SPEC REU R Resources: Reporting Regression Results with texreg – Homework

Gaea Morales, Claudia Salas Gimenez, and Ben Graham

March 2025

Welcome to the final assignment of Module 8: Regression II. In this module, you have learned how to run multivariate regression models (both linear and binomial) and create publication-quality tables to present regression results. This homework assignment will test these skills by having you conduct a regression analysis using replication data and presenting your findings in a professional and polished manner.

We will work with the `ProductivePacifists_Data.RDATA` dataset from the paper "Productive Pacifists: The Rise of Production-Oriented States and Decline of Profit-Motivated Conquest," published in International Studies Quarterly (2020). This assignment explores whether a country's land-oriented economic structure affects its propensity to initiate or escalate territorial disputes. The goal is to estimate multiple regression models, test different specifications, and present results in a comprehensive regression table.

Save your responses in your personal subfolder in the `412_413` shared Google Drive folder. The R script should be titled `HW_DM3_[YOUR INITIALS]`.

Initial Setup

Before beginning your analysis, ensure your working directory is correctly set up, and load the necessary libraries along with the dataset.

For reference, the `ProductivePacifists_Data.RDATA` data was introduced in the research paper *"Productive Pacifists: The Rise of Production-Oriented States and Decline of Profit-Motivated Conquest"* by Markowitz, Mulesky, Graham and Fariss (2020). You can find the research paper and the online appendix here.

```r
# Set working directory
#setwd("YourFolderPath")

# Load required libraries
library(dplyr)
library(texreg)
library(broom)
## We will use this package to tidy model results for plotting

# Load the dataset
load("ProductivePacifists_Data.RDATA")
```

Exercise 1: Exploring the Relationship Between Economic Orientation and Territorial Disputes

Does a country's land-oriented economic structure increase its likelihood of engaging in territorial disputes? Some scholars argue that land-dependent economies are more likely to seek territorial expansion, especially for resource-rich areas. Others suggest that domestic institutions, such as regime type and military capacity, may mediate this effect. We will assess whether economic orientation alone drives conflict behavior or if broader political and strategic factors shape these decisions.

To test this idea, we will use two dependent variables representing different types of territorial conflict:

- `tc_dummy_ICOW`: Binary measure indicating whether a country initiates a territorial claim in a given year. This captures the initial stage of territorial disputes, reflecting a country's willingness to assert control over contested land.
- `tc_mid_resource_dummy_ICOW`: Binary measure indicating whether a country participates in a Militarized Interstate Dispute (MID) over economically valuable territory in a given year. This focuses on conflicts that escalate into military disputes, emphasizing the role of economic resources in territorial disputes.

The key independent variables measuring economic orientation are:

- `land_oriented_medium_continuous`: A continuous measure of economic orientation.
- `land_oriented_medium_binary`: A binary indicator of whether a country is land-oriented.

Exercise 1.1: Running Simple Regressions

Estimate four simple regression models to assess the relationship between economic orientation and territorial conflict initiation, using both dependent variables and the binary and continuous measures of economic orientation as independent variables.

Given that the dependent variables are binary, which statistical model is most appropriate? Although both OLS and logistic regression can be used for binary outcomes, logistic regression is generally preferred. It constrains predicted probabilities to remain between 0 and 1, effectively capturing the nonlinear relationship between predictors and the likelihood of an event occurring. In contrast, OLS might yield predictions outside this range and may not accurately reflect the underlying probability structure of the data.

```r
# Estimate models
## DV: binary measure of whether a country initiates a territorial claim
model1 <- glm(tc_dummy_ICOW ~ land_oriented_medium_continuous,
              family = binomial(link = "logit"), data = dat)

summary(model1)
```

```
##
## Call:
## glm(formula = tc_dummy_ICOW ~ land_oriented_medium_continuous,
##     family = binomial(link = "logit"), data = dat)
##
## Coefficients:
##                                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)                     -3.22280    0.08363 -38.534   <2e-16 ***
## land_oriented_medium_continuous -0.05914    0.02982  -1.983   0.0473 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3977.7  on 13570  degrees of freedom
## Residual deviance: 3973.4  on 13569  degrees of freedom
##   (2028 observations deleted due to missingness)
## AIC: 3977.4
##
```

```
## Number of Fisher Scoring iterations: 6
```

```r
## DV: binary measure of whether a country initiates a territorial claim
model2 <- glm(tc_dummy_ICOW ~ land_oriented_medium_binary,
              family = binomial(link = "logit"), data = dat)

summary(model2)
```

```
##
## Call:
## glm(formula = tc_dummy_ICOW ~ land_oriented_medium_binary, family = binomial(link = "logit"),
##     data = dat)
##
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -4.6375     0.1899 -24.422  < 2e-16 ***
## land_oriented_medium_binary   1.5895     0.1947   8.162  3.3e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4992.2  on 15302  degrees of freedom
## Residual deviance: 4884.0  on 15301  degrees of freedom
##   (296 observations deleted due to missingness)
## AIC: 4888
##
## Number of Fisher Scoring iterations: 7
```

```r
## DV: binary measure of participation in militarized disputes over economically valuable
## territory
model3 <- glm(tc_mid_resource_dummy_ICOW ~ land_oriented_medium_continuous,
              family = binomial(link = "logit"), data = dat)

summary(model3)
```

```
##
## Call:
## glm(formula = tc_mid_resource_dummy_ICOW ~ land_oriented_medium_continuous,
##     family = binomial(link = "logit"), data = dat)
##
## Coefficients:
##                                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)                     -3.52155    0.09082 -38.776   <2e-16 ***
## land_oriented_medium_continuous -0.02529    0.03089  -0.819    0.413
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3374.0  on 13570  degrees of freedom
## Residual deviance: 3373.3  on 13569  degrees of freedom
##   (2028 observations deleted due to missingness)
## AIC: 3377.3
##
## Number of Fisher Scoring iterations: 6
```

```r
## DV: binary measure of participation in militarized disputes over economically valuable
## territory
model4 <- glm(tc_mid_resource_dummy_ICOW ~ land_oriented_medium_binary,
              family = binomial(link = "logit"), data = dat)

summary(model4)
```

```
##
## Call:
## glm(formula = tc_mid_resource_dummy_ICOW ~ land_oriented_medium_binary,
##     family = binomial(link = "logit"), data = dat)
##
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -3.8814     0.1315 -29.512   <2e-16 ***
## land_oriented_medium_binary   0.3527     0.1422   2.481   0.0131 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3791.4  on 15302  degrees of freedom
## Residual deviance: 3784.7  on 15301  degrees of freedom
##   (296 observations deleted due to missingness)
## AIC: 3788.7
##
## Number of Fisher Scoring iterations: 6
```

Exercise 1.2: Estimating Multivariate Models

Next, let's deepen the analysis by introducing control variables to better isolate the effect of economic orientation on territorial disputes. These variables help account for broader political and strategic factors that could influence conflict behavior beyond economic incentives alone. Specifically, regime type (`autocracy_BX`) could affect a country's propensity to initiate conflicts, and military expenditures (`milex_constant2010us_AFM`) capture military capacity, influencing the feasibility and likelihood of engaging in militarized disputes. Other relevant factors include the number of neighboring countries sharing a land border (`land_CONT`) and the overall measure of military capabilities (`cinc_MC`). Controlling for these factors ensures that the associations we observe between economic orientation and territorial disputes aren't confounded by regime characteristics or military capabilities.

Run four additional regression models using different subsets of the variables described above. You may use either linear or logistic regression models, as appropriate.

Note: Use the logged version of military expenditures to reduce skewness, normalize distribution, and facilitate clearer comparisons across countries.

```r
# Create logged military expenditures variable
dat$milex_logged <- log(dat$milex_constant2010us_AFM + 1)
## Adding 1 to avoid log(0)

## DV: binary measure of whether a country initiates a territorial claim
model5 <- glm(tc_dummy_ICOW ~ land_oriented_medium_continuous  + autocracy_BX +
                milex_logged + land_CONT + cinc_MC,
              family = binomial(link = "logit"), data = dat)
```

```r
summary(model5)
```

```
##
## Call:
## glm(formula = tc_dummy_ICOW ~ land_oriented_medium_continuous +
##     autocracy_BX + milex_logged + land_CONT + cinc_MC, family = binomial(link = "logit"),
##     data = dat)
##
## Coefficients:
##                                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)                     -8.121087   1.802421  -4.506 6.62e-06 ***
## land_oriented_medium_continuous  0.054257   0.033420   1.624   0.1045
## autocracy_BX                     0.587060   0.134292   4.372 1.23e-05 ***
## milex_logged                     1.238990   0.582112   2.128   0.0333 *
## land_CONT                        0.049879   0.004853  10.277  < 2e-16 ***
## cinc_MC                          7.360776   0.817390   9.005  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3285.7  on 10466  degrees of freedom
## Residual deviance: 2997.2  on 10461  degrees of freedom
##   (5132 observations deleted due to missingness)
## AIC: 3009.2
##
## Number of Fisher Scoring iterations: 7
```

```r
## DV: binary measure of whether a country initiates a territorial claim
model6 <- glm(tc_dummy_ICOW ~ land_oriented_medium_binary + autocracy_BX +
                milex_logged + land_CONT + cinc_MC,
              family = binomial(link = "logit"), data = dat)

summary(model6)
```

```
##
## Call:
## glm(formula = tc_dummy_ICOW ~ land_oriented_medium_binary + autocracy_BX +
##     milex_logged + land_CONT + cinc_MC, family = binomial(link = "logit"),
##     data = dat)
##
## Coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 -4.723818   0.571017  -8.273   <2e-16 ***
## land_oriented_medium_binary  2.025144   0.230502   8.786   <2e-16 ***
## autocracy_BX                 0.156065   0.117754   1.325   0.1851
## milex_logged                -0.289341   0.171593  -1.686   0.0918 .
## land_CONT                    0.059805   0.005191  11.521   <2e-16 ***
## cinc_MC                      7.698385   0.745846  10.322   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
##     Null deviance: 3989.4  on 11578  degrees of freedom
## Residual deviance: 3570.0  on 11573  degrees of freedom
##   (4020 observations deleted due to missingness)
## AIC: 3582
##
## Number of Fisher Scoring iterations: 7
```

```r
## DV: binary measure of participation in militarized disputes over economically
## valuable territory
model7 <- glm(tc_mid_resource_dummy_ICOW ~ land_oriented_medium_continuous +
                autocracy_BX + milex_logged + land_CONT + cinc_MC,
              family = binomial(link = "logit"), data = dat)

summary(model7)
```

```
##
## Call:
## glm(formula = tc_mid_resource_dummy_ICOW ~ land_oriented_medium_continuous +
##     autocracy_BX + milex_logged + land_CONT + cinc_MC, family = binomial(link = "logit"),
##     data = dat)
##
## Coefficients:
##                                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)                    -17.020513   1.972702  -8.628  < 2e-16 ***
## land_oriented_medium_continuous   0.079902   0.030670   2.605  0.00918 **
## autocracy_BX                      0.369155   0.132311   2.790  0.00527 **
## milex_logged                      4.282918   0.635723   6.737 1.62e-11 ***
## land_CONT                         0.006236   0.008217   0.759  0.44786
## cinc_MC                           0.762750   1.239684   0.615  0.53837
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2931  on 10466  degrees of freedom
## Residual deviance: 2858  on 10461  degrees of freedom
##   (5132 observations deleted due to missingness)
## AIC: 2870
##
## Number of Fisher Scoring iterations: 6
```

```r
## DV: binary measure of participation in militarized disputes over economically
## valuable territory
model8 <- glm(tc_mid_resource_dummy_ICOW ~ land_oriented_medium_binary +
                autocracy_BX + milex_logged + land_CONT + cinc_MC,
              family = binomial(link = "logit"), data = dat)

summary(model8)
```

```
##
## Call:
## glm(formula = tc_mid_resource_dummy_ICOW ~ land_oriented_medium_binary +
##     autocracy_BX + milex_logged + land_CONT + cinc_MC, family = binomial(link = "logit"),
##     data = dat)
##
## Coefficients:
```

```
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -2.162e+01  2.011e+00 -10.754  < 2e-16 ***
## land_oriented_medium_binary  9.611e-01  1.715e-01   5.605 2.08e-08 ***
## autocracy_BX                2.261e-01  1.305e-01   1.733   0.0831 .
## milex_logged                5.648e+00  6.423e-01   8.793  < 2e-16 ***
## land_CONT                  -8.704e-06  8.999e-03  -0.001   0.9992
## cinc_MC                     2.452e-01  1.203e+00   0.204   0.8386
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3276.2  on 11578  degrees of freedom
## Residual deviance: 3146.2  on 11573  degrees of freedom
##   (4020 observations deleted due to missingness)
## AIC: 3158.2
##
## Number of Fisher Scoring iterations: 7
```

Exercise 2: Create a Publication-Grade Regression Table

Create a publication-quality regression table summarizing the results of your regression models.

```r
# Create a regression table in HTML format. This might different for everyone because the
# variables you choose may be different, but the functions and process should be very
# similar.
htmlreg(list(model1, model2, model3, model4, model5, model6, model7, model8),
        caption = "Regression Results",
        custom.model.names = c("Model 1", "Model 2", "Model 3", "Model 4",
                               "Model 5", "Model 6", "Model 7", "Model 8"),
        custom.coef.names = c("Intercept", "Economic Orientation (Binary)",
                              "Economic Orientation (Continuous)",
                              "Autocracy", "Military Expenditures (Logged)",
                              "Land Borders", "Military Capabilities"),
        file = "regression_results.html")
```

Bonus Exercise

For this bonus assignment, extend one of your regression models from Exercise 1.2 by controlling for unobserved time heterogeneity using fixed effects, and present the results in a publication-quality HTML table. Incorporating time fixed effects helps control for time-specific factors (e.g., economic crises), thereby providing a clearer picture of the relationship between economic orientation and territorial disputes.

To do so, update one of your logistic regression models to treat the year variable as a factor (using `as.factor(year)`). This will create dummy variables for each year, capturing time-specific influences on the dependent variable. Because including dummy variables for every year can result in an overwhelming number of coefficients in your output, omit these individual time coefficients from your table using the `omit.coef` function. Instead, add a custom row indicating that time fixed effects have been incorporated with the `custom.gof.rows` function.

Note: There are specialized packages (such as `plm`) for fixed effects analysis, we will focus on the basic approach of controlling for time heterogeneity by introducing dummy variables via `as.factor(year)`.

```r
# Run a logistic regression with time fixed effects using as.factor(year)
## DV: binary measure of participation in militarized disputes over economically
## valuable territory
model_time_fe <- glm(tc_mid_resource_dummy_ICOW ~ land_oriented_medium_binary +
                     autocracy_BX + milex_logged + land_CONT + cinc_MC +
                     as.factor(year), family = binomial(link = "logit"),
                     data = dat)
summary(model_time_fe)

# Create regression table in HTML format
htmlreg(list(model1, model2, model3, model4, model5, model6, model7, model8, model_time_fe),
        caption = "Regression Results with Time Fixed Effects",
        caption.above = TRUE,
        custom.model.names = c("Model 1", "Model 2", "Model 3", "Model 4",
                               "Model 5", "Model 6", "Model 7", "Model 8",
                               "Time FE Model"),
        custom.coef.names = c("Intercept", "Economic Orientation (Binary)",
                              "Economic Orientation (Continuous)",
                              "Autocracy", "Military Expenditures",
                              "Land Borders", "Military Capabilities"),
        omit.coef = "as.factor\\(year\\)",
        ## Omit year dummies from the table
        custom.gof.rows = list("Time Fixed Effects" = c("No", "No", "No", "No",
                                                        "No", "No", "No", "No",
                                                        "Yes")),
        file = "regression_results_fe.html")
```