Regions of Exception: Appendix

# Modeling Village Electrification in Indonesia

The first column of Table 1 (Model 1, equation (1)) is a simple regression of village population, the natural logarithm of distance in kilometers from the village to the office of the district head (a measure of village remoteness), and a dummy variable capturing whether the village is rural or urban (*desa* or *kelurahan*), on the percentage of a village’s households with access to state-provided electricity from PLN. Villages are indexed with *i*, and provinces with *k.* Province fixed effects $D\_{k}$ are included but not reported.

|  |  |  |
| --- | --- | --- |
|  | $$Electricity\_{ik}=β\_{0}+β\_{0}Village Population\_{ik}+β\_{0}Distance to village head\_{ik}+β\_{0}Rural\_{ik}+D\_{k}+ε\_{ik}$$ | (1) |

There is, not surprisingly, a strong and highly statistically significant relationship between each independent variable and the percentage of households with access to state-provided electricity. Due to exogeneity concerns, it is not possible to give a causal interpretation to results for village population—perhaps people migrate to those villages where electricity is available, especially in the most remote regions—but distance to the office of the district head and the rural dummy are almost certainly strictly exogenous to the processes that generate electrification.[[1]](#endnote-1)

Table 1: Determinants of Village Electrification (2011)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Model 1** | **Model 2** | **Model 3** | **Model 4** |
| *Fixed Effects* |  |  |  |  |
| Village population (ln) | 0.07\*\*\* | 0.07\*\*\* | 0.07\*\*\* | 0.07\*\*\* |
| (0.01) | (0.01) | (0.01) | (0.01) |
| Distance to village head’s office (ln) | -0.06\*\*\* | -0.06\*\*\* | -0.06\*\*\* | -0.07\*\*\* |
| (0.00) | (0.01) | (0.01) | (0.01) |
| Rural dummy | -0.06\*\*\* | -0.11\*\*\* | -0.11\*\*\* | -0.11\*\*\* |
| (0.01) | (0.02) | (0.02) | (0.02) |
| Distance to Jakarta (ln) | -- | -- | -0.00 | -0.05\* |
| -- | -- | (0.01) | (0.02) |
| Percent Muslim Villages | -- | -- | -0.03 | 0.04 |
|  | -- | -- | (0.05) | (0.09) |
| Village population \* Distance to Jakarta | -- | -- | -- | 0.01 |
| -- | -- | -- | (0.01) |
| Distance to village head’s office \* Distance to Jakarta | -- | -- | -- | -0.01 |
| -- | -- | -- | (0.00) |
| Rural dummy \* Distance to Jakarta | -- | -- | -- | -0.03 |
| -- | -- | -- | (0.02) |
| Village population \* Percent Muslim | -- | -- | -- | -0.02 |
| -- | -- | -- | (0.03) |
| Distance to village head’s office \* Percent Muslim | -- | -- | -- | -0.03 |
| -- | -- | -- | (0.02) |
| Rural dummy \* Percent Muslim | -- | -- | -- | 0.08 |
| -- | -- | -- | (0.07) |
| Constant | 0.57\*\*\* | 0.54\*\*\* | 0.54\*\*\* | 0.55\*\*\* |
| (0.01) | (0.03) | (0.03) | (0.03) |
| *Random Effects* |  |  |  |  |
| St Dev[Village population (ln)] | -- | 0.06\*\*\* | 0.06\*\*\* | 0.06\*\*\* |
| -- | (0.01) | (0.01) | (0.01) |
| St Dev[Distance to village head’s office (ln)] | -- | 0.03\*\*\* | 0.03\*\*\* | 0.03\*\*\* |
| -- | (0.00) | (0.00) | (0.00) |
| St Dev[Rural dummy] | -- | 0.12\*\*\* | 0.12\*\*\* | 0.11\*\*\* |
| -- | (0.02) | (0.02) | (0.01) |
| St Dev[Constant] | -- | 0.15\*\*\* | 0.16\*\*\* | 0.14\*\*\* |
| -- | (0.02) | (0.02) | (0.02) |
| Corr[Village population, Distance to village head’s office] | -- | -0.13 | -0.13 | -0.13 |
| -- | (0.18) | (0.18) | (0.18) |
| Corr[Village population, Rural dummy)] | -- | 0.14 | 0.14 | 0.30 |
| -- | (0.18) | (0.18) | (0.19) |
| Corr[Distance to village head’s office, Rural dummy] | -- | -0.22 | -0.23 | -0.11 |
| -- | (0.19) | (0.19) | (0.19) |
| Corr[Village population, Constant] | -- | -0.39\* | -0.39\* | -0.29 |
| -- | (0.18) | (0.18) | (0.18) |
| Corr[Distance to village head’s office, Constant] | -- | 0.23 | 0.21 | 0.31 |
| -- | (0.18) | (0.19) | (0.19) |
| Corr[Rural dummy, Constant] | -- | 0.77\*\*\* | 0.79\*\*\* | 0.72\*\*\* |
| -- | (0.18) | (0.19) | (0.18) |
| St Dev[Residual] | -- | 0.24\*\*\* | 0.24\*\*\* | 0.24\*\*\* |
| -- | (0.00) | (0.00) | (0.00) |
| Provinces | 33 | 33 | 33 | 33 |
| Villages | 77961 | 77961 | 77961 | 77961 |

Standard errors appear in parentheses. Model 1 is an OLS regression with province-level fixed effects (not reported). Models 2-4 are multilevel models. Each independent variable is centered at its grand mean to facilitate interpretation. \* p < .05, \*\* p < .01, \*\*\* p < .001.

But are these relationships the same across provinces? The remaining three columns in Table 1 model the heterogeneity explicitly using a multilevel modeling approach.[[2]](#endnote-2) Model 2 (equation (2)) is a random-slope model, in which the relationship between each independent variable and the proportion of households with electricity is allowed to vary across provinces. For village-level independent variables $P=\left\{Village population, Distance to village head, Rural\right\}$, the model is

|  |  |  |
| --- | --- | --- |
|  | $$Electricity\_{ik}=γ\_{00}+\sum\_{p=1}^{P}γ\_{p0}x\_{pik}+ δ\_{0k}+\sum\_{p=1}^{P}δ\_{pk}x\_{pik}+ε\_{ik}$$ | (2) |

This will illustrate whether, for example, larger villages in Papua and West Papua provinces have less access to electricity than do larger villages in other provinces. Model 3 (equation (3)) includes two province-level covariates $Q=\left\{Distance to Jakarta, Percent Muslim\right\}$ to determine whether they explain the provincial-level variation in the average level of electrification for typical villages.

|  |  |  |
| --- | --- | --- |
|  | $$Electricity\_{ik}=γ\_{00}+\sum\_{q=1}^{Q}γ\_{0q}z\_{qk}+\sum\_{p=1}^{P}γ\_{p0}x\_{pik}+ δ\_{0k}+\sum\_{p=1}^{P}δ\_{pk}x\_{pik}+ε\_{ik}$$ | (3) |

Finally, Model 4 (equation (4)) interacts each province-level predictor with each village-level predictor to account for differences in province-specific slopes:

|  |  |  |
| --- | --- | --- |
|  | $$Electricity\_{ik}=γ\_{00}+\sum\_{q=1}^{Q}γ\_{0q}z\_{qk}+\sum\_{p=1}^{P}γ\_{p0}x\_{pik}+\sum\_{q=1}^{Q}\sum\_{p=1}^{P}γ\_{pq}z\_{qk}x\_{pik}$$$$+ δ\_{0k}+\sum\_{p=1}^{P}δ\_{pk}x\_{pik}+ε\_{ik}$$ | (4) |

By proceeding in this fashion, the analysis first acknowledges the presence of unobserved province-level heterogeneity (Model 1), then also the possibility of unit heterogeneity across provinces (Model 2), and finally, attempts to model both using observables (Models 3 and 4).

# Modeling Support for Malaysia’s BN

Table 2 investigates state-level heterogeneity in the relationship between *bumiputera* population share and BN vote share in 2013. First I estimate a simple fixed effects regression of *bumiputera* population and its square (Model 1, equation (5)).

|  |  |  |
| --- | --- | --- |
|  | $$BN Vote Share\_{ik}=β\_{0}+β\_{1}Pct Bumiputera\_{ik}+ β\_{2}\left(Pct Bumiputera\_{ik}\right)^{2}+ D\_{k}+ε\_{ik}$$ | (5) |

Model 2 modifies equation (5) by replacing the fixed effects$ D\_{k}$ with a dummy for East Malaysia. In Model 3 (equation (6)) I estimate a random slope model, allowing the coefficient on *bumiputera* to vary across states.

|  |  |  |
| --- | --- | --- |
|  | $$BN Vote Share\_{ik}=γ\_{00}+γ\_{0}Pct Bumiputera\_{ik}+ δ\_{0k}+δ\_{k}Pct Bumiputera\_{ik}+ε\_{ik}$$ | (6) |

Finally, Model 4 adds to equation (6) a variable capturing the land area of each electoral district, on the intuition that this will capture some of the competing urban-rural cleavage in Malaysian politics.

Table 2: Determinants of BN Vote (2013)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Model 1** | **Model 2** | **Model 3** | **Model 4** |
| *Fixed Effects* |  |  |  |  |
| Percent Bumiputera | 0.45\*\*\* | 0.29\*\*\* | 0.49\*\*\* | 0.44\*\*\* |
| (0.05) | (0.07) | (0.04) | (0.05) |
| Percent Bumiputera Squared | -0.00\*\* | -0.01\*\* | -- | -- |
| (0.00) | (0.00) | -- | -- |
| East Malaysia Dummy | -- | 11.81\*\* | -- | -- |
| -- | (3.63) | -- | -- |
| District Land Area (ln) | -- | -- | -- | 1.93\*\*\* |
| -- | -- | -- | (0.42) |
| Constant | 54.23\*\*\* | 53.76\*\*\* | 53.30\*\*\* | 54.29\*\*\* |
| (0.47) | (2.41) | (2.10) | (2.02) |
| *Random Effects* |  |  |  |  |
| St Dev[Percent Bumiputera] | -- | -- | 0.11\*\*\* | 0.14\*\*\* |
| -- | -- | (0.05) | (0.05) |
| St Dev[Constant] | -- | -- | 7.88\*\*\* | 7.55\*\*\* |
| -- | -- | (1.78) | (1.80) |
| Corr[Percent Bumiputera, Constant] | -- | -- | 1.14\* | 1.71\* |
| -- | -- | (0.50) | (0.70) |
| St Dev[Residual] | -- | -- | 6.82\*\*\* | 6.57\*\*\* |
| -- | -- | (0.35) | (0.33) |
| States and Federal Territories | 16 | 16 | 16 | 16 |
| Electoral Districts | 221 | 221 | 221 | 221 |

Standard errors appear in parentheses. Model 1 is an OLS regression with state-level fixed effects (not reported). Model 2 is an OLS regression without fixed effects and a dummy for East Malaysia. Models 3-4 are multilevel models. Each independent variable is centered at its grand mean to facilitate interpretation. \* p < .05, \*\* p < .01, \*\*\* p < .001.

# *Reflect*, *Theorize*, *Model*, and *Adjust*: A Checklist

In the main text, I discuss ways in which comparative methods may be employed for purposes other than causal inference in the mainstream positivist mode. Here, I offer a four-step checklist of suggests for researchers who may confront regions of exception in their own work and who wish to adhere more strictly to mainstream techniques for causal inference using comparative methods.

The first step is to *reflect* on the population of units or regions that comprise the analysis. Obvious candidates for regions of exception include regions with distinct colonial histories, social structures, or demographic or geographical features, as noted above. Regions of exception are also likely to be those regions where an outcome of interest is present when it is not present anywhere else in the country. A second, closely related step is to *theorize* this variation. If a regional exceptionalness is something to be explained—why only southern Thailand features an insurgency, or why poverty is so persistent in Papua—then this requires elaborating those causes that might uniquely predict these outcomes in general terms, and also the potential causal interactions.

From there, the next step is to *model* heterogeneity. In quantitative analysis, standard model checking procedures may be useful for revealing outliers or influential observations. Hierarchical data may be amenable to multilevel analyses to uncover heterogeneous causal effects, as shown above; region fixed effects can adjust for unobserved heterogeneity if that is the central inferential challenge. Qualitative comparisons may select cases based on a theory of the regions in which particular causes are plausible.

The fourth step is, then, is to *adjust causal claims* where necessary. Where small-n or complex interactive causality renders research designs indeterminate, then causal claims must depend on a different logic of inference. Where nonrandom missing data prevent generalization across all subnational units, then the limits of generalization should be made clear. Where unit heterogeneity assumptions are untenable, average causal effects should not be reported. Scope conditions may rescue internal validity while explicitly acknowledging that external validity is provisional. Together, these four steps—reflect, theorize, model, and adjust—place a burden on researchers to understand both their cases and the inferential tools that they employ. Yet following each of them will produce more rigorous causal findings while also revealing more clearly the limits of the subnational comparative method. Used properly, they will make legible once again those regions of exception and the politics that they produce.

# Notes

# References

Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

Jones, Bradford S., and Marco Steenbergen. 2002. "Modeling Multilevel Data Structures." *American Journal of Political Science* 46(1): 218-37.

Martinez-Bravo, Monica. 2014. "The Role of Local Officials in New Democracies: Evidence from Indonesia." *American Economic Review* 104(4): 1244-87.

1. See Martinez-Bravo 2014 on the exogeneity of the village-level urban/rural distinction to contemporary political economy. [↑](#endnote-ref-1)
2. Gelman and Hill 2007, 251-99; Jones and Steenbergen 2002. [↑](#endnote-ref-2)