Section Note 08: Midterm Review

JJ Chen[*]

True or False

Some questions are for concepts, some for notations. Almost all of them are false. This list is by no means complete, but quite a few of them come from my grading experience. As usually, bugs might exist.

LS Method

1. For two $n \times 1$ vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, $\|\boldsymbol{u}+\boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2$.

- **Correct if $\boldsymbol{u} \perp \boldsymbol{v}$. In general, $\|\boldsymbol{u}+\boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 - 2 < \boldsymbol{u}, \boldsymbol{v} >$.**

2. $\hat{\beta}$ can be written as $\boldsymbol{X}'\boldsymbol{y}/\|\boldsymbol{X}\|^2$.

- **I think it would be no problem if we write X as x. That is, x is a vector.**

3. For two $n \times 1$ vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, $\mathrm{Var}(\boldsymbol{u}\boldsymbol{v}') = \boldsymbol{u}' \mathrm{Var}(\boldsymbol{v})\boldsymbol{u}$.

- **Should be $\mathrm{Var}(\boldsymbol{u}\boldsymbol{v}') = \boldsymbol{u}\, \mathrm{Var}(\boldsymbol{v})\boldsymbol{u}'$.**

4. For a $n \times 1$ vector $\boldsymbol{u}$, $\mathrm{E}(\boldsymbol{u})$ and $\mathrm{Var}(\boldsymbol{u})$ are $n \times 1$ vector.

[*] jchen215@uic.edu

- $\text{Var}(\boldsymbol{u})$ **is** $n \times n$**.**

5. For two random variables u and v, $\text{cov}(u, v) = \text{E}(uv)$.

- **Correct if** $\text{E}(u) = 0$ **or** $\text{E}(v)$ **= 0 or both. In general,** $\text{cov}(u, v) = \text{E}(uv) - \text{E}(u)\,\text{E}(v)$**. Try relate this to** $\text{E}(\boldsymbol{X}\boldsymbol{\varepsilon}) = 0$**.**

GM Model

1. Estimates are parameters.

- **Think about** $\boldsymbol{\beta}$**,** $\hat{\boldsymbol{\beta}}$**, and** $\hat{\boldsymbol{\beta}} = 12$**. Try use standard notations.**

2. Residuals are random errors.

- **For some authors the difference is blurry. But make sure you distingush** $\boldsymbol{y} - \text{E}(\boldsymbol{y}|\boldsymbol{X})$**,** $\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$**, and** $\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$**.**

3. Data are random variables in a regression model.

- **Data are realization of random variables.**

4. Run a regression in Stata using the command `reg y x`. The degrees of freedom of the residual sum of squares is $n - 1$.

- **Stata automatically add an intercept, so it would be** $n - 1 - 1$**.**

5. $\boldsymbol{\varepsilon} \perp \boldsymbol{x}$, and $\boldsymbol{e} \perp\!\!\!\perp \boldsymbol{x}$.

- **Better to write** $e \perp x$ **and** $\boldsymbol{\varepsilon} \perp\!\!\!\perp x$**. The notation** $\perp$ **is for orthogonality and** $\perp\!\!\!\perp$ **for stochastic independence. One idea is that they both mean "orthoganality", but they are defined differently in different vector spaces. Another way to think of it is that** $\perp$ **means that the angle between two deterministic vectors are 90 degrees, and** $\perp\!\!\!\perp$ **means that the conditional distribution of** ε **given** x **doest not depend on** $\boldsymbol{x}$**.**

6. $\Sigma_i e_i = 0$.

- **If you add an intercept.**

7. If you run a regression, the sum of the residuals is 0.

- **Intercept.**

7. $\mathrm{E}\left(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\right) = \sigma^2 I_{n\times n}$ for the original GM model.

- $\mathrm{E}\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right) = \sigma^2 I_{n\times n}$ **and** $\mathrm{E}\left(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\right) = n\sigma^2$**.**

8. $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$.

- $\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ **or** $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{\varepsilon}$**.**

9. $\sigma^2 = \Sigma_{i=1}^n \varepsilon_i^2/n$ and $\hat{\sigma^2} = s^2 = \Sigma_{i=1}^n e_i^2/n$.

- **The first one is probably ok if we know all ε_i. The second one should adjust for degrees of freedom if we want an unbiased estimate.**

10. $\hat{\sigma}$ (or s) is an unbiased estimate of σ.

- $\hat{\sigma}^2$ **(or s^2, and I mean $\boldsymbol{e}'\boldsymbol{e}/n-k$) is an unbiased estimate of σ^2.**

11. $\|\boldsymbol{e}\|^2 = \boldsymbol{e}'\boldsymbol{e} = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$.

- $\boldsymbol{e}'\boldsymbol{e} = \boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon}$**, where $\boldsymbol{M}$ is a residual maker.**

12. $\mathrm{E}(\boldsymbol{y}|\boldsymbol{X}) = \boldsymbol{X}\hat{\boldsymbol{\beta}}$.

- **We assume** $\mathrm{E}(\boldsymbol{y}|\boldsymbol{X}) = \boldsymbol{X}\boldsymbol{\beta}$**.**

13. $\mathrm{Var}(\boldsymbol{y}|\boldsymbol{X}) = \sigma^2$.

- $\operatorname{Var}(\boldsymbol{y}|\boldsymbol{X}) = \sigma^2 \boldsymbol{I}_n$ **or** $\operatorname{Var}(y_i|X_i) = \sigma^2$**.**

14. $\boldsymbol{y} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \varepsilon$.

- $\boldsymbol{y} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{e}$ **or** $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$**.**

15. $\operatorname{Var}(\boldsymbol{\beta}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$ and $\operatorname{Var}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$ in the original GM model.

- $\operatorname{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$ **and** $Est. \operatorname{Var}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$**.**

16. $\operatorname{E}(\hat{\boldsymbol{y}}|\boldsymbol{X}) = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ in the modified GM model.

- $\operatorname{E}(\hat{\boldsymbol{y}}|\boldsymbol{X}) = \operatorname{E}(\boldsymbol{P}\boldsymbol{y}|\boldsymbol{X}) = \boldsymbol{P}\operatorname{E}(\boldsymbol{y}|\boldsymbol{X}) = \boldsymbol{P}\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}\boldsymbol{\beta}$**.**

17. If there is no variation in $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k$, then their estimated coefficients are zero.

- **They won't have estimated coefficients because of the rank condition. If there is no variation in y, then estimated coefficients would be zero.**

18. In the GM model, residuals are independent from one observation to another.

- **Better to replace residuals with random errors. And it is an assumption.**

Hypothesis Testing

1. Run a regression of `wage` on `grade` and get a t-value of 2. The probability that $\beta_{\text{grade}} \neq 0$ is about 95%.

- **From a frequentlists' point of view, $\beta_{\textbf{grade}}$ is either zero or not zero.**

2. We are about 95% confidence that $\beta_{\text{grade}} \neq 0$ for the above t-value.

- **This is not the idea, see the last question.**

2. If the above model is right and $\beta_{\text{grade}} = 0$, there is about a 5% of chance getting $|t| < 2$.

- **Should be $|t| > 2$. But this is the idea for hypothesis testing.**

3. A statistical significant overall F-value suggests the model is correct.

- **There are several possibilities: an unlikely event occurred; or the model is right and the coefficients differ from 0; or the model is wrong.**

4. Assume the error term is normally distributed. We will not be able to use the t-test if we only have 4 degrees of freedom.

- **The normal assumption allows us to have exact distribution for the test statistic. It's okay to use it even if we only have 4 degrees of freedom.**

Specification Issues

1. A specification says what variables go into a model, what the functional form is, and what should be assumed about the disturbance term. If the data are generated some other way, we misspecify the model. We can correct the specification error.

- **Teach me if you know how to correctly specify the model (aka, the underlying data generating process).**

2. In practice, we can add control variables to reduce omitted variable bias.

- **Not necessarily. Bad controls (controls that are outcome variables of the explanatory variable of interest) would probably add bias. Control variables that have impacts on y_i but do not correlated with x_i would not reduce bias also. (But they are still "useful" as they could help to increase precisions of other estimates.) Finally, the logic of control variables is problematic in reality, see for example Clarke (2005) (consume carefully though).**

3. An estimated coefficient on the quadratic term measures the effect of the quadratic term on the dependent variable, control for other regressors.

- **It would be impossible to hold the linear term fixed.**

4. Consider a log-linear specification: $\log(y_i) = \alpha + \beta x_i + \epsilon_i$. The interpretation of β is that a one unit change in x is associated with $100 \times \beta\%$ change in y.

- **Better to use $ln(y_i)$.**

5. We can't run a model $\ln(y_i) = \alpha + \beta x_i + \epsilon_i$, where x_i is a categorical variable.

- **We can, but we are imposing restriction. It would be better to define some dummies based on x_i and run a model with dummies.**

6. I don't have good review topics for specification issues. On the one hand, things like interpretation of dummy variable and interaction terms, OVB, and functional forms and so on are pretty basic. On the other hand, I feel frustrated because I don't know how to correctly specify a regression model. In textbooks and problem sets we talk about "the true model", but in reality we do not have the luxury. Below are some responses to the fundamental specification issue. Perhaps all of them are bad responses. I don't know.

   1. "Nothing is perfect."
   2. "I know."
   3. "Linearity is a good approximation."
   4. "The assumptions are reasonable."
   5. "The assumptions don't matter."

6. "The assumptions are conservative."
7. "You can't prove the assumptions are wrong."
8. "The biases will cancel."
9. "We can model the biases."
10. "We're doing what everybody else does."
11. "Now we use more sophisticated techniques. (We have GLS, IVLS, SEM, MM, GMM, MLE, ANOVA, Logits, Probits, Tobits, AIC, BIC, AR, ARIMA, ARCH, LASSO, Neural nets, DAGs, POM, IV, ATE, ATET, ATEN, LATE, Split sample IV, DID, DDD, Sharp RD, Fussy RD, RKD, ...)"
12. "If we don't do it, someone else will."
13. "What would you do?"
14. "Policy makers are better off with us than without us."
15. "We're estimating a lower bound."
16. "Not using a model is still a model."
17. "The model is still useful."
18. "You have to do the best you can with the data."
19. "You have to make assumptions in order to make progress."
20. "Where's the harm?"

Appendix

Clarke, Kevin A. 2005. "The phantom menace: Omitted variable bias in econometric research." *Conflict Management and Peace Science* 22 (4): 341–352.