Supplemental information for
"The Demand for Insurance: Incorporating the Severity of Losing Office into the Insurance Model of Judicial Independence"
Journal of Law and Courts

Brad Epperly[*] Yuleng Zeng[†]

The expected utility model of independence-as-insurance

At heart, the logic of the insurance model is an expected utility framework consisting of the demand for insurance (itself a function of both the likelihood of risk and its severity) and the costs it imposes. Before addressing these two terms, however, a handful of additional parameters require discussion.

The first are the utilities leaders derive from each possible outcome: remaining in office, losing power and being unpunished, and finally being punished after leaving office, ordered respectively such that $x > y > z$. Following the insurance framework's logic, leaders can take one of two actions, A or B, to affect the probability of punishment after leaving office: pursuing independence to insure against the future (A) or refraining from insurance, to avoid paying the costs associated with independent judiciaries while in power (B).[1] This choice should affect the relative probabilities of outcomes y and z, but not the likelihood of x (remaining in office). With these values set, we turn to the parameters of the utility function: the likelihood of losing office, l; the risks associated with losing office, r; a tuning parameter, k, capturing the additional effect[2] of pursuing an independent court on reducing the risks; the demand for insurance (the

[*]University of British Columbia Okanagan, brad.epperly@ubc.ca
[†]University of Groningen, yuleng.zeng@rug.nl

[1]While really a range, here we dichotomize for analytical clarity.

[2]That is, $k < 1$ indicates a highly independent court further reducing risk. $k \geq 1$ indicates the risks are increased (remain the same).

product of the risks associated with losing office and the likelihood of losing office), d; and the "insurance premium" leaders must pay if they choose independent courts (regardless of the outcome), c.

Depending on whether a leader chooses option A (high independence) or option B (low independence), the parameters to determine the probability of each outcome are as follows:

$$Pr_A(x) = 1 - l \qquad Pr_A(y) = l * (1 - kr) \qquad Pr_A(z) = l * kr$$

$$Pr_B(x) = 1 - l \qquad Pr_B(y) = l * (1 - r) \qquad Pr_B(z) = l * r$$

These parameters are inserted into the utility function for leaders for each strategy, which are as follows:

$$U(A) = Pr_A(x) * u(x) + Pr_A(y) * u(y) + Pr_A(z) * u(z) - c$$

$$U(B) = Pr_B(x) * u(x) + Pr_B(y) * u(y) + Pr_B(z) * u(z)$$

This expected utility framework makes determining if insurance is worth the constraints imposed by independent courts easy, as all one need do is determine if $U(A) > U(B)$. When this is the case, the greater expected utility means independence is net beneficial.

Comparative statics

Before addressing the two factors' effects, we make two minimal assumptions. First, that across the three possible outcomes (remaining in office, removal, removal with punishment) leaders derive maximal utility from remaining in power and minimal from being removed and then punished. Therefore, normalizing to the standard 0–1 range, the utility of remaining in office is $u(x) = 1$, removal with punishment $u(z) = 0$, and removal without punishment $0 < u(y) < 1$. These are reasonable values for the utilities leaders should expect, as leaders would obviously derive maximal utility remaining in office and minimal from being imprisoned, exiled, or killed after leaving office. Second, the probability of being punished after leaving office is lower when courts are independent (i.e. $k < 1$).

Leaders choose an independent judiciary if they expect the choice will give them a higher payoff than the alternative. To examine this, we define $U(AB)$ as the difference between the two utility functions (i.e. $U(A) - U(B)$). We have

$$\begin{aligned} U(AB) &\equiv U(A) - U(B) \\ &= l(1-kr)u(y) - l(1-r)u(y) + lkru(z) - lru(z) - c \\ &= lr(1-k)u(y) - c \\ &= d * (1-k)u(y) - c \end{aligned}$$

Under the above two assumptions, the effect of d is always positive ($\partial U(AB)/\partial d = (1-k)u(y) > 0$): as the value of d increases, leaders are more likely to choose independent courts (we further develop this via proof in the appendix). That is, when the the demand for insurance is greater, pursuing higher levels of independence maximizes a leader's expected utility.

Proof

The two utility functions are reproduced as follows:

$$\begin{aligned} U(A) &= Pr_A(x) * u(x) + Pr_A(y) * u(y) + Pr_A(z) * u(z) - c \\ U(B) &= Pr_B(x) * u(x) + Pr_B(y) * u(y) + Pr_B(z) * u(z) \end{aligned}$$

We also know that

$$\begin{aligned} &Pr_A(x) = 1 - l; Pr_A(y) = l * (1 - kr); Pr_A(z) = l * kr \\ &Pr_B(x) = 1 - l; Pr_B(y) = l * (1 - r); Pr_B(z) = l * r \end{aligned}$$

To examine which choice is better, we define $U(AB)$ as the difference between the two utility functions:

$$\begin{aligned} U(AB) &\equiv U(A) - U(B) \\ &= l(1-kr)u(y) - l(1-r)u(y) + lkru(z) - lru(z) - c \\ &= d(1-k)u(y) - d(1-k)u(z) - c \end{aligned}$$

where $d = l * r$. We make two assumptions:

Assumption 1. $u(z) = 0$ *and* $u(x) = 1$.

Assumption 2. $k < 1$.

Under these two assumptions, we can obtain the following results.

Lemma 1. *As l increases, both $U(A)$ and $U(B)$ decrease.*

Proof. Take the derivative of $U(A)$ with respect to l, we have

$$\frac{\partial U(A)}{\partial l} = (1 - kr)u(y) - 1$$

which is always negative since $1-kr < 1$ and $u(y) < u(x) = 1$. Analogously, take the derivative of $U(B)$ with respect to l, we have

$$\frac{\partial U(B)}{\partial l} = (1 - r)u(y) - 1$$

which is always negative since $1 - r < 1$ and $u(y) < u(x) = 1$. □

Lemma 2. *As d increases, $U(AB)$ always increases. For lower values of d (i.e., when $d < c/\left((1-k)u(y)\right)$), choice B is optimal. Otherwise, choice A is optimal.*

Proof. Take the derivative of $U(AB)$ with respect to d, we have

$$\frac{\partial U(AB)}{\partial d} = (1 - k)u(y)$$

which is always positive under the assumptions. If $d < ((1-k)u(y))$, then $U(AB) < 0$ which suggests choice B is optimal since it gives the player a higher payoff. Conversely, choice A is optimal. □

Lemma 3. *Write $d^* \equiv c/\left((1-k)u(y)\right)$. As c increases, d^* increases.*

Proof. Take the derivative of d^* with respective to c, we have

$$\frac{\partial l^*}{\partial c} = 1/\left((1-k)u(y)\right)$$

which is always positive since $k < 1$ and $u(y) > 0$. □

Robustness

Cross validation

Appendix Figure 1 illustrates the predictive accuracy of our main observational model of de facto independence (Model 2 from Table 2 of the article) using cross-validation to assess the

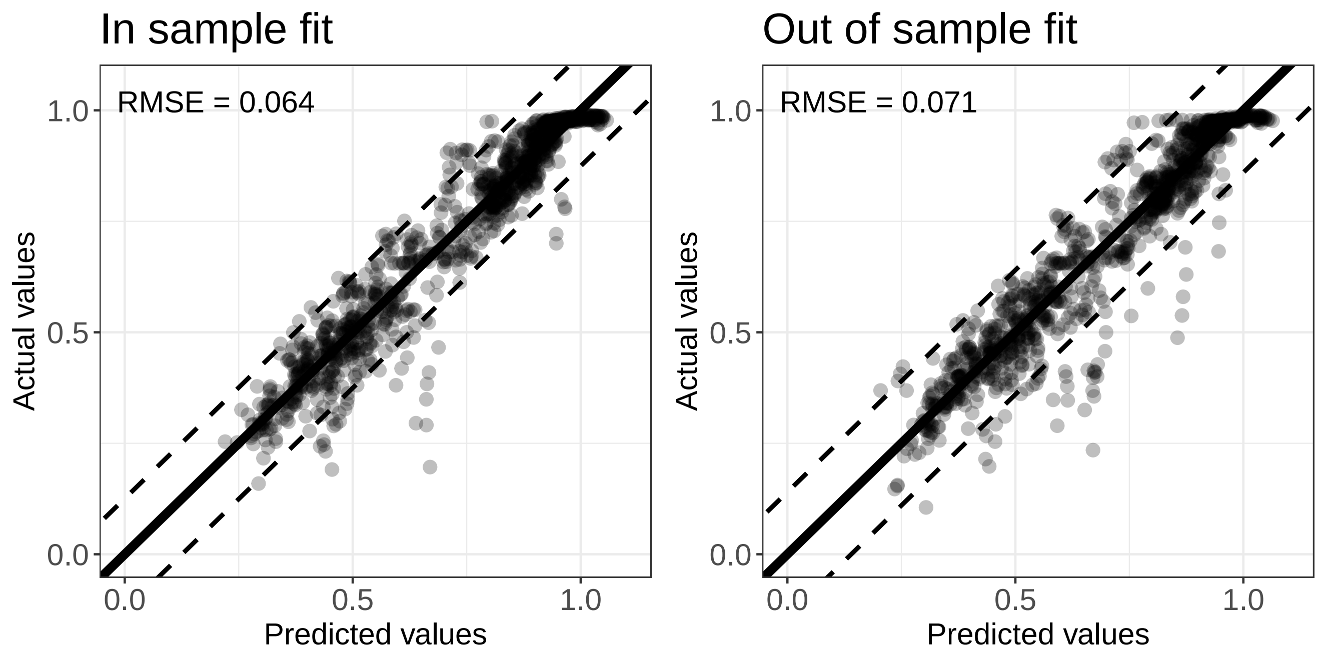

Figure 1: **Model fit via cross-validation.** Cross-validation of Model 2 in the main article (panel model with potential confounders), showing both the model fit in the training set (in-sample) and the out-of-sample fit.

degree to which the model captures the data-generating process. Data are partitioned into training and test sets, and then the model estimates from the training set are applied to the test set, allowing us to make ”out-of-sample” predictions.[3] As can be seen from the two plots in Appendix Figure 1, there is negligible decrease in model fit: the root mean-square error of the out-of-sample fit is 0.069, compared to 0.063 for the in-sample fit. This is an increase in error of 10%, on data not used to create the model parameters used to make the out-of-sample predictions.

Breaking out demand for insurance into constituent terms

Appendix Table 1 presents the second-stage results of four 2SLS models. Models 1 and 2 reproduce Models 3 and 4 from the main article (i.e. the two 2SLS models using one and two instruments, respectively), although in this instance Model 1 is fit to 1,804 observations rather than the 1,854 observations it is fit to in the main article. This is done so we can directly com-

[3]This is done after removing two observations (i.e. less than one-tenth of one percent of observations, with no substantive or statistical changes in results), each of which is the sole democratic year for the respective country.

pare the models, as they are fit to the same data.[4] Note: in Table 2 and all subsequent tables, a p-value reported as (0.00) is one in which $p < 0.005$, and thus when rounded to two decimal places (presented for ease of interpretation) reports as (0.00); the same logic applies to goodness-of-fit tests.

Recall that our demand variable is the product of electoral competition and the percent of previously-punished leaders, proxies for the likelihood and severity of losing office, respectively. It is thus, in effect, the interaction term between these two constituent terms; per standard practice, interaction terms should include their constituents. As we argue in the main article, for theoretical reasons this is *not* the case here: we are not actually looking at an interaction, but rather a measure of demand for insurance. As this demand is the theoretical quantity of interest, we should not include constituent terms. However, to demonstrate that our preferred specification is not only theoretically but also empirically superior, we do just this in Models 3–6. Model 3 therefore includes the "constituent" terms of the "interaction" that is our demand variable, as well as the interaction (reported as *Demand for insurance*). Model 4 ignores our demand specification, and does not report the product (i.e. interaction) of *Percent previously punished* and *Electoral competition*. Looking at the results for Model 3, we see that while other covariates have similar estimates and errors, breaking out the demand variable and including its components leads to covariates for all three (demand, percent previously punished, and competition) with massive standard errors, far from reaching statistical significance.

While Model 4 (ignoring the demand variable entirely in favor of just its two component parts) returns to covariates reaching conventional levels of statistical significance, Model 4 also fails the Sargan test for the exogeneity of the instruments (recall, for a Sargan test, that one does *not* want a significant result). While the Sargan test leads us to reject this model, it does not tell us which of the two instruments might have been identified as non-exogenous. As such, we include Models 5 and 6, which instrument competition with only the value of oil production and foreign aid, respectively. We are, of course, unable to test for the exogeneity of the instruments here, which should cause us some worry compared to Model 2 (where we can, and which is supported by the results of the Sargan test). Further worry is to be found in the results of each model: the Wu-Hausman test fails in Model 5 (suggesting it is not consistent with a plain OLS version of the model), and the covariates of interest for both Models 5 and 6 have massive standard errors and do not come close to being statistically significant. Put simply, all the results from appendix Table 1 provide strong support for using our demand

[4]Note: Model 2 is fit to the same observations as it is in the main article, a function of the fact that there are missing values for the second instrument in 2.7% of the observations for which we have the first instrument.

Table 1: **Separating the components of demand for insurance.** Models of de facto independence separating the demand for insurance measure. Models 1 and 2 are fit to the 2SLS models in the main article (there Models 3 and 4), that is fit two one and two instruments, respectively. Model 3 includes the components of demand estimated as separate variables (and instruments for these). Model 4 looks not at the product (i.e. does not use demand for insurance), but only the components individually (also instrumented). Models 5 and 6 break out Model 4, each using only one rather than both of the instruments for competition.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Intercept	-1.49^{***}	-1.44^{***}	-3.85^{*}	0.44	-0.70	-2.11
	(0.31)	(0.30)	(1.90)	(0.32)	(0.44)	(2.28)
Demand for insurance	0.46^{**}	0.44^{**}	-2.34			
	(0.17)	(0.17)	(1.85)			
log(GDP/capita)	0.29^{***}	0.29^{***}	0.20^{***}	0.21^{***}	0.24^{***}	0.41^{*}
	(0.02)	(0.02)	(0.06)	(0.02)	(0.02)	(0.17)
Years democratic (logged)	0.15^{***}	0.15^{***}	0.17^{***}	0.11^{***}	0.12^{***}	0.45
	(0.02)	(0.02)	(0.03)	(0.01)	(0.01)	(0.30)
Percent previously punished			2.78	-0.55^{***}	-0.06	1.11
			(1.98)	(0.14)	(0.19)	(1.44)
Electoral competition			8.31	-0.56^{*}	0.50	-2.34
			(5.69)	(0.26)	(0.37)	(1.73)
Num. obs.	1804	1804	1804	1804	1804	1804
Weak instr. test: Demand	0.00	0.00	0.00			
Weak instr. test: Punished			0.00	0.00	0.00	0.01
Weak instr. test: Elec. comp.			0.00	0.00	0.00	0.00
Wu-Hausman test	0.00	0.00	0.00	0.00	0.59	0.00
Sargan test		0.08		0.00		

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

measure rather than its components, or its components and their product.

In the manuscript, we argue that fidelity to the causal logics of an insurance model of judicial independence means scholars should use the *demand for insurance*, that is the product of the risks of losing office and the likelihood of losing office, rather than simply the latter of the two components, as is done in existing work. We further note that this theoretical justification for using the product—in itself a good enough reason to shift empirical work—as opposed to the component terms is corroborated by the results of instrumental variable analysis. Indeed, the above makes clear that regardless of the model specification or the instrument and data used, our demand for insurance measure is also preferable at the empirical level, which again strongly supports our conceptual argument, and suggests we should prefer an operationalization of the insurance model that assesses the demand for insurance.

Table 2: **Alternative measures of demand.** Models of de facto independence using alternative demands for independence. Models 1–4 use the product of the square of the likelihood of losing office (electoral competition) and the risks associated with losing office (percent previous leaders punished).

	Model 1	Model 2	Model 3	Model 4
Intercept	0.33^{***}	-0.91^{***}	-0.96^{***}	-1.00^{***}
	(0.01)	(0.06)	(0.13)	(0.13)
Demand for insurance	0.04^{***}	0.03^{***}	0.40^{**}	0.44^{**}
	(0.01)	(0.00)	(0.15)	(0.15)
log(GDP/capita)		0.30^{***}	0.24^{***}	0.25^{***}
		(0.02)	(0.02)	(0.02)
Years democratic (logged)		-0.00	0.13^{***}	0.12^{***}
		(0.01)	(0.02)	(0.02)
Num. obs.	2217	2216	1854	1804
Weak instruments test			0.00	0.00
Wu-Hausman test			0.00	0.00
Sargan test				0.42

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

Alternative measures of likelihood component of demand for insurance

As we note in the main manuscript, our results are robust to alternative specifications of our demand for insurance measure. As pointed out by an astute reviewer, it is possible that the logic of the demand for insurance only kicks in when some threshold of the likelihood of losing office is reached. In other words, at very low levels of competition, there should be no demand for insurance, regardless of the perceived severity of threat of losing office. There is reason to suspect this is a possibility (and thus examine alternative specifications of demand), as there is some experimental evidence that individuals demand for insurance is low under the condition of low likelihood/high severity of risk, whereas it is comparably higher under conditions of high likelihood/low severity (Ganderton et al. 2000).[5]

To account for this, we use three alternative ways of modeling the demand for insurance, in each aiming to prioritize the importance of losing office over the downside risk should such a loss be realized. In all three, this means modifying our component of demand that captures this likelihood of losing office, i.e. the level of electoral competition. The first way we modify this component is by looking at the square of competition, rather than competition itself. Because the Henisz measure of competition is measured 0–1, we simply use a linear transformation, multiplying this value by 10; our observed values thus range from 0–7.2, rather than 0–0.72.

[5] Where severity is still a meaningful threat, of course, as these experiments were uninterested in examining why individuals fail to purchase insurance for "risks" of trivially low levels.

Table 3: **Alternative measures of demand.** Models of de facto independence using alternative demands for independence. Models 1–4 use the product of the likelihood of losing office (electoral competition) and the risks associated with losing office (percent previous leaders punished) using a threshold for competition. Here, all instances of competition lower than the mean level (0.4) are set to zero.

	Model 1	Model 2	Model 3	Model 4
Intercept	0.34^{***}	-0.91^{***}	-0.73^{***}	-0.75^{***}
	(0.00)	(0.06)	(0.08)	(0.09)
Demand for insurance	0.02^{***}	0.01^{**}	0.50^{*}	0.56^{*}
	(0.01)	(0.00)	(0.21)	(0.22)
log(GDP/capita)		0.30^{***}	0.21^{***}	0.21^{***}
		(0.02)	(0.03)	(0.03)
Years democratic (logged)		-0.00	0.13^{***}	0.12^{***}
		(0.01)	(0.02)	(0.03)
Num. obs.	2217	2216	1854	1804
Weak instruments test			0.01	0.02
Wu-Hausman test			0.00	0.00
Sargan test				0.42

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

We then follow our previous procedure for producing the demand for insurance, taking the product of this measure and the percent of those in office previously punished (adding to the latter a constant of 1, so there still exists some demand even absent any former leader being punished).[6] The results of this alternative specification as presented in Appendix Table 2, which show that our model results hold regardless of this alternative model specification which gives greater weight to the likelihood of losing office.

The second and third alternative ways of modeling the demand for insurance do so by setting some threshold for competition, when not met there is no demand for insurance because the likelihood of losing office is set to zero, and thus the product—our *Demand for insurance* variable–is zero. The first of these we set at 40% of legislative seats being held by a single opposition party, i.e. when the Henisz measure of competition is at least 0.4. That is, if the score for competition is less than 0.4, it is recoded as 0, to suggest the threshold for meaningful electoral threat was not reached. This presents a major challenge to the incumbent, and is a high bar: in one-quarter of U.S. observations, this level is not achieved (recall this is total legislative seats across all relevant legislative bodies, so here the House and Senate). As we do not think there was no risk of losing office for the U.S. President in those years where the

[6]Indeed, to some degree this original formulation already prioritized competition, as by construction when competition was zero there was zero demand, whereas when punishment was zero there was some chance for demand.

Table 4: **Alternative measures of demand.** Models of de facto independence using alternative demands for independence. Models 1–4 use the product of the likelihood of losing office (electoral competition) and the risks associated with losing office (percent previous leaders punished) using a threshold for competition. Here, all instances of competition lower than the mean level (0.4) are set to zero.

	Model 1	Model 2	Model 3	Model 4
Intercept	0.34^{***}	-0.91^{***}	-0.73^{***}	-0.75^{***}
	(0.00)	(0.06)	(0.08)	(0.09)
Demand for insurance	0.02^{***}	0.01^{**}	0.50^{*}	0.56^{*}
	(0.01)	(0.00)	(0.21)	(0.22)
log(GDP/capita)		0.30^{***}	0.21^{***}	0.21^{***}
		(0.02)	(0.03)	(0.03)
Years democratic (logged)		-0.00	0.13^{***}	0.12^{***}
		(0.01)	(0.02)	(0.03)
Num. obs.	2217	2216	1854	1804
Weak instruments test			0.01	0.02
Wu-Hausman test			0.00	0.00
Sargan test				0.42

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

score ranges from 0.35–0.39, this is a tough test for our framework — indeed these years did in fact see alternative in the party of the executive.[7]

As 0.4 is also the observed mean for the competition measure in our data, we employ the first quartile value of competition (0.347) in our third alternative measure of demand, following the same procedure whereby any observed value below that threshold is recoded as zero (and the resulting demand measure takes the product of this value and the percent of previously punished leaders). Results for these two alternative specifications of the demand measure are reported in Appendix Tables 4 and 5, respectively. The results remain similar, further corroborating our focus on the product of the likelihood of losing office and its severity, and providing some evidence that the role of the likelihood of losing office might be the stronger of the two factors. We say only *some* evidence that it *might* be, however, because the correlations between these measures are very high, to such a degree that it is as likely that the results are being driven by the significant overlap between the alternative model specifications. These correlations are reported in Appendix Table 6, where one can observe that the lowest correlation found is that between the original measure and that using a minimum competition value of 0.40, but even here the two measures are still correlated at 0.82.

[7]These are the bulk of the 1960s and the late Carter and early Reagan years.

Table 5: **Alternative measures of demand.** Models of de facto independence using alternative demands for independence. Models 1–4 use the product of likelihood risks of losing office (electoral competition) and the risks associated with losing office (percent previous leaders punished) using a threshold for competition. Here, all instances of competition lower than the first quartile level (0.347) are set to zero.

	Model 1	Model 2	Model 3	Model 4
Intercept	0.33^{***}	-0.92^{***}	-1.25^{***}	-1.19^{***}
	(0.00)	(0.06)	(0.35)	(0.22)
Demand for insurance	0.03^{***}	0.02^{***}	0.79	0.69^{*}
	(0.00)	(0.00)	(0.47)	(0.27)
log(GDP/capita)		0.30^{***}	0.25^{***}	0.25^{***}
		(0.02)	(0.03)	(0.02)
Years democratic (logged)		-0.00	0.10^{*}	0.10^{**}
		(0.01)	(0.04)	(0.03)
Num. obs.	2217	2216	1854	1804
Weak instruments test			0.07	0.05
Wu-Hausman test			0.00	0.00
Sargan test				0.68

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

Table 6: **Correlations between alternative measures of demand.** Correlation matrix of original measure of demand for insurance and three alternative specifications

	Original demand	Competition squared	Competition > 0.40	Competition > 0.347
Original demand	1.00	0.95	0.82	0.90
Competition squared	0.95	1.00	0.87	0.90
Competition > 0.40	0.82	0.87	1.00	0.83
Competition > 0.347	0.90	0.90	0.83	1.00

Alternative weighting for differing forms of punishment

We can additionally examine whether it is the case that our results are being driven by our particular construction of our *Demand for insurance* variable, this time not as a function of how we operationalize the likelihood of losing office, but rather the severity of risk. Recall that we operationalize the severity of losing office as the percent of post-World War 2 leaders who were punished after leaving office, using the Archigos data (Goemans, Gleditsch and Chiozza 2009). In doing so, we treat the three possible punishments contained in these data (imprisoned, exiled, or killed) as having equal weight. We do so in our main manuscript for two reasons, the first prosaic and the second theoretically-driven. The first reason is the ease of presentation: as our goal is to introduce a new way of thinking about the insurance model, and the construction of our measure already takes significant space in our manuscript, keeping this part of the demand variable simple allows for a more straightforward presentation and easy interpretation. The second reason is that there is no clear and obvious reason to think that a given leader would be paying greater attention to the nature of punishment, rather than the incidence of punishment itself. That is, there is no clear reason to assume the clarity from the differential signal of form of punishment is meaningful, or greater than the signal from punishment in general. Nonetheless, the space provided here allows us to break down this aspect of our *Demand for insurance* measure, and assess whether our results are being driven by some particular form of punishment.

Here, we make the plausible assumption that for any leader, there is a preference to being free in exile rather than in prison domestically, and in prison over being killed. As such, we can then reweight the post-tenure fates of former leaders, decreasing the weights applied to imprisonment compared to punishment, and decreasing even more the weights applied to exile compared to punishment. Given our component of demand assessing the severity of risk is the percentage of former leaders who were punished writ large, we can just apply some fractional weight to those who suffered a lesser punishment than being killed. If, for example, we weight imprisonment at three-quarters the severity of being killed, then it would take four former leaders being imprisoned to produce the same percent punished as three former leaders being killed. Similarly, if it is the case that a life as Professor and Senior Researcher at the Prigogine Institute for Mathematical Investigations of Complex Systems at Moscow State University (the comparably cozy fate of Askar Akayev after his exile from Kyrgyzstan in 2005) is far less onerous than imprisonment or death, we could weight such an outcome as half as severe as being killed.

This is precisely what we do in appendix Table 7. It might, however, be the case that we are not downweighing the comparable lack of severity of such outcomes as much as we could. To address this, we also weight imprisonment as half as severe as being killed, and exile as one-fourth as severe. As such, it would require two former leaders imprisoned to weigh as equally as one former leader killed, and/or four former leaders exiled to weigh as equally as one former leader killed. The results of this second alternative specification are presented in appendix Table 8. Note: in both specifications, we also comparably reweight the instrument for the severity of losing office (the percent of former leaders of neighboring countries punished) following the same process. As can be seen looking at Tables 7 and 8, the results of models using both alternative specifications closely resemble those with our original measure, suggesting our results are robust to moderate respecification of our *Demand for insurance* measure. Much like the above respecification of the likelihood component of demand, however, the respecifications for the severity component show high levels of correlation with our original measure, with the lowest of the pairwise correlations for the three measures of demand using varying severity

Table 7: **Reweighting the severity of losing office.** Four models of de facto independence and the demand for insurance. Each model reproduces the comparable model in the main manuscript, but this times weights the imprisonment of former leaders as being three-quarters as severe as being killed, and exile as half as severe.

	Model 1	Model 2	Model 3	Model 4
Intercept	0.31^{***}	-0.94^{***}	-1.37^{***}	-1.46^{***}
	(0.01)	(0.06)	(0.27)	(0.27)
Demand for insurance	0.03^{***}	0.03^{***}	0.46^{**}	0.51^{**}
	(0.01)	(0.00)	(0.17)	(0.17)
log(GDP/capita)		0.30^{***}	0.27^{***}	0.27^{***}
		(0.02)	(0.02)	(0.02)
Years democratic (logged)		-0.01	0.10^{***}	0.09^{**}
		(0.01)	(0.03)	(0.03)
Num. obs.	2217	2216	1854	1804
Weak instruments test			0.00	0.00
Wu-Hausman test			0.00	0.00
Sargan test				0.57

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

Table 8: **Reweighting the severity of losing office.** Four models of de facto independence and the demand for insurance. Each model reproduces the comparable model in the main manuscript, but this times weights the imprisonment of former leaders as being half as severe as being killed, and exile as one-fourth as severe.

	Model 1	Model 2	Model 3	Model 4
Intercept	0.31^{***}	-0.94^{***}	-1.41^{***}	-1.50^{***}
	(0.01)	(0.06)	(0.29)	(0.29)
Demand for insurance	0.03^{***}	0.03^{***}	0.45^{**}	0.50^{**}
	(0.01)	(0.00)	(0.17)	(0.17)
log(GDP/capita)		0.30^{***}	0.28^{***}	0.28^{***}
		(0.02)	(0.02)	(0.02)
Years democratic (logged)		-0.01	0.12^{***}	0.11^{***}
		(0.01)	(0.02)	(0.02)
Num. obs.	2217	2216	1854	1804
Weak instruments test			0.00	0.00
Wu-Hausman test			0.00	0.00
Sargan test				0.32

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

being $r = 0.96$.

Table 9: **Considering the past ten years of former leader fates.** Four models of de facto independence and the demand for insurance. These reproduce the models in main manuscript, but consider only the past decade of leader punishment rather than all (to date) post-war leader fates.

	Model 1	Model 2	Model 3	Model 4
Intercept	0.33^{***}	-0.92^{***}	-1.69^{**}	-1.38^{***}
	(0.01)	(0.06)	(0.57)	(0.24)
Demand for insurance	0.01^{*}	0.01^{*}	0.72	0.50^{**}
	(0.01)	(0.00)	(0.39)	(0.17)
log(GDP/capita)		0.30^{***}	0.26^{***}	0.26^{***}
		(0.02)	(0.03)	(0.02)
Years democratic (logged)		-0.00	0.13^{***}	0.13^{***}
		(0.01)	(0.03)	(0.02)
Num. obs.	2217	2216	1854	1804
Weak instruments test			0.05	0.01
Wu-Hausman test			0.00	0.00
Sargan test				0.29

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

Alternative weighting for differing time periods for previous punishment

To further examine how robust our model of demand for insurance is to alternative specifications, we additionally reexamine our proxy for the severity of the risks associated with losing office by considering alternative time periods of former leader punishment. In the main manuscript, we consider the percentage of total (to date) previous post-World War 2 leaders who were punished (for reasoning regarding this specification, see the main article). It is possible, however, that executives consider the recent past rather than the entirety of this time span. To assess this, we reproduce the four models presented in the main manuscript, this time considering alternative time spans of former leader punishment. Note: in both specifications, we also comparably reweight the instrument for the severity of losing office (the percent of former leaders of neighboring countries punished) following the same process.

Appendix Tables 9 and 10 use only the previous 10 and 20 years, respectively, of former leaders in the creation of the proxy for the severity of losing office. We consider each a quite difficult test of the robustness of our results, because it takes an already-uncommon event (punishment of democratic executives after leaving office) and constrains it to a very limited amount of time. As such, we expect more difficulty in finding significant results, which is to a limited extent true. In both appendix Tables 9 and 10, the instrumented *Demand for insurance* variable falls just short of statistical significance in Model 3. Nonetheless, it remains statistically significant with highly comparable point estimates in the linear models with country fixed ef-

Table 10: **Considering the past twenty years of former leader fates.** Four models of de facto independence and the demand for insurance. These reproduce the models in main manuscript, but consider only the past two decades of leader punishment rather than all (to date) post-war leader fates.

	Model 1	Model 2	Model 3	Model 4
Intercept	0.32^{***}	-0.94^{***}	-1.95^{**}	-1.72^{***}
	(0.01)	(0.06)	(0.74)	(0.41)
Demand for insurance	0.02^{***}	0.02^{***}	0.77	0.63^{*}
	(0.01)	(0.00)	(0.45)	(0.24)
log(GDP/capita)		0.30^{***}	0.29^{***}	0.28^{***}
		(0.02)	(0.03)	(0.03)
Years democratic (logged)		-0.00	0.14^{***}	0.14^{***}
		(0.01)	(0.03)	(0.03)
Num. obs.	2217	2216	1854	1804
Weak instruments test			0.07	0.03
Wu-Hausman test			0.00	0.00
Sargan test				0.47

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

fects (Models 1 and 2), as well as our preferred Model 4, which uses multiple instruments for demand to allow for a Sargan test of the exogeneity of the instruments.

While these results show our model is arguably less robust than previous robustness checks, we find this to be entirely explicable given the data differences: the mean value of percentage of previously punished leaders in the post-war era is 0.16, while considering only the previous 10 and 20 years this mean drops to 0.08 and 0.11, respectively. Naturally, decreasing the salience of one of the key components of the demand for insurance should make it more difficult to estimate significant effects.

Combining weights for time period and form of punishment

As a final robustness check, we combine the previous two alternative ways of operationalizing the severity of punishment, assessing both 10- and 20-year time periods for previous leaders and reweighting the severity based on the form of punishment. The logic here is that if one thinks that the immediate past is all that matters, and so too does the form of punishment, then these two factors should both be considered when constructing our proxy for the severity of punishment. Note: in each instance, we again respecify our instrument for severity by matching the procedure used to create the severity measure itself. Therefore, if we assess, for example, low weights for imprisonment and exile (0.5 and 0.25 compared to being killed, respectively) *and* do so only over the previous ten years, when we create the instrument using

Table 11: **Considering both time and form of punishment in measuring severity of losing office.** Four models of de facto independence and the demand for insurance, each comparable to the models used in the main manuscript. Here, however, only the previous 10 years of former leaders fates are considered, and in each the weight of imprisonment and exile are 0.75 and 0.5, respectively.

	Model 1	Model 2	Model 3	Model 4
Intercept	0.32^{***}	-0.93^{***}	-1.41^{***}	-1.32^{***}
	(0.01)	(0.06)	(0.33)	(0.21)
Demand for insurance	0.02^{***}	0.01^{**}	0.55^{*}	0.48^{**}
	(0.01)	(0.00)	(0.23)	(0.15)
log(GDP/capita)		0.30^{***}	0.25^{***}	0.25^{***}
		(0.02)	(0.02)	(0.02)
Years democratic (logged)		-0.00	0.11^{***}	0.11^{***}
		(0.01)	(0.03)	(0.03)
Num. obs.	2217	2216	1854	1804
Weak instruments test			0.01	0.00
Wu-Hausman test			0.00	0.00
Sargan test				0.50

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

the percentage of former leaders in neighboring countries punished after leaving office, then we also reweight imprisonment and exile, and look only at the previous ten years.

The results of these are presented across appendix Tables 11, 12, 13, and 14, which respectively assess: the previous 10 years with weights for imprisonment and exile of 0.75/0.5, the previous 10 years with weights of 0.5/0.25, the previous 20 years with weights of 0.75/0.5, and the previous 20 years with weights of 0.5/0.25. The robustness of these results across all sixteen of these model specifications is impressive: in every combination, the demand for insurance remains substantively and statistically significant, and in all versions of Model 4 across Tables 11, 12, 13, and 14, which employ multiple instruments the model excels at weak instruments tests, Wu-Hausman tests, and Sargan tests.

Table 12: **Considering both time and form of punishment in measuring severity of losing office.** Four models of de facto independence and the demand for insurance, each comparable to the models used in the main manuscript. Here, however, only the previous 10 years of former leaders fates are considered, and in each the weight of imprisonment and exile are 0.5 and 0.25, respectively

	Model 1	Model 2	Model 3	Model 4
Intercept	0.32^{***}	-0.93^{***}	-1.30^{***}	-1.27^{***}
	(0.01)	(0.06)	(0.25)	(0.18)
Demand for insurance	0.02^{***}	0.02^{***}	0.49^{**}	0.46^{***}
	(0.01)	(0.01)	(0.19)	(0.13)
log(GDP/capita)		0.30^{***}	0.25^{***}	0.25^{***}
		(0.02)	(0.02)	(0.02)
Years democratic (logged)		-0.00	0.09^{**}	0.09^{***}
		(0.01)	(0.03)	(0.03)
Num. obs.	2217	2216	1854	1804
Weak instruments test			0.00	0.00
Wu-Hausman test			0.00	0.00
Sargan test				0.64

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

Table 13: **Considering both time and form of punishment in measuring severity of losing office.** Four models of de facto independence and the demand for insurance, each comparable to the models used in the main manuscript. Here, however, only the previous 20 years of former leaders fates are considered, and in each the weight of imprisonment and exile are 0.75 and 0.5, respectively.

	Model 1	Model 2	Model 3	Model 4
Intercept	0.32^{***}	-0.94^{***}	-1.52^{***}	-1.48^{***}
	(0.01)	(0.06)	(0.37)	(0.27)
Demand for insurance	0.03^{***}	0.02^{***}	0.55^{*}	0.52^{**}
	(0.01)	(0.00)	(0.24)	(0.17)
log(GDP/capita)		0.30^{***}	0.27^{***}	0.27^{***}
		(0.02)	(0.02)	(0.02)
Years democratic (logged)		-0.00	0.11^{***}	0.11^{***}
		(0.01)	(0.03)	(0.03)
Num. obs.	2217	2216	1854	1804
Weak instruments test			0.01	0.00
Wu-Hausman test			0.00	0.00
Sargan test				0.68

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

Table 14: **Considering both time and form of punishment in measuring severity of losing office.** Four models of de facto independence and the demand for insurance, each comparable to the models used in the main manuscript. Here, however, only the previous 20 years of former leaders fates are considered, and in each the weight of imprisonment and exile are 0.5 and 0.25, respectively

	Model 1	Model 2	Model 3	Model 4
Intercept	0.31^{***}	-0.93^{***}	-1.34^{***}	-1.35^{***}
	(0.01)	(0.06)	(0.26)	(0.21)
Demand for insurance	0.03^{***}	0.02^{***}	0.47^{**}	0.47^{***}
	(0.01)	(0.00)	(0.18)	(0.14)
log(GDP/capita)		0.30^{***}	0.26^{***}	0.26^{***}
		(0.02)	(0.02)	(0.02)
Years democratic (logged)		-0.00	0.10^{***}	0.09^{***}
		(0.01)	(0.03)	(0.03)
Num. obs.	2217	2216	1854	1804
Weak instruments test			0.00	0.00
Wu-Hausman test			0.00	0.00
Sargan test				0.82

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

Supplemental information references

Ganderton, Philip T, David S Brookshire, Michael McKee, Steve Stewart and Hale Thurston. 2000. "Buying Insurance for Disaster-Type Risks: Experimental Evidence." *Journal of Risk and Uncertainty* 20(3):271–289.

Goemans, Hein, Kristian S. Gleditsch and Giacomo Chiozza. 2009. "Introducing Archigos: A Data Set of Political Leaders." *Journal of Peace Research* 46(2):269–83.