# Appendix – Extreme Boundary Analysis

*Procedure*

Extreme Boundary Analysis (EBA), proposed in its discussed view by Sala-i-Martin (1997) subsequent modifications proposed by other authors (Hegre and Sambanis 2006), is a method for testing the robustness of an estimate of a variable by running models with all possible combinations of other variables. This paper employs EBA for two purposes. The first aim is to assess the sensitivity of income level estimates. The second is to examine the extent of bias associated with "omitted variables" or "selection bias" problems. To simplify, if both linear and quadratic terms of GDP per capita have theoretically predicted significant effect (positive and negative correspondingly to produce inverted “U”-shape), one can conclude that the result is robust and is unlikely to be highly affected by endogeneity.

Assume we have a following model:

where *y* is dependent variable (in our case – binary variable for unarmed revolutions), *x* - main independent variable under consideration (GDP per capita), *z* - control variable from a set of *k* variables, and - dummy variables for fixed effects (years).

I run models with all possible combinations of control variables *z* and have number of models (*m* – number of variables in the model, *n* – overall number of available variables). To slightly reduce computational problem, I include to each equation constant controls that theoretically and empirically have proven their significance. Thus, the final model is model:

where *c* - constant control variables from a set of *j* variables that are included in each model. As constant variables I consider population and level of democracy (with its squared version).

After running *m* models one can calculate and its variance, , by using following formulas (they are weighted means):

Then overall *p-value* can be calculated as usual:

In turn, - weight of each model that is calculated as:

Where *L* is some statistic that can be calculated differently based on model goodness of fit. In case of linear model, it is R-squared, but in case of other models this question is more complicated. For the discussed classification problem of “rare-events”, I use G-mean that does not suffer much from imbalanced class problem (O’Brien and Ishwaran 2019). It is calculated as that is geometric mean of true positive rate and true negative rate that tends to catch “rare events” data better then pseudo-R-squares.

One more issue is combining multiple imputation technic with extreme boundary analysis. I have never seen research where authors did it, but the idea in my view is straightforward: firstly, I estimate each combination using all datasets and then aggregate estimates via “Rubin combination rules” (King et al. 2001, 53); and secondly, aggregate that estimates among combinations[[1]](#footnote-1) using rules formulated above (weighted means). In other words, multiple imputation goes first. In this research, 50 imputed datasets are exploited.

The whole other estimation procedure is the same as in the main text for regression analysis: one-way FE for year, clustered standard errors on country, bias-reduction estimator and 1-year lag for all independent variables.

Additionally to the main specification with polynomial main independent variable (linear and quadratic terms), I also run the same procedure with only linear term of GDP per capita to test possible significance of linear effect of income level.

*Data*

As potential confounding variables, I introduce a number of factors, which are presented in Table 1 below. The right column depicts the role of each factor in the models. The designation "control" is applied to variables that can be incorporated into a given specification. The designation "constant control" is applied to variables that are included in each model. Consequently, for each dependent variable, 23,772 model specifications are estimated (ranging from four to 12 control variables with two constant controls).

|  |  |  |
| --- | --- | --- |
| Table 1. Variables used in the analysis | | |
| Variable | Source | Role in the model |
| Nonviolent campaign (NAVCO 1.3) | NAVCO 1.3 (Chenoweth and Christopher 2020) | Dependent variable |
| Unarmed revolution (Beissinger's + Goldstone et al. data) | Beissinger (2022), Goldstone et al. (2023) | Dependent variable |
| GDP pc (2017$, PPP), ln | Gapminder (2024) | Main independent variable |
| Population (in thousands), ln | World Population Prospects, UN (2022) | Constant control variable |
| Polity score | Polity-V data (Marshall and Keith 2020) | Constant control variable |
| Growth of GDP pc, 5-year average | Calculated from GDP per capita data | Control variable |
| Regime durability (in years), ln | Polity-V data (Marshall and Keith 2020) | Control variable |
| Urbanization (in %) | Beissinger (2022) and World Bank (2023) | Control variable |
| Executive corruption index | V-Dem (Coppedge et al. 2021) | Control variable |
| Mean years of schooling | United Nations Development Program (UNDP 2022) | Control variable |
| Oil pruduction pc (kWh) | Our World in Data project (2024) | Control variable |
| Nonviolent revolutions in the same region (NAVCO 1.3 or Beissinger's + Goldstone et al. data) | Calculated from dependent variable | Control variable |
| Nonviolent revolutions in the same region, lag (NAVCO 1.3 or Beissinger's + Goldstone et al. data) | Calculated from dependent variable | Control variable |
| Number of revolutions in the country's history (since 1950) at year t (NAVCO 1.3 or Beissinger's + Goldstone et al. data) | Calculated from dependent variable | Control variable |
| Discriminated population (% of total population) | Ethnic Power Relations Dataset (Vogt et al. 2015) | Control variable |
| Share of people in age 15-24 in adult (15+) population | Calculated based on World Population Prospects, UN (2022) | Control variable |
| Equal distribution of resources index | V-Dem (Coppedge et al. 2021) | Control variable |

*Results*

The graph below depicts the distributions of the z-statistic from all models where the dependent variable is the unarmed revolution from NAVCO 1.3. It can be observed that both linear and quadratic terms of GDP per capita are highly resilient to different combinations of variables. The coefficients are statistically significant, with an overall two-tailed p-value for both terms of less than 0.001. The z-statistic for the linear component is 5, while that for the quadratic component is -4.9. Conversely, the GDP per capita term as only one term, which models a linear link, is insignificant in most models. Its overall p-value is 0.44 (z-statistic is 0.77), and there is a right tail that crosses the critical area.

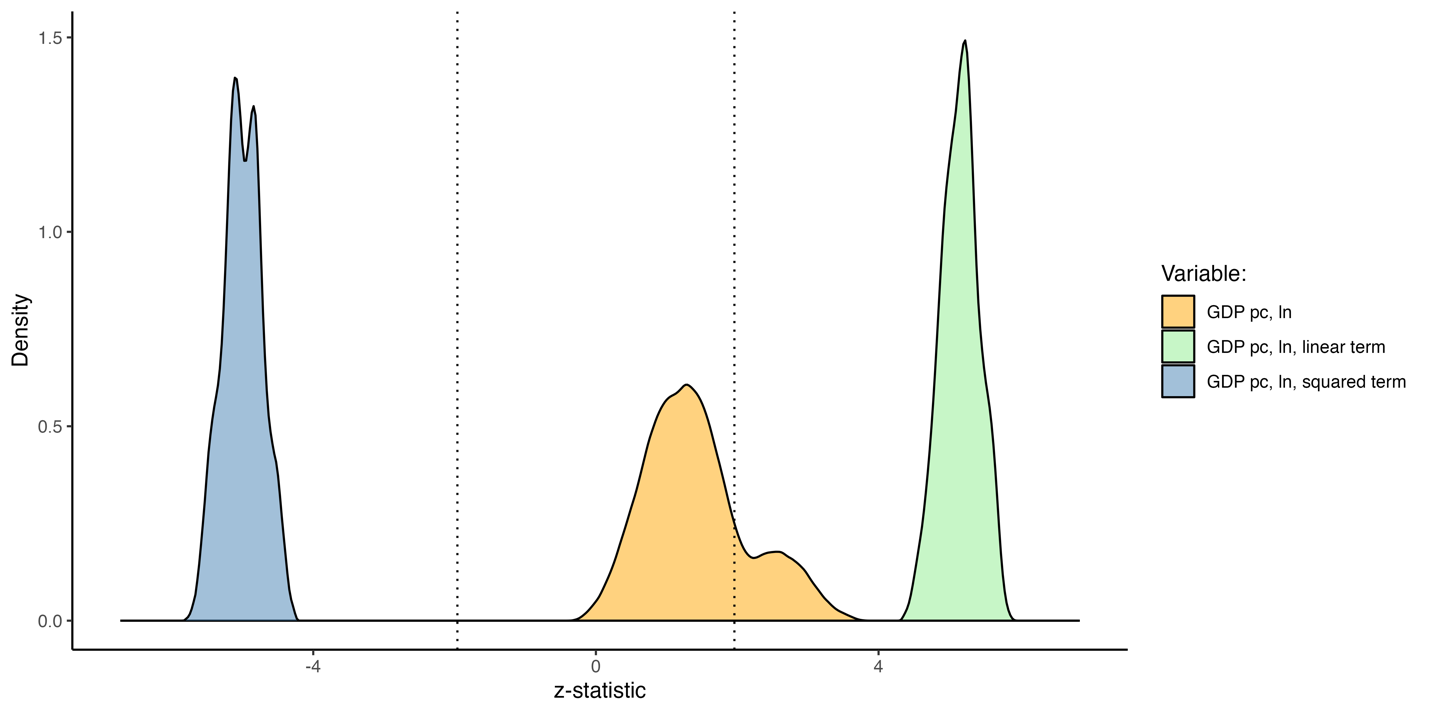


Figure 1. The distributions of z-statistic from EBA for both parts of quadratic term and separated linear term with NAVCO 1.3 as source for dependent variable.

In turn, in the graph below distributions of *z-statistic* from all models where unarmed revolutions from Beissinger’s extended data is dependent variable are plotted. It can be seen both linear and quadratic terms of GDP per capita are extremely resilient to different combinations of variables: coefficients are highly significant (overall p-value for both terms are <<0.001 and statistic is 4.77 and -4.8 respectively for linear and quadratic components). Meanwhile, GDP per capita as one term (which models only linear link) in most models’ specification is insignificant, its overall p-value is 0.93 (statistic is 0.09), while there is right tail that crosses critical area.

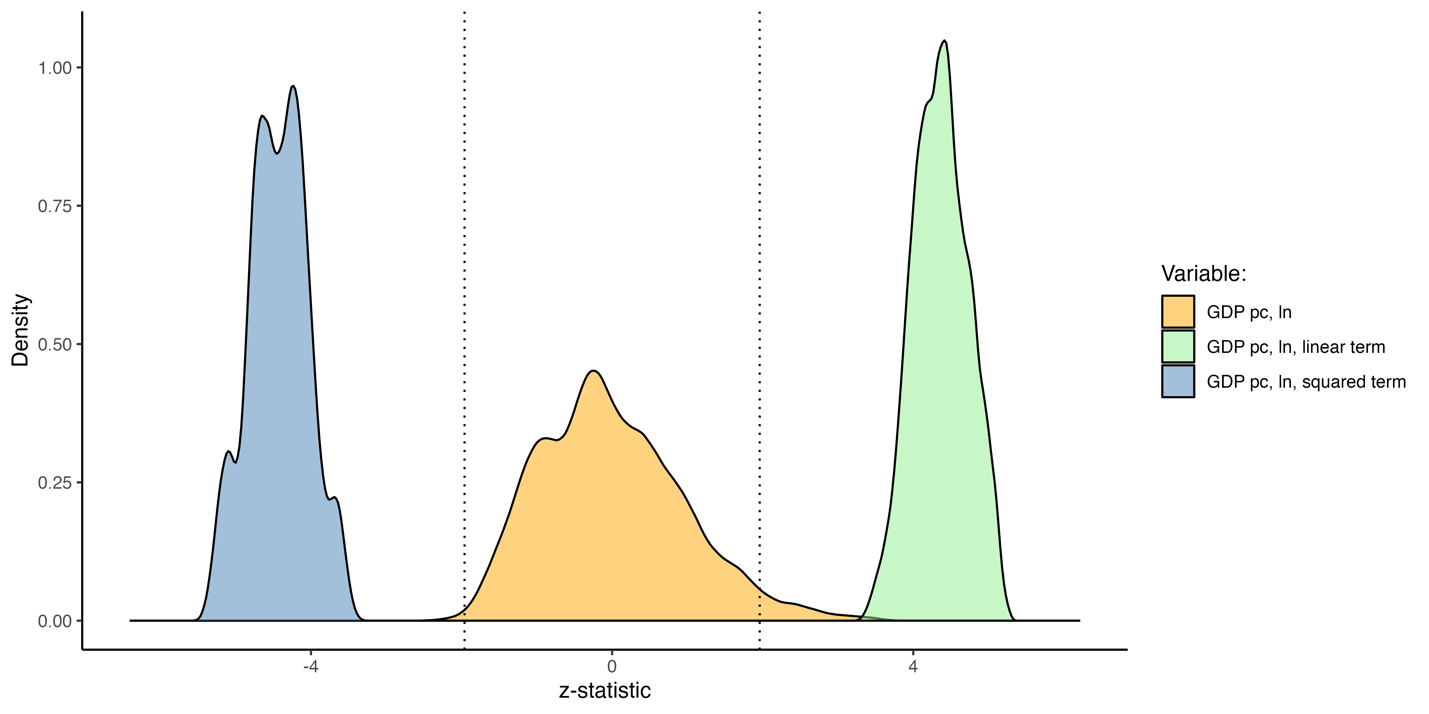


Figure 2. The distributions of z-statistic from EBA for both parts of quadratic term and separated linear term with Beissinger’s extended data as source for dependent variable.

Table 2 below with aggregated results of EBA shows overall coefficients and variances for both dependent variables.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Table 2. Extreme boundary analysis aggregated results. | | | | | | |
| Main independent variable | NAVCO 1.3 | | | Beissinger’s extended data | | |
| b | se | p | b | se | p |
| GDP pc, ln, linear term | 5.061 | 0.982 | <0.001 | 4.719 | 1.08 | <0.001 |
| GDP pc, ln, quadratic term | -0.284 | 0.057 | <0.001 | -0.271 | 0.061 | <0.001 |
| GDP pc, ln | 0.157 | 0.114 | 0.168 | -0.003 | 0.144 | 0.986 |
| *Note: results are based on 23 772* *unique specifications with at least 4 control variables from 12 possible and 2 constant controls; fixed-effects for years are included; standard errors are clustered on countries; each model is estimated on 50 imputed datasets; all variables are lagged for 1 year.* | | | | | | |

# References

Beissinger, Mark. 2022. *The Revolutionary City: Urbanization and the Global Transformation of Rebellion*. Princeton University Press.

Chenoweth, Erica, and W. S. Christopher. 2020. *List of Campaigns in NAVCO 1.3*. Cambridge, MA: Harvard Dataverse.

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, Nazifa Alizada, David Altman, et al. 2021. “V-Dem Dataset V11. 1.”

Gapminder. 2024. “GDP per Capita in Constant PPP Dollars (V30).” https://www.gapminder.org/data/documentation/gd001/.

Goldstone, J. A., L.E. Grinin, V. Ustyuzhanin, and A. Korotaev. 2023. “Revolutionary Events of the 21st Century: A Preliminary Quantitative Analysis.” (4): 54–71. doi:https://doi.org/10.17976/jpps/2023.04.05.

Hegre, Håvard, and Nicholas Sambanis. 2006. “Sensitivity Analysis of Empirical Results on Civil War Onset.” *Journal of Conflict Resolution* 50(4): 508–35. doi:10.1177/0022002706289303.

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(1): 49–69. doi:10.1017/S0003055401000235.

Leyrat, Clémence, Shaun R Seaman, Ian R White, Ian Douglas, Liam Smeeth, Joseph Kim, Matthieu Resche-Rigon, James R Carpenter, and Elizabeth J Williamson. 2019. “Propensity Score Analysis with Partially Observed Covariates: How Should Multiple Imputation Be Used?” *Statistical Methods in Medical Research* 28(1): 3–19. doi:10.1177/0962280217713032.

Marshall, Monty G., and Jaggers Keith. 2020. “Polity V Project: Political Regime Characteristics and Transitions 1800–2018.” https://www.systemicpeace.org/polityproject.html.

O’Brien, Robert, and Hemant Ishwaran. 2019. “A Random Forests Quantile Classifier for Class Imbalanced Data.” *Pattern recognition* 90: 232–49.

Our World in Data. 2024. “Oil Production per Capita.” https://ourworldindata.org/grapher/oil-prod-per-capita?time=1901..latest#sources-and-processing.

Sala-i-Martin, X. 1997. “I Just Ran Two Million Regressions.” 87(2): 178–83. doi:10.3386/w6252.

The World Bank. 2023. “The World Bank Databank: World Development Indicators.” https://databank.worldbank.org/ (May 25, 2022).

UNDP. 2022. “UNDP: Human Development Reports.” *UNDP*. http://hdr.undp.org/en/indicators/103006 (August 5, 2021).

United Nations. 2022. “World Population Prospects.” https://population.un.org/wpp.

Vogt, Manuel, Nils-Christian Bormann, Seraina Rüegger, Lars-Erik Cederman, Philipp Hunziker, and Luc Girardin. 2015. “Integrating Data on Ethnicity, Geography, and Conflict: The Ethnic Power Relations Data Set Family.” *Journal of Conflict Resolution* 59(7): 1327–42.

1. The similar strategy is used, for example, with matching procedure that uses imputations (Leyrat et al. 2019). [↑](#footnote-ref-1)