Supplemental information for:
Brad Epperly, "Political competition and de facto judicial independence in non-democracies," *European Journal of Political Research*.

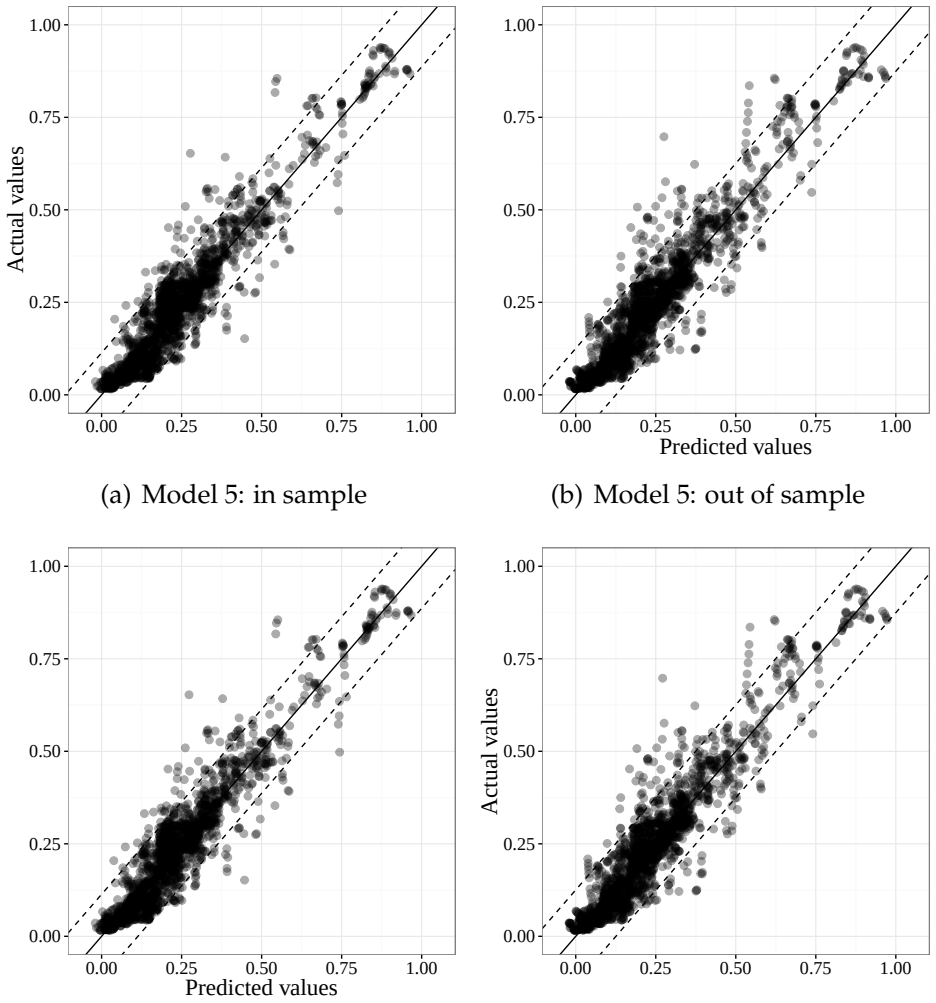
Supplemental information

Cross validation

Cross validation assesses model fit based on one of the most important metrics: out of sample predictive accuracy. Rather than collect new data (often difficult, and in the case of work on political institutions, often impossible) and test models on these data, cross validation procedures call for partitioning the data into multiple data sets. The is used as a 'training' or 'fit' set. Here, a model is fit according to standard practice. Estimates from this model are then used on the other 'test' data set(s). As the data from the test set(s) were themselves not used to derive the model estimates, the model is effectively being tested out of sample. One can then examine any differences between the predictive accuracy of the model in and out of sample and determine how well the model captures the underlying data-generating process (Ward, Siverson and Cao 2007). If out of sample fit is comparable to in sample fit, there is persuasive evidence that the model fits the underlying, empirical process of interest generating the data, and is not merely a collection of statistically significant effects that actually fail to accurately capture the data-generating process.

Figure 2 illustrates the in and out of sample fit for Models 5 and 6 from Table 1. Each of the four plots show the actual values of judicial independence plotted against the predicted values. Plot (a) shows Model 5's in sample accuracy, whereas plot (b) shows how well it fits out of sample. Plots (c) and (d) show the same for Model 6. In each, the solid black line is the line of best fit, where $y = x$. The upper dashed black lines show the product of the root mean square error (RMSE) and 1.96, effectively showing the upper 95% confidence interval of the distribution of errors. The lower dashed black lines are the product of the RMSE and -1.96, showing the lower interval. Each observation is plotted with transparency to better convey the distribution where data overlap.

Comparing plots (a) and (b) it is clear that the model fits nearly as well out of sample as it does in sample. The RMSE of the in sample fit is 0.06 (for reference, the RMSE of Model 5 in Table 1, fit on all the data, is also 0.06). The RMSE of the out of sample fit (which is not a model, but rather predicted values using the coefficients of the model estimated on the 'training' data) is 0.06. It is not until the third decimal place that we observe any variation between the in and out of sample fit: the in sample RMSE is 0.058 whereas out of sample it is 0.064. In other words, there is almost no discernible difference between how good the model fits on the data used in its creation and new, out of sample data not used. An almost identical relationship is observed for plots (c) and (d), which show cross validation of Model 6, with country fixed effects. The fact that these models fit so well out of



(a) Model 5: in sample

(b) Model 5: out of sample

(c) Model 6: in sample

(d) Model 6: out of sample

Figure 2: **Cross validation of Models 5 and 6.** Four plots showing in and out of sample predictive accuracy of Models 5 and 6 from Table 1. Each plots the actual values of judicial independence against the values predicted by the model. Solid black lines show where $y = x$, and dashed black lines show the product of the root mean square error and 1.96 and -1.96, effectively showing 95% confidence intervals around the line of best fit.

sample provides another piece of evidence in support of the contention that they accurately capture the underlying data-generating process, and for the hypothesis that competition is a strong predictor of judicial independence in non-democracies.

Freedom House democracy

Though arguably one of the most widespread, the Cheibub, Gandhi and Vreeland (2010) classification of democracy and dictatorship is not the only possible classification I could use in the analysis. Two others in widespread use are the Polity IV democracy scores and the Freedom House Freedom in the World measure. Un-

fortunately, both of these measures explicitly consider whether a country has an independent judiciary when rating a country's regime characteristics, making them inappropriate as means of classification when one is also interested in analyzing judicial independence. For Polity, the culprit is the executive constraints score. While Gibler and Randazzo (2011) simply remove this component and used the remaining Polity score in their analysis of judicial independence, this is potentially problematic, as Gleditsch and Ward (1997) show that it is this very component that most strongly determines a country's overall Polity score in a given year.

The aspect of the Freedom House rating that includes judicial independence is found in its civil liberties score (specifically the rule of law sub-component), which is one of two components used to create the overall democracy score. The other is political rights, which judges to what degree people are able to participate freely in the political process, with sub-components for the electoral process, political pluralism, and the functioning of government (Freedom House 2012). As a robustness check on the model results presented in Table 1 of the main article, I use this Freedom House Political Rights measure as a means of classifying whether a country year is democratic or non-democratic. The measure is a 7 point ordinal scale, with countries with the most political rights scored 1 and those with the least 7. Following standard practice, I use 3 as the cut-off point: countries scoring 3 or less are rated democracies, and those countries scoring 4 or more are considered non-democracies, and thus included in the analysis.

Table 4 presents the results of six models of de facto judicial independence that use this classification scheme for non-democracy, using the same covariates and model structure as the Models 1–6 presented in Table 1 of the article (dates and the number are observations are not identical due to Freedom House data having less temporal coverage than the primary measure used, and each obviously containing different country-years due to classificatory scheme). Similar to the results in the article, the models in Table 4 show significant agreement, with minimal differences in coefficient estimates between the hierarchical and fixed effects models. Even more persuasive is the overwhelming overlap in the results not only within but also between the two data sets using different classifications of democracy: the estimates for electoral competition between models within and across the two data sets are nearly identical. Although there is significant overlap in the observations within each data set, differences remain. Despite this fact, the estimated coefficients show substantial similarities, and there is very little change in levels of statistical significance between the data sets. These results provide further support for the key argument forwarded in this article, with competition once again being consistently associated with higher levels of judicial independence.

Excluding the most closed of autocracies

There is potential that the lack of competition (as well as judicial independence) in the most closed of autocratic regimes—those that Roessler and Howard (2009, 107) define as closed authoritarian, i.e. those autocracies in which the "citizenry

Table 4: **Freedom House Political Rights classification of democracy.** Six models of de facto judicial independence using the Freedom House Political Rights classification of democracy. Models 1*a*–6*a* replicate the results of Table 1 of the article using this alternative classification scheme.

	Model 1*a*	Model 2*a*	Model 3*a*	Model 4*a*	Model 5*a*	Model 6*a*
Intercept	0.26^{***}	0.28^{***}	0.24^{***}	0.10^{***}	0.24^{***}	0.12^{***}
	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)
Electoral competition	0.14^{***}	0.13^{***}	0.16^{***}	0.16^{***}	0.19^{***}	0.18^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(GDP/capita)	0.03^{***}	0.02^{*}	0.03^{***}	0.03^{***}	0.03^{***}	0.03^{***}
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Common law heritage	0.09^{**}	0.09^{**}	0.12^{***}	0.26^{***}	0.15^{***}	0.23^{***}
	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)
Year	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Years in office	-0.00^{***}	-0.00^{***}	-0.00^{***}	-0.00^{***}		
	(0.00)	(0.00)	(0.00)	(0.00)		
Government share of GDP	-0.00	-0.00				
	(0.00)	(0.00)				
Foreign direct investment	-0.08	-0.10				
	(0.05)	(0.05)				
Conflict	-0.00	-0.00				
	(0.00)	(0.00)				
AICc	-4212.82	-4570.84	-6181.73	-6621.03	-7458.71	-7954.93
N	1671	1671	2404	2404	2950	2950

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

is constitutionally excluded from participating in the selection" of executives—is driving the results of the analyses in the main manuscript.[1] To address this concern, I employ two strategies to remove these observations from the data analyzed. First, I removed all country-years considered autocratic by the Cheibub, Ghandi, and Vreeland measure that score -6 or less on the Polity measure (while recognizing the problems with Polity, it remains one clear and commonly-used option). As a second method, I look to the Wahman and Hadenius. (2013) Types of Authoritarian Regimes measure.[2] Here, I removed all those country-years considered autocratic by Cheibub, Ghandi, and Vreeland that Wahman, Teorell, and Hadenius classified as either a monarchy, no-party regime, or one-party regime.

Table 5 replicates the models of the manuscript, removing the most closed of autocratic regimes to ensure the lack of competition within these regimes is not driving the key relationship. Models 1 and 2, denoted *p*, exclude those autocratic country-years measured as full autocracies by Polity (a score of -6 or lower); Models

[1] While their conceptualization of forms of autocratic regimes is theoretically useful ,Roessler and Howard (2009) rely heavily on existing Polity and Freedom House scores to create a measure of their concept. For this reason, as well as the limited (less than 20 year) temporal coverage they offer, I refrain from using this measure in the analysis below, instead using two alternative ways of measuring fully autocratic regimes.

[2] The "regtype1" measure, specifically.

Table 5: **Excluding fullest autocracies.** Four models excluding the fullest/most closed autocratic regimes. Models subscripted *p* exclude those autocratic country-years measured as full autocracies by Polity, Models subscripted *wth* exclude monarchies, one-party, and no-party regimes according to Wahman, Teorell, and Hadenius.

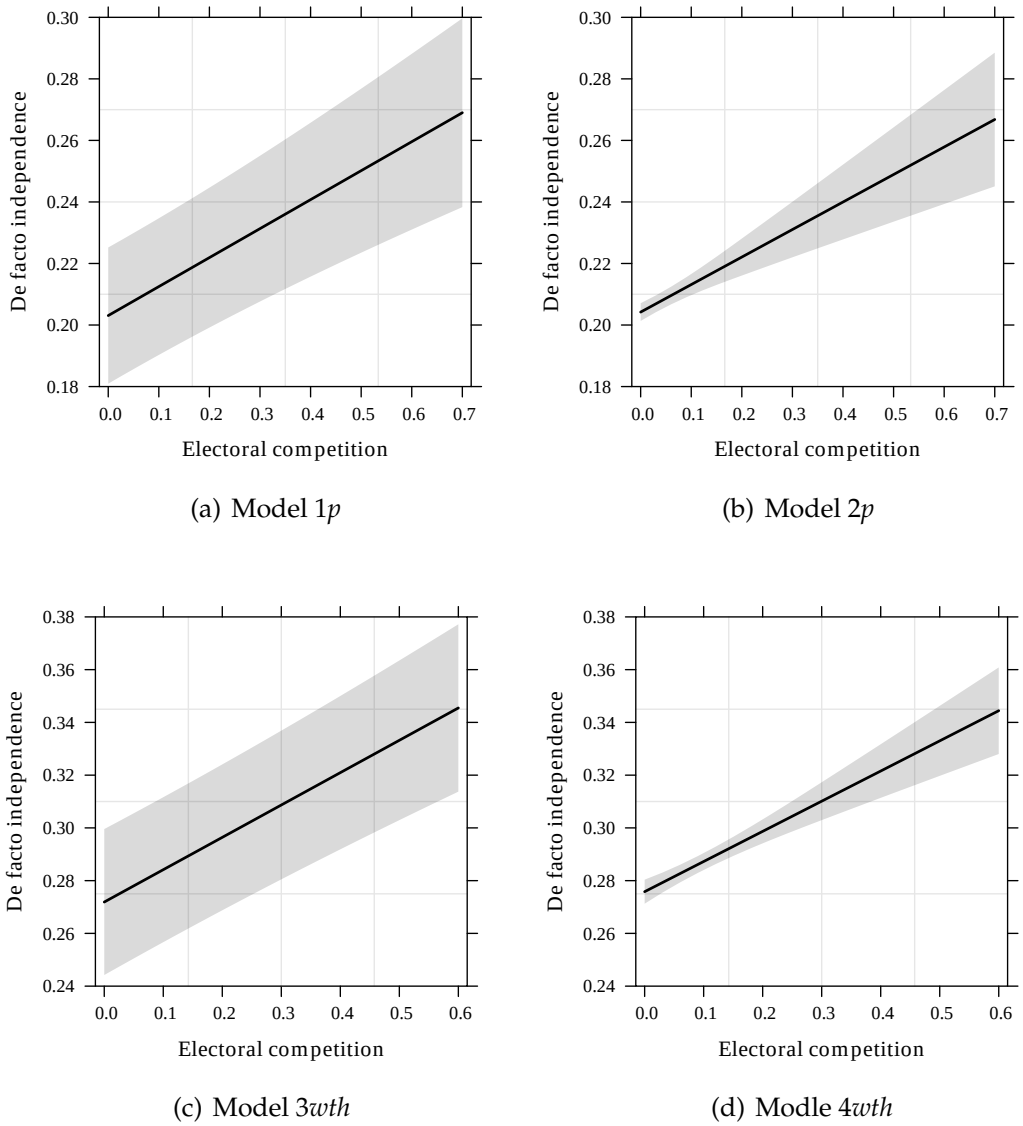
	Model 1*p* Random effects	Model 2*p* Fixed effects	Model 3*wth* Random effects	Model 4*wth* Fixed effects
Intercept	0.20^{***}	0.15^{***}	0.22^{***}	0.13^{***}
	(0.01)	(0.01)	(0.02)	(0.01)
Electoral competition	0.09^{***}	0.09^{***}	0.12^{***}	0.11^{***}
	(0.02)	(0.02)	(0.02)	(0.02)
log(GDP/capita)	0.01	0.00	0.06^{***}	0.06^{***}
	(0.01)	(0.01)	(0.01)	(0.01)
Common law heritage	0.03	0.29^{***}	0.15^{***}	0.23^{***}
	(0.03)	(0.02)	(0.03)	(0.02)
Year	0.00^{***}	0.00^{***}	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Years in office	0.00^{***}	0.00^{***}	-0.00^{***}	-0.00^{**}
	(0.00)	(0.00)	(0.00)	(0.00)
Government share of GDP	-0.00^{*}	-0.00^{*}	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Foreign direct investment	-0.02	-0.01	0.07	0.07
	(0.15)	(0.15)	(0.05)	(0.05)
Conflict	-0.01^{**}	-0.01^{**}	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
AICc	-3030.57	-3261.80	-3233.71	-3495.41
N	965	965	1265	1265

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

3 and 4, denoted *htw*, exclude monarchies, one-party, and no-party regimes according to Wahman, Teorell, and Hadenius' classification scheme.

The results are highly similar to the results presented in Table 1 of the main article: in both random and fixed effects specifications the coefficient estimate for electoral competition remains highly significant, and the change in small, as the random and fixed effects estimates in the main manuscript are -0.13 and -0.12, respectively. Excluding the extreme autocracies from Polity leads to a slightly lessened effect, whereas excluding those regimes according to the Wahman, Teorell, and Hadenius scheme produces no discernible change in the estimate. Figure 3 illustrates the predicted value of de facto independence as electoral competition varies, with each plot showing the predictions from the listed model in Table 5.

Comparing Table 5 and Figure 3 to the results of the main article, there is substantial support for the inferences drawn. Regardless of the fact that the models excluding according to Polity contain only 56% the number of observations as found in the main manuscript analyses, the results are highly similar. The same is true for restricting according to Wahman, Teorell, and Hadenius' classification scheme, where the number of observations are only 73% that of the manuscript analyses analyzing all autocratic regimes.



(a) Model 1*p*

(b) Model 2*p*

(c) Model 3*wth*

(d) Modle 4*wth*

Figure 3: **Predicted independence by non-democratic regime type.** Four figures showing the predicted value of de facto judicial independence as electoral competition varies. Each plot shows predicted values from the listed model in Table 5.

Model averaging

Table 1 of the main article, as discussed, includes models where covariates that some might think relevant due to their lack of importance according to model averaging criteria. Table 6 reports the results of this model averaging (using AICc), including both the full averages, conditional averages, and importance of each covariate in a hierarchical specification. The results of Table 6 further illustrate the importance of the key covariates in the "full" data, justifying the exclusion of the unimportant covariates so as to fit the model to a larger number of observations.

Table 6: **Model averaging.** Results of model averaging using a hierarchical linear specification with country-varying intercepts, averaged according to AICc.

	Full average	Conditional average	Importance
Intercept	0.24***	0.24***	—
	(0.02)	(0.02)	
Electoral competition	0.13***	0.13***	1.00
	(0.01)	(0.01)	
log(GDP/capita	0.03***	0.03***	1.00
	(0.01)	(0.01)	
Common law heritage	0.11***	0.11***	1.00
	(0.03)	(0.03)	
Year	0.00***	0.00***	1.00
	(0.00)	(0.00)	
Years in office	−0.00**	−0.00**	0.93
	(0.00)	(0.00)	
Government share of GDP	−0.00	−0.00	0.46
	(0.00)	(0.00)	
Foreign direct investment	−0.02	−0.05	0.36
	(0.04)	(0.05)	
Conflict	0.00	0.00	0.29
	(0.00)	(0.00)	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Averaged competition

As noted in the main text, I average the level of electoral competition over the half decade prior to the year in question to ascertain whether it is the overall levels of competition in the recent past that might be critical rather than the immediate levels facing autocrats. As Table 7 demonstrates, results presented in Table 1 of the main text are robust to changing this specification of the critical independent variable. In fact, the key difference between the models in Tables 1 and 7 is that competition is *more* important when averaging the previous half decade of competition.

Temporal restrictions

As noted in the article, results are robust when data are partitioned by decade, and analyses restricted to each temporal subset. In other words, results are consistent across time periods and not the function of patterns observed, for example, before or after the end of the Cold War. Table 8 presents six hierarchical models (fixed effects models are omitted for space, similarities are analogous to prior specifications). Models 5*c*–5*f* use a similar specification to Model 5 in Table 1, limiting covariates to take full advantage of data back to 1960. These models cover the decades of the 1960s, 1970s, 1980s, and 1990s. Models 1*e* and 1*f*, fit to 1980–90 and 1990–2000, respectively, include all the covariates used in Model 1 from Table 1. Table 8 demonstrates that results presented in the article are strikingly robust: though the coefficient for competition fluctuates somewhat over the decade subsets, it consis-

Table 7: **Five year averages of electoral competition.** Results of six models of judicial independence when electoral competition is averaged over the previous half decade. Each model specification is similar to that in Table 1, with Model 1*b* similar to Model 1, etc.

	Model 1*b*	Model 2*b*	Model 3*b*	Model 4*b*	Model 5*b*	Model 6*b*
Intercept	0.24^{***}	0.18^{***}	0.23^{***}	0.10^{***}	0.21^{***}	0.14^{***}
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Electoral competition	0.18^{***}	0.16^{***}	0.17^{***}	0.16^{***}	0.24^{***}	0.23^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
log(GDP/capita)	0.03^{***}	0.02^{**}	0.03^{***}	0.03^{***}	0.03^{***}	0.03^{***}
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Common law heritage	0.112^{***}	0.197^{***}	0.152^{***}	0.262^{***}	0.159^{***}	0.210^{***}
	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)
Year	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Years in office	-0.00^{**}	-0.00^{*}	-0.00^{***}	-0.00^{***}		
	(0.00)	(0.00)	(0.00)	(0.00)		
Government share of GDP	-0.00	-0.00				
	(0.00)	(0.00)				
Foreign direct investment	-0.04	-0.06				
	(0.05)	(0.05)				
Conflict	0.00	0.00				
	(0.00)	(0.00)				
AICc	-4413.74	-4753.18	-6704.87	-7148.08	-10135.86	-10638.51
N	1731	1731	2518	2518	3933	3933

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

tently remains highly statistically significant and associated with increased levels of judicial independence.

Multiple imputation

Another potential concern regarding the robustness of the model is that missingness in the data is not random, but systematic, which produces biases the results of the model (King et al. 2001). Multiple imputation has been shown to be a helpful mechanism in overcoming the problem of non-random missingness in data, and produces better results than listwise deletion (where each observation that has a missing value on one or more covariates is not included in the modeling process). Unfortunately, it is often difficult to determine whether missing data is missing at random. To overcome this limitation, I use multiple imputation, creating *m* data sets (in this case 15) and using the known information contained in the original data to predict the values of the missing information.[3] The difference between any one imputed cell across the *m* data sets is a result of how well the model predicts the given cell; the use of multiple data sets is to include the uncertainties around the predicted quantities into the modeling process. Here, I employ the approach advocated by Honaker and King (2010), which explicitly deals with the issues sur-

[3]All multiple imputation was conducted using the time-series cross-sectional imputation technique in the `Amelia` package in *R* (Honaker, King and Blackwell 2011).

Table 8: **Electoral competition and judicial independence by decade.** Results of six models of judicial independence when the analysis is restricted to one decade at a time.

	Model 5*c* (1960s)	Model 5*d* (1970s)	Model 5*e* (1980s)	Model 5*f* (1990s)	Model 1*e* (1980s)	Model 1*f* (1990s)
Intercept	0.16^{***}	0.19^{***}	0.22^{***}	0.22^{***}	0.28^{***}	0.24^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Electoral competition	0.21^{***}	0.09^{***}	0.12^{***}	0.06^{***}	0.12^{***}	0.05^{***}
	(0.01)	(0.02)	(0.02)	(0.01)	(0.03)	(0.01)
Common law heritage	0.14^{***}	0.14^{***}	0.14^{***}	0.18^{***}	0.08^{*}	0.12^{***}
	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)
log(GDP/capita)	0.03^{***}	0.03^{***}	0.03^{***}	0.03^{***}	0.03^{**}	0.04^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Year	-0.00^{***}	0.00	0.00^{***}	-0.00	0.00^{***}	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Years in office					0.00	-0.00
					(0.00)	(0.00)
Government share of GDP					-0.00	-0.00
					(0.00)	(0.00)
Foreign direct investment					-0.13	0.05
					(0.10)	(0.06)
Conflict					0.00	-0.00
					(0.01)	(0.00)
AICc	-2962.51	-3654.10	-3217.62	-2988.17	-1953.58	-2202.68
N	749	976	970	840	616	616

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

rounding multiple imputation in time-series cross-sectional data.

Table 9 presents the results of hierarchical and fixed effects models averaged over the 15 imputed data sets. Models *1g* and 2*e* show further support for the findings presented in Table 1: in both the results across covariates are similar to the models fit to the non-imputed data.[4] The key difference is that, once again, the effects of competition are even greater in the robustness checks than in the primary models, with the covariate for competition increasing by over 50% in both random and fixed effects specifications.

[4]The estimate of the intercept has changed noticeably, however. This is a function of the imputed models using covariates that are not mean-centered as in the other models. Doing so does not effect the estimates of the covariates or any other quantities of interest.

Table 9: **Models using multiple imputation data.** Results of two models of judicial independence when the analysis is conducted on 15 multiply-imputed data sets (results presented are averaged over all 15.

	1*g*	2*e*
Intercept	-0.85^{**}	-0.67^{*}
	(0.29)	(0.29)
Electoral competition	0.20^{***}	0.20^{***}
	(0.01)	(0.01)
log(GDP/capita)	0.03^{***}	0.03^{***}
	(0.00)	(0.00)
Common law heritage	0.30^{***}	0.16^{***}
	(0.01)	(0.03)
Year	0.00^{*}	0.00^{*}
	(0.00)	(0.00)
Years in office	-0.00^{*}	-0.00^{**}
	(0.00)	(0.00)
Government share of GDP	0.00	-0.00
	(0.00)	(0.00)
Foreign direct investment	-0.01	-0.01
	(0.02)	(0.02)
Conflict	-0.00	-0.00
	(0.00)	(0.00)
AIC	$-11,152.86$	$-10,649.61$
Imputed observations	$4,321$	$4,321$

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Supplemental information references

Cheibub, Jose A., Jennifer Gandhi and James R. Vreeland. 2010. "Democracy and Dictatorship Revisted." *Public Choice* 143(1):67–101.

Freedom House. 2012. "Freedom In the World." `http://www.freedomhouse.org/report-types/freedom-world`.

Gibler, Douglas and Kirk Randazzo. 2011. "Testing the Effects of Independent Judiciaries on the Likelihood of Democratic Backsliding." *American Journal of Political Science* 55(3).

Gleditsch, Kristian S and Michael D Ward. 1997. "Double Take A Reexamination of Democracy and Autocracy in Modern Polities." *Journal of Conflict Resolution* 41(3):361–383.

Honaker, James and Gary King. 2010. "What to Do about Missing Values in Time–Series Cross–Section Data." *American Journal of Political Science* 54(2):561–581.

Honaker, James, Gary King and Matt Blackwell. 2011. "Amelia II: A Program for Missing Data." *Journal of Statistical Software* 45(7):1–47.

King, Gary, James Honaker, Anne Joseph and Kenneth Scheve. 2001. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation." *American Political Science Review* 95(01):49–69.

Roessler, Philip G and Marc M Howard. 2009. Post-Cold War Political Regimes: When Do Elections Matter? In *Democratization By Elections*, ed. Staffan I. Lindberg. The Johns Hopkins University Press Baltimore, MD p. 432.

Wahman, Michael, Jan Teorell and Axel Hadenius. 2013. "Authoritarian Regime Types Revisited: Updated Data in Comparative Perspective." *Comparative Politics* 19(1).

Ward, Michael D., Randolph M. Siverson and Xun Cao. 2007. "Disputes, Democracies, and Dependencies: A Reexamination of the Kantian Peace." *American Journal of Political Science* 51(3):583–601.