Quantitative Methods Exam

UC Berkeley · Travers Department of Political Science

Summer 2011

There are three section to this exam: quantitative Methods, formal modeling, and Data and Design. Some of the questions are easier than others. Please do your best to answer them; partial grades will be given. You do not need to answer every question perfectly to do well on the exam. If you need to make additional assumptions not stated in a given question in order to obtain an answer, please clearly state these assumptions and the reasons for them.

1 Quantitative Methods

1. Let X and Y be random variables with finite means. Prove the law of iterated expectations, $E(E(Y|X)) = E(Y)$.

2. Consider a simple random sample of N observations of Y and X from some joint distribution.

   (a) For the least squares regression of sample y's on sample x's, $\widehat{y} = \widehat{\alpha} + \widehat{\beta}x$, derive the slope $\widehat{\beta}$ and intercept $\widehat{\alpha}$ by minimizing the mean squared error of the regression.

   (b) Show that this solution implies that sample x's are uncorrelated with the residuals.

3. Suppose $Y = \alpha + \beta x + \varepsilon$, where ε is a random variable with $E(\varepsilon) = 0$ and $Var(\varepsilon|x) = \sigma^2$ for all x.

   (a) Suppose that α and β are estimated by ordinary least squares regression. What is the expected value of $\widehat{\beta}$ in repeated simple random sampling?

   (b) Identify a condition such that $E(\widehat{\beta}) = \beta$, and interpret its substantive meaning.

4. Prove that if one conditions non-parametrically (e.g., by matching) on the true propensity score, asymptotically, the observed covariates, X, will be balanced between treatment and

control groups, where the propensity score is a function of X.

5. In the potential outcomes framework, the causal effect of treatment $T \in \{0, 1\}$ on unit i is $\tau_i = Y_i(1) - Y_i(0)$, where $Y_i(1)$ denotes the potential outcome of i under treatment and $Y_i(0)$ the potential outcome under control. In a regression discontinuity (RD) design, if the potential outcomes are distributed smoothly at the cut-point, the design estimates the average causal effect of treatment at the cut-point, $Z_i = c$:

$$\tau_{RD} \equiv \mathbb{E}[Y_i(1) - Y_i(0)|Z_i = c] = \lim_{Z_i \downarrow c} \mathbb{E}[Y_i(1)|Z_i = c] - \lim_{Z_i \uparrow c} \mathbb{E}[Y_i(0)|Z_i = c]. \quad (1)$$

   This equation is for the sharp RD design. In Lee's (2008) application to U.S. House election it has been shown that baseline covariates, such as campaign contributions, are not balanced in a neighborhood of the cutpoint. What RD assumptions does this violate if any?

   Formally define a fuzzy RD design, as done in Equation 1. Could a fuzzy design overcome the problems found in using RD with U.S. House elections? Why or why not?

2 Formal modeling

Answer 2 of the following 5 questions. Each question has equal weight.

1. Discordia is in the throes of what has become known as the Bougainvillea Revolution. The Plaza of Perpetual Passivity in the center of the capital city is packed with protesters calling for the ouster of the long-entrenched President. The President has ordered the military to clear the square. The military's leaders must choose whether to clear the square by shooting tear gas canisters or by shooting bullets. The next day, the protest leaders will have to decide whether to escalate the protests to a nationwide uprising or to back down. If the movement backs down the game ends. If the protests escalate, the military will then have to decide whether to support the President in instituting a nationwide crackdown or to change sides and place the President under arrest. The protest movement is uncertain whether the military is fundamentally loyal to the President or if it is sympathetic to the regime change. If the military is loyal, the cost it incurs from clearing the square with tear gas is 0 and its cost of clearing with bullets is 2. If it is sympathetic these costs are 0 and 6. If the movement is cleared from the square with tear gas, the movement incurs a loss of d, and if it is cleared with bullets it incurs a loss of $D > d$. If there is no escalation, the military gets a payoff of 5. Supporting the crackdown gives the military a benefit of 2 if it is loyal to the President but costs it 2 if it is sympathetic to the movement. If it does not support the crackdown following escalation by the movement it incurs no costs or benefits. At the start of the interaction the protest movement believes the military is loyal with probability $\alpha = \frac{1}{2}$. The protest movement can be treated as a unitary decision-maker, as can the military. The President makes no decisions in this game. The movement's payoff for backing down is 0. The movement's payoff for escalation followed by a crackdown is -100. The movement's payoff for escalation that does not result in a crackdown is 50.

   (a) Write out the game tree for this interaction.

   (b) Let $\hat{\alpha}$ designate the movement's beliefs about the probability the military is loyal fol-

lowing the initial clearing of the square. How high must $\hat{\alpha}$ be to ensure that the movement will not choose to escalate?

(c) Is there a separating equilibrium in which only a loyal military will clear the square with bullets?

(d) Is there an equilibrium in which the military clears with bullets regardless of its loyalty?

(e) Is there an equilibrium in which the military clears with tear gas regardless of its loyalty?

(f) In what ways do the values of d and D affect equilibrium play?

2. Continuing with the analysis of the Revolution game, do the following:

(a) Find the mixed-strategy equilibrium of this game in which the military, if loyal, uses tear gas with probability μ, and the movement backs down after tear gas is used with probability λ.

(b) Let $\alpha = \frac{1}{4}$. Characterize a pooling Perfect Bayesian Equilibrium of this game that would violate the Intuitive Criterion (invented by Cho and Kreps) and demonstrate that it does so.

(c) Now suppose that the military can clear the square with any level of brutality $b \geq 0$, incurring cost $c(b) = \frac{b}{\theta}$, where $\theta = 1$ if the military is sympathetic and $\theta = 2$ if the military is loyal. Construct a pooling equilibrium in which the military chooses a strictly positive level of brutality $b^* > 0$ irrespective of its type. Make sure to specify the opposition's beliefs for all possible off-path behavior by the military.

3. Two candidates, A and B, each propose (and can commit to) a platform with a level of taxes τ and a level of public goods g. The government's budget constraint is $\tau \geq g + r$, where r is rents from holding office. Candidate j's utility is $p_j r$, where p_j is candidate j's probability of winning. Losing candidates receive utility of 0. There is a continuum of citizens indexed

i. A citizen i will vote for candidate A if

$$1 - \tau_B + h(g_A) > 1 - \tau_B + h(g_B) + \sigma_i$$

and will vote for B otherwise. σ^i is distributed uniformly on $[-1, 1]$. $h(g) = \ln g$.

(a) Solve for the equilibrium proposed levels of taxes and rents. Hint: Solve for the optimal proposal g_A first.

(b) Explain the following in a way that would be understandable to someone who has never studied game theory:

i. What does $h(g)$ represent?

ii. What does σ_i represent?

iii. Why do g and r take the values you found, and what does the equilibrium tell us about politics?

(c) What assumption(s) might be problematic or too unrealistic, and how would you conjecture that the results would differ if these assumption(s) changed?

4. Consider a game between a dictator D and an opposition O. Let policy be a unidimensional variable $x \in x\mathbb{R}$. The dictator's ideal point is 0 and the opposition's ideal point is 1, which also happens to be the preferences of the median person in this country. Ruling the country generates rents in the amount x, because ordinary people in the country are more motivated to produce when policy is closer to their ideal. Both players receive additively separable utility that derives from the distance of policy to their ideal point and from whatever share of the rents they receive. Thus, with a policy of x and a share of rents $s \in [0, 1]$ granted to the opposition, the dictator has a payoff of $(1 - s)\, x\, -x^2$ and the opposition has a total payoff of $sx - (1 - x)^2$. The game has two moves. First, the dictator offers the opposition a policy proposal $x \in [0, 1]$ and a share of the rents $s \in [0, 1]$. Second, the opposition either accepts or rejects this offer and attempts to overthrow the dictator, succeeding with

probability $q \in (0,1)$. Whoever wins that conflict can choose any policy and keep all the rents. Furthermore, if the dictator wins he imposes a punishment of L on the opposition, where $L = 1$ is exogenously fixed. Thus, the utility of the opposition for a successful overthrow is $\max_x x - (1-x)^2$, and the utility of a failed overthrow is -1. The utility to the dictator of a successful overthrow is 0 and the utility of a failed overthrow is $\max_x x - x^2$.

(a) What is the expected utility to the opposition of an overthrow attempt?

(b) Characterize the set of offers (x, s) that will be accepted by the opposition.

(c) What is the expected utility to the dictator if the opposition attempts an overthrow?

(d) If $q = \frac{1}{4}$ what is the equilibrium proposal (x^*, s^*) from the dictator? Will it be accepted?

(e) If $q = \frac{3}{4}$, what is the equilibrium proposal (x^*, s^*) from the dictator? Will it be accepted?

(f) Find the conditions on q necessary for the dictator to make an offer $s^* > 0$.

5. How important is empirical testing to validate a formal model? Is a formal model with no testable implications valuable? To what extent are any formal models testable? How does the endeavor of testing a formal model differ from other empirical work? Discuss with reference to specific published models. You may restrict attention to one substantive area/subfield or range more widely in your answer.

3 Data and Design

This portion of the exam consists of two different sections, both of which are equally weighted. Please make sure you have read all of the articles in the reference list carefully. Most of these articles were previously assigned.

3.1 Fearon and Laitin 2003

This section is based on Fearon and Laitin (2003). This is a prominent article which examines if ethnic and religious antagonisms are the cause of the civil wars which proliferated as the cold war ended. Please read the article before proceeding.

Table 1 of this article can be replicated using the following R file: `http://sekhon.berkeley.edu/qe/FearonLaitin_replication_qe1.R`. The output of this R file can be found here: `http://sekhon.berkeley.edu/qe/FearonLaitin_replication_qe1.Rout`. Reading the R source code and comparing the output with Table 1 suffices to define the variables in the dataset.

Please answer the following questions:

1. Many researchers have questioned the findings of Fearon and Laitin. One concern has been that the countries which undergo civil wars are very different from those which do not so there is a lack of common support. The parametric models which Fearon and Laitin use have to rely upon extrapolation. So, let's use matching to see if we recover the same substantive results as Fearon and Laitin.

   (a) Take Model 1 from Table 1, and estimate the causal effect of ethnic fractionalization using matching. The estimand is ATT. For this question, match on all of variables which Fearon and Laitin include in their Model 1. In order to use matching, make the ethnic fractionalization variable discrete. For this question: call control all countries with ethnic fractionalization levels below the median of this variable, and call treated all countries above the median. Note that we only need to match *countries* since their

level of ethnic fractionalization does not change over time. After matching, consider what to do with the time dimension of the data. *Hint:* consider how matching was used in both Gordon and Huber (2007) and Galiani, Gertler, and Schargrodsky (2005). Both articles have data across units and over time but still use matching.

(b) Redo question 1, but make ethnic fractionalization discrete but not dichotomous—e.g., make it have at least three different categories. Discretize the variable with an eye towards maximizing the chances of finding a significant effect. Please redo the matching.

(c) Analyze the research design used in this article. What identification strategy is used? What are the inferential problems? Is there a way forward? What is the probative value of the evidence for *any* causal question?

2. Using this dataset, someone has estimated an OLS where the dependent variable is per capita income. In particular, the analyst estimates the following model which is based on Model 1 in Table 1 of Fearon and Laitin (2003) (but with a different dependent variable). The dependent variable is per capital income at time t (variable name: $gdpen$). And this is regressed on: per capital income at $t-1$ ($gdpenl$), war at $t-1$ ($warl$), log(population) at $t-1$ ($lpopl1$), log(% mountainous) ($lmtnest$), noncontigous state ($ncontig$), oil exporter (Oil), new state ($nwstate$), instability at $t-1$ ($instab$), democracy at $t-1$ ($polity2l$), ethnic fractionalization ($ethfrac$) and religious fractionalization ($relfrac$). Note that the time-series cross-section nature of the dataset is maintained.

   The following R code estimates this model: `http://sekhon.berkeley.edu/qe/FearonLaitin_income_qe1.R`.

   (a) Please provide bootstrap confidence intervals for every parameter of this least squares model in the manner you think is most appropriate. For this question write your own bootstrap code—e.g., you cannot use functions in the `boot` and related libraries. Given the complex structure of the dataset, the `boot` function will not work anyways.

   (b) Under what assumptions does your bootstrap provided correct coverage?

(c) Please provide bootstrap confidence intervals for this regression model using a "better bootstrap"—e.g., a bootstrap based on a pivot statistic like the studentized version of the difference of means bootstrap (case 3 in the lecture notes). If you have already done this for (a), ignore this question.

(d) Please provide bootstrap confidence intervals for the logistic regression Model 1 in Table 1 of Fearon and Laitin.

3.2 Regression Discontinuity and Close House Elections

This question concerns Lee (2008) and Caughey and Sekhon (2011). The following URL provides David Lee's dataset: `http://sekhon.berkeley.edu/stuff1/LeeRDdata.zip`

Note that Lee's dataset is different from that used by Caughey and Sekhon (2011). For example, it includes far fewer variables and it contains some errors and missing values that were imputed. Both the Caughey and Sekhon article and an appendix that includes details about their dataset are available on Sekhon's webpage.

Questions:

1. Please use David Lee's replication files to replicate the tables and figures in Lee's article.

2. To the extent possible please use Lee's dataset to replicate the key tables and figures in Caughey and Sekhon (2011). Which key findings differ between Lee (2008) and Caughey and Sekhon (2011) because of data differences and which findings are consistent even if one uses Lee's original dataset?

References

Caughey, D. and J. S. Sekhon (2011). Elections and the regression-discontinuity design: Lessons from close u.s. house races, 1942–2008. *Political Analysis*.

Fearon, J. D. and D. D. Laitin (2003). Ethnicity, insurgency, and civil war. *American Political Science Review 97*(1), 75–90.

Galiani, S., P. Gertler, and E. Schargrodsky (2005). Water for life: The impact of the privatization of water services on child mortality. *Journal of Political Economy 113*(1), 83–120.

Gordon, S. and G. Huber (2007). The effect of electoral competitiveness on incumbent behavior. *Quarterly Journal of Political Science 2*(2), 107–138.

Lee, D. S. (2008). Randomized experiments from non-random selection in u.s. house elections. *Journal of Econometrics 142*(2), 675–697.