Quantitative Empirical Methods Exam

Yale Department of Political Science, January 2017

You have seven hours to complete the exam. This exam consists of three parts.

Back up your assertions with mathematics where appropriate and show your work. Good answers will provide a direct answer that illustrates an understanding of the question, and calculations or statistical arguments to validate the answer. Where applicable, exceptional answers will include all of these *as well as* proofs that are technically complete, including formally articulating sufficient assumptions and regularity conditions. Questions will not be weighted equally, and a holistic score will be assigned to the exam, and thus it is important to demonstrate your understanding of the material to the best of your ability.

Part 1 (Short Answer Section) consists of seven short answer questions. *Advice*: Note there are multiple correct answers to some questions, and we encourage you to give the most complete (but still succinct) solution possible. Do not leave sub-parts of questions unanswered.

Part 2 (Essay Section) contains a recent, well-regarded empirical article. We will ask you to offer an evaluation of its methodological approach and presentation of results. In particular, we will advise you to pay particular attention to the identification conditions (either explicit or implicit), the associated estimation strategy, and possible threats to inference. Your response may be anywhere from 500 to 1500 words.

The only aids permitted for Parts 1 and 2 are (i) one page of double-sided notes, (ii) a word processor on one of the Statlab computers to write up your answers (you may also write up your answers to Part 1 using pencil/pen and paper). After handing in your answers for Parts 1 and 2 of the exam, you may begin Part 3 (though feel free to look ahead). You may hand in Parts 1 and 2 whenever you wish, but we recommend spending no longer than five hours on Parts 1 and 2.

Part 3 (Computer Assisted Section) will involve using statistical software to answer one longer exercise with three associated questions. A complete answer to Part 3 will include code and output, as well as your written answers. *Advice*: We recommend that you explain what you are trying to do in comments. Even if you are not able to execute your program correctly, you can receive partial credit for explaining clearly what you wanted to do and why.

For Part 3, you are permitted (i) unrestricted use of your own computer with access to the internet or (ii) use of a Statlab computer with access to the internet. The only restriction for Part 3 is that you may not interact with anyone, online or otherwise. For Part 3 (Computer Assisted Portion) of the exam, please turn in a hard copy of your code to Colleen, and also email a digital copy of the code to colleen.amaro@yale.edu.

1 Short Answer Section

1. Prof. Smedley makes the following claim: "If $\rho(X, Y) = 0$ and $\rho(Y, Z) = 0$, then it must be the case that $\rho(X, Z) = 0$, where ρ is the correlation operator." Is Smedley correct? Either prove the claim or provide a counterexample.

2. Smedley intended to run a linear regression of Y on X and Z. He accidentally ran a linear regression of X on Y and Z. Please answer (with explanation) the following true/false questions about how the estimated coefficient on X from the intended regression ($\hat{\beta}_1$) relates to the estimated coefficient on Y ($\hat{\beta}_2$) from the second regression.
   (a) True or false? $\hat{\beta}_1$ is equal to $\hat{\beta}_2$ with probability 1.
   (b) True or false? The sign of $\hat{\beta}_1$ is equal to the sign of $\hat{\beta}_2$ with probability 1.
   (c) True or false? The "usual" Wald-type p-values against the null that the coefficient is zero using the classical standard error will be identical for $\hat{\beta}_1$ and $\hat{\beta}_2$ with probability 1.

3. Assume the following data generating process for some outcome y:

$$y = \alpha + \beta x + \epsilon.$$

   However, a researcher only observes a measurement of y that contains error, denoted y^* where the error is defined as $e_y = y - y^*$.

   (a) If the researcher estimates the following model, $y^* = a + bx + d$, under what nontrivial assumption(s) will $\hat{b}$ be an unbiased estimate for β? Will $\hat{b}$ be a consistent estimate of β under such assumption(s)? Show why or why not.
   (b) Now consider a case where the researcher measures x with error denoted as x^*, where the error is defined as $e_x = x - x^*$. The researcher (with an accurate measure y) estimates the following model: $y = f + gx^* + h$. Assume that $E[x^* \times e_x] = 0$ and $E[e_x] = 0$. Will $\hat{g}$ be an unbiased estimate of β? Show why or why not.
   (c) The researcher manages to take two such measurements of x, x_1^* and x_2^*. Denote the respective error as $e_{x1}^* = x - x_1^*$ and $e_{x2}^* = x - x_2^*$. Assume the following conditions hold:

$$E[x_1^* \times e_{x1}] = 0; E[e_{x1}] = 0$$

$$E[x_2^* \times e_{x2}] = 0; E[e_{x2}] = 0$$

   Using just y, x_1^* and x_2^*, can you present an estimator that will provide an unbiased estimate of β (i.e., the marginal effect of x on y)?

4. Suppose that we are trying to conduct inference on $\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[X_i]$. Assume that $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[X_i] = \mu$, such that μ is finite. Let $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Further suppose we computed a Wald-type confidence interval as $\hat{\mu} \pm \frac{1.96}{n}\sqrt{\sum_{i=1}^{n}(X_i - \hat{\mu})^2}$. Give at least one example of a data generating process such that the asymptotic coverage of the resulting confidence interval will not be (at least) 95%. [Note: this question is *not* identical to August 2016's question.]

5. Smedley convinced the local government of a small city in a low-income country to conduct a field experiment to study the effects of providing public goods on tax compliance. The study randomly assigned city blocks to one of two experimental conditions: an expansion in street pavement supply, and a control group, where the city had no street pavement projects. The full breakdown of assigned and actual treatment is presented in the table below, along with observed rates of later tax compliance in the two experiment arms.

		Assigned treatment	
		Street paving	No street paving
Actual treatment	Street paving	420	0
	No street paving	80	350
	Total N	500	350
Subsequent outcome	Tax compliance	.74	.70

(a) Since assignment to treatment was randomized, we can be confident that the independence assumption hold. Articulate a set of additional (nontrivial) conditions such that Smedley can estimate the causal effect of the provision of street pavement on tax compliance.

(b) Assume the conditions you listed in (a) hold, and calculate the $\widehat{ITT_D}$ (i.e., the effect of assignment to treatment on treatment), $\widehat{ITT}$ (i.e., the reduced form effect), and $\widehat{CACE}$ (i.e., the Complier Average Causal Effect). Interpret your results.

(c) Which of the conditions you listed in (a) can be evaluated empirically to ascertain their plausibility in this case? Given the evidence, how plausible is the case that such condition(s) hold?

2 Essay section

Read the article attached to your exam. Offer a critical evaluation of its methodological approach and presentation of results. Note: "critical" does not imply that you should only criticize – where praise is warranted, or where the authors' claims are well-justified, it is recommended that you give credit to the authors when their arguments are convincing and/or novel with respect to standard practice. Your response may be anywhere from 500 to 1500 words.

We advise you to pay particular attention to the identification conditions (either explicit or implicit), the associated estimation strategy, and possible threats to inference. Justify each of your claims and, where applicable, suggest ways in which this line of research might be improved. (We do not expect you to have special expertise in the topic area, but we do expect you to bring to bear your general analytical skills as a political scientist).

Article (with SI): Danziger, Shai, Jonathan Levav, and Liora Avnaim-Pessoa. 2011. Extraneous factors in judicial decisions. *PNAS*. 6889–6892.

3 Computer Assisted Portion

Scholars often discuss the bias-variance tradeoff. Scholars are often willing to accept the potential for some or greater bias in an estimator if it means that it will be more efficient under some circumstances. We would like you to conduct some simulations to illustrate this idea, in the context of estimators that tend to "favor" some parameter values over others.

Suppose we have n i.i.d. observations, $X_i \sim N(\mu, 1)$, and we are interested in estimating μ. In this setting, the sample mean $\hat{\mu}^0 = \frac{1}{n}\sum X_i$ is known to be:

- unbiased. An estimator $\hat{\mu}$ is unbiased if and only if $\mathrm{E}\left[\hat{\mu}\right] = \mu$ for any true value of μ.
- consistent. An estimator $\hat{\mu}$ is consistent if and only if $\hat{\mu} \xrightarrow{p} \mu$ for any true value of μ.
- the minimum variance unbiased estimator. An estimator $\hat{\mu}$ is unbiased if $\mathrm{Var}\left[\hat{\mu}^0\right] \leq \mathrm{Var}\left[\hat{\mu}^U\right]$ for any true value of μ and any unbiased estimator $\hat{\mu}^U$.

Thus we cannot improve on $\hat{\mu}^0$ without introducing the possibility of some bias. We want you to consider the behavior of three alternative estimators, $\hat{\mu}^1, \hat{\mu}^2, \hat{\mu}^3$, given different values of n and different true values of μ. Recall that the mean squared error of an estimator,

$$MSE[\hat{\mu}] = \mathrm{E}\left[(\hat{\mu} - \mu)^2\right] = \mathrm{E}\left[\hat{\mu} - \mu\right]^2 + \mathrm{Var}\left[\hat{\mu}\right].$$

The three estimators, which are all functions of the sample mean, are:

- A shrinkage estimator: $\hat{\mu}^1 = (1 - n^{-1/4}) \times \hat{\mu}^0$
- A fixed thresholding estimator: $\hat{\mu}^2 = \begin{cases} 0 & : \quad |\hat{\mu}^0| \leq 200^{-1/4} \\ \hat{\mu}^0 & : \quad |\hat{\mu}^0| > 200^{-1/4} \end{cases}$
- A sieve-type thresholding estimator: $\hat{\mu}^3 = \begin{cases} 0 & : \quad |\hat{\mu}^0| \leq n^{-1/4} \\ \hat{\mu}^0 & : \quad |\hat{\mu}^0| > n^{-1/4} \end{cases}$

Please answer the following three questions and all subquestions to the best of your ability. There is a (truly) optional bonus. Use at least 2500 draws for all simulations, and remember to set a seed.

1. Calculate the bias, variance and MSE of each of $\hat{\mu}^0, \hat{\mu}^1, \hat{\mu}^2, \hat{\mu}^3$ using simulations when:
   (a) $\mu = 0$ and $n = 10$.
   (b) $\mu = 1/4$ and $n = 10$.
   (c) $\mu = 1$ and $n = 10$.
   (d) $\mu = 0$ and $n = 1000$.
   (e) $\mu = 1/4$ and $n = 1000$.

(f) $\mu = 1$ and $n = 1000$.

2. Suppose that $n = 100$. Create four plots consisting of the MSE of each of $\hat{\mu}^0, \hat{\mu}^1, \hat{\mu}^2$ and $\hat{\mu}^3$ (on the y-axis) against the true value of μ (on the x-axis, with values ranging from -1 to 1) using simulations. [Note: this may be computationally intensive if you do not code efficiently.]

3. What do you conclude about each of the proposed estimators $\hat{\mu}^1, \hat{\mu}^2$ and $\hat{\mu}^3$?

   (a) Are any of the proposed estimators unbiased (i.e., $\mathrm{E}[\hat{\mu}] = \mu$ for any true value of μ)? For each estimator, provide evidence (simulations, theory, or intuition) for your logic.

   (b) Based on your simulations and/or theoretical calculations, which of the estimators are consistent (i.e., $\hat{\mu} \xrightarrow{p} \mu$ for any true value of μ)?

   (c) Would you ever recommend any of the proposed estimators? In what circumstances might you do so?

Bonus. $\hat{\mu}^0$ is the maximum likelihood estimator of μ, and thus should be "asymptotically efficient." One of the proposed estimators has asymptotically lower MSE than $\hat{\mu}^0$ when $\mu = 0$, and otherwise has the asymptotic MSE of $\hat{\mu}^0$. Which estimator (of $\hat{\mu}^1, \hat{\mu}^2$ and $\hat{\mu}^3$) is this? Explain why it can have this property. [Note: you are encouraged to use the Internet to answer this, but please cite your sources.]