# Appendix – Matching procedure

*Procedure*

The matching procedure is designed to balance observations of the "treated" and "control" groups in order to avoid model dependencies issues. In this paper, I utilize this technique to create a balanced case-control sample (where the "case" is revolution) and to address the "rare-events" problem by reducing the number of nonevents in the original dataset and isolating the effect of income level from other covariates.

As matching is not parametric method, it is advised to use several techniques for getting balanced sample. I use two different approaches: (1) nearest neighbor matching with robust rank-based Mahalanobis distance; (2) optimal matching with robust rank-based Mahalanobis distance that optimizes the overall measure of distances between matched pairs neither than nearest neighbor procedure where each revolution is assigned a control unit separately (Ho et al. 2011). I do not use matching based on propensity scores because of the considerations reflected in King and Nielsen paper (2019): propensity scores may lead to more unbalanced covariates distribution and, as a consequence, considerably biased estimates.

To address the issue of missing data and ensure the results are comparable with those of the main analysis, I also employ the multiple imputations technique, which was utilized in the main regression models with the same 50 imputed datasets. I follow approach described in Leyrat et al. (2019) that suggests to run identical analysis (making matched data and estimate logistic regression) on each imputed dataset and then to combine estimate using “Rubin combination rules” (King et al. 2001, 53).

I use the same covariates as in the main analysis with exact matching on region and year to block variance from cross-sectional (region) or time (year) dimensions that might affect effect of income level on revolutions. As number of revolutions is relatively small, so I try to find 2 cases for each revolution from NAVCO and 4 from Beissinger’s extended dataset that increases efficiency of estimation. Matching was performed using the “MatchIt” package (Ho et al. 2011) in R.

*Covariates balance*

Firstly, in the Table below covariate balance among unadjusted (original) and matched data is shown, where mean and range is shown across all imputed datasets. As balance metrics Standardized Mean Difference (SMD) and Variance Ration (VR) factor are chosen to catch distributions of covariates comprehensively. One can see that initial data was highly unbalanced: distributions of covariates have huge differences across revolutionary and non-revolutionary cases, while after the matching procedure the data becomes relatively balanced. Meanwhile, in case of NAVCO 1.3 cases (see Table 1) with nearest matching, regime durability and revolutions in region have high SMD even after balancing, that shows the protentional imbalance in the resulted datasets. Nevertheless, another test, VR, does not indicate imbalance in them which indicates a positive result in balancing the data. The same situation exists with optimal procedure.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Table 1. Balance of NAVCO 1.3 cases. | | | | | |
| Variable | Mathing | SMD before | SMD after | VR before | VR after |
| Economic growth, 5-year average | nearest | 0.01 | -0.02 | 0.74 | 0.92 |
| Population, log | nearest | 0.59 | 0.2 | 0.84 | 0.93 |
| Polity | nearest | -0.05 | -0.02 | 0.73 | 0.8 |
| Regime durability, log | nearest | -0.27 | **-0.27** | 0.96 | 1.16 |
| Urbanization | nearest | -0.03 | -0.05 | 0.7 | 0.83 |
| Corruption | nearest | 0.55 | 0.14 | 0.67 | 0.92 |
| Education | nearest | 0.14 | -0.01 | 0.83 | 0.9 |
| Revolutions in the same region | nearest | 0.26 | **-0.46** |  |  |
| Revolutions in the past | nearest | 0.52 | 0.17 | 2.62 | 1.35 |
| Economic growth, 5-year average | optimal | 0.01 | -0.02 | 0.74 | 1.24 |
| Population, log | optimal | 0.59 | **0.25** | 0.84 | 0.95 |
| Polity | optimal | -0.05 | 0 | 0.73 | 0.81 |
| Regime durability, log | optimal | -0.27 | **-0.31** | 0.96 | 1.23 |
| Urbanization | optimal | -0.03 | -0.06 | 0.7 | 0.81 |
| Corruption | optimal | 0.55 | 0.16 | 0.67 | 0.89 |
| Education | optimal | 0.14 | -0.02 | 0.83 | 0.89 |
| Revolutions in the same region | optimal | 0.26 | **-0.46** |  |  |
| Revolutions in the past | optimal | 0.52 | 0.18 | 2.62 | 1.38 |
| Note: as statistic for balance metrics are gotten averages across 50 imputed datasets; statistic the exceeds critical point is bolded | | | | | |

In case of Beissinger’s extended data cases (see Table 2) with nearest matching, the overall balance is worse comparing with NAVCO 1.3 data that is a product of a smaller number of cases. However, there is no crossing of the decisive boundary for any variable by two indicators (SMD and VR) at once.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Table 2. Balance of Beissinger’s extended data cases. | | | | | |
| Variable | Mathing | SMD  before | SMD  after | VR  before | VR  after |
| Economic growth, 5-year average | nearest | 0.17 | -0.1 | 0.55 | 0.72 |
| Population, log | nearest | 0.48 | 0.03 | 0.78 | 0.9 |
| Polity | nearest | 0.05 | -0.11 | 0.69 | 0.86 |
| Regime durability, log | nearest | -0.29 | **-0.24** | 0.99 | 1.17 |
| Urbanization | nearest | 0.1 | -0.07 | 0.7 | 0.89 |
| Corruption | nearest | 0.67 | **0.29** | 0.63 | 0.87 |
| Education | nearest | 0.32 | -0.09 | 0.8 | 0.93 |
| Revolutions in the same region | nearest | 0.38 | **-0.47** |  |  |
| Revolutions in the past | nearest | 0.53 | **0.22** | 3.79 | 1.4 |
| Economic growth, 5-year average | optimal | 0.17 | -0.1 | 0.55 | 0.86 |
| Population, log | optimal | 0.48 | 0.03 | 0.78 | 0.93 |
| Polity | optimal | 0.05 | -0.1 | 0.69 | 0.84 |
| Regime durability, log | optimal | -0.29 | **-0.29** | 0.99 | 1.2 |
| Urbanization | optimal | 0.1 | -0.05 | 0.7 | 0.9 |
| Corruption | optimal | 0.67 | **0.3** | 0.63 | 0.86 |
| Education | optimal | 0.32 | -0.09 | 0.8 | 0.92 |
| Revolutions in the same region | optimal | 0.38 | **-0.47** |  |  |
| Revolutions in the past | optimal | 0.53 | **0.22** | 3.79 | 1.53 |
| Note: as statistic for balance metrics are gotten averages across 50 imputed datasets; statistic the exceeds critical point is bolded | | | | | |

*Results*

In the Table 3 models for both dependent variables, from NAVCO 1.3 and Beissinger’s extended data, with implementation of nearest and optimal matching procedure are presented. Due to balanced data on covariates, all models have only one explanatory variable – GDP per capita with its quadratic form. One can see, that income level is significant across exploiting different dependent variables and procedures, whereas its significance lower in models with Beissinger’s data that is connected with relatively small number of observations in comparison with NAVCO 1.3. Meanwhile, both linear and quadratic terms have theoretically expected signs, producing inverted “U”-shape relationship between the probability of unarmed revolution and income level.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 3. Logistic regression models on unarmed revolutions occurrence with different matching procedures | | | | | | | | |
| Variable | NAVCO 1.3 | | | | Beissinger’s extended data | | | |
| Nearest | | Optimal | | Nearest | | Optimal | |
| b | se | b | se | b | se | b | se |
| (Intercept) | -17.2\*\* | 5.53 | -17.43\*\* | 5.54 | -16.16\* | 6.76 | -16.1\* | 6.74 |
| GDP per capita, ln | 3.84\*\* | 1.29 | 3.9\*\* | 1.29 | 3.55\* | 1.56 | 3.54\* | 1.55 |
| GDP per capita, squared, ln | -0.22\*\* | 0.07 | -0.22\*\* | 0.07 | -0.21\* | 0.09 | -0.21\* | 0.09 |
| Case-control ratio | 1:2 | | 1:2 | | 1:4 | | 1:4 | |
| Note: \*\*\*p<0.001, \*\*p<0.01, \*p<0.05; predictors are at t-1; standard errors are heteroscedasticity-consistent; all models are estimated on 50 imputed models. | | | | | | | | |

In the Figures 1 and 2 marginal effect of GDP per capita across its values and adjusted predictions of unarmed revolutions conditional on GDP per capita are presented respectively for models with NAVCO 1.3 as dependent variable. It becomes evident that there is a significant inverted "U"-shaped relationship between income level and the probability of an unarmed revolution occurring. Initially, the probability of an unarmed revolution increases with income level, reaching a peak at a certain point in GDP per capita. Subsequently, the probability declines as income level continues to increase, reaching a negative significant effect. Moreover, this outcome is fully replicated for two matching procedures the results for which are indistinguishable, indicating the robustness of the estimates.

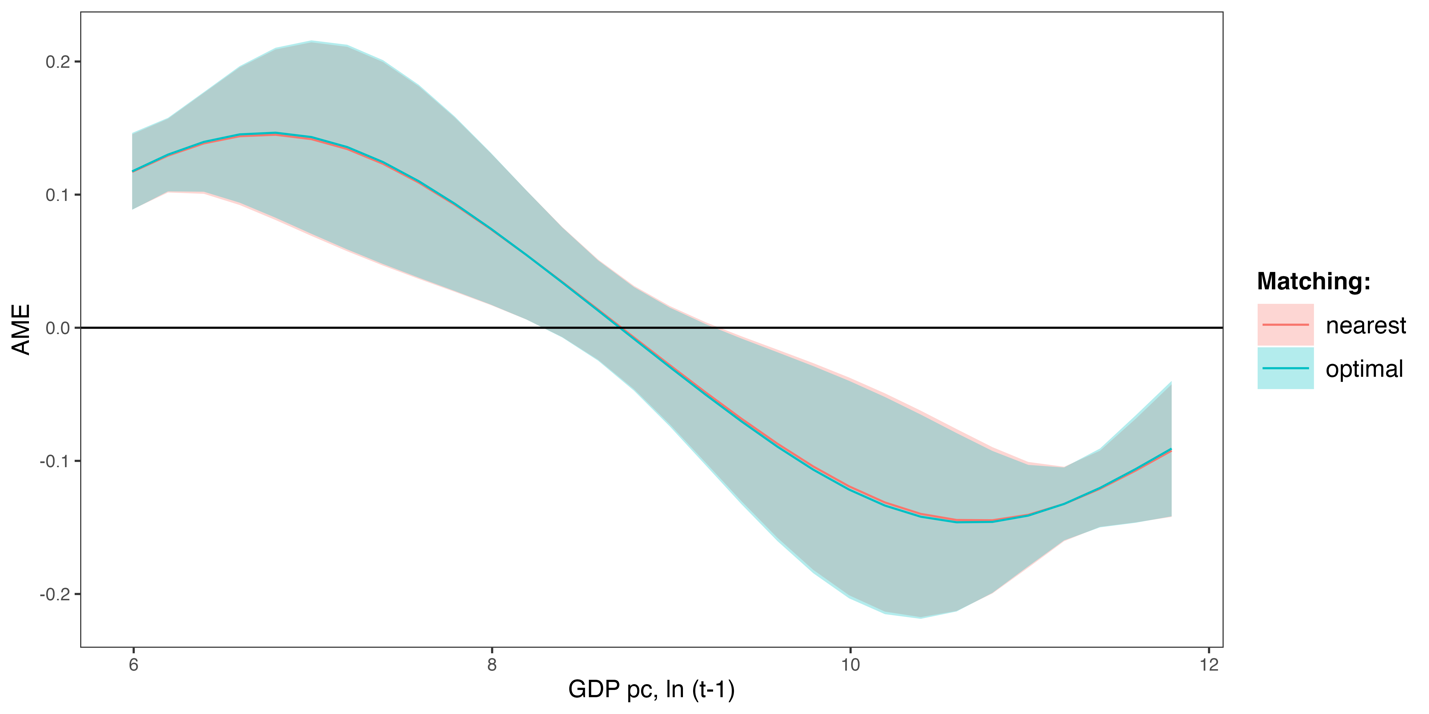


Figure 1. Marginal effect of GDP per capita conditional on its values with NAVCO 1.3 as dependent variable. Note: 95% CI is plotted.

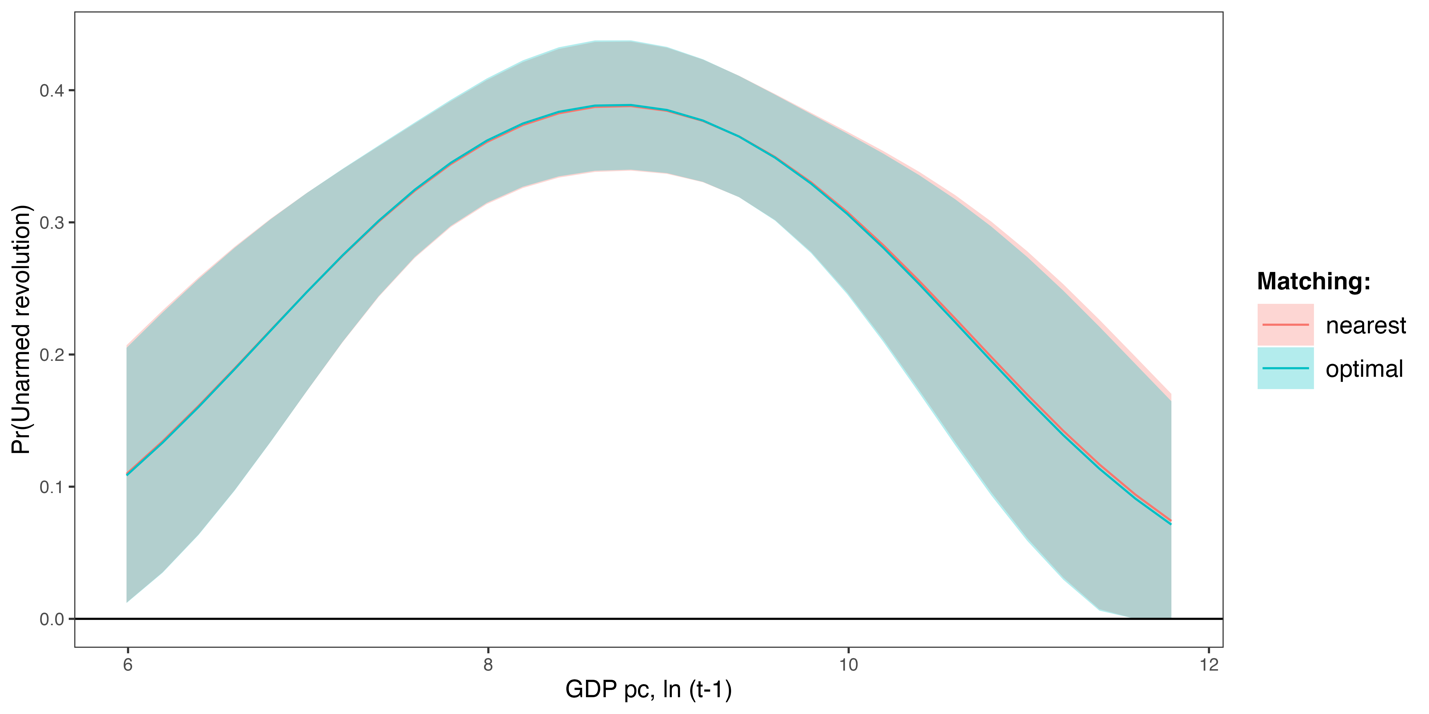


Figure 2. Adjusted predictions of probability of unarmed conditional on GDP per capita with NAVCO 1.3 as dependent variable. Note: 95% CI is plotted.

In the Figures 3 and 4 marginal effect of GDP per capita across its values and adjusted predictions of unarmed revolutions conditional on GDP per capita are presented respectively for models with Beissinger’s extended data as dependent variable. It becomes evident that there is a significant inverted "U"-shaped relationship between income level and the probability of an unarmed revolution occurring that replicates the results with NAVCO 1.3 described above.

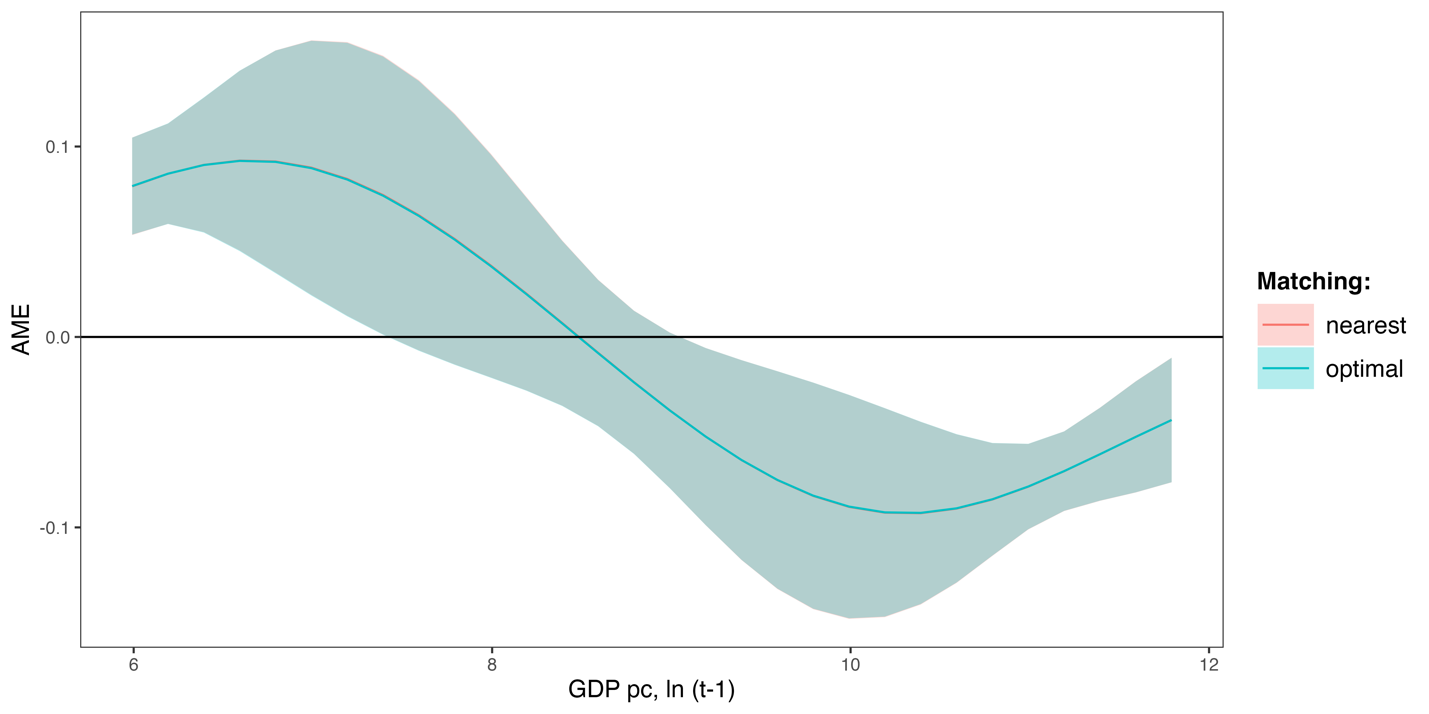


Figure 3. Marginal effect of GDP per capita conditional on its values with Beissinger’s extended data as dependent variable. Note: 95% CI is plotted.

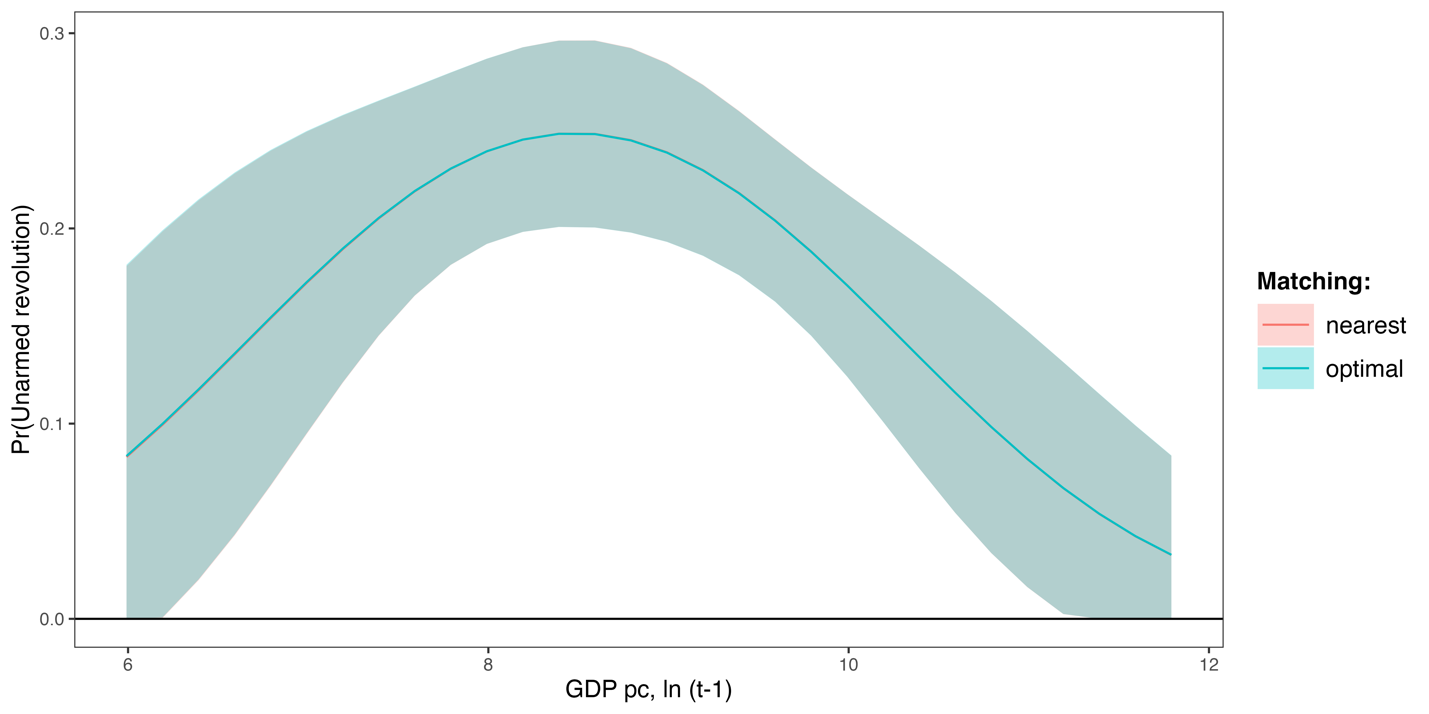


Figure 4. Adjusted predictions of probability of unarmed conditional on GDP per capita with Beissinger’s extended data as dependent variable. Note: 95% CI is plotted.

# References

Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart. 2011. “MatchIt: Nonparametric Preprocessing for Parametric Causal Inference.” *Journal of Statistical Software* 42(8). doi:10.18637/jss.v042.i08.

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.

King, Gary, and Richard Nielsen. 2019. “Why Propensity Scores Should Not Be Used for Matching.” *Political Analysis* 27(4): 435–54. doi:10.1017/pan.2019.11.

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.