Annex F: State Size and Democracy
Usage of Alternative Indicator for State Size: V-Dem Sample

Wee Chin Hin, Winston
National University of Singapore, Department of Political Science

November 5, 2023

1 Introduction

We recall that the composite measure for state size comprises two time-varying variables: population size and land area. I construct this measure to account for the high collinearity in any models that include both population size and land area as predictors. This measure thus ensures that any models used will control for both population size and geographic land area. The formula for the composite measure is formally:

$$\text{StateSize}_{\text{composite}} = \ln \sqrt{\{\frac{\text{Population}_{i,t} - \overline{\text{Population}_i}}{\sigma_{\text{Population}}} - \text{floor}[\min(\frac{\text{Population}_{i,t} - \overline{\text{Population}_i}}{\sigma_{\text{Population}}})\} \times \{\frac{\text{LandArea}_{i,t} - \overline{\text{LandArea}_i}}{\sigma_{\text{LandArea}}} - \text{floor}[\min(\frac{\text{LandArea}_{i,t} - \overline{\text{LandArea}_i}}{\sigma_{\text{LandArea}}})]\}}$$

We subtract the floor values of the minimum values of the Z-score of each component in this composite indicator to shift the Z-scores of both components above zero, since the natural log function typically cannot take any negative values, which can only be defined in terms of complex numbers. Because some states exhibit odd characteristics of having large population sizes over a compact area, or minute population dispersed over a vast area, we take the geometric mean of both components so that the composite measure is not biased towards either component. To account for skewness, we then take the natural log of the resultant value. This annex uses the composite measure for state size as the primary predictor in the following models.

Consequently, the results will be expressed in terms of 10% changes in this composite indicator. Let A be the shifted Z-score of population size and B be to the shifted Z-score of land area. The changes in the dependent variable is formally expressed as:

$$\begin{gathered} \mathrm{Y}_1 = \beta_1 \ln \sqrt{\mathrm{AB}} \\ \mathrm{Y}_2 = \beta_1 \ln(1.1 \times \sqrt{\mathrm{AB}}) = \beta_1 \ln \sqrt{1.21 \times \mathrm{AB}} \\ \Delta \mathrm{Y} = \mathrm{Y}_2 - \mathrm{Y}_1 = \beta_1(\ln(1.1 \times \sqrt{\mathrm{AB}}) - \ln \sqrt{\mathrm{AB}}) = \beta_1(\ln 1.1 + \ln \sqrt{\mathrm{AB}} - \ln \sqrt{\mathrm{AB}}) = \beta_1 \ln 1.1 \end{gathered}$$

Given the set of equations above, a 10% increase in the non log-transformed measurement is actually analogous to the increase in the product of the two shifted Z-scores by 21%. For interpretability, this is also analogous to the increase in the product of both population size and land area by 21%. This annex will, however, only parse changes in the dependent variable in terms of 10% increases in the non log-transformed composite state size measure (i.e. the square-root portion of the indicator).

As an aside, this composite measure is nothing but running the regression on both logged population and logged land area, and yielding the standardised coefficients for both. As a proof, we first regress Y on the natural log of the root of AB. By concatenating logged population and logged land area into a single predictor, we eliminate collinearity arising from the two variables. We observe that:

$$\begin{gathered} \mathrm{Y} = \beta \ln \sqrt{\mathrm{AB}} = \beta \ln(\sqrt{\mathrm{A}}\sqrt{\mathrm{B}}) \\ \mathrm{Y} = \beta(\ln \sqrt{\mathrm{A}} + \ln \sqrt{\mathrm{B}}) = \beta(0.5 \ln \mathrm{A} + 0.5 \ln \mathrm{B}) \\ \therefore \mathrm{Y} = \beta_1 \ln \mathrm{A} + \beta_2 \ln \mathrm{B} \end{gathered}$$

2 Empirical Strategy 1: Regression

Table F1: Relationship between Composite Measure for State Size and Participatory Democracy

	v2x_partipdem					
	Exposure-Outcome, Pooled			Pooled (No Interaction)	Panel FE (No Interaction)	Pooled (Interaction)
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Composite: State Size	0.1334***	0.1214***	0.0970***	0.0881***	0.3343***	0.0921*
	(0.0128)	(0.0131)	(0.0124)	(0.0127)	(0.0973)	(0.0454)
	p = 0.0000	p = 0.0000	p = 0.0000	p = 0.0000	p = 0.0006	p = 0.0428
Clientelism		−0.2686***		−0.2275***	−0.2869***	−0.1167
		(0.0538)		(0.0470)	(0.0410)	(0.0632)
		p = 0.000001		p = 0.000002	p = 0.0000	p = 0.0648
Military Participation Rate			−0.0143***	−0.0138***	−0.0001	0.0110*
			(0.0017)	(0.0017)	(0.0008)	(0.0046)
			p = 0.0000	p = 0.0000	p = 0.9052	p = 0.0178
Urbanisation (% of Population)	−0.1841***	−0.1357**	−0.0391	−0.0033	0.5224***	0.2435***
	(0.0484)	(0.0493)	(0.0439)	(0.0435)	(0.0699)	(0.0642)
	p = 0.0002	p = 0.0060	p = 0.3721	p = 0.9391	p = 0.0000	p = 0.0002
Per Capita GDP (Logged)	0.1010***	0.0413**	0.0946***	0.0443***	0.0071	0.0393*
	(0.0109)	(0.0158)	(0.0101)	(0.0146)	(0.0120)	(0.0156)
	p = 0.0000	p = 0.0090	p = 0.0000	p = 0.0025	p = 0.5526	p = 0.0119
Government Expenditure (% of GDP)	0.0242	−0.0417	0.0002	−0.0547	−0.1491**	−0.1170
	(0.1060)	(0.0995)	(0.0930)	(0.0909)	(0.0574)	(0.0868)
	p = 0.8194	p = 0.6754	p = 0.9980	p = 0.5474	p = 0.0095	p = 0.1779
Foreign Aid Received (% of GNI)	0.0638	−0.1084	0.0337	−0.1111	−0.0597	0.1146
	(0.1675)	(0.1639)	(0.1558)	(0.1559)	(0.0548)	(0.1607)
	p = 0.7031	p = 0.5085	p = 0.8289	p = 0.4761	p = 0.2760	p = 0.4758
Resource Dependence	−0.0010	0.0012	−0.0020**	−0.0001	−0.0016*	0.0004
	(0.0009)	(0.0011)	(0.0008)	(0.0010)	(0.0006)	(0.0012)
	p = 0.2277	p = 0.2832	p = 0.0097	p = 0.9508	p = 0.0134	p = 0.7084
Ethnic Fractionalization	−0.0784*	−0.1355***	−0.0366	−0.0865*		−0.0702*
	(0.0395)	(0.0401)	(0.0357)	(0.0372)		(0.0345)
	p = 0.0475	p = 0.0008	p = 0.3056	p = 0.0202		p = 0.0418
Islamic	−0.2254***	−0.2307***	−0.1975***	−0.2029***		−0.2009***
	(0.0287)	(0.0305)	(0.0263)	(0.0276)		(0.0290)
	p = 0.0000	p = 0.0000	p = 0.0000	p = 0.0000		p = 0.0000
Population × Urbanization						−0.0289
						(0.0605)
						p = 0.6326
Clientelism × Urbanization						−0.2127*
						(0.0985)
						p = 0.0309
MPR × Urbanization						−0.0345***
						(0.0058)
						p = 0.0000
Constant	−0.4010***	0.2180	−0.3787***	0.1449	−0.0011	0.0391
	(0.0882)	(0.1524)	(0.0795)	(0.1397)	(0.0178)	(0.1445)
	p = 0.00001	p = 0.1527	p = 0.000002	p = 0.2996	p = 0.9495	p = 0.7868
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	No	No	No	No	Yes	No
N	1204	1204	1204	1204	1204	1204
R-squared	0.5003	0.5279	0.6037	0.6234	0.5456	0.6563
Adj. R-squared	0.4764	0.5049	0.5844	0.6047	0.5128	0.6383
Residual Std. Error	0.1425 (df = 1148)	0.1386 (df = 1147)	0.1270 (df = 1147)	0.1238 (df = 1146)	0.0516 (df = 1122)	0.1185 (df = 1143)
F Statistic	20.8982*** (df = 55; 1148)	22.9036*** (df = 56; 1147)	31.2032*** (df = 56; 1147)	33.2792*** (df = 57; 1146)	16.6295*** (df = 81; 1122)	36.3821*** (df = 60; 1143)

***p < .005; **p < .01; *p < .05

Table F2: Relationship between Composite Measure for State Size and Military Participation Rate

	milrate		
	Pooled (No Interaction)	**Panel FE (No Interaction)**	**Pooled (Interaction)**
	Model 1	**Model 2**	**Model 3**
Composite: State Size	−2.4146***	4.8697	−0.3519
	(0.3244)	(3.2117)	(1.0995)
	p = 0.0000	p = 0.1295	p = 0.7490
Clientelism	2.9770	11.6418***	0.1341
	(2.1384)	(2.7363)	(2.2951)
	p = 0.1639	p = 0.00003	p = 0.9535
Urbanisation (% of Population)	9.5976***	−3.2474	7.2753***
	(1.4806)	(3.3127)	(2.1087)
	p = 0.0000	p = 0.3270	p = 0.0006
Per Capita GDP (Logged)	0.2147	1.0858	0.3166
	(0.4600)	(0.7696)	(0.4584)
	p = 0.6408	p = 0.1583	p = 0.4897
Government Expenditure (% of GDP)	−0.9452	1.5073	−0.9565
	(3.7470)	(2.4459)	(3.6265)
	p = 0.8009	p = 0.5378	p = 0.7920
Foreign Aid Received (% of GNI)	−0.1982	0.4059	1.7761
	(4.2331)	(3.7209)	(4.2720)
	p = 0.9627	p = 0.9132	p = 0.6776
Resource Dependence	−0.0922**	−0.0784*	−0.0707*
	(0.0339)	(0.0389)	(0.0335)
	p = 0.0066	p = 0.0440	p = 0.0348
Ethnic Fractionalization	3.5540**		2.9871*
	(1.2932)		(1.3150)
	p = 0.0060		p = 0.0232
Islamic	2.0111***		1.1739
	(0.6235)		(0.7121)
	p = 0.0013		p = 0.0993
Population × Urbanization			−2.8287
			(1.7885)
			p = 0.1138
Clientelism × Urbanization			5.2307
			(3.8310)
			p = 0.1722
Constant	−5.3001	−0.3095	−4.5061
	(5.1305)	(0.7754)	(4.9887)
	p = 0.3016	p = 0.6899	p = 0.3664
Year FE	Yes	Yes	Yes
Country FE	No	Yes	No
N	1204	1204	1204
R-squared	0.2819	0.1853	0.2857
Adj. R-squared	0.2469	0.1273	0.2495
Residual Std. Error	4.5191 (df = 1147)	2.8489 (df = 1123)	4.5112 (df = 1145)
F Statistic	8.0421*** (df = 56; 1147)	3.1933*** (df = 80; 1123)	7.8957*** (df = 58; 1145)

***p < .005; **p < .01; *p < .05

Table F3: Relationship between Composite Measure for State Size and Clientelism

	v2xnp_client		
	Pooled (No Interaction)	**Panel FE (No Interaction)**	**Pooled (Interaction)**
	Model 1	**Model 2**	**Model 3**
Composite: State Size	−0.0391***	0.0938	−0.1077*
	(0.0113)	(0.1017)	(0.0469)
	p = 0.0006	p = 0.3565	p = 0.0218
Military Participation Rate	0.0023	0.0051***	0.0050
	(0.0015)	(0.0010)	(0.0045)
	p = 0.1278	p = 0.0000004	p = 0.2617
Urbanisation (% of Population)	0.1574***	0.1035	0.1906***
	(0.0434)	(0.0647)	(0.0542)
	p = 0.0003	p = 0.1095	p = 0.0005
Per Capita GDP (Logged)	−0.2210***	−0.0822***	−0.2258***
	(0.0089)	(0.0147)	(0.0099)
	p = 0.0000	p = 0.0000001	p = 0.0000
Government Expenditure (% of GDP)	−0.2415***	−0.5974***	−0.2620***
	(0.0791)	(0.1186)	(0.0836)
	p = 0.0023	p = 0.0000005	p = 0.0018
Foreign Aid Received (% of GNI)	−0.6365***	−0.1266	−0.6559***
	(0.1791)	(0.1484)	(0.1910)
	p = 0.0004	p = 0.3936	p = 0.0006
Resource Dependence	0.0085***	−0.0004	0.0086***
	(0.0011)	(0.0014)	(0.0012)
	p = 0.0000	p = 0.7887	p = 0.0000
Ethnic Fractionalization	−0.2193***		−0.2091***
	(0.0273)		(0.0286)
	p = 0.0000		p = 0.0000
Islamic	−0.0240		−0.0062
	(0.0278)		(0.0298)
	p = 0.3872		p = 0.8346
Population × Urbanization			0.1030
			(0.0634)
			p = 0.1045
MPR × Urbanization			−0.0038
			(0.0056)
			p = 0.4943
Constant	2.3013***	−0.0122	2.3226***
	(0.0738)	(0.0150)	(0.0790)
	p = 0.0000	p = 0.4173	p = 0.0000
Year FE	Yes	Yes	Yes
Country FE	No	Yes	No
N	1204	1204	1204
R-squared	0.7729	0.4492	0.7743
Adj. R-squared	0.7618	0.4100	0.7629
Residual Std. Error	0.1244 (df = 1147)	0.0596 (df = 1123)	0.1241 (df = 1145)
F Statistic	69.7094*** (df = 56; 1147)	11.4487*** (df = 80; 1123)	67.7386*** (df = 58; 1145)

***p < .005; **p < .01; *p < .05

3 Empirical Strategy 2: Causal Mediation Analysis

3.1 H_2: The long-term operation of clientelism

Figure F1 shows the relationship between the new state size measure used on democracy which is operated by clientelism, denoted by the ACME. Unlike the other models in the main thesis and in the supplementary materials for the V-Dem sample, the ACME for the composite state size measure is significant in Figure F1 at 0.00869. This means that a 10% increase in the composite state size measure is associated with, on average, an increase in the V-Dem Participatory Democracy Score by 0.000854 units. Recalling that the relationship between the composite measure for state size and clientelism is negative (Model 1, Table F3) and the relationship between clientelism and democracy is negative (Model 4, Table F1), we can thus observe that both population size and the geographic size of the state diminish the negative impact of clientelism on democracy. Consequently, higher levels of democracy are thus achieved in more populous and geographically larger states.

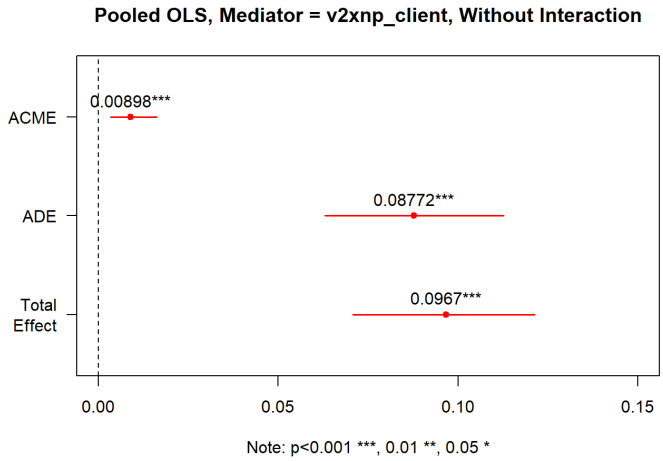

Figure F1: Long-term operation of clientelism.

3.2 H_{2a}: The short-term operation of clientelism

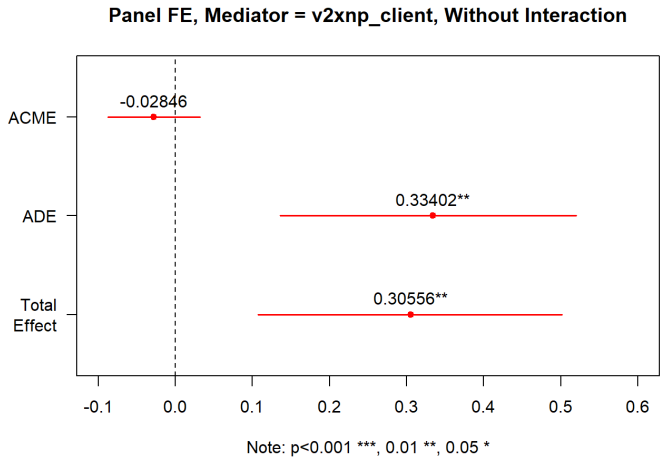

Figure F2: Short-term operation of clientelism.

Figure F2 shows the short-term effect of state size on democracy which is operated by clientelism, denoted by the ACME. Unlike models where only population size is used, the ACME in this model is

not significant. We cannot reject the null hypothesis that clientelism does not operate the joint effect of both population size and geographic area on democracy.

3.3 H_{2b}: The conditional operation of clientelism

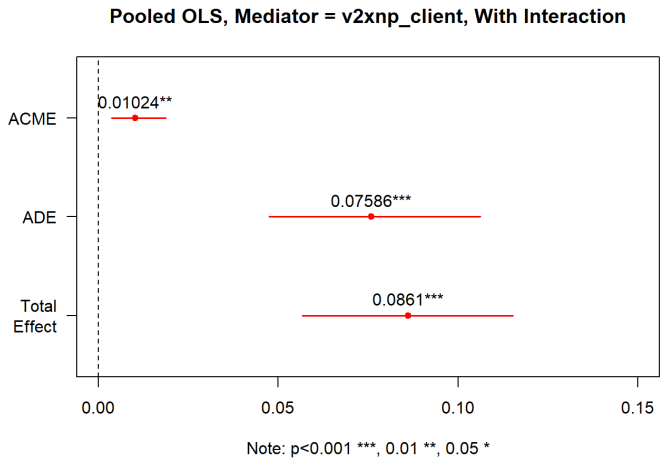

Figure F3: Conditional operation of clientelism.

From Figure F3, we observe that the ACME is 0.01024, indicating that a 10% increase in the composite measure is associated with, on average, an increase in the V-Dem Participatory Democracy score by 0.000976 units. However, further inspecting the model by modelling the less urban sample and the more urban sample shows that having controlled for land area through the composite measure, there is no heterogeneity in the results. Among less urban states, with an ACME of 0.01425 significant at the 0.01 level, a 10% increase in the composite measure is associated with, on average, an increase in the V-Dem Participatory Democracy score by 0.00136 units (see Fig. F4). However, there is a positive but non-significant ACME among the more urban states, leading us to conclude that there is not enough evidence of an effect of state size on democracy among the more urban states (see Fig. F5).

Again, this may indicate that less urban states with larger populations and geographic areas place resource strains on clientelistic linkages, thereby diminishing the undemocratic effects of clientelism and increasing the levels of democracy as a result. However, in the more urban states, the joint effect of land area on the population may not be as significant since the population is already pooled in urbanised polities. While the main results (Fig. 4 of the thesis) show a negative relationship between population size and democracy among the more urban states, owing to the fact that more urban states tend to be more economically developed and can therefore bankroll informal, clientelistic practices, the effect of land area seems to not be significant anyway; clientelistic practices that are being bankrolled through a greater population size and stronger economy are unlikely to be affected by geographic areas, especially when populations and political power of patrons are concentrated in urban centres.

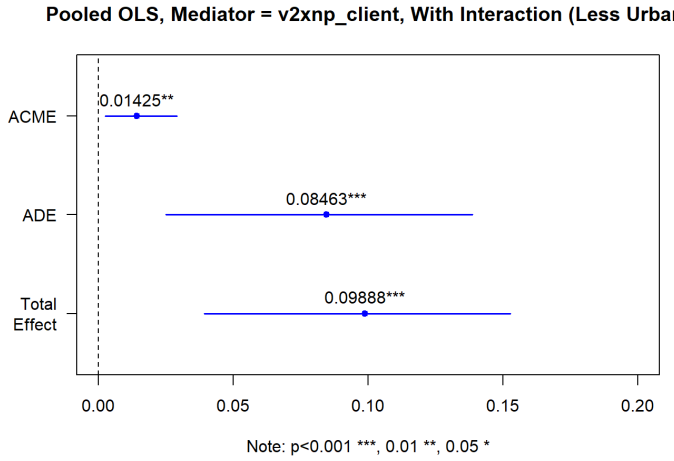

Figure F4: Conditional operation of clientelism in less urban states.

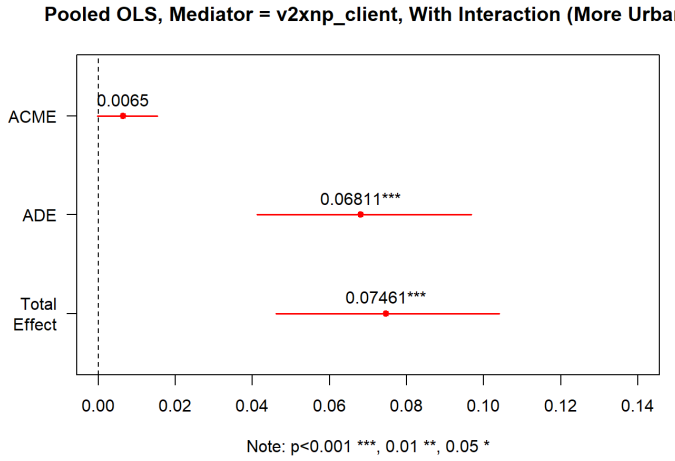

Figure F5: Conditional operation of clientelism in more urban states.

3.4 H_3: The long-term operation of coercive capacity

Figure F6 shows the effect of state size on democracy which is operated by coercive capacity, denoted by the ACME. The ACME is positive at 0.03316, significant at the 0.01 level. This means that for every 10% increase in the composite measure of state size, the V-Dem Participatory Democracy score increases by 0.00316 units, with this effect being operated by military participation rate. Furthermore, Model 1 of Table F2 shows a negative and significant relationship between state size and military participation rate. We can thus reject the null hypothesis and conclude that there is a joint effect of population size and geographic area on democracy that is being operated by coercive capacity. That is, less populous and geographically smaller states, being more likely to possess larger coercive capacities relative to their population size, are likely to have lower levels of democracy.

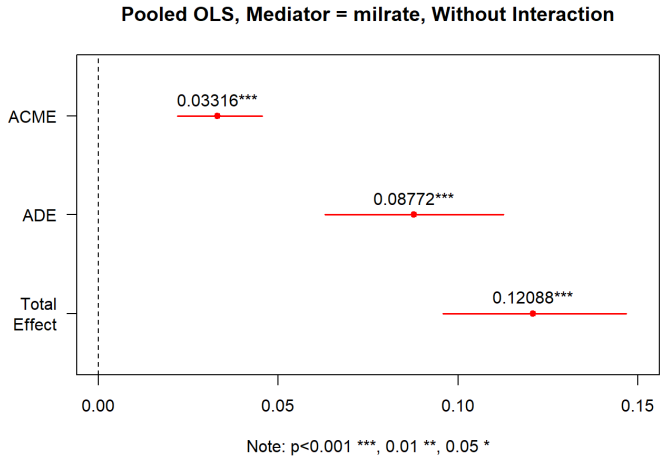

Figure F6: Long-term operation of coercive capacity.

3.5 H_{3a}: The short-term operation of coercive capacity

Figure F7 shows the short-term effect of state size on democracy which is operated by coercive capacity, denoted by the ACME. The ACME is, however, not significant. We cannot reject the null hypothesis that there is no operation of the joint effect of population size and geographic area on democracy through coercive capacity on the short term. The potential explanation for this has been detailed in Chapter 5 of the thesis.

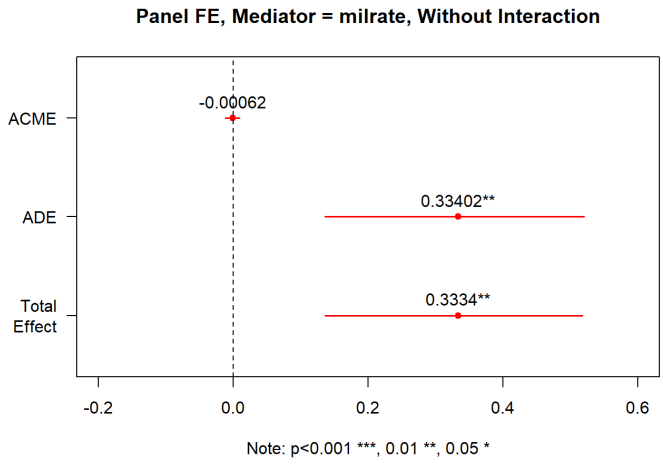

Figure F7: Short-term operation of coercive capacity.

3.6 H_{3b}: The conditional operation of coercive capacity

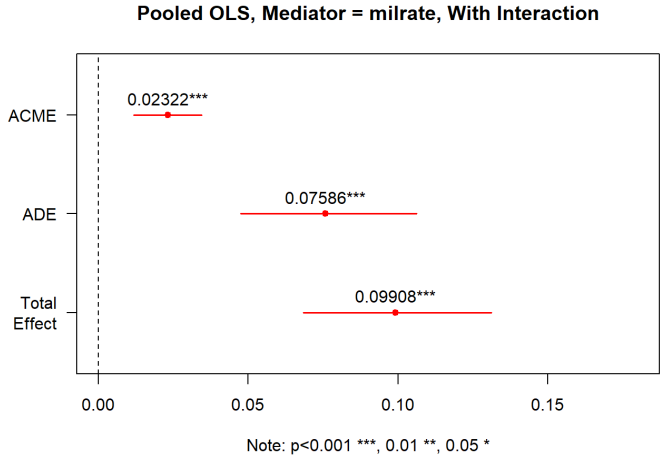

Figure F8: Conditional operation of coercive capacity.

The ACME in Figure F8 shows a positive and significant result in the ACME at 0.02322 when the interaction term is included, and there does seem to be a difference among less urban and more urban states. This ACME indicates that a 10% increase in the composite measure of state size is associated with, on average, an increase in the V-Dem Participatory Democracy score by 0.00221 units. We can thus reject the null hypothesis that there is no joint effect of population size and geographic area that is operated by coercive capacity.

On further inspection, we observe that the relationship between state size and democracy operated by coercive capacity is not significant among less urban states but significant in the more urban states. Turning first to the less urban states (see Fig. F9), a non-significant ACME leads us to conclude that the null hypothesis, that the joint effect of population size and geographic area on democracy is not operated by coercive capacity, cannot be rejected.

On the second count, the ACME of the more urban states indicated that a 10% increase in the composite measure for state size is associated with, on average, an increase in the V-Dem Participatory Democracy score by 0.00444 units (see Fig. F10). We can thus conclude that there is a joint effect of population size and geographic area on democracy that is operated by coercive capacity. In all likelihood, having established in the thesis that less urban states are also less economically developed, one key mechanistic insight is that less urban states are constrained by economic resources to fund their militaries, or are unable to pool manpower for their militaries, that it ebbs the effect of population size or geographic area on coercive capacity and consequently democracy.

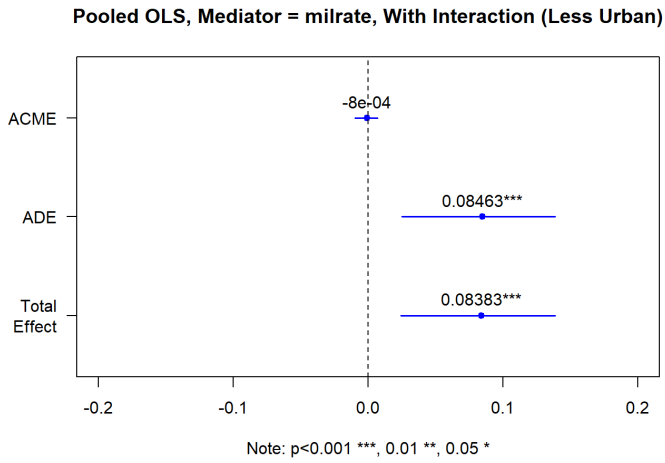

Figure F9: Conditional operation of coercive capacity inless urban states.

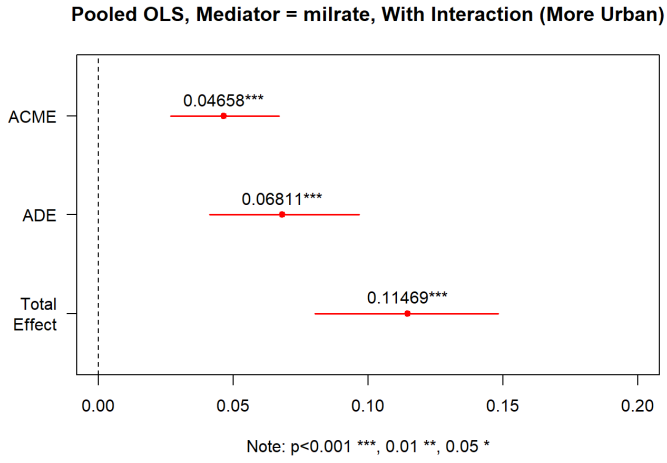

Figure F10: Conditional operation of coercive capacity in more urban states.