EXAMINATION:
EMPIRICAL ANALYSIS AND RESEARCH METHODOLOGY

Yale University

Department of Political Science

August 2013

You have seven hours (and fifteen minutes) to complete the exam. You can use the points assigned to each question as a (rough) guide to allocation of your time. The exam is 100 points long, which is approximately 15 points per hour. This exam consists of two parts.

First Part (9:00am - 2:15pm; 70 points)
The First Part of the exam will be distributed at 9am. Your answers for the First Part will be collected at 2:15pm. The only aids permitted for the First Part are (1) one page of double-sided notes, (2) a calculator, and (3) a word processor on one of the Statlab computers to write up your answers (you may also write up answers using pencil/pen and paper). At 2:15 all your answers for the First Part must be submitted to the proctor. Hold on to the instructions for the First Part of the exam.

Advice: Use algebra to back up your assertions wherever appropriate, and remember to show your work. Do not leave sub-parts of questions unanswered. Your answers should be succinct and to the point.

Second Part (2:30pm - 4:30pm; 30 points)
The Second Part of the exam will be distributed at 2:30pm. Your answers will be collected at 4:30pm. The Second Part involves using statistical software. A complete answer to the Second Part will include the code and output, as well as your written answers. For the Second Part you are permitted (1) unrestricted use of your own computer with access to the internet or (2) use of a Statlab computer with access to the internet. The only restriction for the Second Part is that you may not interact with anyone, online or otherwise. In addition, you must credit in your answer any sources (code or references) that help you.

Advice: Explain what you are trying to do in comments. Even if you are not able to execute your program correctly, you can receive partial grades for explaining clearly what you wanted to do and why. We recommend you use a program such as *knitr* (see here) for R or *log* for Stata to easily record the input and output from your analysis.

1 Part One of EARM Exam. 9:00-2:15. (70 points)

Rules: No references or aids permitted except one page of notes and a calculator. Answers may be written up by hand or word processor.

1.1 (4 points)

We have a random sample of size n from some population. Give an example of:

1. a consistent and unbiased estimator of the mean of the population
2. an inconsistent and unbiased estimator of the mean of the population
3. a consistent and biased estimator of the mean of the population
4. an inconsistent and biased estimator of the mean of the population

1.2 (10 points)

Consider the model $Y_i = \alpha + \epsilon_i$, where Y_i is an observation, α is some scalar parameter, and ϵ_i is a disturbance term. Only Y_i is observed. Let k be some estimate, and r_i the residuals from estimate k: $r_i = Y_i - k$.

1. (3 points) Derive the estimator that minimizes the sum of the squared residuals (denote this $\hat{\alpha}$):

$$\hat{\alpha} = \arg\min_k \left(\sum_{i=1}^{N} (Y_i - k)^2 \right)$$

2. (4 points) State the assumptions necessary to prove $\hat{\alpha}$ is an unbiased estimator of α. Prove that given those assumptions, $\hat{\alpha}$ is an unbiased estimator of α.
3. (3 points) Given the assumptions you stated in part 2, derive the variance of $\hat{\alpha}$.

1.3 (12 points)

Consider two experiments. The first, E_1, gives us:

$$\frac{P(d_{E_1}|H_a)}{P(d_{E_1}|H_0)} = 5$$

$$p_{E_1} = 0.1$$

The second, E_2, gives us:

$$\frac{P(d_{E_2}|H_a)}{P(d_{E_2}|H_0)} = 1.5$$

$$p_{E_2} = 0.001$$

$P(d_{E_k}|H_i)$ denotes the probability of the observed data from experiment k given hypothesis i. (Implicit to this is that the alternative is specific, so that H_a implies a $P(d|H_a)$.)

p_{E_k} denotes the p-value from a test of the null hypothesis (H_0) against the alternative hypothesis (H_a) using the data from experiment k.

1. (2 points) Explain what is $\frac{P(d_{E_k}|H_a)}{P(d_{E_k}|H_0)}$.
2. (2 points) Explain what is a p-value.
3. (4 points) Explain why the above scenario may seem puzzling. Is this scenario possible? If not, explain why not. If so, explain how it is possible.
4. (4 points) Which experiment provides stronger evidence in favor of H_a, and against H_0? (Explain what you mean by "stronger evidence".)

1.4 (4 points)

What is a local average treatment effect (LATE)? In particular, what is the LATE for:

1. a laboratory experiment
2. a regression discontinuity design
3. one-to-one matching on treated units with a caliper
4. OLS regression with fixed effects (fixed effects = a dummy variable for every cross-sectional unit)

1.5 (4 points)

Consider the regression equation $Y_i = \alpha + X_i\beta + \epsilon_i$. Assume $\epsilon_i \sim N(0, \sigma^2)$. You use OLS to estimate the parameters, and find that $\hat{\beta}$ is 6 with an estimated standard error of 3. You then do a t-test with the null hypothesis that β equals 0. Are the following statements true or false, or is there insufficient information to answer? Explain.

1. $\hat{\beta}$ is statistically significant.

2. The probability that $\beta = 0$ is about 5%.

3. You can be about 95% confident that $\beta \neq 0$.

4. The results from this t-test provides evidence that the model is correct.

1.6 (4 points)

A paper reports some survey experiments of American citizens. The outcome of interest is their support for US military interventions. The causal factors of interest are variables related to how the military intervention is presented: the stakes of the conflict (T_S), whether elites from both political parties endorsed the intervention (T_E), and the expected fatalities that the US will suffer (T_F). The paper articulates, and briefly justifies, some theoretical predictions about how greater stakes of conflict (T_S) should lead educated men to increase their support for intervention, but otherwise have no effect; bipartisan endorsement (T_E) should increase support of uneducated Republicans, but otherwise no effect; and expected fatalities (T_F) should decrease support of educated Republican women, but otherwise have no effect.

The following table reports, for each experiment, the n, p-value, and subgroup for each hypothesis test of the null of no ATE against the alternative of a positive or negative ATE, and the subgroup for the hypothesis test. p-values are one-sided, calculated according to the above predictions. If you only saw this information, what concerns might you have about the reported results? Specifically, Smedley recalls hearing a methods scholar express concern about experiments with small-n, but he can't recall the reason why this would be a concern given that the results are significant. Can you explain why one might be legitimately concerned about small-n (even when results are significant), or explain why in this case it is not a concern?

Table 1: Results from Paper (p-values are one-sided)

Treatment	p-value	n	Sub-Group
T_S	0.04	55	educated men
T_E	0.02	62	uneducated Republicans
T_F	0.03	43	educated Republican women

1.7 (5 points)

You are asked to review a paper that claims that having daughters makes legislators vote more liberally on women's issues, such as equal pay and abortion rights. The authors write,

> Among members of Congress who have two children, those with two daughters voted liberally on women's issues 70 percent of the time, while those with two sons voted liberally only 50 percent of the time. Similarly, among members of Congress who have three children, those with three daughters voted liberally 80 percent of the time, and those with three sons voted liberally 40 percent of the time.

Are you convinced by this evidence? Are there other comparisons you would like to see before recommending publication? What, if anything, would you advise the authors to do differently?

1.8 (3 points)

Evaluate the following claim:

> The matching process mimics random assignment. Applying matching to observational data modifies the data such that they approximate experimental data. Thus, we can analyze observational data through the lens of experimental design and differentiate statistical association from causal effect.

1.9 (12 points)

Professor Smedley theorizes that participating in the political process makes citizens more knowledgeable about politics generally. In order to evaluate his theory, he got the principal of Augusta High School in Maine (where voters may register at the polling place on election day) to let him conduct an experiment on the senior class, all 200 of whom would be 18 years old and eligible to vote by election day that year. Smedley made a list of the seniors and randomly chose 50 of them to be in a treatment group. Realizing he couldn't actually force anyone to participate in politics, he instead decided to try to persuade the students in the treatment group to vote by giving them a one hour lecture about why voting is essential to a well-functioning democracy. After election day, he found out which of the seniors voted by looking at the official record. Smedley reasoned that since he randomly assigned his lecture, he could analyze the effect of voting on political knowledge without worrying about endogeneity. For his dependent variable, Smedley then gave all 200 students a quiz after election day that measured their knowledge of current political events.

Using a difference in means test, Smedley found that the students who voted scored 15 points higher on the quiz than the students who didn't vote. The difference was significant. He concluded "As I expected, political participation has a positive effect on political knowledge."

1. What has Smedley actually estimated, and is there any risk that it might be a biased estimate of the causal effect of voting on quiz score?

2. How would you analyze his data differently? Be explicit about what you would do, what kind of causal effect you're estimating, and what assumptions are required to identify that effect.

3. Smedley proudly shows you a note from one of the Augusta students that reads:

   > Dear Prof Smedley, Thank you so much for coming to our school. Your lecture on voting was so inspiring and really peaked my interest in politics. I told all my girlfriends who didn't attend all about what you said. You're the best, Suzie.

   What does this note imply about the strengths and weaknesses of different estimation strategies? Explain.

4. If you had been able to help Smedley at the design stage of his project, would you have advised him to do anything differently? Explain.

1.10 (12 points)

Suppose we have a dataset that consists of the following variables for each observation:

$$\{Y_{i,t}, Y_{i,t-1}, D_{i,t}\}$$

where i denotes the cross-sectional unit, the second subscript denotes the time-unit (either t or $t-1$), $Y_{i,t}$ denotes the outcome, $Y_{i,t-1}$ denotes the lagged outcome, $D_{i,t}$ denotes the treatment variable. Assume that we only observe one set of these variables for each cross-sectional unit (so we don't measure $D_{i,t-1}$).

Consider three possible data generating processes (DGP): $\mathcal{D}_G$, $\mathcal{D}_1$, $\mathcal{D}_2$. $\mathcal{D}_G$ is the most general. For $\mathcal{D}_G$ we assume that the lagged outcome $Y_{i,t-1}$ can affect the future outcome but cannot affect the treatment variable, the treatment variable can affect the outcome (but obviously not the lagged outcome, which is in the past), the treatment variable and lagged outcome may be confounded (may share a common unobserved cause), and the lagged outcome and outcome may be confounded (share a common unobserved cause).

$\mathcal{D}_1$ is a special case of $\mathcal{D}_G$ in which we assume that there is no unobserved confounding between the lagged outcome and the outcome. $\mathcal{D}_2$ is a special case of $\mathcal{D}_G$ in which we assume that there is no effect of the lagged outcome on the outcome.

Formally we can write this as follows, where $\nu_{i,k}$, ϵ_i, and μ_i denote unobserved independent random variables from an unknown distribution, $\leftarrow$ denotes a deterministic causal relationship (it is like an $=$, except it clarifies that causality only goes one way), and f_k denote unknown functions.

$$\mathcal{D}_G : \begin{cases} Y_{i,t-1} & \leftarrow f_1(\epsilon_i, \mu_i, \nu_{i,1}) \\ D_{i,t} & \leftarrow f_2(\epsilon_i, \nu_{i,2}) \\ Y_{i,t} & \leftarrow f_3(\mu_i, D_{i,t}, Y_{i,t-1}, \nu_{i,3}) \end{cases}$$

$$\mathcal{D}_1 : \begin{cases} Y_{i,t-1} & \leftarrow f_1(\epsilon_i, \mu_i, \nu_{i,1}) \\ D_{i,t} & \leftarrow f_2(\epsilon_i, \nu_{i,2}) \\ Y_{i,t} & \leftarrow f_3(D_{i,t}, Y_{i,t-1}, \nu_{i,3}) \end{cases}$$

$$\mathcal{D}_2 : \begin{cases} Y_{i,t-1} & \leftarrow f_1(\epsilon_i, \mu_i, \nu_{i,1}) \\ D_{i,t} & \leftarrow f_2(\epsilon_i, \nu_{i,2}) \\ Y_{i,t} & \leftarrow f_3(\mu_i, D_{i,t}, \nu_{i,3}) \end{cases}$$

1.10.1 Non-parametric Identification

Suppose each of these variables has only a few levels. Is it possible to identify (create a consistent estimator for) the effect of $D_{i,t}$ on $Y_{i,t}$? If so, how? If not, why not? You may assume for the sake of simplicity that there is a constant treatment effect, which we will denote as β. Explain for each of $\mathcal{D}_G$, $\mathcal{D}_1$, $\mathcal{D}_2$.

1.10.2 Parametric Identification

Suppose now that each of these variables are continuous variables. Suppose we also know that each of f_k are linear additive functions (e.g. $f_1 = \alpha_1\epsilon_i + \alpha_2\mu_i + \alpha_3\nu_{i,1}$, where α_k are unknown parameters). Is it possible to identify (create a consistent estimator for) the effect of $D_{i,t}$ on $Y_{i,t}$? If so, how? If not, why not? Explain for each of $\mathcal{D}_G$, $\mathcal{D}_1$, $\mathcal{D}_2$.

1.10.3 Signing Bias

Suppose as in 1.10.2 that each of the f_k are linear additive functions. In addition, suppose that all of the parameters (other than β) are weakly positive ($\alpha_k \geq 0 \quad \forall k$). Denote your preferred estimator under $\mathcal{D}_1$ for β as E_1 and your preferred estimator under $\mathcal{D}_2$ for β as E_2. Is it possible to say anything about the direction of the bias if we use E_2 under $\mathcal{D}_1$? Is it possible to say anything about the direction of the bias if we use E_1 under $\mathcal{D}_2$? Explain.

2 Part Two of EARM Exam. 2:30-4:30. (30 points)

Rules: Software and Internet Permitted. All online resources permitted. You may use your own computer. You must cite any resource that you use in your answer. Functioning code must be provided as part of your answer.

Advice: Explain what you are trying to do in comments or otherwise. Even if you are not able to execute your program correctly, you can receive partial grades for explaining clearly what you wanted to do and why.

Recall from question 1.10 the problem of identifying the causal effect under different DGP: $\mathcal{D}_G$, $\mathcal{D}_1$, $\mathcal{D}_2$. For this question you will evaluate using simulations different estimators for these DGP. Assume a simple linear additive $\mathcal{D}_G$ as follows:

$$Y_{i,t-1} \leftarrow \epsilon_i + \mu_i + \nu_{i,1}$$

$$D_{i,t} \leftarrow \epsilon_i + \nu_{i,2}$$

$$Y_{i,t} \leftarrow \gamma\mu_i + \beta D_{i,t} + \theta Y_{i,t-1} + \nu_{i,3}$$

With each exogeneous variable drawn i.i.d from a standard normal: $\nu_{i,k} \underset{i.i.d}{\sim} N(0,1)$, $\epsilon_i \underset{i.i.d}{\sim} N(0,1)$, $\mu_i \underset{i.i.d}{\sim} N(0,1)$. Assume that the true causal effect is: $\beta = 1.4$.

For $\mathcal{D}_1$ we assume that $\gamma = 0$ and set $\theta = 0.33$ (this implies that $Cov(Y_{i,t}, Y_{i,t-1}) = 2.4$).

For $\mathcal{D}_2$ we assume that $\theta = 0$ and set $\gamma = 1$ (this implies that $Cov(Y_{i,t}, Y_{i,t-1}) = 2.4$).

2.1 Estimators (2 points)

State your preferred estimator for β for $\mathcal{D}_1$. Denote this E_1. State your preferred estimator for β for $\mathcal{D}_2$; denote it E_2. We will then evaluate these two estimators using Monte Carlos.

2.2 Monte Carlos (13 points)

Evaluate the bias and mean squared error of E_1 and E_2 when used in $\mathcal{D}_1$ and $\mathcal{D}_2$, for $n = 30$. You should therefore estimate the bias and MSE for four different scenarios. To avoid losing time waiting for the computer, you may confine yourself to $m = 1000$ simulations, or smaller if need be.

2.3 Conclusion (5 points)

Reflect on what 1.10 and Part 2 demonstrates about the appropriate use of lagged dependent variables (LDV) to estimate causal effects. What other issues may arise or lessons might one draw for the use of LDV that were not part of these questions (1.10 and Part 2)?

2.4 (Advanced) Generalizing to $\mathcal{D}_G$ (10 points)

Can you say anything more general about when E_1 or E_2 is preferred under $\mathcal{D}_G$ (so $\gamma > 0$ and $\theta > 0$)? Use simulations to investigate the conditions under $\mathcal{D}_G$ when each of these estimators has smaller MSE. Try to articulate a general rule for when each is preferred. (Feel free to modify other features of $\mathcal{D}_G$ if it would help.)