

**GSERM 2017 – Regression III**  
*Final Examination*

**Instructions**

1. This examination has two sections. **You are to choose one question from each section and answer it.** *Do not attempt to answer both questions in either section.*
2. Your answer to the question in Section I should take the form of a short (< 800 word) essay.
3. Your answer to Section II should take the form of a brief empirical analysis of the data that answers the question(s) asked. That may include plots, tables, and any other techniques you think will be useful in answering the question.
4. You have two options for completing the exam:
  - (a) You may choose to complete the exam “in class” and submit it (following the instructions below) no later than 3:30 p.m. St. Gallen time on Friday, June 23, 2017. All exams submitted at or before this time will be graded together.
  - (b) Alternatively, you may submit the exam as a “take home” exam, no later than 5:00 p.m. St. Gallen time on Monday, June 26, 2017. All exams submitted at or before this time will be graded together.
5. When you have completed the exam, you will submit your answer to the instructor electronically as a PDF file by e-mailing it to [zorn@psu.edu](mailto:zorn@psu.edu).

**Section I**

1. Nonlinear relationships present a unique set of challenges for empirical analysts. Broadly speaking, two approaches exist to addressing potential nonlinearity: transforming data so that linear models can be used, or fitting explicitly nonlinear models. In your essay discuss these two approaches. What are their relative advantages and disadvantages? Under what conditions might one choose to adopt one approach over the other, and why?
2. “Outliers” are a persistent concern in the analysis of observational data. In your essay, discuss the conceptual, statistical, and substantive importance of outliers. How do we know them when we see them? How can we identify them in the context of regression models? What are strategies for dealing with outlying observations?

## Section II

1. In December 1998, the members of the U.S. House of Representatives voted on four “articles of impeachment” of President William Clinton. Articles which passed would then be sent to the U.S. Senate for trial; if convicted, President Clinton would have been removed from office. The articles, therefore, were widely (and correctly) viewed as a referendum on President Clinton’s legitimacy and leadership.

We will examine data on the House’s impeachment votes on President Clinton, in an effort to assess the important predictors of those votes. The data are available in the course github repository’s “Exam” folder in .CSV format as `impeachment.csv`. The units of observation are the 433 voting members of the U.S. House of Representatives. The variables in the data are as follows:

- `votescum` is the number (from zero to four) of articles in favor of which a given member voted. That is, a `votescum` of 1 means that the member in question voted in favor of one of the four articles of impeachment; a `votescum` of two means s/he voted for two of those articles, and so forth. Since a vote in favor of an article was a vote against President Clinton, higher numbers denote stronger opposition to Clinton’s leadership. All 433 members voted (“yea” or “nay”) on all four articles.
- `pctbl96` equals the percentage of that member’s House district that was African-American, as of 1996.
- `unionpct` is the proportion of that member’s House district that were members of organized labor. Both African-Americans and union members are generally more supportive of the Democratic party, and of Clinton.
- `clint96` equals the percentage of the two-party presidential vote that President Clinton received in that district in the 1996 election.
- `GOPmember` is a dummy variable, coded 1 if the representative was a member of the Republican party and 0 if they were a Democrat.
- `ADA98` is the member’s 1998 Americans for Democratic Action score, a measure of how liberal that member’s voting behavior is; it ranges from a low of zero (most conservative) to a high of 100 (most liberal).

Summary statistics for the data are as follows:

```
> summary(impeachment)
```

name	state	district	votesum	pctbl96
Length:433	Length:433	Min. : 1	Min. :0.00	Min. : 0.0
Class :character	Class :character	1st Qu.: 3	1st Qu.:0.00	1st Qu.: 2.0
Mode :character	Mode :character	Median : 6	Median :2.00	Median : 5.4
		Mean :10	Mean :1.85	Mean :11.9
		3rd Qu.:13	3rd Qu.:4.00	3rd Qu.:14.0
		Max. :52	Max. :4.00	Max. :74.0

  

unionpct	clint96	GOPmember	ADA98
Min. :0.0257	Min. :26.0	Min. :0.000	Min. : 0.0
1st Qu.:0.0930	1st Qu.:42.0	1st Qu.:0.000	1st Qu.: 5.0
Median :0.1690	Median :48.0	Median :1.000	Median : 30.0
Mean :0.1636	Mean :50.3	Mean :0.527	Mean : 46.3
3rd Qu.:0.2150	3rd Qu.:57.0	3rd Qu.:1.000	3rd Qu.: 90.0
Max. :0.3733	Max. :94.0	Max. :1.000	Max. :100.0

### *Directions*

1. Choose a statistical model to assess the relationship between the five covariates and `votesum`. Discuss briefly the reason(s), statistical and/or substantive, for your choice.
2. Fit the model, and discuss in substantive terms the marginal impact of each of the covariates on the response variable `votesum`, using whatever methods you deem appropriate.
3. Finally, compare your findings using your chosen method to an alternative (and presumably less plausible) statistical approach. What, if any are the differences between the results, and which do you feel are more accurate/correct?

2. The research question in this exercise is the influence of ideology on Supreme Court voting over time. In particular, the conventional wisdom states that political actors often become more politically moderate over time. To test this hypothesis, you'll examine data on the voting patterns of justices sitting on the Supreme Court during the Vinson, Warren, Burger, and Rehnquist courts (1946-1994) ( $N = 32$ ,  $T = 49$ , unbalanced). The data are available on the course github repository, in the Final Exam folder, as `court.csv`. The variables are:

- `justice` (the justice identifier variable),
- `year` (the year identifier),

- `civlib` (the percentage of liberal votes cast by that justice in civil rights and liberties decisions in that year),
- `econs` (the percentage of liberal votes cast by that justice in economics decisions in that year),
- `score` (the normed “Segal/Cover” (1989) ideology score of the justice, ranging from -1 (most right/conservative) to 1 (most left/liberal) and
- `tenure` (the number of years the justice has served on the Court, as of that year).

If the conventional wisdom is correct, one possible manifestation is that the effect of `score` on voting liberalism should be positive, but the interaction of `score` and `tenure` should be negative (as justices moderate their extremism later in their careers).

Your assignment is as follows:

1. First, examine voting liberalism in civil rights and liberties cases (`civlib`).
  - Estimate both fixed- and random-effects models for this outcome, and discuss your results, both substantively and statistically.
  - Evaluate the statistical and substantive plausibility of random effects, and compare the two sets of results in terms of model specification.
2. Repeat the above steps for the variable on economics cases (`econs`).
3. Talk in general terms about which model(s) you prefer for these analyses, and why.