" just as much, on average.

We next turn to an analysis of mayoral effects on crime, an outcome over which we might expect mayors to have greater influence. After all, mayors directly appoint police chiefs and shape law enforcement policy within their jurisdictions. Nevertheless, we find no evidence that mayors affect either property or violent crime rates in their cities.

Our results reveal no evidence of mayoral effects for two of the most important outcomes in a city, the economy and crime rates. Such results are generally consistent with the argument that mayors simply lack control over governance and service provision within their jurisdictions (e.g., Yates 1977). In future analyses, we plan to explore the variation in mayoral powers across cities—particularly the distinction between strong and weak mayor forms of government—to discover whether there is greater evidence of mayor effects in cities where the mayor is a more powerful position in city government.

Discussion and Conclusion

We have proposed a new approach for estimating leader effects. The primary innovation relies on random permutations for inference. Relative to methods previously used in the literature, ours makes use of more variation in the data and relies on fewer and weaker assumptions. In Monte Carlo simulations, our method performs well even in relatively small samples. In addition, our method does not require knowledge of the particular circumstances surrounding leader transitions or the cause of leader deaths in office, making our approach easier to generalize to subnational or other less prominent offices where leader biographies are unlikely to be available.

Applying our method, we show that world leaders matter for economic growth, consistent with prior research in the same tradition. Our results, however, challenge the conclusion that leaders matter more in autocracies than in democracies; if anything, our results point in the opposite direction. We also show that leaders matter most in countries that transition between autocracy and democracy, which is to our knowledge a novel result.

We also apply our method to settings where leader effects had not previously been estimated. We find no evidence that U.S. governors and mayors affect income and employment in their jurisdictions. We do find that governors, but not mayors, influence crime rates. While the extant literature has been focused on the economy, the latter results suggest the importance of matching different offices to relevant outcomes when estimating leader effects.

This paper is part of a larger, ongoing project on leadership. We are in the process of collecting data for a wide range of political and non-political leadership positions, and hope, through applying our method in different contexts, to improve our general understanding of whether leaders matter and why.

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2005. Institutions as the Fundamental Cause of Long-Run Growth. *Handbook of Economic Growth*, Philippe Aghion, and Steven N. Durlauf, eds., Amsterdam.

Bertrand, Marianne, and Antoinette Schoar. 2003. Managing with Style: The Effect of Managers on Firm Policies. *Quarterly Journal of Economics* 118(4):1169-1208

Timothy Besley, Jose G. Montalvo and Marta Reynal-Querol. 2011. Do Educated Leaders Matter? *The Economic Journal* 121(554):F205-227.

Bolt, Jutta and Jan Luiten van Zanden. 2014. The Maddison Project: Collaborative Research on Historical National Accounts. *Economic History Review* 67(3):627-651.

Canes-Wrone, Brandice. 2006. *Who Leads Whom? Presidents, Policy, and the Public.* Chicago University Press, Chicago.

Carlyle, Thomas. 1859. *On Heroes, Hero Worship and the Heroic in History.* Wiley and Halsted, New York.

Corman, Hope, and Naci Mocan. 2005. Carrots, Sticks, and Broken Windows. *Journal of Law and Economics* 48(1):235-266.

Dewan, Torun, and David Myatt. 2008. The Qualities of Leadership: Communication, Direction and Obfuscation. *American Political Science Review* 102(3):351–368.

Dewan, Torun, and Francesco Squintani. 2015. On Good Leaders and Their Associates. Working paper <bit.ly/2kbHxn3>.

Easterly, William, and Steven Pennings. 2016. Shrinking dictators: how much economic growth can we attribute to national leaders? Working paper <bit.ly/2lz8TAs>.

Glaeser, Edward L., Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2004. Do Institutions Cause Growth? *Journal of Economic Growth* 9(3):271-303.

Goemans, H.E., Kristian Skrede Gleditch, and Giacomo Chiozza. 2009. Introducing *Archigos*: A Data Set of Political Leaders, 1875-2003. *Journal of Peace Research* 46(2):269-283.

Jones, Benjamin F., and Benjamin A. Olken, (2005) Do Leaders Matter? National Leadership and Growth since World War II. *Quarterly Journal of Economics* 120(3):835-864.

Kennedy, Sean, and Parker Abt. 2016. "Trump is right about violent crime: It's on the rise in major cities." *The Washington Post*, August 5.

Levi, Margaret, and Josh Ahlquist. 2011. Leadership: What it means, what it does, and what we want to know about it. *Annual Review of Political Science* 14:1-24.

Logan, John, and Harvey Molotch. 1987. *Urban Fortunes: The Political Economy of Place*. University of California Press, Berkeley.

Marx, Karl. 1852. The Eighteenth Brumaire of Louis Napoleon. *Die Revolution*, New York.

North, Douglas. 1990. *Institutions, Institutional Change, and Economic Performance.* Cambridge University Press, Cambridge.

Peterson, Paul. 1981. *City Limits*. University of Chicago Press, Chicago.

Sanburn, Josh. 2016. Chicago Is Responsible for Almost Half of the Increase in U.S. Homicides. *Time Magazine*, September 19.

Tiebout, Charles. 1956. A Pure Theory of Local Government Expenditures. *Journal of Political Economy* 64(5):416-424.

Yao, Yang and Muyang Zhang. 2015. Subnational Leaders and Economic Growth: Evidence from Chinese Cities. *Journal of Economic Growth* 20(4):405-436.

Yates, Douglas. 1977. *The Ungovernable City: The Politics of Urban Problems and Policy Making*. Cambridge, MA: MIT Press.

Table 1. Results for World Leaders and GDP, 1876-2010

subset	r^2	$E[r^2 \mid null]$	difference	p-value
All Countries	.286	.251	.035	.000
Autocracies	.175	.153	.022	.120
Democracies	.441	.394	.047	.075
Transitional	.274	.240	.034	.009

Table 2. Results for U.S. Governors

outcome	years	r^2	$E[r^2 \mid null]$	difference	p-value
Income	1930-2015	.166	.154	.013	.219
Employment	1970-2015	.450	.500	-.051	.994
Property Crime	1961-2012	.182	.167	.016	.044
Violent Crime	1961-2012	.173	.159	.014	.120

Table 3. Results for U.S. Mayors

outcome	years	r^2	$E[r^2 \mid null]$	difference	p-value
Income	1970-2015	.174	.183	-.009	.767
Employment	1970-2015	.608	.606	.002	.413
Property Crime	1986-2012	.191	.198	-.007	.743
Violent Crime	1986-2012	.188	.193	-.005	.683

Figure 1. Monte Carlos with Random Noise, Serial Correlation, but No Leader Effects

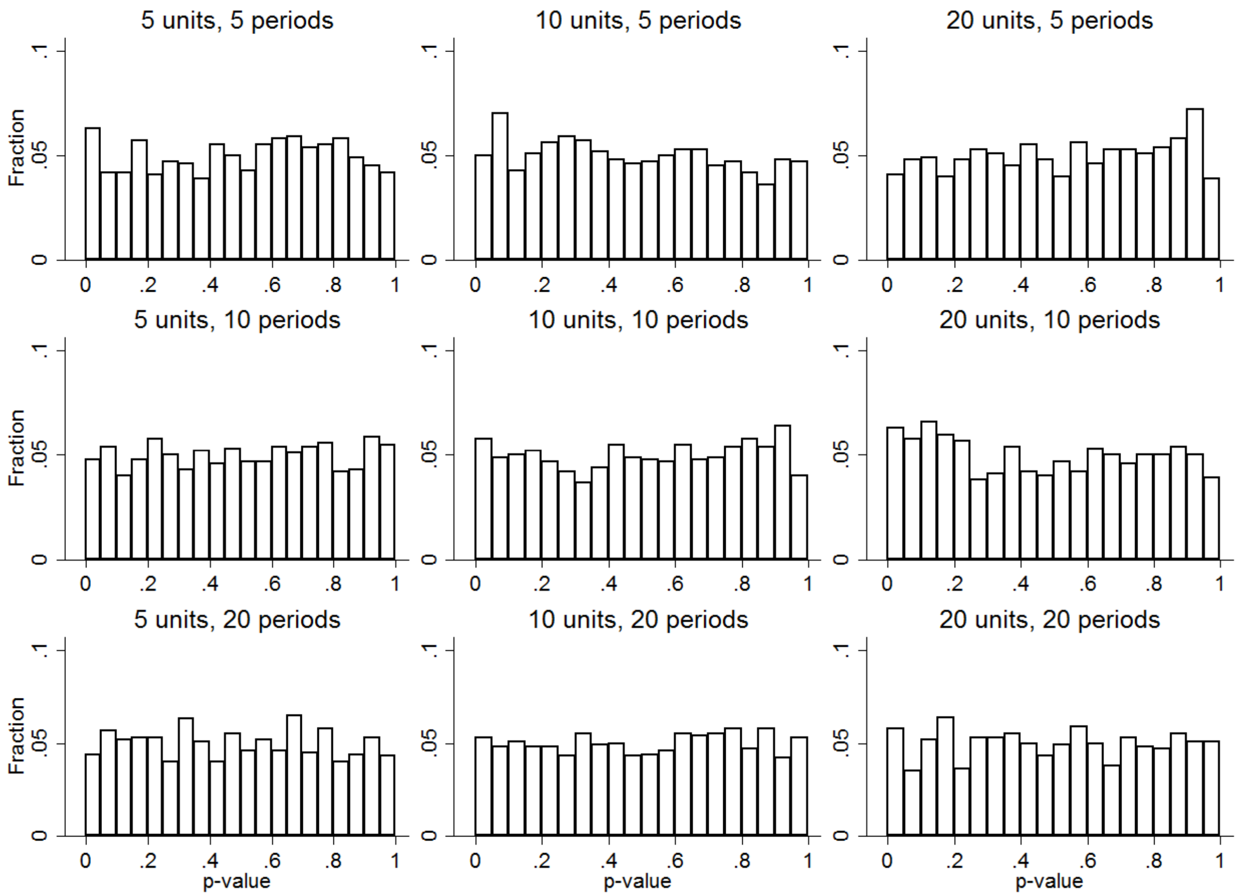

Figure 2. Simulated Distributions of r^2 under Null, World Leaders

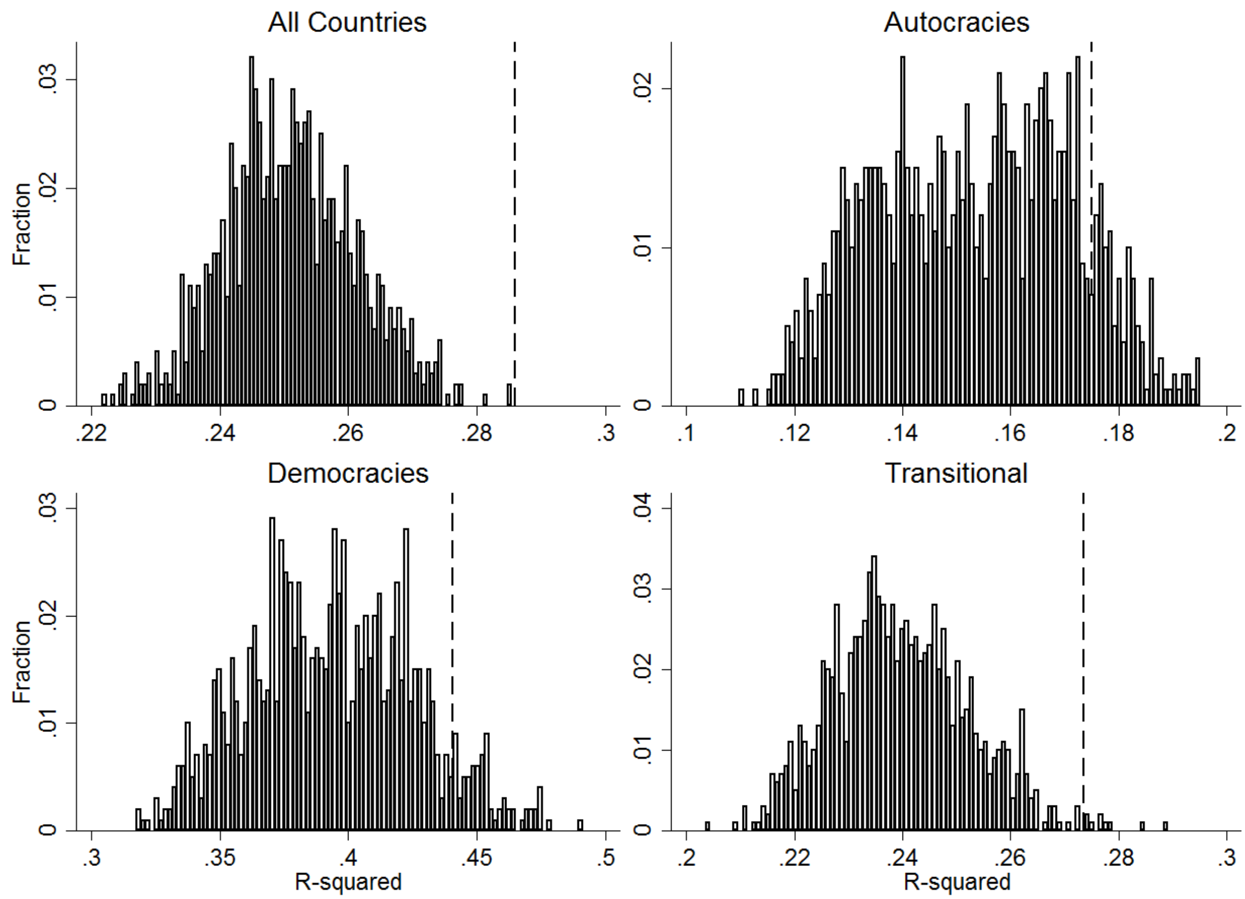

Figure 3. Simulated Distributions of r^2 under Null, U.S. Governors

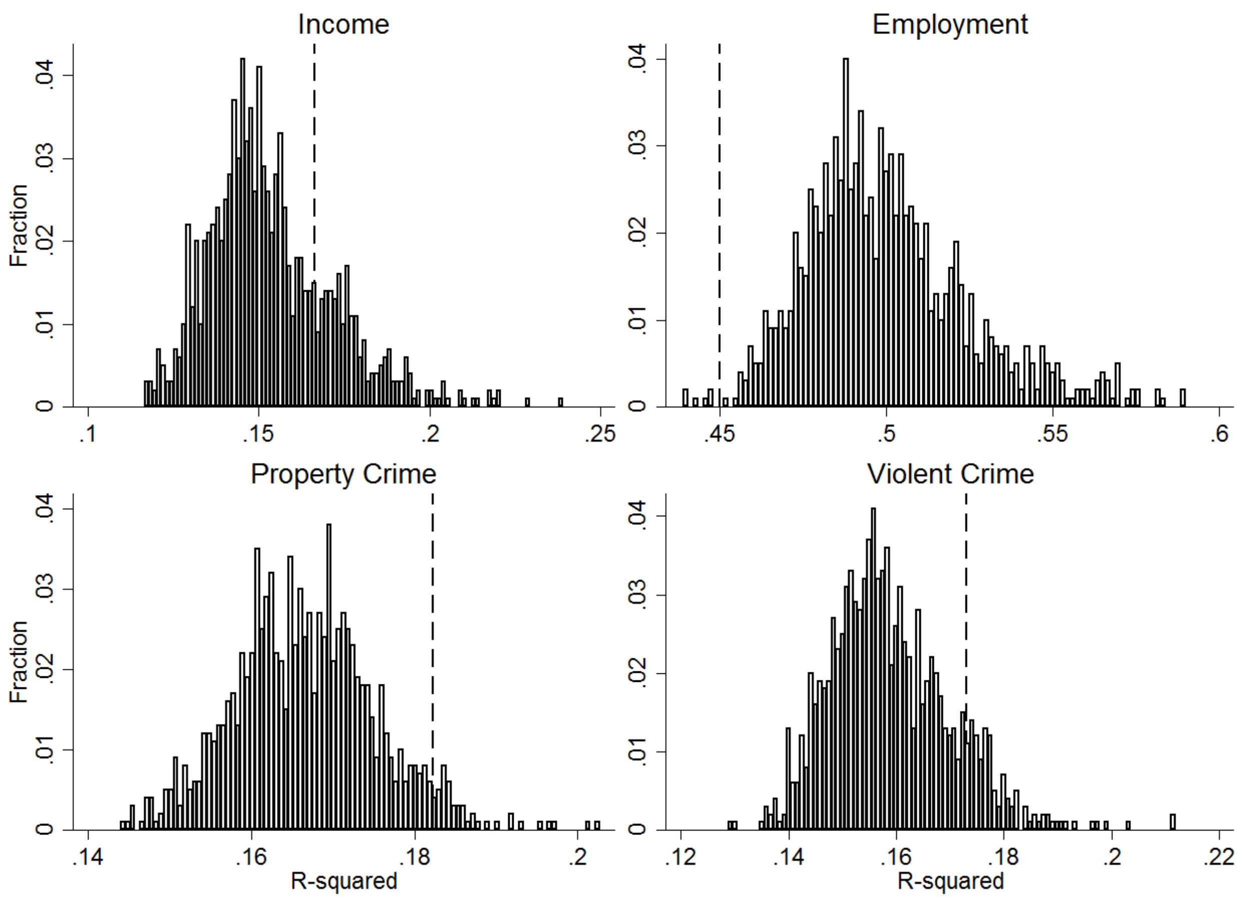

Figure 4. Simulated Distributions of r^2 under Null, U.S. Mayors