Econ 672 – Midterm Exam

Matias D. Cattaneo
Department of Economics
University of Michigan

February 23, 2018

Contents

Notes:

- Duration of examination: *3 hours.*
- Please answer *all three* questions. All parts are equally weighted (5 points).
- Please start each question on a separate page.
- Please provide as much detail as possible in your answers.
- Answers without proper justification will not receive (partial) credit.

1 Question 1: Scheffé Simultaneous Confidence Intervals (15 points)

Suppose the classical Normal OLS assumptions hold: $\mathbf{y}|\mathbf{X} \sim \mathsf{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$, where $\mathbf{y} = (y_1, \ldots, y_n)' \in \mathbb{R}^n$, $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]' \in \mathbb{R}^{n\times d}$ with $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{id})' \in \mathbb{R}^d$, and $\boldsymbol{\beta} \in \mathbb{R}^d$ and $\sigma^2 \in \mathbb{R}_{++}$ denote the parameters of the model, $d \in \mathbb{Z}_{++}$. Let $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_d)' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ be the ordinary least squares (OLS) estimator of $\boldsymbol{\beta}$, and $s^2 = \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2/(n-d)$ an unbiased estimator of σ^2.

Consider the null hypothesis $\mathsf{H}_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, with $\boldsymbol{\beta}_0 = (\beta_{01}, \beta_{02}, \cdots, \beta_{0d})'$. Scheffé's celebrated result for simultaneous confidence intervals for the coefficients $\beta_{0\ell}$, $\ell = 1, 2, \ldots, d$, is:

$$\mathbb{P}\left[\, |\hat{\beta}_\ell - \beta_{0\ell}| \leq \sqrt{s^2 \cdot (\mathbf{X}'\mathbf{X})^{-1}_{\ell\ell} \cdot d \cdot \mathcal{F}^{-1}_{d,n-d}(1-\alpha)} \ :\ \ell = 1, 2, \ldots, d \,\right] \geq 1 - \alpha,$$

where $(\mathbf{X}'\mathbf{X})^{-1}_{\ell\ell}$ denotes the ℓ-th diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$, $\ell = 1, 2, \ldots, d$, $\mathcal{F}^{-1}_{d,n-d}(1-\alpha)$ denotes the $(1-\alpha)$ quantile of the F distribution with d and $n-d$ degrees of freedom, denoted by $\mathcal{F}_{d,n-d}$, and $\alpha \in (0, 1)$.

1. Show that under H_0,
$$\frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)}{d \cdot s^2} \sim \mathcal{F}_{d,n-d}.$$

   - **Solution**: Let $\mathbf{U} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0$. Since
$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}, \qquad s^2 = \mathbf{y}'\mathbf{M}_\mathbf{X}\mathbf{y}/(n-d) = \mathbf{U}'\mathbf{M}_\mathbf{X}\mathbf{U}/(n-d)$$
where $\mathbf{M}_\mathbf{X} = \mathbf{I} - \mathbf{P}_\mathbf{X}$ and $\mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, we can rewrite the statistic as
$$\frac{\mathbf{U}'\mathbf{P}_\mathbf{X}\mathbf{U}/d}{\mathbf{U}'\mathbf{M}_\mathbf{X}\mathbf{U}/(n-d)} \sim \mathcal{F}_{d,n-d}.$$
Here we use the fact that $\mathbf{U}|\mathbf{X} \sim \mathsf{N}(0, \sigma^2\mathbf{I}_n)$, $\mathbf{P}_\mathbf{X}\mathbf{M}_\mathbf{X} = 0$, $rank(\mathbf{P}_\mathbf{X}) = d$ and $rank(\mathbf{M}_\mathbf{X}) = n-d$. Hence $\mathbf{U}'\mathbf{P}_\mathbf{X}\mathbf{U} \sim \chi^2_d$, $\mathbf{U}'\mathbf{M}_\mathbf{X}\mathbf{U} \sim \chi^2_{n-d}$, and they are independent of each other.

2. Show that
$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \max_{\mathbf{a}\in\mathbb{R}^d} \frac{[\mathbf{a}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)]^2}{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}$$

   Hint: you can use Cauchy-Schwarz Inequality, $(\mathbf{u}'\mathbf{v})^2 \leq (\mathbf{u}'\mathbf{u})(\mathbf{v}'\mathbf{v})$ for conformable vectors $\mathbf{u}$ and $\mathbf{v}$, to establish an (achievable) maximum.

- **Solution**: By Cauchy-Schwarz inequality, for any $\mathbf{a} \in \mathbb{R}^d$,

$$\begin{aligned}[\mathbf{a}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)]^2 &= [\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1/2}(\mathbf{X}'\mathbf{X})^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)]^2 \\ &\leq \left(\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}\right)\left((\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\right)\end{aligned}$$

which implies that for any $\mathbf{a} \in \mathbb{R}^d$

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \geq \frac{[\mathbf{a}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)]^2}{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}.$$

Thus we establish an upper bound. To show this upper bound is achievable, simply let $\mathbf{a} = (\mathbf{X}'\mathbf{X})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$.

3. Show that

$$\mathbb{P}\left[|\mathbf{a}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)| \leq \sqrt{s^2 \cdot \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a} \cdot d \cdot \mathcal{F}^{-1}_{d,n-d}(1-\alpha)} \;:\; \forall \mathbf{a} \in \mathbb{R}^d\right] = 1 - \alpha,$$

and use this result to establish the lower bound on the coverage probability of the simultaneous confidence intervals given by Scheffé.

- **Solution**: From part 1, we already have

$$\mathbb{P}\left[\frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)}{d \cdot s^2} \leq \mathcal{F}^{-1}_{d,n-d}(1-\alpha)\right] = 1 - \alpha$$

Then it follows from the conclusion in part 2 that the event inside the probability is equivalent to

$$|\mathbf{a}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)| \leq \sqrt{s^2 \cdot d \cdot \mathcal{F}^{-1}_{d,n-d}(1-\alpha) \cdot \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}, \qquad \forall \mathbf{a} \in \mathbb{R}^d,$$

which implies that

$$|\hat{\beta}_\ell - \beta_{0\ell}| \leq \sqrt{s^2 \cdot d \cdot \mathcal{F}^{-1}_{d,n-d}(1-\alpha) \cdot (\mathbf{X}'\mathbf{X})^{-1}_{\ell\ell}}, \qquad \text{for } \ell = 1, 2, \ldots, d.$$

(simply let $\mathbf{a}$ be a vector with ℓth element equal to 1 and others equal to 0). By monotonicity of probability measure, the probability of this implied event is no less than $1 - \alpha$.

2 Question 2: 2SLS with Possibly Many Instruments (30 points)

Consider a textbook example of IV model with non-random instruments and all the usual regularity conditions. Let $\{(u_i, v_i) : 1 \le i \le n\}$ be random sample of mean-zero and finite fourth moments random variables, and let the data generating process be:

$$\mathbf{y} = \mathbf{x}\beta + \mathbf{u}, \qquad \mathbb{E}[u_i] = 0, \qquad \mathbb{V}[u_i] = \sigma_u^2$$
$$\mathbf{x} = \mathbf{Z}\boldsymbol{\pi} + \mathbf{v}, \qquad \mathbb{E}[v_i] = 0, \qquad \mathbb{V}[v_i] = \sigma_v^2, \qquad \mathbb{E}[u_i v_i] = \sigma_{uv},$$

where $\mathbf{y} = (y_1, \ldots, y_n)' \in \mathbb{R}^n$, $\mathbf{x} = (x_1, \ldots, x_n)' \in \mathbb{R}^n$, $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n]' \in \mathbb{R}^{n \times K}$ with $\mathbf{z}_i = (z_{i1}, z_{i2}, \ldots, z_{iK})' \in \mathbb{R}^K$, $\{\mathbf{z}_i : i \ge 1\}$ a non-random sequence, are observed, while $\mathbf{u} = (u_1, \ldots, u_n)' \in \mathbb{R}^n$ and $\mathbf{v} = (v_1, \ldots, v_n)' \in \mathbb{R}^n$ are not observed. The (structural) parameter of interest is $\beta \in \mathbb{R}$, while $\boldsymbol{\pi} \in \mathbb{R}^K$ denotes the first-stage coefficients. Note that $\mathbb{E}[\mathbf{u}|\mathbf{Z}] = \mathbb{E}[\mathbf{u}] = \mathbf{0}$, because $\mathbf{Z}$ is assumed non-random. The only twist relative to the classical IV model is that we assume $K = K_n \to \infty$ as $n \to \infty$. Specifically, we consider "many" instruments asymptotics, where

$$\frac{K}{n} \to \rho \in [0, 1) \qquad \text{and} \qquad \frac{\boldsymbol{\pi}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\pi}}{n} \to \mu > 0,$$

where the second condition captures the idea of "strong" instruments (can you see why?).

The 2SLS estimator of β is $\hat{\beta}_{\mathsf{2SLS}} = (\mathbf{x}'\mathbf{P}\mathbf{x})^{-1}\mathbf{x}'\mathbf{P}\mathbf{y}$, where $\mathbf{P} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' = [p_{ij} : 1 \le i, j \le n]$ is the (orthogonal) projection matrix with typical element (i, j)-th element p_{ij}. Recall the following properties of $\mathbf{P}$: $p_{ij} = p_{ji}$ (symmetry), $p_{ij} = \sum_{\ell=1}^n p_{i\ell}p_{\ell j}$ (idempotency), $\sum_{i=1}^n p_{ii} = K$ (full column rank), $p_{ii} \in (0, 1)$ for all i, and $p_{ij} \in [-1/2, 1/2]$ for all $i \ne j$.

1. Show that $\mathbb{E}[\frac{\mathbf{u}'\mathbf{u}}{n}] = \sigma_u^2$, $\mathbb{E}[\frac{\mathbf{v}'\mathbf{v}}{n}] = \sigma_v^2$, $\mathbb{E}[\frac{\mathbf{x}'\mathbf{u}}{n}] = \sigma_{uv}$, $\mathbb{E}[\mathbf{x}'\mathbf{P}\mathbf{u}] = K\sigma_{uv}$, and $\mathbb{E}[\mathbf{u}'\mathbf{P}\mathbf{u}] = K\sigma_u^2$.

   - **Solution**: The first two results follow immediately, while the third one follows from $\mathbb{E}[\mathbf{x}'\mathbf{u}/n] = \mathbb{E}[\mathbf{v}'\mathbf{u}/n] = \sigma_{uv}$ (since $\mathbf{Z}$ is non-random). For the fourth result, note that $\mathbb{E}[\mathbf{x}'\mathbf{P}\mathbf{u}] = \mathbb{E}[\mathbf{v}'\mathbf{P}\mathbf{u}] = K\sigma_{uv}$ because $\mathbb{E}[v_i u_j] = 0$ for all $i \ne j$ and $\sum_{i=1}^n P_{ii} = K$. Finally, the last result follows analogously.

2. Show that $\frac{\mathbf{x}'\mathbf{x}}{n} \to_{\mathbb{P}} \mu + \sigma_v^2$, $\frac{\mathbf{x}'\mathbf{P}\mathbf{x}}{n} \to_{\mathbb{P}} \mu + \rho\sigma_v^2$, and $\frac{\mathbf{x}'\mathbf{P}\mathbf{u}}{n} \to_{\mathbb{P}} \rho\sigma_{uv}$.

- **Solution**: Using the first-stage equation, we have

$$\mathbf{x}'\mathbf{x}/n = \mathbf{v}'\mathbf{v}/n + 2\mathbf{v}'\mathbf{Z}\boldsymbol{\pi}/n + \boldsymbol{\pi}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\pi}/n \to_{\mathbb{P}} \mu + \sigma_v^2,$$

using the LLN, and because $\mathbb{E}[\mathbf{v}'\mathbf{Z}\boldsymbol{\pi}/n] = 0$ and $\mathbb{V}[\mathbf{v}'\mathbf{Z}\boldsymbol{\pi}/n] = O(n^{-1})$, which gives the first result (use Markov inequality to show the second term is $o_{\mathbb{P}}(1)$).

For the second result, note that

$$\mathbf{x}'\mathbf{P}\mathbf{x}/n = \mathbf{v}'\mathbf{P}\mathbf{v}/n + 2\mathbf{v}'\mathbf{Z}\boldsymbol{\pi}/n + \boldsymbol{\pi}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\pi}/n \to_{\mathbb{P}} \mu + \rho\sigma_v^2,$$

which follows from $\mathbb{E}[\mathbf{v}'\mathbf{P}\mathbf{v}/n] = \sum_{i=1}^n p_{ii}\sigma_v^2/n = K\sigma_v^2/n \to \rho\sigma_v^2$, because $\mathbb{E}[v_i v_j] = 0$ for all $i \neq j$, and because $\mathbb{V}[\mathbf{v}'\mathbf{P}\mathbf{v}/n] \to 0$ after some calculations and using basic properties of projection matrices. Specifically,

$$\begin{aligned}
\mathbb{V}[\mathbf{v}'\mathbf{P}\mathbf{v}/n] &= \frac{1}{n^2}\mathbb{V}\left[\sum_{i=1}^n p_{ii}v_i^2 + 2\sum_{i<j} p_{ij}v_iv_j\right] \\
&= \frac{1}{n^2}\left(\mathbb{V}\left[\sum_{i=1}^n p_{ii}v_i^2\right] + \mathbb{V}\left[2\sum_{i<j} p_{ij}v_iv_j\right]\right) \\
&= \frac{1}{n^2}\left(\mathbb{V}[v_i^4]\sum_{i=1}^n p_{ii}^2 + 4\sum_{i<j} p_{ij}^2\mathbb{V}\left[v_iv_j\right]\right) \\
&= \frac{1}{n^2}\left(\mathbb{V}[v_i^4]\sum_{i=1}^n p_{ii}^2 + 4\sigma_v^4\sum_{i<j} p_{ij}^2\right) \\
&\leq \frac{C}{n^2} trace(\mathbf{P}'\mathbf{P}) \\
&\leq \frac{CK}{n^2} \to 0
\end{aligned}$$

where $C > 0$ is some universal constant.

For the third result,

$$\mathbf{x}'\mathbf{P}\mathbf{u}/n = \boldsymbol{\pi}'\mathbf{Z}'\mathbf{u}/n + \mathbf{v}'\mathbf{P}\mathbf{u}/n \to_{\mathbb{P}} 0 + \rho\sigma_{uv}$$

which follows from the similar argument for the second result and the conclusion proved in part 1.

3. Show that

$$\hat{\beta}_{\mathtt{2SLS}} \to_{\mathbb{P}} \beta + \frac{\rho\sigma_{uv}}{\mu + \rho\sigma_v^2}.$$

Under which conditions is $\hat{\beta}_{\mathtt{2SLS}}$ consistent (for β)? Provide intuition for each case. Give conditions and intuition under which $\hat{\beta}_{\mathtt{2SLS}}$ is "biased" (i.e., inconsistent) upwards or downwards.

- **Solution**: It follows directly by previous results because, as usual,

$$\hat{\beta}_{\mathtt{2SLS}} = \beta + (\mathbf{x}'\mathbf{P}\mathbf{x}/n)^{-1}\mathbf{x}'\mathbf{P}\mathbf{u}/n,$$

and the result follows by the continuous mapping theorem. Clearly, $\hat{\beta}_{\mathtt{2SLS}} \to_{\mathbb{P}} \beta$ if either $\rho = 0$ ("few" instruments) or $\sigma_{uv} = 0$ (exogeneity), but otherwise $\hat{\beta}_{\mathtt{2SLS}}$ is not a consistent estimator of β. Since $\mu + \rho\sigma_v^2 > 0$, the direction of bias depends on the sign of the numerator. When $\rho \neq 0$ and $\sigma_{uv} > 0$ ($\mathbf{u}$ and $\mathbf{x}$ are positively correlated), the two-stage estimator is upwards biased. When $\rho \neq 0$ and $\sigma_{uv} < 0$ ($\mathbf{u}$ and $\mathbf{x}$ are negatively correlated), then the two-stage estimator is downwards biased.

4. Consider the following *infeasible* bias corrected 2SLS estimator:

$$\tilde{\beta}_{\mathtt{2SLS-BC}} = (\mathbf{x}'\mathbf{P}\mathbf{x})^{-1}(\mathbf{x}'\mathbf{P}\mathbf{y} - \mathbf{v}'\mathbf{P}\mathbf{u}).$$

Show that $\tilde{\beta}_{\mathtt{2SLS-BC}} \to_{\mathbb{P}} \beta$. Why is it not possible to replace $\mathbf{v}$ and $\mathbf{u}$ by their estimated counterparts?

- **Solution**:

$$\tilde{\beta}_{\mathtt{2SLS-BC}} = \beta + (\mathbf{x}'\mathbf{P}\mathbf{x})^{-1}(\mathbf{x}'\mathbf{P}\mathbf{u} - \mathbf{v}'\mathbf{P}\mathbf{u}) = \beta + (\mathbf{x}'\mathbf{P}\mathbf{x})^{-1}\boldsymbol{\pi}'\mathbf{Z}'\mathbf{u}.$$

By the argument in the solution to part 2, the second term is $o_{\mathbb{P}}(1)$. Clearly, by subtracting $\mathbf{v}'\mathbf{P}\mathbf{u}$ from $\mathbf{x}'\mathbf{P}\mathbf{y}$, the non-negligible bias (the correlation between $\mathbf{u}$ and $\mathbf{v}$) is removed.

It is impossible to replace $\mathbf{v}$ and $\mathbf{u}$ by their estimated counterparts since the residuals from the first stage, $\hat{\mathbf{v}}$, is exactly orthogonal to $\mathbf{Z}$, i.e., $\hat{\mathbf{v}}'\mathbf{Z} = 0$.

5. It can be shown that $\sqrt{n}(\tilde{\beta}_{\texttt{2SLS-BC}} - \beta) \to_d \mathsf{N}(0, V_1(\rho))$, where

$$V_1(\rho) = \lim_{n\to\infty} \frac{n\mathbb{V}[\mathbf{x}'\mathbf{P}\mathbf{u} - \mathbf{v}'\mathbf{P}\mathbf{u}]}{(\mathbb{E}[\mathbf{x}'\mathbf{P}\mathbf{x}])^2}.$$

Characterize precisely the limiting variance $V_1(\rho)$. Under which conditions it coincides with the classical asymptotic variance of 2SLS estimators studied in class?

- **Solution**: From part (a), we have

$$\sqrt{n}(\tilde{\beta}_{\texttt{2SLS-BC}} - \beta) = \left(\frac{\mathbf{x}'\mathbf{P}\mathbf{x}}{n}\right)^{-1} \cdot \frac{1}{\sqrt{n}}\boldsymbol{\pi}'\mathbf{Z}'\mathbf{u}.$$

By part 2, $\mathbf{x}'\mathbf{P}\mathbf{x}/n \to_{\mathbb{P}} \mu + \rho\sigma_v^2$. Note that $\mathbb{E}[\boldsymbol{\pi}'\mathbf{z}_i\mathbf{u}] = 0$ and $\mathbb{V}[\frac{1}{\sqrt{n}}\boldsymbol{\pi}'\mathbf{Z}'\mathbf{u}] = \frac{\boldsymbol{\pi}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\pi}}{n}\sigma_u^2 \to \mu\sigma_u^2$. Then

$$\sqrt{n}(\tilde{\beta}_{\texttt{2SLS-BC}} - \beta) \to_d \mathsf{N}\left(0, \frac{\mu\sigma_u^2}{(\mu + \rho\sigma_v^2)^2}\right).$$

Clearly, when $\rho = 0$, it coincides with the classical asymptotic variance of 2SLS.

6. Consider now the following *feasible* bias corrected 2SLS estimator:

$$\hat{\beta}_{\texttt{2SLS-BC}} = (\mathbf{x}'\hat{\mathbf{P}}\mathbf{x})^{-1}\mathbf{x}'\hat{\mathbf{P}}\mathbf{y}, \qquad \hat{\mathbf{P}} = \mathbf{P} - \frac{K}{n}\mathbf{I}_n,$$

which can be interpreted as a kind of jackknife IV estimator (JIVE).

Show that $\hat{\beta}_{\texttt{2SLS-BC}} \to_{\mathbb{P}} \beta$.

Note: Under additional regularity conditions, it can also be shown that

$$\sqrt{n}(\tilde{\beta}_{\texttt{2SLS-BC}} - \beta) \to_d \mathsf{N}(0, V_2(\rho)), \qquad V_2(\rho) = \lim_{n\to\infty} \frac{n\mathbb{V}[\mathbf{x}'\hat{\mathbf{P}}\mathbf{u}]}{(\mathbb{E}[\mathbf{x}'\hat{\mathbf{P}}\mathbf{x}])^2},$$

with $V_2(\rho) \neq V_1(\rho)$ and more complicated in general, but establishing these results is beyond the scope of this exam.

- **Solution**: Using the previous results we have

$$\begin{aligned}
\hat{\beta}_{\texttt{2SLS-BC}} &= \beta + \left(\mathbf{x}'\left(\mathbf{P} - \frac{K}{n}\mathbf{I}_n\right)\mathbf{x}/n\right)^{-1}\mathbf{x}'\left(\mathbf{P} - \frac{K}{n}\mathbf{I}_n\right)\mathbf{u}/n \\
&= \beta + \left(\mu + \rho\sigma_v^2 - \rho[\mu + \sigma_v^2]\right)^{-1}(\rho\sigma_{uv} - \rho\sigma_{uv}) + o_{\mathbb{P}}(1),
\end{aligned}$$

  which gives the consistency result.

3 Question 3: ARMA(1,1) Estimation and Inference (35 points)

Let $\{y_t : 1, 2, \ldots, T\}$ be a sample from the stationary ARMA(1,1) process:

$$y_t = \alpha y_{t-1} + \varepsilon_t + \beta \varepsilon_{t-1},$$

where $\{\varepsilon_t : t \geq 1\}$ are i.i.d. $\mathsf{N}(0, \sigma^2)$, $|\alpha| < 1$, $|\beta| < 1$, $\alpha \neq -\beta$, $\sigma^2 \in \mathbb{R}_{++}$, and appropriate initial conditions. Consider the family of (instrumental variables) least squares estimators of α:

$$\hat{\alpha}_p = \frac{\sum_{t=p+1}^T y_{t-p} y_t}{\sum_{t=p+1}^T y_{t-p} y_{t-1}}, \qquad p \geq 1.$$

1. Find $\mathbb{E}[y_t]$ and show that

$$\gamma_0 = \mathbb{E}[y_t^2] = \frac{(1 + 2\alpha\beta + \beta^2)\sigma^2}{1 - \alpha^2}, \qquad \gamma_s = \mathbb{E}[y_t y_{t-s}] = \alpha^{s-1} \cdot (\alpha\gamma_0 + \beta\sigma^2), \qquad s \geq 1.$$

- **Solution**: First, note that $\mathbb{E}[y_t] = \alpha\mathbb{E}[y_{t-1}] + \mathbb{E}[\varepsilon_t] + \beta\mathbb{E}[\varepsilon_{t-1}] = \alpha\mathbb{E}[y_{t-1}]$, and therefore

$$\mathbb{E}[y_t] = 0.$$

Next, $\mathbb{E}[y_t^2] = \alpha^2\mathbb{E}[y_{t-1}^2] + \mathbb{E}[\varepsilon_t^2] + \beta^2\mathbb{E}[\varepsilon_{t-1}^2] + 2\alpha\mathbb{E}[y_{t-1}\varepsilon_t] + 2\alpha\beta\mathbb{E}[y_{t-1}\varepsilon_{t-1}] + 2\beta\mathbb{E}[\varepsilon_t\varepsilon_{t-1}] = \alpha^2\gamma_0 + \sigma^2 + \beta^2\sigma^2 + 2\alpha\beta\sigma^2$, and the result for γ_0 follows. For $s = 1$, we have

$$\gamma_1 = \mathbb{E}[y_t y_{t-1}] = \alpha\mathbb{E}[y_{t-1}y_{t-1}] + \mathbb{E}[\varepsilon_t y_{t-1}] + \beta\mathbb{E}[\varepsilon_{t-1}y_{t-1}] = \alpha\gamma_0 + \beta\sigma^2$$

and the result for γ_1 follows. Finally, the general case $s \geq 2$, follows because

$$\gamma_s = \mathbb{E}[y_t y_{t-s}] = \alpha\mathbb{E}[y_{t-1}y_{t-s}] + \mathbb{E}[\varepsilon_t y_{t-s}] + \beta\mathbb{E}[\varepsilon_{t-1}y_{t-s}] = \alpha\gamma_{s-1}.$$

2. Show that, if $p = 1$, then $\hat{\alpha}_1 \to_{\mathbb{P}} \alpha + \frac{\beta\sigma^2}{\gamma_0}$.

   Give conditions under which the estimator $\hat{\alpha}_1$ is consistent for α, and provide intuition.

- **Solution**: For $p = 1$,

$$\begin{aligned}\hat{\alpha}_1 &= \frac{\sum_{t=2}^T y_{t-1} y_t}{\sum_{t=2}^T y_{t-1}^2} = \frac{\sum_{t=2}^T y_{t-1}(\alpha y_{t-1} + \varepsilon_t + \beta\varepsilon_{t-1})}{\sum_{t=2}^T y_{t-1}^2} \\ &= \alpha + \frac{\sum_{t=2}^T y_{t-1}\varepsilon_t}{\sum_{t=2}^T y_{t-1}^2} + \frac{\beta\sum_{t=2}^T y_{t-1}\varepsilon_{t-1}}{\sum_{t=2}^T y_{t-1}^2}.\end{aligned}$$

  By Law of Large Numbers, $\frac{1}{T-1}\sum_{t=2}^T y_{t-1}^2 \to_{\mathbb{P}} \gamma_0$, $\frac{1}{T-1}\sum_{t=2}^T y_{t-1}\varepsilon_t \to_{\mathbb{P}} \mathbb{E}[y_{t-1}\varepsilon_t] = 0$, and $\frac{1}{T-1}\sum_{t=2}^T y_{t-1}\varepsilon_{t-1} \to_{\mathbb{P}} \mathbb{E}[y_{t-1}\varepsilon_{t-1}] = \sigma^2$. The desired result follows. Clearly, if $\beta = 0$, ARMA(1,1) reduces to AR(1), and contemporaneous exogeneity condition is satisfied. So $\hat{\alpha}_1$ will be consistent.

3. Show that, if $p \geq 2$, then $\hat{\alpha}_p \to_{\mathbb{P}} \alpha$.

- **Solution**: For $p \geq 2$,

$$\begin{aligned}\hat{\alpha}_p &= \frac{\sum_{t=p+1}^T y_{t-p} y_t}{\sum_{t=p+1}^T y_{t-p} y_{t-1}} = \frac{\sum_{t=p+1}^T y_{t-p}(\alpha y_{t-1} + \varepsilon_t + \beta\varepsilon_{t-1})}{\sum_{t=p+1}^T y_{t-p} y_{t-1}} \\ &= \alpha + \frac{\sum_{t=p+1}^T y_{t-p}\varepsilon_t}{\sum_{t=p+1}^T y_{t-p} y_{t-1}} + \frac{\beta\sum_{t=p+1}^T y_{t-p}\varepsilon_{t-1}}{\sum_{t=p+1}^T y_{t-p} y_{t-1}}.\end{aligned}$$

  By Law of Large Numbers,

$$\begin{aligned}&\frac{1}{T-p}\sum_{t=p+1}^T y_{t-p} y_{t-1} \to_{\mathbb{P}} \mathbb{E}[y_{t-p} y_{t-1}] = \gamma_{p-1}, \\ &\frac{1}{T-p}\sum_{t=p+1}^T y_{t-p}\varepsilon_t \to_{\mathbb{P}} \mathbb{E}[y_{t-p}\varepsilon_t] = 0, \\ &\frac{1}{T-p}\sum_{t=p+1}^T y_{t-p}\varepsilon_{t-1} \to_{\mathbb{P}} \mathbb{E}[y_{t-p}\varepsilon_{t-1}] = 0.\end{aligned}$$

  Thus $\hat{\alpha}_p \to_{\mathbb{P}} \alpha$.

4. Assuming $\sum_{t=1}^T y_{t-p}\varepsilon_t/\sqrt{T}$ convergences in law to a Gaussian distribution, show that

$$\sqrt{T}(\hat{\alpha}_p - \alpha) \to_d \mathsf{N}(0, V(p)), \qquad p \geq 2,$$

   and give the formula of the asymptotic variance $V(p)$.

- **Solution**: From part 3, we already know

$$\hat{\alpha}_p - \alpha = \frac{\sum_{t=p+1}^T y_{t-p}\varepsilon_t}{\sum_{t=p+1}^T y_{t-p}y_{t-1}} + \frac{\beta\sum_{t=p+1}^T y_{t-p}\varepsilon_{t-1}}{\sum_{t=p+1}^T y_{t-p}y_{t-1}}$$

Then it follows that

$$\begin{aligned}
&\sqrt{T}(\hat{\alpha}_p - \alpha) \\
&= \left(\frac{1}{T}\sum_{t=p+1}^T y_{t-p}y_{t-1}\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=p+1}^T y_{t-p}(\varepsilon_t + \beta\varepsilon_{t-1})\right) \\
&= \left(\frac{1}{T}\sum_{t=p+1}^T y_{t-p}y_{t-1}\right)^{-1}\left(\frac{1}{\sqrt{T}}\left(\sum_{t=p+1}^T (y_{t-p} + \beta y_{t-p+1})\varepsilon_t - \beta y_{T-p+1}\varepsilon_T + \beta y_1\varepsilon_p\right)\right) \\
&= \left(\frac{1}{T}\sum_{t=p+1}^T y_{t-p}y_{t-1}\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=p+1}^T (y_{t-p} + \beta y_{t-p+1})\varepsilon_t\right) + o_{\mathbb{P}}(1)
\end{aligned}$$

By Law of Large Numbers and Continuous Mapping Theorem, $(\frac{1}{T}\sum_{t=p+1}^T y_{t-p}y_{t-1})^{-1} \to_{\mathbb{P}} \gamma_{p-1}^{-1}$. Define $v_t = (y_{t-p} + \beta y_{t-p+1})\varepsilon_t$. Note that $\mathbb{E}[v_t] = 0$ and $\mathbb{E}[v_t v_{t'}] = 0$ for $t \neq t'$. Then it directly follows from the assumption in the question that

$$\frac{1}{\sqrt{T}}\sum_{t=p+1}^T v_t \to_d \mathsf{N}(0, \mathbb{E}[v_t^2])$$

where

$$\begin{aligned}
\mathbb{E}[v_t^2] &= \mathbb{E}[y_{t-p}^2\varepsilon_t^2 + \beta^2 y_{t-p+1}^2\varepsilon_t^2 + 2\beta y_{t-p}y_{t-p+1}\varepsilon_t^2] \\
&= (1+\beta^2)\sigma^2\gamma_0 + 2\beta\sigma^2\gamma_1.
\end{aligned}$$

Then it follows that

$$\sqrt{T}(\hat{\alpha}_p - \alpha) \to_d \mathsf{N}\left(0, \gamma_{p-1}^{-2}((1+\beta^2)\gamma_0 + 2\beta\gamma_1)\sigma^2\right).$$

5. Show that $V(p)$ is monotonically increasing in p, and use this result to rank the estimators in the class $\{\hat{\alpha}_p : p \geq 2\}$. Give intuition for this ranking.

- **Solution**: By part 1, for $p \geq 2$, γ_p is monotonically decreasing in p. Then from part 4, the asymptotic variance is monotonically increasing in p, which implies that $\hat{\alpha}_2$ should be the most efficient one in this class of estimators. As p increases, the correlation between y_{t-p} and y_{t-1} decreases for this ARMA(1,1) process. In other words, as an instrument for y_{t-1}, y_{t-p} becomes weaker as p increases, which decreases the efficiency.

6. Propose a consistent estimator of $V(p)$, $p \geq 2$, and use it to form an asymptotically valid 95% confidence interval for α.

   - **Solution**: A natural estimator of $V(p)$ is

   $$\hat{V}(p) = \hat{\gamma}_{p-1}^{-2} \left(\frac{\hat{\gamma}_0}{T-1} \sum_{t=2}^{T} (y_t - \hat{\alpha}_p y_{t-1})^2 + 2\hat{\gamma}_1 \cdot \frac{1}{T-2} \sum_{t=3}^{T} (y_t - \hat{\alpha}_p y_{t-1})(y_{t-1} - \hat{\alpha}_p y_{t-2}) \right)$$

   where $\hat{\gamma}_p = \frac{1}{T-p} \sum_{t=p+1}^{T} y_{t-p} y_t$. Then an asymptotically valid 95% confidence interval is

   $$\left[\hat{\alpha}_p - z_{0.975} \cdot \sqrt{\frac{\hat{V}(p)}{T}}, \hat{\alpha}_p + z_{0.975} \cdot \sqrt{\frac{\hat{V}(p)}{T}} \right]$$

   where $z_{0.975}$ is 0.975 quantile of standard normal.

7. Suppose that $\beta = 0$. Compute the asymptotic relative efficiency of the standard least squares estimator $\hat{\alpha}_1$ relative to the (asymptotically) best estimator in the class $\{\hat{\alpha}_p : p \geq 2\}$.

   - **Solution**: When $\beta = 0$, it reduces to AR(1) process. Then

   $$\sqrt{T}(\hat{\alpha}_1 - \alpha) \rightarrow_d \mathsf{N}(0, \gamma_0^{-1}\sigma^2).$$

   By part 5, we know $\hat{\alpha}_2$ is the best in the class $\{\hat{\alpha}_p : p \geq 2\}$. It is clear that

   $$\sqrt{T}(\hat{\alpha}_2 - \alpha) \rightarrow_d \mathsf{N}(0, \gamma_1^{-2}\gamma_0\sigma^2).$$

   The relative efficiency of $\hat{\alpha}_1$ relative to $\hat{\alpha}_2$ is $\gamma_0^2/\gamma_1^2 = \alpha^{-2} > 1$. So $\hat{\alpha}_1$ is asymptotically more efficient than $\hat{\alpha}_2$.