Auxiliary Material

'Democracy doesn't always happen over night: regime change in stages and economic growth' (Boese-Schlosser and Eberhardt, forthcoming in *The Review of Economics and Statistics*)

This document contains additional analysis on the determinants of episode failure mentioned in the 'Empirical Results' section (paragraph on 'Successful and Failed Episodes') of the published paper.

An Early Warning System for Episode Failure

We study the determinants of episode failure with an 'early warning system' approach borrowed from the literature on banking crises (e.g. Eberhardt & Presbitero 2021). Our empirical results derive from a random effects logit model augmented with country-specific means of all covariates (Mundlak 1978, Chamberlain 1982) — this allows us to interpret coefficients as estimate 'within-country' estimates. We estimate a latent 'crisis' model, where the observed variable is a realized systemic event when the latent variable exceeds some threshold: we code the dependent variable as equal to one in the year a country exits an episode without transition to a democratic regime, and zero otherwise. With reference to our PCDID analysis, our sample includes all 'treated' countries and those in 'control sample 2'.

Determinants Our analysis focuses on economic and social conditions alongside 'upheaval' in an early warning system for episode failure.[1] We consider determinants from six different groups (for data sources see Online Appendix A of the paper): First, we control for *changes in economic conditions* using per capita GDP growth and the change in export/trade. Second, given the relevance of the *military* in the context of regime fragility, we capture the buildup of military personnel using data from COW. Third, *natural disasters* may devastate economic

[1] Since all regressions capture 'within-country' effects, levels variables (e.g. per capita income) are discounted in favour of changes or 'trigger' crisis variables (see Eberhardt & Presbitero 2021).

and social conditions in a country and foster the return to a 'strong' (autocratic) leader or accelerate a transition to democracy if incumbent systems are shown to be failing. Fourth, *financial crises* (banking or currency crises) can similarly undermine episodes if the uncertainty created by the political process is blamed for financial fragility. Fifth, the relationship between *natural resources* and political institutions is a fraught one, and we include an indicator for oil booms to capture this variant of the natural resource curse. Finally, since *coups attempts* relate to democratic failure they make a natural candidate to study episode failure.[2]

Variable transformation An important aspect of our empirical investigation is how we model the pre-failure period: simply lagging all regressors by one time period may ignore slow-moving processes contributing to episode failure, e.g. lacklustre economic performance *over several years*. We therefore provide additional results following the practice in Eberhardt & Presbitero (2021), among others, using moving averages of regressors to capture pre-failure dynamics: we adopt MA(2) to MA(5) transformations, with the MA(2) defined as the mean of values at $t-1$ and $t-2$, and the higher-order MAs adding further lags accordingly. Higher-order MA-transformations come with the potential caveats that they 'wash out' spikes, while for shorter episodes this transformation may conflate events *during* and *before* the episode.[3]

Results Our results in Table 1 are presented as average marginal effects (in %): for continuous variables we multiply the margins with the standard deviation while for binary indicators we report magnitudes for a shift from 0 to 1. Standard errors are computed using the Delta method, those in the underlying RE-Logit regressions are clustered at the country-level. Wald tests uniformly confirm the inclusion of within-country averages.

Across specifications the effect of military buildup has a consistent and statistically significant attenuating effect on episode failure: a one standard deviation increase in military

[2]Our regressions do not include a measure for episodes since this would per fectly predict failure.

[3]We estimate these models for the most elaborate specification. Using a comparison of ROC areas suggests that this model has borderline ($p = .10$) higher predictive power than the model excluding coup attempts.

personnel is associated with a 0.2-0.3 percentage point decrease in the propensity of a failure. A second consistent and statistically significant trigger of episode failure is the presence of an oil boom, underlining the detrimental effects of natural resources in the political process. Finally, in models allowing for a longer time horizon in effects to accumulate — the MA(3)-MA(5) models — banking crises seem to reduce the propensity of episode failure. In economic terms the effects of banking crises and oil booms are an order of magnitude larger than that of a change in military personnel. Economic effects of all other variables studied are small and/or imprecisely estimated. Most notably, although with the correct positive sign, the presence of coup attempts has a comparatively small effect and is estimated imprecisely.

References

Chamberlain, G. (1982), 'Multivariate regression models for panel data', *Journal of Econometrics* **18**(1), 5–46.

Eberhardt, M. & Presbitero, A. F. (2021), 'Commodity prices and banking crises', *Journal of International Economics* **131**, 103474.

Mundlak, Y. (1978), 'On the pooling of time series and cross section data', *Econometrica* **46**(1), 69–85.

Table 1: Determinants of Episode Failure (RE Mundlak Logit)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Transformation	Lagged by one period					MA(2)	MA(3)	MA(4)	MA(5)
Dep. Variable	Failure of Episode at time t								
Continuous Covariates (magnitudes in percent for 1sd increase)									
Δ Military	-0.180	-0.190	-0.172	-0.183	-0.171	-0.138	-0.228	-0.285	-0.361
personnel	(1.89)*	(1.95)*	(1.86)*	(2.05)**	(1.95)*	(1.17)	(1.70)*	(1.83)*	(2.34)**
Δ Export/Trade	-0.161	-0.161	-0.160	-0.153	-0.148	-0.111	-0.110	0.049	-0.013
	(0.86)	(0.87)	(0.88)	(0.84)	(0.86)	(0.59)	(0.51)	(0.24)	(0.06)
GDPpc	0.213	0.227	0.198	0.200	0.221	0.104	0.083	0.026	0.007
growth rate	(1.27)	(1.31)	(1.21)	(1.20)	(1.32)	(0.41)	(0.37)	(0.13)	(0.03)
Binary Indicators (magnitudes in percent)									
Natural disaster		0.269	0.251	0.278	0.255	-0.104	-0.624	-0.047	0.429
		(0.19)	(0.18)	(0.20)	(0.18)	(0.12)	(0.76)	(0.08)	(1.07)
Banking crisis			-2.596	-2.556	-2.577	-1.579	-2.738	-2.823	-2.149
			(1.03)	(1.02)	(1.03)	(1.07)	(1.85)*	(2.27)**	(2.43)**
Currency crisis			-0.277	-0.382	-0.354	-0.342	0.395	0.21	0.662
			(0.23)	(0.31)	(0.29)	(0.36)	(0.55)	(0.34)	(1.38)
Oil boom				3.676	3.655	1.768	2.770	2.329	2.858
				(1.89)*	(1.90)*	(0.94)	(1.84)*	(1.71)*	(2.14)**
Coup Attempt					0.646	0.329	0.315	0.250	0.292
					(0.86)	(0.67)	(0.88)	(0.83)	(1.09)
Observations	4,905	4,905	4,905	4,905	4,905	5,008	5,111	5,214	5,316
Countries	102	102	102	102	102	102	102	102	102
Events	125	125	125	125	125	127	130	132	135
LogL	-577.11	-575.04	-571.53	-569.86	-566.82	-577.34	-586.11	-595.23	-605.27
AUROC	0.584	0.591	0.615	0.628	0.647	0.641	0.664	0.663	0.668
se(AUROC)	0.025	0.026	0.024	0.025	0.024	0.024	0.023	0.022	0.023
Wald chi^2 (FE)	7.45	12.22	19.10	22.55	26.57	25.82	27.09	26.20	22.19
Wald p-value	0.059	0.016	0.004	0.002	0.001	0.001	0.001	0.001	0.005
Uncond. Prob.	2.55%	2.55%	2.55%	2.55%	2.55%	2.55%	2.55%	2.55%	2.55%

Notes: The table presents results (marginal effects) of a Random Effects 'Mundlak-Chamberlain' Logit regression, see text for details. We run 'early warning system' regressions akin to the analysis in the literature on banking crises. Our dependent variable is a dummy for the failure of a democratic episode in year t. All regressors in columns (1) to (5) are lagged one period $(t-1)$, those in columns (6)-(9) are MA-tranformed as indicated at the top of the table. All results for continuous variables are marginal effects (in percent) for a 1 standard deviation increase in the covariate, e.g. -.180 in column (1) indicates that a 1sd increase in military personnel is associated with a 0.18 percentage point reduction in the propensity of an episode failure. The results for binary indicators are also in percent and for a shift from 0 to 1. The unconditional propensity is 2.55%. 'Events' refers to the number of episode failures in the sample, which comprises the treatment sample and control sample 2 from our main analysis. LogL reports the maximised log likelihood. AUROC is the Area under the receiver operating curve (a higher value indicates higher predictive power; 0.5 is the lower benchmark, 1 the maximum), with the associated standard error also reported. The Wald test analyses the within-country average terms and implicitly tests whether country FE 'matter' (H_0: FE jointly insignificant). Absolute t-ratios in parentheses. Statistical significance is indicated using * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.