Econometrics II: Assignment 2

Due: Thursday, February 20th

1 The r.v. Y is distributed according to the uniform distribution on $[a, b]$,

$$Y \sim \text{unif}[a, b].$$

Assume we observe n i.i.d. realizations $Y_1, \ldots, Y_n$. We are interested in $\mu = \mathbb{E}(Y)$.

For this question, you will need to look up expected value, variance, and covariances of order statistics. I do not expect you to know these by heart, but I leave looking them up as part of the exercise. But to be clear, you may quote these results without proof.

1.1 Show: The OLS estimator is $\hat{\mu}_{\text{OLS}} = \overline{Y}$, the sample average.

(Relatedly, recall that any OLS routine can be [ab]used to computed simple sample averages. How?)

1.2 Consider the alternative estimator that first orders all observations by size, say in increasing order, and then sets

$$\hat{\mu}_{\text{order}} = \frac{Y_1^* + Y_n^*}{2},$$

where $(Y_1^*, \ldots, Y_n^*)$ is the rearranged data set.

Show: (i) This estimator is unbiased. (ii) This estimator has weakly lower variance than $\overline{Y}$, strictly so if $n \geq 3$.

1.3 Does the Gauss-Markov Theorem apply to $\overline{Y}$? If not, why not? If yes, why is it not contradicted?

2 Consider simple linear regression (i.e., X is a scalar) without constant, i.e., the true model is:

$$Y = \beta \cdot X + \varepsilon,$$

though we will also consider estimation of the simple linear regression model

$$Y = \alpha + \beta \cdot X + \varepsilon$$

on the same data. In the following, the slope estimator *without* constant is denoted $\tilde{\beta}$, the estimator with constant is $\hat{\beta}$. We assume i.i.d. sampling and existence of moments etc. as needed.

2.1 Solve for $\tilde{\beta}$ in closed form.

2.2 Characterize and compare mean and variance of $\tilde{\beta}$ and $\hat{\beta}$.

2.3 Clarify that the Gauss-Markov theorem is not contradicted by doing two things:

- Characterize $\mathbb{E}(\tilde{\beta}|\boldsymbol{X})$ *without* assuming that $\alpha \neq 0$.
- State and prove an adapted Gauss-Markov Theorem for the case where we truly know that there is no constant.

3 This question is a reminder that most confidence regions invert hypothesis tests and that the concept of confidence interval is intellectually distinct from the CLT. The fact that most confidence intervals you'll ever see are "estimator ± 2 standard errors" can tend to obscure this.

We want to estimate the efficacy of a vaccine that has been randomly administered to (about) half of a trial sample. Outcomes (i.e., infection of not) were observed a few months later.

3.1 We begin by estimating $\pi \equiv \Pr\{vaccinated \mid infected\}$ from the corresponding sample frequency. Suppose that 178 infections occurred in sample, of these 9 in vaccinated participants. Compute

- the estimator $\hat{\pi}$,
- a 95% Wald (t-statistic based) confidence interval,
- a 95% confidence interval by inverting a Poisson approximation,
- a 95% confidence interval by inverting the binomial distribution.

Note that the last two intervals will be asymmetric. if $F_\pi(x)$ is a hypothesized (approximate) sampling distribution of the sample average as function of true parameter π, you will have to find π^* s.t. $F_{\pi^*}(\overline{x})$ takes on specific values, where $\overline{x}$ is the empirical sample average. Explain.

(You can check your answer to the last part by using an off-the-shelf implementation of Clopper-Pearson testing, e.g. MATLAB's binofit, but please work from scratch first.)

3.2 Prove: If $[\underline{\theta}, \overline{\theta}]$ is a $(1-\alpha)$-confidence interval for θ and $f(\cdot)$ is a known, strictly monotonic function, then $[f(\underline{\theta}), f(\overline{\theta})]$ or $[f(\overline{\theta}), f(\underline{\theta})]$ (depending on direction of monotonicity of $f(\cdot)$) is a $(1-\alpha)$-confidence interval for $f(\theta)$.

3.3 Vaccine efficacy is defined as relative risk reduction, that is,

$$VE \equiv 1 - \frac{\Pr(infected \mid vaccinated)}{\Pr(infected \mid unvaccinated)}.$$

Show: If $\rho \equiv \Pr(vaccinated)$ is known, vaccine efficacy is a strictly monotonic function of π.

3.4 Suppose now that the above numbers came from a trial with 18559 person-years of exposure in the vaccinated group and 18708 years in the control group. (Assignment was originally by flipping fair coins, but this adjusts for attrition and the like. Note that we ignore randomness in this number, which is negligible in this example.) Compute estimators and confidence intervals for VE that correspond to the three methods above.

3.5 The example is not hypothetical. Can you find the numbers we worked with and that you computed in Table 8, page 22, of the document provided with this homework?

Quick Answers

1.1 This is a standard OLS model except that $\text{var}(\varepsilon) = (1-\pi)\sigma^2$. We conclude that $\mathbb{E}(\hat{\beta}|\boldsymbol{X}) = \beta$ and $\text{var}(\hat{\beta}|\boldsymbol{X}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}(1-\pi)\sigma^2$.

1.2 Here, the core observation is that all those realizations (Y_i, X_i) where $\varepsilon_i = 0$ will lie on the same hyperplane. This means two things: (i) Because with normally distributed errors, this is a zero probability event, as soon as $k+1$ such realizations occurred, we can discover them. (ii) At that point, we can perfectly learn β. The implication is that $\Pr(\tilde{\beta} = \beta \to 1$ rapidly (the complementary probability decays exponentially).

In short, the estimator defined by:

- Computing $\hat{\beta}$ only from $k+1$ linearly dependent observations if those exist (and the ones occurring earliest in the data if the set is not unique),
- Computing OLS otherwise,

is unbiased in this d.g.p. and converges faster than any polynomial rate.

1.3 The estimator is not linear and also not in general unbiased 9see next question). regarding the nonlinearity, the estimator is linear in the data used but notice that selecting a subset of data to use is not linear.

1.4 The true issue is that the estimator is biased. To see this, recall that the quality of being unbiased means to be unbiased for all true d.g.p.'s that are consistent with maintained assumptions. (Even the no-data estimator is unbiased for some true d.g.p.!) Consider a linear homoskedastic model in which ε equals 1 with probability .5 and $N(-1,1)$ with remaining probability, then $\Pr(\tilde{\beta} = \beta + (1,0,\ldots,0)) \to 1$ and $\tilde{\beta}$ is therefore biased.

2.1 I here derive $\tilde{\beta}_1$ as least-squares estimator of β_1. Residuals due to an arbitrary estimator b_1 are

$$Y_i - b_1 X_i$$

and $\tilde{\beta}_1$ is defined as minimizer of the sum of squared residuals:

$$\tilde{\beta}_1 = \arg\min_{b_1} \sum_{i=1}^{n} (Y_i - b_1 X_i)^2.$$

Thus $\tilde{\beta}_1$ is characterized by FOC

$$\begin{aligned} -2\sum_{i=1}^{n} X_i(Y_i - \hat{\beta}_1 X_i) = -2\sum_{i=1}^{n}(X_i Y_i - \widehat{\beta}_1 X_i^2) = 0 \\ \Longrightarrow \sum_{i=1}^{n} X_i Y_i = \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 \\ \Longrightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}. \end{aligned}$$

Of course, a method of moments derivation starting from $\sum X_i(Y_i - \tilde{\beta}_1 X_i) = 0$ or $E(X(Y - \beta_1 X)) = 0$ is acceptable as well. Note also that the second order condition is easily verified in the example: the second derivative of the objective is $2\sum_{i=1}^{n} b_1 X_i^2 > 0$.

2.2

$$\begin{aligned} \tilde{\beta}_1 &= \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2} = \frac{\sum_{i=1}^{n} X_i(\beta_1 X_i + \varepsilon_i)}{\sum_{i=1}^{n} X_i^2} \\ &= \frac{\beta_1 \sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n} X_i \varepsilon_i}{\sum_{i=1}^{n} X_i^2} \\ &= \frac{\beta_1 \sum_{i=1}^{n} X_i^2}{\sum_{i=1}^{n} X_i^2} + \frac{\sum_{i=1}^{n} X_i \varepsilon_i}{\sum_{i=1}^{n} X_i^2} \\ &= \beta_1 + \frac{\sum_{i=1}^{n} X_i \varepsilon_i}{\sum_{i=1}^{n} X_i^2}, \end{aligned}$$

thus

$$\begin{aligned} \mathbb{E}\left(\tilde{\beta}_1 | X_1, \ldots, X_n\right) &= \mathbb{E}\left(\beta_1 + \frac{\sum_{i=1}^{n} X_i \varepsilon_i}{\sum_{i=1}^{n} X_i^2} | X_1, \ldots, X_n\right) \\ &= \beta_1 + \frac{1}{\sum_{i=1}^{n} X_i^2} \sum_{i=1}^{n} X_i E(\varepsilon_i | X_1, \ldots, X_n) \\ &= \beta_1, \end{aligned}$$

using $\mathbb{E}(\varepsilon_i | X_1, ..., X_n) = 0$.

$$\begin{aligned} \operatorname{var}(\tilde{\beta}_1 | X_1, \ldots, X_n) &= \operatorname{var}\left(\frac{\sum_{i=1}^{n} X_i \varepsilon_i}{\sum_{i=1}^{n} X_i^2} | X_1, \ldots, X_n\right) \\ &= \frac{\sum_{i=1}^{n} X_i^2 \operatorname{var}(\varepsilon_i | \cdot)}{(\sum_{i=1}^{n} X_i^2)^2} \\ &= \frac{\sigma^2 \sum_{i=1}^{n} X_i^2}{(\sum_{i=1}^{n} X_i^2)^2} \\ &= \frac{\sigma^2}{\sum_{i=1}^{n} X_i^2}. \end{aligned}$$

Direct comparison of algebraic expressions reveals that $var(\tilde{\beta}_1) \geq var(\widehat{\beta}_1)$. (This would also be expected from Gauss-Markov because $\hat{\beta}_1$ is BLUE in the model $y_i = \beta_1 x_i + u_i$. But without seeing a proof of that theorem, which I did not provide, you wouldn't be able to know that that is true, i.e. that the Gauss-Marlov theorem applies to $\hat{\beta}_1$ in the model without constant.)

This finding might at first be puzzling: Isn't $\tilde{\beta}_1$ shown to be BLUE by Gauss-Markov and must therefore have lower variance than $\hat{\beta}_1$? Well, $\tilde{\beta}_1$ is BLUE in the more general model and has the advantage over $\hat{\beta}_1$ of being unbiased whatever the true value of β_1. The proof of Gauss-Markov, which we omitted, uses the fact that an estimator to be compared to the OLS estimator must be unbiased *for all true values of* β_1. Indeed, a moment's reflection reveals that the Gauss-Markov theorem cannot hold with respect to all linear estimators that are unbiased for *some* true parameter values: The silly estimator that always "guesses" 0, no matter what data were observed, is linear in y_i, is unbiased in the special case that $\beta_1 = 0$, and has zero variance.)

2.3 Gauss-Markov is not contradicted because (i) if we actually know the intercept equals 0, then $\tilde{\beta}_1$ is the least squares estimator and a carbon copy of the GM theorem from class applies, (ii) if we do not know this, the $\tilde{\beta}_1$ is biased as casn be seen here:

$$\begin{aligned}
\tilde{\beta}_1 &= \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2} \\
&= \frac{\sum_{i=1}^n X_i(\beta_0 + \beta_1 X_i + \varepsilon_i)}{\sum_{i=1}^n X_i^2} \\
&= \beta_0 \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n X_i^2} + \beta_1 \frac{\sum_{i=1}^n X_i^2}{\sum_{i=1}^n X_i^2} + \frac{\sum_{i=1}^n X_i \varepsilon_i}{\sum_{i=1}^n X_i^2} \\
\implies \mathbb{E}\left(\tilde{\beta}_1 | \cdot\right) &= \beta_0 \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n X_i^2} + \beta_1 + \frac{\sum_{i=1}^n X_i E(\varepsilon_i | X_i)}{\sum_{i=1}^n X_i^2} = \beta_0 \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n X_i^2} + \beta_1 \neq \beta_1.
\end{aligned}$$

3.1 The estimator $\hat{\pi} = 9/178 \approx .05$.

The sample variance is given by the binomial variance, i.e. $(9/178 \cdot 169/178)/178) \approx .00027$, leading to standard error of .0164. We conclude that the Wald confidence interval is

$$[.05 - 1.96 \cdot .0164, .05 + 1.96 \cdot .0164] = [.0184, .0827].$$

Next, we can use the Poisson approximation for the number of vaccinated among the infected. Manual tinkering reveals that the number 9 is the .025-quantile of the Poisson distribution with parameter 17.1, whereas 8 is the .975-quantile if the parameter is 4.1. Thus the Poisson confdcience interval is

$$[4.1/178, 17.1/178] = [.023, .096].$$

We can similarly invert the binomial distribution to get exact ("Clopper-Pearson") confidence intervals. I just used the binofit command in MATLAB to get

$$[.023, .094].$$

But note that you can verify the relevant binomial quantiles with these parameter values.

3.2 For monotone increasing f:

$$\Pr\big(f(\underline{\theta}) \leq f(\theta) \leq f(\overline{\theta})\big) = \Pr\big(\underline{\theta} \leq \theta \leq \overline{\theta}\big) = 1 - \alpha.$$

3.3 Using $\wedge$ for logical AND and repeatedly using Bayes' Theorem,

$$\begin{aligned}
VE &\equiv 1 - \frac{\Pr(infected \mid vaccinated)}{\Pr(infected \mid unvaccinated)} \\
&= 1 - \frac{\Pr(infected \wedge vaccinated)/\Pr(vaccinated)}{\Pr(infected \wedge unvaccinated)/\Pr(unvaccinated} \\
&= 1 - \frac{\Pr(infected \wedge vaccinated)/\rho}{\Pr(infected \wedge unvaccinated)/(1-\rho)} \\
&= 1 - \frac{\Pr(vaccinated \mid infected)\Pr(infected)/\rho}{\Pr(unvaccinated \mid infected)\Pr(infected)(1-\rho)} \\
&= 1 - \frac{\pi/\rho}{(1-\pi)/(1-\rho)},
\end{aligned}$$

which is decreasing in π by inspection.

3.4 Assuming ρ takes the known value $18559/(18559 + 18708) \approx .498$, we can use our findings from 3.2 to pass all previous CI's through the function VE. In order, the resulting confidence intervals are:

$$\begin{aligned}
CI^{Wald} &= [.909, .981] \\
CI^{Poisson} &= [.893, .976] \\
CI^{bino} &= [.896, .976]
\end{aligned}$$

In sum, we see that the Normal approximation is far from perfect here/ The Poisson approximation is very good. While we ignored that ρ is estimated, quantitative negligibility of this effect is illustrated by the fact that we precisely recover the reported CI.