

How oil income and missing hydrocarbon rents data influence autocratic survival: A response to Lucas and Richter (2016)

Research and Politics
July–September 2017: 1–6
© The Author(s) 2017
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2053168017719794
journals.sagepub.com/home/rap

Joseph Wright¹ and Erica Frantz²

Abstract

This paper re-examines the findings from a recently published study on hydrocarbon rents and autocratic survival by Lucas and Richter (LR hereafter). LR introduce a new data set on hydrocarbon rents and use it to examine the link between oil income and autocratic survival. Employing a placebo test, we show that the authors' strategy for dealing with missingness in the new hydrocarbon rents data set – filling in missing data with zeros – creates bias in the reported estimates of interest. Addressing missingness with multiple imputation shows that the LR findings linking oil rents to democratization do not hold. Instead, we find that hydrocarbon rents reduce the chances of transition to a new dictatorship, consistent with the conclusions of Wright et al.

Keywords

Oil rents, autocratic stability, multiple imputation

Introduction

In “Oil and autocratic regime survival” (*British Journal of Political Science*, 2015), Wright, Frantz and Geddes (2015) (WFG hereafter) identify a new mechanism linking natural resource wealth to autocratic regime survival. To do so, they evaluated the impact of resource wealth in dictatorships on democratization, transition to subsequent dictatorship, or both, looking at within country changes over time and cross-country differences: increases in resource wealth improve the survival of dictatorships primarily by lowering the risk of ouster by rival autocratic groups. Lucas and Richter (2016) (LR hereafter) re-evaluate these empirical tests using new data on resource wealth from the Global State Revenues and Expenditures (GSRE) data set.

The new data leverage recently released budget data collected from the International Monetary Fund to measure government income from resources, while accounting for a range of factors, such as the cost of resource production and taxation income from resource corporations. LR re-estimate some WFG empirical models with the GSRE measure of resource rents. Like WFG, LR find strong evidence linking greater resource wealth to more durable dictatorships. They find little evidence in support of WFG's

central finding, however, that increases in resource wealth within dictatorships over time lower the risk of transition to a new dictatorship. Instead, LR conclude that resource wealth bolsters dictatorships by hindering their chance of democratization. LR attribute the differences in findings to the use of the GSRE resource rents data. In sum, the central message from the LR analysis is that once more accurate and widely reported resource rent data are used, WFG's key finding disappears.

In this response, we do not challenge the quality of the GSRE data set nor delve into how it compares to existing data on natural resource wealth.¹ The GSRE data set offers scholars a more fine-grained picture of government income from resources and enables more detailed analyses of a variety of dynamics in the resource curse literature.

¹Department of Political Science, Pennsylvania State University, University Park, PA, USA

²Department of Political Science, Michigan State University, East Lansing, MI, USA

Corresponding author:

Joseph Wright, Department of Political Science, Pennsylvania State University, 203 Pond Lab, University Park, PA 16802, USA.

Email: josephwright@gmail.com



The empirical strategy LR use to deal with missingness in the GSRE resource rents data, however, may be problematic. We illustrate how a potentially flawed approach explains the differences in the results that LR obtain. Once a less biased method of addressing missingness is used, we find that the central WFG result holds even with the GSRE data: increases in resource wealth within dictatorships over time lower the risk that they will transition to new dictatorships.

In what follows, we explain the LR strategy for accounting for missing observations in the GSRE data and show empirically how it distorts our understanding of the relationship between autocratic regime survival and resource wealth.

The GSRE data and missingness

The GSRE data set offers 39 unique indicators that cover many aspects of state finances for 161 countries from 1946 to 2006. LR use these data to construct a measure of resource rents (which they will refer to as *GSRE rents*). Though LR assert that one of the positives of the GSRE data set is increased data availability for non-democratic regimes (Lucas and Richter, 2016: 2), 84% of the observations in the Geddes, Wright and Frantz (2014) (GWF hereafter) sample of dictatorships that they use in their analysis are missing *GSRE rents* data. In the original study, by contrast, WFG measured resource rents using data from Haber and Menaldo (2011) (which we will refer to as *HM rents*). Less than 3% of the observations in the GWF sample of dictatorships with data on level of development are missing *HM rents* data.²

To increase coverage of *GSRE rents* in the sample, LR add 0s to *GSRE rents* if two conditions occur: (1) the observation has a value of 0 in *HM rents*; and (2) the observation is missing in *GSRE rents*. LR write that this decision is “appropriate because zero hydrocarbon income is not randomly distributed across the population of countries since it exists only for those countries without known oil and gas production from which state income could be generated” (Lucas and Richter, 2016: 8).

Using *GSRE rents* with 0s added to measure resource rents, LR estimate all of WFG’s main models. Their findings contradict WFG’s in two key ways. LR find that: (1) increases in resource wealth within dictatorships over time do not affect the chance of a transition to new dictatorship; and (2) increases in resource wealth within dictatorships over time lower the chance of democratization.

Lucas and Richter acknowledge that using *GSRE rents* with 0s added to capture resource rents instead of *HM rents* reduces the sample size (from 4,138 observations to 2,731). To evaluate whether the particular subsample of 2,731 observations is driving their results, LR estimate the relationship between resource rents and autocratic regime survival on this subsample using *HM rents*, as well. The results from this exercise are consistent with WFG’s

results, except that the within-country coefficient of *HM rents* loses statistical significance in the model of autocratic-to-autocratic transitions. Because of this difference, LR assess that it “remains unclear” (Lucas and Richter, 2016: 4) whether the non-finding for the autocrat-to-autocratic model is due to the superiority of the *GSRE rents* data or the reduced sample size. With the democratic transition model, by contrast, the results are the same as WFG’s even in the subsample, suggesting (according to LR) that “the negative significant relationship between GSRE hydrocarbon rents and the onset of democratic transition relates back to the new GSRE data on hydrocarbon rents” (Lucas and Richter, 2016: 4).

The primary reason that LR’s findings differ from WFG’s, however, is not because the values of the GSRE data differ from the HM data, but rather because of the bias introduced by LR’s strategy for dealing with missingness in *GSRE rents*. By only filling in missing values for dictatorships that have 0 resource income, LR’s sample – by construction – over-represents dictatorships without resource wealth and under-represents those with it. This is problematic because the dictatorships excluded from the sample – which have resource income according to the *HM rents* measure but for which there are no GSRE data – may be systematically different from the other resource-rich dictatorships in the sample in their propensity to transition to new dictatorship, democratize, or both. This means that regardless of whether *GSRE rents* with 0s added or *HM rents* is used to measure resource wealth, the estimates may be biased.

To alleviate the concern that adding 0s to *GSRE rents* is driving their results, LR estimate the main models using *GSRE rents* without 0 values and their results hold (Lucas and Richter, 2016: 4). Yet, this does little to tackle the underlying issue if the resource-rich dictatorships excluded from the sample have different baseline rates of regime transition than those that are in it.

Dissecting the LR analysis

Next, we compare features of the observations LR use in their analysis with those that they exclude from it. If the latter are systematically more/less likely to have resource wealth and more/less likely to transition to new dictatorship or democracy, then the results in the LR study may be biased.

There are three subsets of observations from the sample used in the original WFG study:

1. Observations with non-missing *GSRE rents* data (*s1*); the white space in Figure 1 (678 observations; 16% of the sample);
2. Observations with missing *GSRE rents* data and 0 values in *HM rents* (the observations LR filled in with 0 values) (*s2*); the red space in Figure 1 (2,053 observations; 50% of sample); and

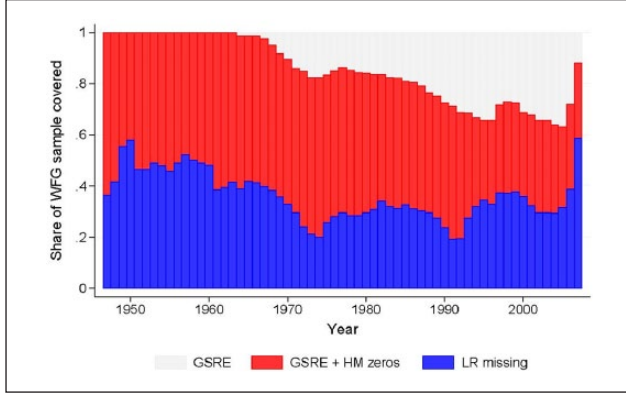


Figure 1. Data coverage in the Lucas–Richter analysis.

3. Observations with missing *GSRE rents* data and positive values in *HM rents* (*s3*); the blue space in Figure 1 (1,407 observations; 34% of sample).

These three subsets of observations – *s1*, *s2*, and *s3* – do not overlap with one another; and together they comprise the full sample of 4,138 observations in the original study. Together, *s1* \cup *s2* make up the sample of 2,731 observations that LR use in theirs (i.e., *GSRE rents with zeros added*).

As Figure 1 illustrates, there is very little white space, indicating that the bulk of the observations in the WFG sample are missing in *s1*. Though adding *s2* to the sample increases the sample size considerably, it is still missing data for over a third of observations throughout most of the period. Looking at the blue bars, it is clear that *s3* – the set of observations excluded from LR’s analysis – comprises a sizable portion of the WFG sample during most years.

There are important trends over time in the pattern of missingness that are worth pointing out. Prior to the start of the third wave of democratization in 1974, roughly 40% of the sample used in WFG is missing in *s2*. During this time (1947–1973), the chance of an autocratic transition was 3.4%, more than double the chance of a democratic transition (1.5%).³ The post-1973 era, by contrast, saw the risk of democratic transition outpace the risk of autocratic transition (2.6% versus 2.0%). These trends are troubling because the pre-1974 period is also one with relatively low oil income, which rose substantially for many dictatorships after the oil price shocks of the early 1970s (Ross, 2015). This suggests that it is possible that the subset of observations LR exclude from their analysis (*s3*) varies in meaningful ways from the subset that they use (*s1* \cup *s2*).

To evaluate this, we examine whether *s1*, *s2* and *s3* differ systematically in the values of the main explanatory variable and the dependent variables in our analysis. The first three columns of Table 1 show basic coverage statistics for each of the three subsets of the WFG sample. Columns

3–5 show the mean values for *HM rents* (denoted as *HMrents*), and the autocratic and democratic transition

probabilities. These reveal that *s1* comprises a high resource income but low transition probability part of the WFG sample, while *s2* represents an extremely low resource income (i.e., no resource income) but higher transition probability part. Compared to the excluded part of the WFG sample (*s3*), the LR sample (*s1* \cup *s2*) has: (1) lower resource rents (1.4 log units versus 3.1 log units); (2) a lower democratic transition probability (2.0 versus 2.8); and (3) a higher autocratic transition probability (2.6 versus 2.2). These statistics suggest that excluding *s3* from the sample may bias results because the observations in *s3* have, on average, higher resource rents, a higher incidence of democratic transition, and a lower incidence of autocratic transition.

In the Online Appendix, we examine the correlates of the missing observations and find that, even after controlling for covariates and time trends, *HM rents* is systematically higher in the part of the WFG sample excluded from the LR analysis (*s3*). This is unsurprising because the LR strategy for dealing with missing *GSRE rents* adds back into the estimating sample those observations with zero *HM rents*, thus only treating as missing those observations that have positive oil and gas income.

Re-analysis of the LR models

To aid our re-analysis of the LR models, let us first define the variables used throughout. *LRrents* is *GSRE rents* with 0s added if *GSRE rents* is missing and *HM rents* is zero. *LRrents* has non-missing observations for *s1* \cup *s2*. *HMrents* is simply the original HM data on oil and gas income used by WFG; these data have coverage over *s1* \cup *s2* \cup *s3*. *Nrents* is the same as *LRrents*, except it uses HM data instead of GSRE data for *s1*. *Nrents* has non-missing observations for *s1* \cup *s2*. Thus, *Nrents* employs the LR procedure for filling in missing observations and truncating the sample using HM data. *Nrents* differs from *HMrents* only in its coverage (*Nrents* excludes *s3*). This means that *Nrents* and *HMrents* will have different values when calculating the country means and deviations for the empirical application.

Following WFG and LR, we use these three variables – *LRrents*, *HMrents*, and *Nrents* – to construct the country-mean values and the deviations from those means. The model specification used throughout, again following WFG and LR, is:

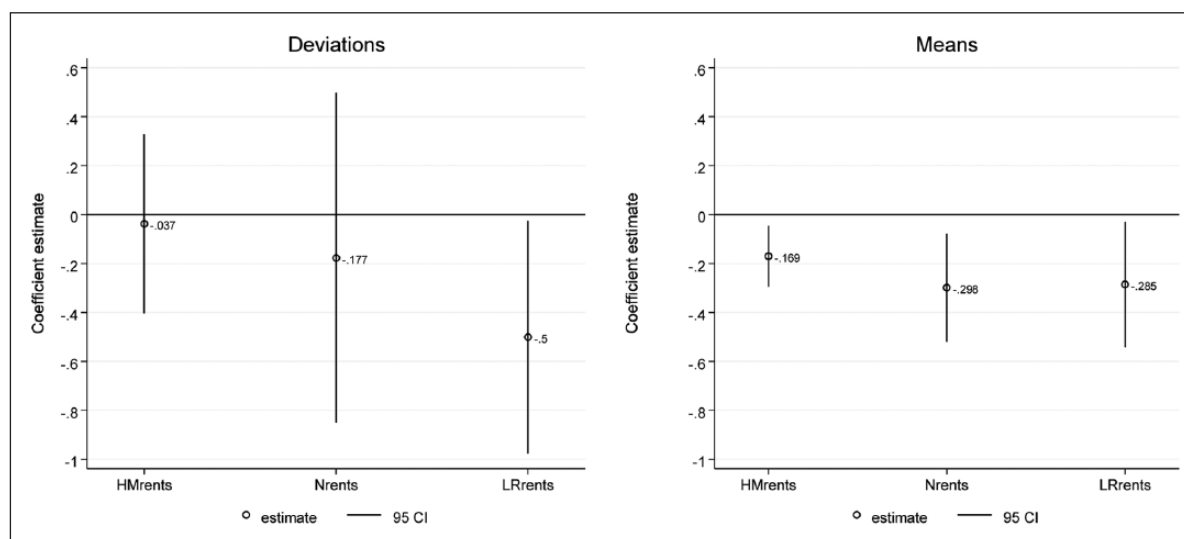
$$Pr(Y_t = 1 | Y_{t-1} = 0) = \alpha_0 + Y_i^- + O_i^- + (O_{i,t-1} - O_i^-) + X_i^- + (X_{i,t-1} - X_i^-) + \vartheta_{i,t} + \zeta_t + \mu_{i,t} \quad (1)$$

where O_i^- is the country-means, which captures the between-country effects, and $(O_{i,t-1} - O_i^-)$ is the deviations, which captures the within-country (over time) effects. WFG use *HMrents* to construct O_i^- and $(O_{i,t-1} - O_i^-)$, while LR use *LRrents*. In some of the models we report, we employ *Nrents* to measure O_i^- and $(O_{i,t-1} - O_i^-)$. The

Table 1. Resource income and transition probabilities, by sample subset.

Sample	Total number of countries	Number of countries with Haber and Menaldo rents (<i>HMrents</i>) >0	Years covered	<i>HMrents</i>	Democratic transition probability	Autocratic transition probability
<i>s1</i>	40	34	1964–2007	5.7	1.2	1.3
<i>s2</i>	70	0	1947–2007	0	2.2	3.1
<i>s1</i> \cup <i>s2</i>	90	34	1947–2007	1.4	2.0	2.6
<i>s3</i>	69	66	1947–2007	3.1	2.8	2.2
<i>s1</i> \cup <i>s2</i> \cup <i>s3</i>	114	67	1947–2007	2.0	2.2	2.5

Data sources: Geddes, Wright and Frantz (2014); Haber and Menaldo (2011); Lucas and Richter (2016). Years: 1946–2008.

**Figure 2.** Placebo tests for democratic transitions.

dependent variable, Y_t , is either autocratic transition or democratic transition.

In our re-analysis, we approach the missing data issue from two directions. First, we conduct a placebo test by employing the LR *procedure* for dealing with missing data but instead of using the GSRE data to measure O for *s1* we use the original HM data on oil income. That is, we mimic the LR *procedure* by estimating the models using a sample restricted to *s1* \cup *s2* with *Nrents* instead of *LRrents*. This analysis helps us understand whether the *procedure* that LR use to conduct their analysis or the new GSRE *data* they employ account for the differing results. In the Online Appendix, we retest the LR specifications but multiply impute the missing GSRE data in *s3*. (That is, we use *LRrents* and multiple imputation to fill in the rest of the WFG observations for which *LRrents* is missing.) This multiple imputation exercise shows us whether the newly available data change the WFG findings when using the entire WFG estimating sample.

Placebo tests

Figures 2 and 3 show the results of the placebo tests for democratic transition and autocratic transition, respectively. The left panel in each figure reports the coefficient estimate for *deviations* ($O_{i,t-1} - \bar{O}_i$) from four models, while the right panel plots the estimates for *means*. The point estimate and 95% confidence interval is reported.

The first model in each panel, labeled “HM rents,” is the reported model in WFG. It uses *HMrents* as the resource rents variable; the sample consists of *s1* \cup *s2* \cup *s3* ($n = 4,137$). The second model, labeled “N rents,” is the same specification using *HMrents* as the resource rents variable but restricts the sample to *s1* \cup *s2* ($n = 2,731$). The third model, labeled “LR rents,” uses *LRrents* instead of *HMrents* with the sample consisting of *s1* \cup *s2* ($n = 2,731$).

The left panel of Figure 2 shows that the WFG democratic transition model estimates a near zero coefficient for

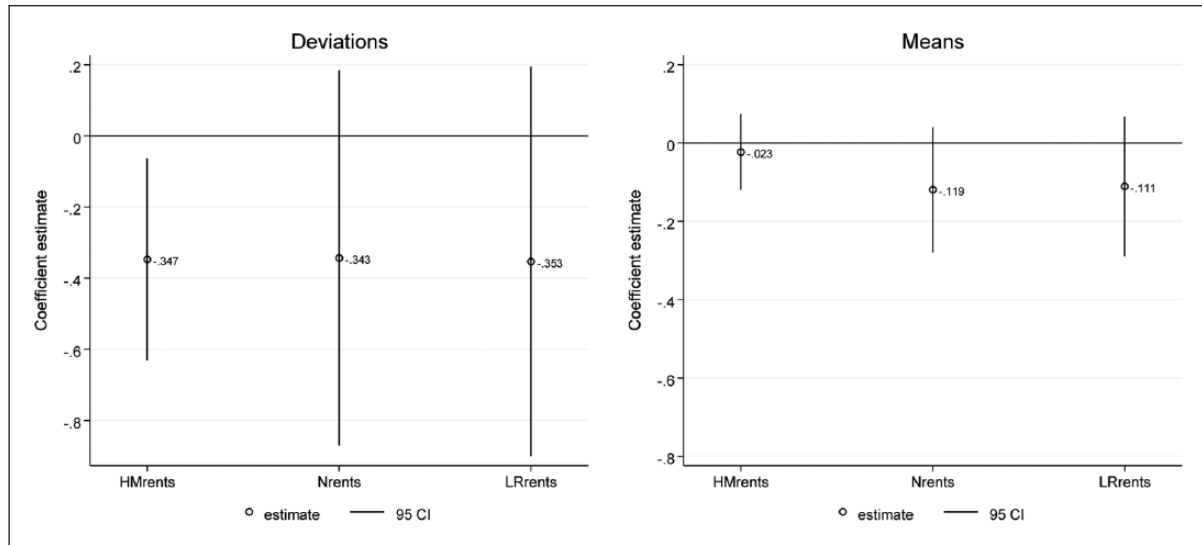


Figure 3. Placebo tests for autocratic transitions.

deviations (-0.037). Using the *Nrents* data on the restricted sample yields a larger negative coefficient (-0.177) but this estimate is not statistically significant. Finally, the LR democratic transition model estimates a large negative coefficient for *deviations* that is statistically significant. The placebo tests in the left panel suggest that the differences between the WFG and LR model estimates are due, in part, to the *procedure* LR employ (-0.037 versus -0.117) and the differing *data* that LR employ (-0.117 versus -0.500). In other words, in the within-country democratic transition model, the placebo tests imply that the GSRE data account for the some but not all of the difference in results between the WFG and LR studies.

The right panel of Figure 2 shows that while the estimates for the *means* in all three democratic transition models yield substantively similar results, changing the sample makes the estimates more negative, indicating a stronger negative cross-country correlation between resource rents and democratic transition, consistent with the expectations in our study.

The left panel of Figure 3 shows that the WFG autocratic transition model estimates a negative and statistically significant coefficient for *deviations* (-0.347). Using the *Nrents* data on the restricted sample yields a negative coefficient (-0.343) almost identical to the WFG estimate (-0.347) but with larger confidence intervals. This suggests that reducing the sample size from over 4100 observations to roughly 2700 observations increases the variance of the estimate. Finally, the LR estimate is more negative than the WFG estimate but again with large confidence intervals. This set of estimates suggests that for the within-country autocratic transition model, the difference between the WFG and LR estimates arises from

reducing the sample size – and not from using the GSRE data to measure hydrocarbon rents. That is, *procedure* and not *data* differences account for the differing WFG and LR estimates of this model.

Finally, the right panel of Figure 3 shows that all three autocratic transition models yield negative but statistically insignificant estimates for the *means*. The main difference in the size of the estimated coefficients appears to be due to changing the sample size and not to using the GSRE data employed in the LR estimate.

These placebo tests help isolate the source of the discrepancies between the WFG and LR models. They imply that the GSRE *data* account for some of the different results LR obtain in the within-country model of democratic transition (left plot of Figure 2). At the same time, the LR *procedure* for dealing with missing data likely accounts for the different results we obtain in the within-country model of autocratic transition. This casts doubts on the LR claim that using the GSRE data “challenges” the central finding that within-country increases in resource rents lower the chance of transition to a new dictatorship. Indeed, dropping a third of the sample increases the variance of the estimate underpinning the main finding in WFG.

Conclusion

In their original study, WFG use Haber and Menaldo’s data on natural resource rents to show that greater resource income boosts the survival prospects of dictatorships. They find that: (1) as resource rents increase within dictatorships over time, the risk of transition to new dictatorship decreases (a within-country effect); and (2) across dictatorships, those with resource wealth are less likely to democratize than

those without it (a cross-country effect). Using new data to measure resource rents – the GSRE data set – LR challenge WFG's key results. They find that when GSRE data are used to capture resource wealth, there is no support for the first finding. This is important given that the bulk of WFG's theoretical argument emphasizes how natural resource rents buffer dictatorships from the threat of overthrow from rival autocrats.

This response shows that the LR strategy for dealing with missingness in the GSRE data set largely accounts for the divergent results. The GSRE data set is missing data for the bulk of the observations in the original WFG sample. LR address this issue by filling in those missing values with 0s any time that there are 0s in the Haber and Menaldo data. This is problematic, however, because the remaining observations that are missing and therefore excluded from their analysis are systematically different from those they include. The excluded observations are more likely to have higher resource income, a lower chance of transitioning to dictatorship, and a higher chance of democratization. This suggests that the LR strategy likely introduces bias into the estimation.

We conduct two types of tests to explore whether this is true and to what extent. First, we use placebo tests to explore whether the *procedure* LR use to conduct their analysis or the new GSRE *data* they employ account for the differing results. Second, we replicate the LR analysis but multiply impute the missing GSRE data. The results from these tests cast doubt on LR's assertion that it is the data that explain the different findings and instead suggest that it is the procedure. Once less biased strategies for dealing with missing data are used, our estimates support the original contention: increases in resource wealth over time help dictatorships hang on to power, not by limiting prospects for democratization, but by helping to avert regime collapses that lead to a new dictatorship.

Declaration of Conflicting Interest

The authors declare that there is no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Supplementary Material

The supplementary files are available at <http://journals.sagepub.com/doi/suppl/10.1177/2053168017719794> and the replication files are available at: <http://dx.doi.org/10.7910/DVN/WMPW7>.

Notes

1. We concur with Lucas and Richter that data sets based on *oil production*, such as the Haber and Menaldo (2011) and (Ross, 2012) data, are better suited to capturing off-budget oil rents than self-reported government budget data, as calculated in the Global State Revenues and Expenditures (GSRE) data. If there are off-budget accounts with which dictators use oil revenue to buy elite support, then government-budget data are unlikely to record these accounts. Further, if the dictator gives *de facto* (private) control over oil resources to elites – perhaps military elites in an effort to buy their support and forestall a coup attempt – post-tax earnings of the oil firm may accrue directly to elites, again circumventing the government budget. Total oil revenue per capita – such as the Haber and Menaldo measure – is based on production data from market oil sales and is therefore more likely to capture off-budget accounts and the private resource and extraction earnings that accrue directly to elites. Thus, in regimes where leaders use off-budget resources or permit elites direct control over resource firms, the GSRE data are likely to underestimate the size of oil rents.
2. This is calculated for the Geddes, Wright and Frantz data set, years 1946–2007 with non-missing data on level of development. Of the 98 observations with missing resource rents data, 20% are missing in the first year the regime enters the data set because the variable is lagged one year in the analysis.
3. See the Online Appendix for a list of the democratic and autocratic transitions missing from the Lucas and Richter analysis.

Carnegie Corporation of New York Grant

This publication was made possible (in part) by a grant from Carnegie Corporation of New York. The statements made and views expressed are solely the responsibility of the author.

References

- Geddes B, Wright J and Frantz E (2014) Autocratic breakdown and regime transitions: A new data set. *Perspectives on Politics* 12(2): 313–331.
- Haber S and Menaldo V (2011) Do natural resources fuel authoritarianism? A reappraisal of the resource curse. *American Political Science Review* 105(1):1–26.
- Lucas V and Richter T (2016) State hydrocarbon rents, authoritarian survival, and the onset of democracy: Evidence from a new dataset. *Research and Politics* 3(3):1–9.
- Ross M (2012) *The Oil Curse: How Petroleum Wealth Shapes the Development of Nations*. Princeton, NJ: Princeton University Press.
- Ross M (2015) “What have we learned about the resource curse?” *Annual Review of Political Science* 18(1): 239–259.
- Wright J, Frantz E and Geddes B (2015) Oil and autocratic regime survival. *British Journal of Political Science* 45(2): 287–306.