SUMMER TERM 2021

8-HOUR ONLINE EXAMINATION

ECON0064: Econometrics

All work must be submitted anonymously. Please ensure that you add your **candidate number** and the module code to the template answer sheet provided. Note that the candidate number is a combination of four letters plus a number, e.g. ABCD9. You can find your candidate number in your PORTICO account, under My Studies then the Examinations container. Please, note that the candidate number is NOT the same as your student number (8 digits), which is printed on your UCL ID card. Submitting with your student number will delay marking and when your results might be available.

Word count: 15 pages

The above word count is provided as guidance only on the expected total length of your submitted answer sheets. You will not be penalised if you exceed the word limit.

Answer **ALL questions from Section A** *and* **TWO questions from Section B**.

The total number of points is 100. Section A contributes 60 points and Section B contributes 40 points.

In cases where a student answers more questions than requested by the examination rubric, the policy of the Economics Department is that the students first set of answers up to the required number will be the ones that count (not the best answers). All remaining answers will be ignored.

Allow enough time to submit your work. Waiting until the deadline for submission risks facing technical problems when submitting your work, due to limited network or systems capacity. You must submit your work within **8 hours of the moment you accessed the examination paper**.

By submitting this assessment, **you pledge your honour** *that you have not violated UCLs Assessment Regulations which are detailed in*

https://www.ucl.ac.uk/academic- manual/chapters/chapter-6-student-casework-framework/section-9-student-academic-misconduct-procedure

which include (but are not limited to) plagiarism, self-plagiarism, unauthorised collaboration between students, sharing my assessment with another student or third party, access another students assessment, falsification, contract cheating, and falsification of extenuating circumstances.

Section A

Answer **ALL questions from this Section**.

Question A1 (20 points)

Consider the linear regression model with a constant regressor and one additional regressor

$$y_i = \beta_1 + w_i\beta_2 + u_i.$$

Assume that all variables have finite second moments, and that the errors u_i have mean zero, are independent of w_i, and have unknown variance $\sigma^2 = \mathbb{E}(u_i^2)$. We observe a random sample (y_i, w_i), $i = 1, \ldots, n$, of $n = 100$ observations. For this observed sample we find

$$\frac{1}{n}\sum_{i=1}^{n} w_i = 3, \quad \frac{1}{n}\sum_{i=1}^{n} y_i = 1, \quad \frac{1}{n}\sum_{i=1}^{n} w_i^2 = 10, \quad \frac{1}{n}\sum_{i=1}^{n} y_i w_i = 3, \quad \frac{1}{n}\sum_{i=1}^{n} y_i^2 = 17. \quad (1)$$

(a) Use the information in (1) to calculate the OLS estimates for β_1 and β_2.

(b) Use your result in (a) and the information in (1) to calculate $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \widehat{u}_i^2$, where $\widehat{u}_i = y_i - \widehat{\beta}_1 - w_i\widehat{\beta}_2$.

In the following please use $\widehat{\sigma}^2$ as an estimator for σ^2.

(c) Use your result in (b) and the information in (1) to calculate estimators for the standard errors of $\widehat{\beta}_1$ and $\widehat{\beta}_2$.

For the following subquestions assume that you calculated $\hat{\beta}_1 = 1.2$, $\hat{\beta}_2 = 0.9$, and

$$\widehat{\mathrm{Var}}(\widehat{\beta}) = \begin{pmatrix} 1.69 & -0.37 \\ -0.37 & 0.25 \end{pmatrix},$$

so $\mathrm{se}(\hat{\beta}_1) = 1.3$, and $\mathrm{se}(\hat{\beta}_2) = 0.5$. *Note that these are not the numbers that you should have actually obtained.*

(d) Test the null hypothesis $H_0 : \beta_2 = 0$ vs. $H_a : \beta_2 \neq 0$ using a two-sided t-test. Can you reject H_0 at 90% confidence level?

(e) Construct a 95% confidence interval for $\theta = \beta_1 + 2\beta_2$.

Question A2 (20 points)

Consider the model

$$\begin{aligned} y_i &= x_i^* \beta + u_i, \\ x_i &= x_i^* + \varepsilon_i, \\ z_i &= x_i^* \rho + \eta_i, \end{aligned}$$

For simplicity assume that $x_i^*, u_i, \varepsilon_i$ and η_i are independent scalar random variables with $\mathbb{E}[x_i^*] = \mathbb{E}[u_i] = \mathbb{E}[\varepsilon_i] = \mathbb{E}[\eta_i] = 0$, $\mathrm{Var}[x_i^*] = \mathrm{Var}[u_i] = \mathrm{Var}[\varepsilon_i] = \mathrm{Var}[\eta_i] = 1$. β and ρ are unknown parameters.

(a) Suppose that you have a random sample of size n from the distribution of (y_i, x_i^*). Let $\hat{\beta}^*$ denote the OLS estimator in the regression of y_i on x_i^*, i.e.,

$$\hat{\beta}^* = \frac{\sum_{i=1}^n x_i^* y_i}{\sum_{i=1}^n x_i^{*2}}.$$

Find the asymptotic distribution of $\hat{\beta}^*$.

(b) Suppose that you have a random sample of size n from the distribution of (y_i, x_i) (so you observe x_i instead of x_i^*). Let $\hat{\beta}$ denote the OLS estimator in the regression of y_i on x_i, i.e.,

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}.$$

Find the probability limit of $\hat{\beta}$. Is it consistent for β?

(c) Suppose that you have a random sample of size n from the distribution of (y_i, x_i, z_i). Let $\hat{\beta}_{\mathrm{IV}}$ denote the IV estimator for the regression of y_i on x_i that uses z_i as instrument, i.e.,

$$\hat{\beta}_{\mathrm{IV}} = \frac{\sum_{i=1}^n z_i y_i}{\sum_{i=1}^n z_i x_i}.$$

Find the probability limit of $\hat{\beta}_{\mathrm{IV}}$ under the assumption that $\rho \neq 0$. Is it consistent for β?

(d) Find the asymptotic distribution of $\hat{\beta}_{\mathrm{IV}}$ under the assumption that $\rho \neq 0$.

(e) Discuss what happens to the asymptotic distribution of $\hat{\beta}_{\mathrm{IV}}$ when $\rho = 0$.

Question A3 (20 points)

Suppose that you estimate a linear regression model using families from a (large) sample of villages

$$y_{ij} = x_{ij}\beta + u_{ij},$$

where $i = 1, \ldots, n$ and j index villages and families within a village respectively. For simplicity, assume that the sample consists of only two families per village, so $j \in \{1, 2\}$, and x_{ij} and β are scalar. Also suppose that observations are independent and identically distributed across villages (so, $(x_{i1}, u_{i1}, x_{i2}, u_{i2})$ has the same distribution for all i, and $(x_{i1}, u_{i1}, x_{i2}, u_{i2})$ is independent of $(x_{\ell 1}, u_{\ell 1}, x_{\ell 2}, u_{\ell 2})$ provided that $i \neq \ell$.). Finally, assume that within a village (u_{i1}, u_{i2}) is independent of (x_{i1}, x_{i2}), $\mathbb{E}[x_{i1}] = \mathbb{E}[x_{i2}] = 0$, $\mathbb{E}[u_{i1}] = \mathbb{E}[u_{i2}] = 0$, and

$$\mathrm{Var}\left[\begin{pmatrix} x_{i1} \\ x_{i2} \end{pmatrix}\right] = \begin{pmatrix} 1 & \kappa \\ \kappa & 1 \end{pmatrix}, \quad \mathrm{Var}\left[\begin{pmatrix} u_{i1} \\ u_{i2} \end{pmatrix}\right] = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Let $\hat{\beta}$ denote the OLS estimator in a regression of all the y's on the x's, i.e.,

$$\hat{\beta} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{2} x_{ij} y_{ij}}{\sum_{i=1}^{n} \sum_{j=1}^{2} x_{ij}^2}.$$

(a) Show that $\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{2} x_{ij}^2 \to_p A$ and $\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{2} x_{ij} y_{ij} \to_p B$ and provide expressions for their non-stochastic limits A and B (as functions of the model's parameters).

(b) Use your result in (a) to find the probability limit of $\hat{\beta}$. Is it consistent for β?

(c) Show that $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{2} x_{ij} u_{ij} \Rightarrow \mathcal{N}(0, V)$ and provide an expression for V.

(d) Use your results in (a) and (c) to derive the asymptotic distribution of $\hat{\beta}$ and provide an expression for its asymptotic variance Σ.

(e) For simplicity, first assume that the value of $\rho \neq 0$ is known. Suggest an estimator $\tilde{\beta}$ more efficient than the OLS estimator $\hat{\beta}$ and provide an expression for its asymptotic variance $\tilde{\Sigma}$ (no derivation is required). What would you do if ρ was unknown? (*Hint: This is a harder question. It might be useful to start with thinking about the conditional variances* $\mathrm{Var}(\hat{\beta}|X)$ *and* $\mathrm{Var}(\tilde{\beta}|X)$*, and the form of* $\mathrm{Var}(u|X) = \mathrm{Var}(u)$*, which is a* $2n \times 2n$ *matrix.*)

Section B

Answer **TWO questions from this Section**.

Question B1 (20 points)

We observe an iid sample (y_i, x_i), $i = 1, \dots, n$. We assume that the scalar outcome y_i is a binary variable taking values 0 and 1, and its conditional distribution is given by

$$P(y_i = 1|x_i;\beta) = x_i\beta, \quad P(y_i = 0|x_i;\beta) = 1 - x_i\beta.$$

As usual, $x_i \in \mathbb{R}^K$ is a (row) vector of covariates, and $\beta \in \mathbb{R}^K$ is a (column) vector of the parameters of interest. Additionally, we assume that x_i is bounded (so all relevant moments exist and are finite), $x_i\beta \in (0,1)$ with probability one, and all the relevant rank conditions are satisfied.

(a) Show that the studied model can be represented in the regression form

$$y_i = x_i\beta + u_i, \quad \mathbb{E}[u_i|x_i] = 0,$$

and derive an expression for $\mathbb{E}[u_i^2|x_i] = \mathrm{Var}[u_i|x_i]$.

(b) What is the asymptotic distribution of the OLS estimator $\hat{\beta}_{\mathrm{OLS}}$? Use the result of Part (a) to simplify your expression for its asymptotic variance Σ_{OLS}.

(c) Since the conditional distribution of y_i (given x_i) is fully parameterized, the studied model can also be estimated by maximum likelihood. Write down the log-likelihood for this model.

(d) What is the asymptotic distribution of the MLE estimator $\hat{\beta}_{\mathrm{MLE}}$? Provide an expression for its asymptotic variance Σ_{MLE}. Would you prefer $\hat{\beta}_{\mathrm{MLE}}$ or $\hat{\beta}_{\mathrm{OLS}}$?

(e) Explain why the OLS estimator is not efficient in this context. Provide an alternative least-squares type estimator, which overcomes this problem. Provide an expression for its asymptotic variance and compare it with Σ_{MLE} (no derivation is required).

(f) Consider an alternative estimator $\tilde{\beta}$ minimizing the following criterion function

$$\tilde{\beta} = \arg\min_b \frac{1}{n}\sum_{i=1}^{n} \frac{(y_i - x_i b)^2}{x_i b(1 - x_i b)}.$$

Is it a good idea to estimate β by $\tilde{\beta}$? Explain.

Question B2 (20 points)

We observe an iid sample (y_i, x_i), $i = 1, \ldots, n$. Suppose

$$y_i = \theta_0^2 + \theta_0 x_i + u_i,$$

where both x_i and $\theta_0 > 0$ are scalar. Additionally, assume $\mathbb{E}[x_i] = 0$, $\mathbb{E}[x_i^2] = \sigma_x^2$, $\mathbb{E}[u_i(1, x_i, x_i^2)] = (0, 0, 0)$, $\mathbb{E}[u_i^2(1, x_i, x_i^2)] = (\sigma_u^2, 0, \gamma\sigma_u^2\sigma_x^2)$. Here $(\theta_0, \sigma_x, \sigma_u)$ are unknown parameters but γ is a *known* number. Also, assume that all relevant moments exist and are finite.

(a) Find the asymptotic distribution of the Nonlinear Least Squares estimator

$$\hat{\theta}_{\text{NLLS}} = \arg\min_{\theta \geqslant 0} \frac{1}{n} \sum_{i=1}^{n} (y_i - \theta^2 - \theta x_i)^2.$$

(b) In a regression, we always need variation in covariates in order to identify and estimate the associated coefficients. Is it also a concern in this model? Explain.

For the following subquestions, consider the moment condition $\mathbb{E}[g(y_i, x_i, \theta)] = 0$, where

$$g(y_i, x_i, \theta) = (y_i - \theta^2 - \theta x_i) \begin{pmatrix} 1 \\ x_i \end{pmatrix}.$$

(c) Find the asymptotic distribution of the GMM estimator with the identity weighting matrix and provide an expression for its asymptotic variance.

(d) Find the asymptotic distribution of the optimally weighted GMM estimator and provide an expression for its asymptotic variance.

(e) Suppose the dataset is very large, so naively computing Two-Step GMM or Continuously Updating GMM estimators could be computationally expensive. Can you suggest an estimator that is as efficient as the optimally weighted GMM estimator, but is easier to compute? If you can find more than one such estimator, suggest the one that is computationally least expensive. Briefly explain why the estimator is as efficient as the optimally weighted GMM estimator.

Question B3 (20 points)

Suppose that $y_t = x_t\beta + u_t$, where $u_t = \rho u_{t-1} + \varepsilon_t$, $x_t = e_t + \theta e_{t-1}$, where $|\rho| < 1$, ε_t and e_t are both iid with mean zero and variance σ_ε^2 and σ_e^2, and ε_t and e_τ are independent for all t and τ. Let $\hat{\beta}$ denote the OLS estimator of β based on a sample of size T, and let $\hat{u}_t$ denote the OLS residual.

(a) Consider a process $g_t = x_t u_t$ and let $\{\gamma_j\}_{j=-\infty}^{\infty}$ denote its autocovariances, i.e., $\gamma_j = \mathrm{Cov}(g_t, g_{t-j})$. Express γ_j (for all integers j) in terms of the model's parameters.

(b) Use an appropriate CLT to argue that $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g_t \Rightarrow \mathcal{N}(0, \Omega)$ and provide an expression for Ω.

(c) Use your result in (b) to derive the asymptotic distribution of $\hat{\beta}$ and provide an expression for its asymptotic variance Σ.

(d) Suppose that $T = 100$, $\hat{\beta} = 2.1$, $\frac{1}{T}\sum_{t=1}^{T} x_t^2 = 5$, $\frac{1}{T-1}\sum_{t=2}^{T} x_t x_{t-1} = 2.5$, $\frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^2 = 4$, $\frac{1}{T-1}\sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1} = 3.6$. Use your result in (c) to construct a 95% confidence interval for β.

(e) Now suppose instead $x_t = \phi x_{t-1} + e_t$, where $|\phi| < 1$. Go over steps (a)-(c) again, to derive the asymptotic distribution of $\hat{\beta}$ and provide an expression for its asymptotic variance Σ in this case.

G020: Examination in Econometrics

2011-2012

There are two sections to the exam. Answer **ALL questions from Section A** and **TWO questions from Section B**. The total number of points is 100. Section A contributes 60 points and Section B contributes 40 points.

Fully justified answers are required to obtain high marks. However, answers are not expected to exceed 10-15 lines. In cases where a student answers more questions than requested by the examination rubric, the policy of the Economics Department is that the students first set of answers up to the required number will be the ones that count (Not the best answers). All remaining answers will be ignored. Calculators are permitted.

Time allowed: 3:00 hours

Section A

Answer **ALL questions from this Section**.

Question 1 (20 points)

Comment on the following statements. If the statement is correct, just say so. Otherwise explain in one or two sentences which part of the statement is incorrect and why. Please be brief in your answers to this question.

(a) **(2 pts)** If the regressors are perfectly collinear, then the OLS estimator is still well-defined, but may be inconsistent. Correct?

(b) **(2 pts)** The number of regressors cannot be larger than the sample size when applying OLS. Correct?

(c) **(2 pts)** Heterscedasticity robust standard error estimates cannot be applied if the errors of the model are actually homoscedastic. Correct?

(d) **(2 pts)** Only an estimator that is unbiased in finite samples can be consistent as the sample size becomes large. Correct?

(e) **(2 pts)** Omitting a relevant regressor from a regression can cause a bias in the OLS estimators of the included regressors. Correct?

(f) **(2 pts)** A good instrumental variable needs to satisfy two conditions: it needs to be exogenous, and it needs to be uncorrelated with all endogenous regressors of the model. Correct?

(g) **(2 pts)** The 2SLS estimator is inconsistent if the errors of the model are heteroscedastic. Correct?

(h) **(2 pts)** We can only apply the generalized method of moments when the number of moment conditions is at least as large as the number of parameters to be estimated. Correct?

(i) **(2 pts)** Any positive definite matrix can be chosen as a GMM weight matrix. This choice has no effect on the GMM estimator. Correct?

(j) **(2 pts)** The parameters of an $AR(p)$ model can be consistently estimated by OLS. Correct?

Question 2 (20 points)

Consider the linear regression model

$$h_i = f_i\beta_1 + m_i\beta_2 + u_i,$$

where $i = 1, \ldots, n$ denotes individuals, h_i is the height of individual i (measured in cm), f_i and m_i are dummy variables that indicate whether individual i is female ($f_i = 1$, $m_i = 0$) or male ($f_i = 0$, $m_i = 1$), and u_i is an unobserved error term. We assume that $\mathbb{E}(u_i|m_i, f_i) = 0$ and $\mathbb{E}(u_i^2|m_i, f_i) = 500$. The model can be rewritten as

$$h = X\beta + u,$$

where h and u are n-vectors, X is an $n \times 2$ matrix with rows (f_i, m_i), and $\beta = (\beta_1, \beta_2)'$.

We observe a sample of $n = 125$ individuals, of which $n_f = 25$ are female and $n_m = 100$ are male. The average female height in the sample is 160 and the average male height in the sample is 169.

(a) **(4 pts)** Calculate the 2×2 matrix $(X'X)^{-1}$ and the 2-vector $X'h$ for this sample.

(b) **(4 pts)** Use your result in (a) and the above information on the variance of u_i to calculate the OLS estimator $\hat{\beta}$ and an estimator for the variance-covariance matrix of $\hat{\beta}$.

(c) **(4 pts)** The model can alternatively be written as $y_i = \gamma_1 + \gamma_2 f_i + u_i$. Find the general formula that expresses the parameters γ_1 and γ_2 in terms of β_1 and β_2.

(d) **(4 pts)** Use your results in (b) and (c) to calculate the OLS estimator $\hat{\gamma}_2$ and the estimated standard error of $\hat{\gamma}_2$.

For the following two subquestions assume that you calculated the estimator $\hat{\gamma}_2 = -18$ and that you calculated the estimator for the standard error of $\hat{\gamma}_2$ to be equal to 10. (These are not the actual numbers you should have obtained above.)

(e) **(2 pts)** Consider the hypothesis $H_0 : \gamma_2 = 0$. Can you reject H_0 at 95% confidence level using a two-sided large sample t-test?

(f) **(2 pts)** Consider the hypothesis $H_0 : \gamma_2 \geq 0$. Can you reject H_0 at 95% confidence level using a one-sided large sample t-test?

(Hint: the 97.5% quantile of $\mathcal{N}(0,1)$ is 1.96. The 95% quantile of $\mathcal{N}(0,1)$ is 1.64, the 90% quantile of $\mathcal{N}(0,1)$ is 1.28)

Question 3 (20 points)

Consider the model

$$y_i = \beta x_i + u_i,$$

where x_i is a single regressor (for simplicity we do not consider a constant). In addition to y_i and x_i we observe one more variable z_i. We assume that

$$\begin{pmatrix} u_i \\ x_i \\ z_i \end{pmatrix} \sim iid\, \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_u & 0 \\ \rho_u & 1 & \rho_z \\ 0 & \rho_z & 1 \end{pmatrix}\right],$$

where $-1 < \rho_u < 1$ and $-1 < \rho_z < 1$. Two estimators for β are given by

$$\hat{\beta}_{\text{OLS}} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}, \qquad \hat{\beta}_{\text{IV}} = \frac{\sum_{i=1}^n z_i y_i}{\sum_{i=1}^n z_i x_i}.$$

The first estimator is the OLS estimator, the second is the 2SLS or IV estimator, using z_i as an instrument for x_i.

(a) **(9 pts)** Assume that x_i is exogenous. Translate this assumption into a condition on ρ_u. Prove that as $n \to \infty$ we have $\sqrt{n}(\hat{\beta}_{\text{OLS}} - \beta) \Rightarrow \mathcal{N}(0, \Sigma_{\text{OLS}})$, and calculate the asymptotic variance Σ_{OLS}. (Hint: two random variables that are normally distributed and uncorrelated are also independent.)

(b) **(9 pts)** Assume that z_i is a relevant instrument. Translate this assumption into a condition on ρ_z. Prove that as $n \to \infty$ we have $\sqrt{n}(\hat{\beta}_{\text{IV}} - \beta) \Rightarrow \mathcal{N}(0,\ \Sigma_{\text{IV}})$, and calculate the asymptotic variance Σ_{IV}.

(c) **(2 pts)** Compare the asymptotic variances Σ_{OLS} and Σ_{IV}. Which estimator would you prefer to use if x_i is exogenous.

Section B

Answer **TWO questions from this Section**.

Question 4 (20 points)

Suppose that y_i takes values 0 and 1 and is Bernoulli distributed with paramter $\theta \in [0, 1]$, i.e. $P(y_i = 1) = \theta$ and $P(y_i = 0) = 1 - \theta$. Our goal is to do inference on θ based on the moment condition

$$\mathbb{E} y_i = \theta. \tag{1}$$

We observe a random sample y_i, $i = 1, \ldots, n$, with sample size $n = 100$. In our sample we observe 20 times the outcome $y_i = 1$, and 80 times the outcome $y_i = 0$.

(a) **(4 pts)** Use the moment condition (1) to calculate the method of moments estimate $\hat{\theta}$ for the observed sample.

(b) **(6 pts)** Show that $\sqrt{n}(\hat{\theta} - \theta) \Rightarrow \mathcal{N}(0, \theta(1-\theta))$. (Hint: For the Bernoulli distribution we have $\text{Var}(y_i) = \theta(1 - \theta)$.)

(c) **(4 pts)** Use your result in (b) to provide an estimator for the standard error of $\hat{\theta}$.

(d) **(6 pts)** A consistent estimator for the parameter $\beta = \theta^2$ is given by $\hat{\beta} = \hat{\theta}^2$. Use the delta-method and your result in (c) to provide an estimator for the standard error of $\hat{\beta}$.

Question 5 (20 points)

Suppose we observe independent random draws y_i, $i = 1, \ldots, n$, from the exponential distribution with parameter $\lambda > 0$, i.e. y_i takes values in $[0, \infty)$ and is distributed according to the probability density function

$$f(y_i|\theta) = \lambda e^{-\lambda y_i}.$$

(a) **(8 pts)** Write down the log-likelihood function $Q_n(\lambda) = \frac{1}{n} \log \prod_{i=1}^{n} f(y_i|\lambda)$ for this model, and derive the maximum likelihood estimator $\hat{\lambda} = \operatorname{argmax}_{\theta > 0} Q_n(\lambda)$. Show that $\hat{\lambda}$ can be expressed as a function of the sample mean $\frac{1}{n} \sum_{i=1}^{n} y_i$ only.

(b) **(2 pts)** Suppose you observe a sample with size $n = 100$ and mean $\frac{1}{100} \sum_{i=1}^{100} y_i = 2$. Use this information and your result in (a) to calculate the maximum likelihood estimator $\hat{\lambda}$ for this sample.

(c) **(6 pts)** Calculate the expected Hessian $\mathbb{E}\left[\frac{\partial^2 \log f(y_i|\lambda)}{\partial \lambda^2}\right]$.

(d) **(4 pts)** Use your result in (c) to calculate an estimator for the standard error of the MLE $\hat{\lambda}$ which you calculated in (b).

Question 6 (20 points)

Consider the $MA(1)$ model

$$y_t = \varepsilon_t + \theta \varepsilon_{t-1}, \qquad \varepsilon_t \ \sim \ iid\, \mathcal{N}(0, \sigma^2),$$

where $\theta \in \mathbb{R}$ and $\sigma > 0$ are the parameters of the model. For simplicity we consider an $MA(1)$ process with mean zero, i.e. $\mathbb{E} y_t = 0$.

(a) **(6 pts)** Calculate the variance $\gamma_0 = \mathbb{E}(y_t^2)$ and the first order autocovariance $\gamma_1 = \mathbb{E}(y_t\, y_{t-1})$.

(b) **(10 pts)** For an observed sample $y_1, y_2, \ldots, y_T$ one calculates the sample moments

$$\frac{1}{T}\sum_{t=1}^{T} y_t^2 = 5, \qquad \frac{1}{T-1}\sum_{t=1}^{T-1} y_t y_{t-1} = 2.$$

Use these sample moments and your result in (a) to calculate the method of moments estimators of θ and σ. There may be multiple solutions to this, calculate all of them.

(c) **(4 pts)** Are the parameters θ and σ uniquely identified from the moment conditions $\gamma_0 = \mathbb{E}(y_t^2)$ and $\gamma_1 = \mathbb{E}(y_t\, y_{t-1})$? Is it possible to use the additional moment conditions provided by higher order autocovariances $\gamma_j = \mathbb{E}(y_t\, y_{t-j})$, $j = 2, 3, 4, \ldots$, to identify the parameters of the model uniquely?

G020: Examination in Econometrics

2012-2013

There are two sections to the exam. Answer **ALL questions from Section A** and **TWO questions from Section B**. The total number of points is 100. Section A contributes 60 points and Section B contributes 40 points.

Fully justified answers are required to obtain high marks. However, answers are not expected to exceed 10-15 lines. In cases where a student answers more questions than requested by the examination rubric, the policy of the Economics Department is that the students first set of answers up to the required number will be the ones that count (Not the best answers). All remaining answers will be ignored. Calculators are permitted.

Time allowed: 3:00 hours

Section A

Answer **ALL questions from this Section**.

Question 1 (20 points)

Consider the linear regression model with one constant regressor and two additional regressors

$$y_i = \beta_1 + w_i\beta_2 + z_i\beta_3 + u_i.$$

We assume that the errors u_i are homoscedastic. We observe a sample (y_i, w_i, z_i), $i = 1, \ldots, n$, of $n = 25$ observations. For this sample we have

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^n w_i = \frac{1}{n}\sum_{i=1}^n z_i = 0, \quad \frac{1}{n}\sum_{i=1}^n w_i z_i = 0, \quad \frac{1}{n}\sum_{i=1}^n w_i^2 = 1, \quad \frac{1}{n}\sum_{i=1}^n z_i^2 = \frac{1}{4}, \\ &\frac{1}{n}\sum_{i=1}^n y_i = 2, \quad \frac{1}{n}\sum_{i=1}^n y_i w_i = 3, \quad \frac{1}{n}\sum_{i=1}^n y_i z_i = -1, \quad \frac{1}{n}\sum_{i=1}^n y_i^2 = 53. \end{aligned} \tag{1}$$

(a) Use the information in (1) to calculate the OLS estimators for β_1, β_2 and β_3.

(b) An estimator for $\sigma^2 = \mathbb{E}(u_i^2|x_i)$ is given by $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \hat{u}_i^2$, where $\hat{u}_i = y_i - \hat{\beta}_1 - w_i\hat{\beta}_2 - z_i\hat{\beta}_3$ are the regression residuals. Use your result in (a) and the information in (1) to calculate $\hat{\sigma}^2$.

(c) Use your result in (b) and the information in (1) to calculate estimators for the standard errors of $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$.

(d) Now, define $v_i = 2 + w_i + z_i$, and consider the linear regression model

$$y_i = \gamma_1 + w_i\gamma_2 + v_i\gamma_3 + u_i.$$

Find the OLS estimators for γ_1, γ_2, and γ_3.

For the following subquestion assume that you calculated $\hat{\beta}_3 = -2$ and $\widehat{\text{std}}(\hat{\beta}_3) = 1.2$. (these are not the numbers that you should have actually obtained).

(e) Test the null hypothesis $H_0 : \beta_3 = 0$ using a two-sided t-test. Can you reject H_0 at 95% confidence level? (Hint: the 97.5% quantile of $\mathcal{N}(0,1)$ is 1.96. The 95% quantile of $\mathcal{N}(0,1)$ is 1.64)

Question 2 (20 points)

Consider a linear regression model with a single regressor

$$y_i = x_i\beta + u_i,$$

where β is the true value of the regression coefficient. We assume that (x_i, u_i) are independent and identically distributed across observation $i = 1, \dots n$, and that $\mathbb{E}x_i = 0$, $\mathbb{E}x_i^2 = 1$, $\mathbb{E}x_i^3 = 0$, $\mathbb{E}x_i^4 = 3$, $\mathbb{E}\left(\frac{x_i^2}{1+x_i^2}\right) = 1/3$, $\mathbb{E}(u_i|x_i) = 0$ and $\mathbb{E}(u_i^2|x_i) = 1 + x_i^2$. The OLS and the weighted least squares (WLS) estimator for β are given by

$$\hat{\beta}_{\text{OLS}} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}, \qquad \hat{\beta}_{\text{WLS}} = \frac{\sum_{i=1}^n w_i x_i y_i}{\sum_{i=1}^n w_i x_i^2},$$

where the w_i are non-negative weights, which depend on x_i only. We assume that $\mathbb{E}(w_i x_i^2)$ and $\mathbb{E}[(w_i x_i u_i)^2]$ exist.

(a) Show that as $n \to \infty$ we have $\sqrt{n}(\hat{\beta}_{\text{WLS}} - \beta) \Rightarrow \mathcal{N}(0, \Sigma_{\text{WLS}})$ and derive a formula for the asymptotic variance Σ_{WLS}.

(b) The OLS estimator is a special case of $\hat{\beta}_{\text{WLS}}$ with $w_i = 1$, i.e. your results in (a) also shows that $\sqrt{n}(\hat{\beta}_{\text{OLS}} - \beta) \Rightarrow \mathcal{N}(0, \Sigma_{\text{OLS}})$ as $n \to \infty$. Calculate the asymptotic variance Σ_{OLS} (this should just be a number).

(c) To obtain the exact value of Σ_{OLS} in (b) you used that $\mathbb{E}(u_i^2|x_i) = 1+x_i^2$. In practice the functional form of $\mathbb{E}(u_i^2|x_i)$ and thus Σ_{OLS} are unknown. Provide a formula for a consistent estimator of Σ_{OLS} that could be used in practice.

(d) Given that we know $\mathbb{E}(u_i^2|x_i) = 1 + x_i^2$, what are the optimal weights w_i, which minimize Σ_{WLS} (you do not need to show that these weights are optimal)? What is the value of Σ_{WLS} that is obtained for these optimal weights? Compare the optimal asymptotic variance Σ_{WLS} and the OLS asymptotic variance Σ_{OLS}.

Question 3 (20 points)

Consider a linear regression model

$$y_i = p_i^* \beta + \varepsilon_i,$$

where p_i^* is the only regressor and $\beta \neq 0$. However, p_i^* is not observed, only p_i is observed, which is a noisy version of p_i^* that is contaminated with measurement error v_i, namely

$$p_i = p_i^* + v_i.$$

In addition, we observe a single instrumental variable z_i, i.e. the three observed variables are y_i, p_i, and z_i. We assume that $(z_i, p_i^*, \varepsilon_i, v_i)$ are independent and identically distributed across observations $i = 1, \ldots, n$, and that

$$\begin{pmatrix} z_i \\ p_i^* \\ \varepsilon_i \\ v_i \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho & 0 & 0 \\ \rho & 1 & 0 & 0 \\ 0 & 0 & \sigma_\epsilon^2 & 0 \\ 0 & 0 & 0 & \sigma_v^2 \end{pmatrix}\right],$$

where $0 < \rho < 1$, $\sigma_\epsilon > 0$, and $\sigma_v > 0$ are unknown parameters. Consider the following estimators

$$\hat{\beta}_{\text{OLS}} = \frac{\sum_{i=1}^n p_i y_i}{\sum_{i=1}^n p_i^2}, \qquad \hat{\gamma} = \frac{\sum_{i=1}^n z_i p_i}{\sum_{i=1}^n z_i^2}, \qquad \hat{\pi} = \frac{\sum_{i=1}^n z_i y_i}{\sum_{i=1}^n z_i^2}.$$

These are three OLS estimators obtained from regressing y_i on p_i, p_i on z_i, and y_i on z_i, respectively.

(a) Derive the probability limit of $\hat{\beta}_{\mathrm{OLS}}$ as $n \to \infty$. Is $\hat{\beta}_{\mathrm{OLS}}$ a consistent estimator for β?

(b) Derive the probability limits of $\hat{\gamma}$ and $\hat{\pi}$ as $n \to \infty$.

(c) How can $\hat{\gamma}$ and $\hat{\pi}$ be combined to obtain a consistent estimator for β? Define $\hat{p}_i = \hat{\gamma} z_i$ (the predicted values of the regression of p_i on z_i). How can your proposed estimator for β be obtained by a regression that involves $\hat{p}_i$?

(d) What is the condition on ρ that guarantees that z_i is a relevant instrument for p_i here? Would z_i be an exogenous instrument if $\mathbb{E}(z_i \varepsilon_i) \neq 0$? Would z_i be an exogenous instrument if $\mathbb{E}(z_i v_i) \neq 0$?

(e) Assume that the original model reads

$$y_i = \beta_1 + p_i^* \beta_2 + w_i \beta_3 + \varepsilon_i,$$

where we now also include a constant and one additional regressor w_i, which is assumed to be exogenous (i.e. uncorrelated with ε_i) and uncorrelated with v_i. Otherwise, all distributional assumptions on $(z_i, p_i^*, \varepsilon_i, v_i)$ are unchanged. Describe how one can consistently estimate β_1, β_2 and β_3 in that case (no proof required)?

Section B

Answer **TWO questions from this Section**.

Question 4 (20 points)

Suppose we observe independent random draws y_i, $i = 1, \ldots, n$, from the Poisson distribution with parameter $\lambda > 0$, i.e. y_i takes values in $\{0, 1, 2, 3, 4, \ldots\}$ and is distributed according to the probability density function

$$f(y_i|\lambda) = \frac{\lambda^{y_i}\, e^{-\lambda}}{y_i!},$$

where $y_i! = 1 \cdot 2 \cdot 3 \cdot \cdots \cdot y_i$ is the factorial of y_i (e.g. $0! = 1$, $1! = 1$, $4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24$). The Poisson distribution satisfies $\mathbb{E} y_i = \lambda$.

(a) Write down the log-likelihood function $Q_n(\lambda) = \frac{1}{n} \log \prod_{i=1}^{n} f(y_i|\lambda)$ for this model, and derive the maximum likelihood estimator $\hat{\lambda} = \operatorname{argmax}_{\lambda > 0} Q_n(\lambda)$.

(b) Calculate the expected Hessian $\mathbb{E}\left[\frac{\partial^2 \log f(y_i|\lambda)}{\partial \lambda^2}\right]$, evaluated at the true λ.

(c) Suppose you observe a sample with size $n = 25$ and mean $\frac{1}{n}\sum_{i=1}^{n} y_i = 4$. Use this information and your result in (a) and (b) to calculate the maximum likelihood estimator $\hat{\lambda}$ and an estimator for the standard error of $\hat{\lambda}$.

(d) Assume that the true distribution of y_i is not the Poisson distribution, but still satisfies $\mathbb{E} y_i = \lambda$. In that case, is the above MLE $\hat{\lambda}$ still consistent? Is the estimated standard error of $\hat{\lambda}$ provided above still appropriate in that case?

Question 5 (20 points)

Suppose we observe independent random draws $y_i = \binom{y_{1i}}{y_{2i}} \in \mathbb{R}^2$ of a two component vector y_i, whose mean and variance-covariance matrix read

$$\mathbb{E} y_i = \begin{pmatrix} \theta \\ 2\theta \end{pmatrix}, \qquad \operatorname{Var}(y_i) = \begin{pmatrix} 1 & 1/3 \\ 1/3 & 1 \end{pmatrix},$$

where $\theta \in \mathbb{R}$ is an unknown parameter of interest.

(a) Let $\hat{\theta}_1$ be the method of moments estimator for θ based on the moment condition $\mathbb{E}(y_{1i}) = \theta$. Let $\hat{\theta}_2$ be the method of moments estimator for θ based on the moment condition $\mathbb{E}(y_{2i}) = 2\theta$. What are the asymptotic variances of $\hat{\theta}_1$ and $\hat{\theta}_2$?

(b) Let $g(y_i, \theta) = y_i - \binom{\theta}{2\theta}$. The GMM estimator $\hat{\theta}_{\text{GMM}}$ minimizes the GMM objective function $\left[\frac{1}{n}\sum_{i=1}^{n} g(w_i, \theta)\right]' W \left[\frac{1}{n}\sum_{i=1}^{n} g(w_i, \theta)\right]$, where W is a 2×2 symmetric positive definite weight matrix. Show that $\hat{\theta}_{\text{GMM}}$ can be written as $\hat{\theta}_{\text{GMM}} = w\hat{\theta}_1 + (1-w)\hat{\theta}_2$, where $w \in \mathbb{R}$ is a function of the weight matrix W only. Find an expression for w.

Question 6 (20 points)

We observe data on a scalar outcome variable y_t that is generated from an MA(1) model

$$y_t = \varepsilon_t + \theta \varepsilon_{t-1},$$

where $\theta \in [-1, 1]$ is an unknown parameter and the ε_t are independently distributed over $t = 0, \ldots, T$, with $\mathbb{E}\varepsilon_t = 0$ and $\mathbb{E}\varepsilon_t^2 = \sigma^2$. For some reason it is decided to incorrectly estimate the model as an AR(1) model. The OLS estimator of the AR(1) coefficient reads

$$\hat{\rho} = \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2}.$$

(a) Calculate the autocovariances $\gamma_j = \text{Cov}(y_t, y_{t-j})$ for $j = 0, 1, 2$ and autocorrelations $\rho_j = \gamma_j / \gamma_0$ for $j = 1, 2$ for the above MA(1) model.

(b) How is the probability limit of $\hat{\rho}$ related to the autocovariances and autocorrelations in (a)?

(c) For our sample we find $\hat{\rho} = 0.4$. Use this information to calculate an estimator $\hat{\theta}$ of $\theta \in [-1, 1]$. Here, we have restricted the domain of θ to make this estimator unique.

(d) The OLS regression gives an estimated standard error for $\hat{\rho}$ of $\widehat{\text{std}}(\hat{\rho}) = 0.096$. Use this information and the delta method to calculate an estimated standard error for the estimator $\hat{\theta}$ from (c).

G020: Examination in Econometrics

2013-2014

There are two sections to the exam. Answer **ALL questions from Section A** and **TWO questions from Section B**. The total number of points is 100. Section A contributes 60 points and Section B contributes 40 points.

Fully justified answers are required to obtain high marks. However, answers are not expected to exceed 10-15 lines. In cases where a student answers more questions than requested by the examination rubric, the policy of the Economics Department is that the students first set of answers up to the required number will be the ones that count (Not the best answers). All remaining answers will be ignored. Calculators are permitted.

Time allowed: 3:00 hours

Section A

Answer **ALL questions from this Section**.

Question 1 (20 points)

Consider the linear regression model with a constant regressor and one additional regressor

$$y_i = \beta_1 + w_i\beta_2 + u_i.$$

Assume that all variables have finite second moments, and that the errors u_i have mean zero, are independent of w_i, and have unknown variance $\sigma^2 = \mathbb{E}(u_i^2)$. We observe a random sample (y_i, w_i), $i = 1, \ldots, n$, of $n = 81$ observations. For this observed sample we find

$$\frac{1}{n}\sum_{i=1}^{n} w_i = 2, \quad \frac{1}{n}\sum_{i=1}^{n} y_i = 2, \quad \frac{1}{n}\sum_{i=1}^{n} w_i^2 = 5, \quad \frac{1}{n}\sum_{i=1}^{n} y_i w_i = 5, \quad \frac{1}{n}\sum_{i=1}^{n} y_i^2 = 30. \tag{1}$$

(a) Use the information in (1) to calculate the OLS estimates for β_1 and β_2.

(b) Use your result in (a) and the information in (1) to calculate $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \widehat{u}_i^2$, where $\widehat{u}_i = y_i - \widehat{\beta}_1 - w_i\widehat{\beta}_2$.

In the following please use $\widehat{\sigma}^2$ as an estimator for σ^2.

(c) Use your result in (b) and the information in (1) to calculate estimators for the standard errors of $\widehat{\beta}_1$ and $\widehat{\beta}_2$.

For the following subquestion assume that you calculated $\hat{\beta}_2 = 9$ and $\widehat{\mathrm{std}}(\hat{\beta}_2) = 5$. (these are not the numbers that you should have actually obtained).

(d) Test the null hypothesis $H_0 : \beta_2 \leq 0$ using a one-sided t-test. Can you reject H_0 at 95% confidence level? (Hint: The 97.5% quantile of $\mathcal{N}(0,1)$ is 1.96. The 95% quantile of $\mathcal{N}(0,1)$ is 1.64.)

Question 2 (20 points)

The data generating process for the three scalar variables w_i, z_i and u_i is given by

$$\begin{pmatrix} w_i \\ z_i \\ u_i \end{pmatrix} \sim iid\ \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{wz} & 0 \\ \rho_{wz} & 1 & \rho_{zu} \\ 0 & \rho_{zu} & 1 \end{pmatrix}\right].$$

where $|\rho_{wz}| < 1$ and $|\rho_{zu}| < 1$. The scalar outcome variable y_i is generated from the model $y_i = w_i\,\beta_1 + z_i\,\beta_2 + u_i$, where β_1 and β_2 are two unknown parameters. For simplicity we do not include a constant into the model. We observe y_i, w_i and z_i for a sample of observations $i = 1, \ldots, n$, while u_i is an unobserved error term. Define $\beta = (\beta_1, \beta_2)'$ and $x_i = (w_i, z_i)$. The OLS estimator for β reads

$$\widehat{\beta} = \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix} = \left(\sum_{i=1}^{n} x_i' x_i\right)^{-1} \sum_{i=1}^{n} x_i' y_i.$$

(a) Assume that $\rho_{zu} = 0$. In this case, is $\widehat{\beta}$ a consistent estimator for β? Prove your answer.

(b) For your answer in (a), explain why the two conditions $\rho_{zu} = 0$ and $|\rho_{wz}| < 1$ are important. Would there be a problem with $\widehat{\beta}$ if $\rho_{wz} = 1$?

(c) Assume that $\rho_{zu} \neq 0$ and $\rho_{wz} = 0$. In this case, is $\widehat{\beta}_1$ a consistent estimator for β_1? Prove your answer.

(d) Assume that $\rho_{zu} \neq 0$ and $\rho_{wz} \neq 0$. Show that $\widehat{\beta}_1 \to_p \beta_1^*$ as $n \to \infty$, where β_1^* is some constant. Find an expression for β_1^* as a function of β_1, ρ_{wz} and ρ_{zu}. Is $\widehat{\beta}_1$ a consistent estimator for β_1?

Question 3 (20 points)

We observe a random sample (y_i, x_i, z_i), $i = 1, \ldots, n$, where n is the sample size, and y_i, x_i and z_i are three scalar variables with finite second moments. For simplicity we assume that $\mathbb{E}y_i = \mathbb{E}x_i = \mathbb{E}z_i = 0$. We consider the model

$$y_i = x_i \beta + u_i \ ,$$

where u_i is a mean zero error term and β is the parameter of interest. We also define $\gamma_{xx} = \mathrm{Var}(x_i)$, $\gamma_{xy} = \mathrm{Cov}(x_i, y_i)$, $\gamma_{xz} = \mathrm{Cov}(x_i, z_i)$ and $\gamma_{yz} = \mathrm{Cov}(y_i, z_i)$. We assume that $\gamma_{xx} > 0$.

(a) Write down consistent estimators for γ_{xx}, γ_{xy}, γ_{xz}, and γ_{yz}. Prove for one of these estimators that it is indeed consistent.

(b) Assume that x_i is exogenous, i.e. that $\mathbb{E}(x_i u_i) = 0$. Use this assumption to express β as a function of γ_{xx} and γ_{xy}. Use this expression for β and your result in (a) to provide a consistent estimator for β. Prove that your estimator for β is indeed consistent.

(c) Assume that x_i is endogenous, i.e. $\mathbb{E}(x_i u_i) \neq 0$, but that the instrument z_i satisfies the exclusion restriction $\mathbb{E}(z_i u_i) = 0$. What second assumption on z_i is required if we want to use z_i as an instrumental variable to estimate β? Use these two assumptions on z_i to express β as a function of γ_{xz} and γ_{yz}. Use this expression for β and your result in (a) to write down a consistent estimator for β, which we denote by $\widehat{\beta}_{\mathrm{IV}}$. No consistency proof for $\widehat{\beta}_{\mathrm{IV}}$ is required.

(d) Under the assumptions in (c), show that $\widehat{\beta}_{\mathrm{IV}}$ is asymptotically normal, i.e. show that $\sqrt{n}(\widehat{\beta}_{\mathrm{IV}} - \beta) \Rightarrow \mathcal{N}(0, \Sigma_{\mathrm{IV}})$ as $n \to \infty$. Provide a formula for the asymptotic variance Σ_{IV}. Provide a consistent estimator $\widehat{\Sigma}_{\mathrm{IV}}$ for Σ_{IV}, but no consistency proof for $\widehat{\Sigma}_{\mathrm{IV}}$ is required. (Note that we do not impose homoscedasticity in this question.)

Section B

Answer **TWO questions from this Section**.

Question 4 (20 points)

Suppose we observe independent random draws y_i, $i = 1, \ldots, n$, that are distributed as $y_i \sim \mathcal{N}(\mu, \sigma^2)$. The density function of the normal distribution is $f(y_i|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i-\mu)^2}{2\sigma^2}\right]$. The parameters of this model are $\theta = (\mu, \sigma)'$, where $\mu \in \mathbb{R}$ and $\sigma > 0$.

(a) Write down the log-likelihood function $Q_n(\mu, \sigma) = \frac{1}{n} \log \prod_{i=1}^{n} f(y_i|\mu, \sigma)$ for this model, and take the first partial derivatives to obtain the first order conditions $\frac{\partial Q_n(\widehat{\mu}, \widehat{\sigma})}{\partial \mu} = 0$ and $\frac{\partial Q_n(\widehat{\mu}, \widehat{\sigma})}{\partial \sigma} = 0$.

(b) Derive the maximum likelihood estimator (MLE) for μ by solving the first order condition $\frac{\partial Q_n(\widehat{\mu}, \widehat{\sigma})}{\partial \mu} = 0$. (Your solution for $\widehat{\mu}$ should not depend on $\widehat{\sigma}$. This is a special feature of this model.)

(c) Derive the MLE for σ by solving the first order condition $\frac{\partial Q_n(\widehat{\mu}, \widehat{\sigma})}{\partial \sigma} = 0$. (Your solution for $\widehat{\sigma}$ will also depend on $\widehat{\mu}$, which was already derived in (b)).

(d) The expected Hessian in this model reads

$$\mathbb{E} \begin{pmatrix} \frac{\partial^2 \log f(y_i|\mu,\sigma)}{\partial \mu^2} & \frac{\partial^2 \log f(y_i|\mu,\sigma)}{\partial \mu \partial \sigma} \\ \frac{\partial^2 \log f(y_i|\mu,\sigma)}{\partial \sigma \partial \mu} & \frac{\partial^2 \log f(y_i|\mu,\sigma)}{\partial \sigma^2} \end{pmatrix} = - \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 2/\sigma^2 \end{pmatrix},$$

where the expectation is over the distribution $f(y_i|\mu, \sigma)$. Use this result for the expected Hessian to calculate the asymptotic variance of $\widehat{\theta} = (\widehat{\mu}, \widehat{\sigma})'$, i.e. provide the 2×2 matrix Σ_{MLE}, which satisfies $\sqrt{n}(\widehat{\theta} - \theta) \Rightarrow \mathcal{N}(0, \Sigma_{\text{MLE}})$ as $n \to \infty$. No proof required, you can use results from the lecture.

(e) We now drop the assumption that y_i is normally distributed, and instead only assume that $\mathbb{E}y_i = \mu$ and $\text{Var}(y_i) = \sigma^2$. Is $\widehat{\mu}$ derived in (b) still a consistent estimator for μ? Is $\widehat{\sigma}$ derived in (c) still a consistent estimator for σ? Is Σ_{MLE} derived in (d) still the correct expression for the asymptotic variance of $\widehat{\theta}$? Explain your answers, but no proof required.

Question 5 (20 points)

Suppose that y_i takes values 0 and 1 and is Bernoulli distributed with paramter $\theta \in [0, 1]$, i.e. $P(y_i = 1) = \theta$ and $P(y_i = 0) = 1 - \theta$. Our goal is to do inference on θ based on the moment condition

$$\mathbb{E}y_i = \theta. \tag{2}$$

The Bernoulli distribution furthermore satisfies $\text{Var}(y_i) = \theta(1-\theta)$. We observe a random sample y_i, $i = 1, \ldots, n$, with sample size $n = 400$. In our sample we observe 40 times the outcome $y_i = 1$ and 360 times the outcome $y_i = 0$.

(a) Use the moment condition (2) to calculate the method of moments estimator $\widehat{\theta}$ for the observed sample.

(b) Show that $\sqrt{n}(\widehat{\theta} - \theta) \Rightarrow \mathcal{N}(0, \Sigma_{\text{GMM}})$, and derive a formula for Σ_{GMM}.

(c) Provide a consistent estimator $\widehat{\Sigma}_{\text{GMM}}$ for the asymptotic variance Σ_{GMM} derived in (b). Calculate $\widehat{\Sigma}_{\text{GMM}}$ for the observed sample.

(d) Use your results above to test the hypothesis $H_0 : \theta = \frac{1}{2}$. Can you reject the hypothesis at 95% confidence level in a large sample two-sided t-test? (Hint: The 97.5% quantile of $\mathcal{N}(0, 1)$ is 1.96. The 95% quantile of $\mathcal{N}(0, 1)$ is 1.64.)

(e) Use the delta method to show that $\sqrt{n}(\widehat{\Sigma}_{\text{GMM}} - \Sigma_{\text{GMM}}) \Rightarrow \mathcal{N}(0, \Omega)$, and provide a formula for Ω as a function of θ only.

Question 6 (20 points)

Consider the AR(1) model

$$y_t = \rho\, y_{t-1} + u_t, \qquad u_t \sim iid\, \mathcal{N}(0, \sigma^2).$$

We assume an infinite past history for this process, and we assume that $|\rho| < 1$. Without proof you can use the fact that the time series y_t generated from this model is stationary and ergodic. We observe a sample $y_1, y_2, \ldots, y_T$.

(a) Derive the unconditional mean $\mathbb{E}y_t$, the variance $\gamma_0 = \mathrm{Var}(y_t)$, and the autocovariances $\gamma_1 = \mathrm{Cov}(y_t, y_{t-1})$ and $\gamma_2 = \mathrm{Cov}(y_t, y_{t-2})$. Your results should be functions of ρ and σ only.

(b) Consider the OLS estimator $\widehat{\rho} = \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2}$ obtained from regressing y_t on y_{t-1}. What is the probability limit of $\widehat{\rho}$ as $T \to \infty$? Prove your answer.

(c) Is $\widehat{\rho}$ an unbiased estimator for ρ? Explain your answer, but no proof required.

(d) Consider the OLS estimator $\widehat{\beta} = \frac{\sum_{t=3}^{T} y_t y_{t-2}}{\sum_{t=3}^{T} y_{t-2}^2}$ obtained from regressing y_t on y_{t-2}. What is the probability limit of $\widehat{\beta}$ as $T \to \infty$? Prove your answer.

(e) Let $\widehat{\gamma} = (\widehat{\gamma}_1, \widehat{\gamma}_2)'$ be the OLS estimator from regressing y_t on y_{t-1} and y_{t-2}, i.e. from applying OLS to the regression model $y_t = \gamma_1\, y_{t-1} + \gamma_2\, y_{t-2} + \epsilon_t$. Do you expect $\widehat{\gamma}_1$ to be close to $\widehat{\rho}$ asymptotically? Do you expect $\widehat{\gamma}_2$ to be close to $\widehat{\beta}$ asymptotically? Explain your answer, but no proof required.

G020: Examination in Econometrics

2014-2015

There are two sections to the exam. Answer **ALL questions from Section A** and **TWO questions from Section B**. The total number of points is 100. Section A contributes 60 points and Section B contributes 40 points.

Fully justified answers are required to obtain high marks. However, answers are not expected to exceed 10-15 lines. In cases where a student answers more questions than requested by the examination rubric, the policy of the Economics Department is that the students first set of answers up to the required number will be the ones that count (not the best answers). All remaining answers will be ignored. Calculators are permitted.

Time allowed: 3:00 hours

Section A

Answer **ALL questions from this Section**.

Question 1 (20 points)

Consider the linear regression model

$$y_i = \beta_1 + \beta_2 h_i + u_i,$$

where y_i is the logarithm of the wage of individual i, h_i is a dummy variable which indicates whether individual i graduated from high-school ($h_i = 1$) or not ($h_i = 0$), and u_i is an unobserved error. Assume that $\mathbb{E}(u_i|h_i) = 0$. Let $x_i = (1, h_i)$ and $\beta = (\beta_1, \beta_2)'$. We observe an iid sample of (y_i, h_i), $i = 1, \ldots, n$. In total we observe $n = 80$ individuals, of which $n_0 = \sum_{i=1}^n (1 - h_i) = 60$ have no high-school degree and $n_1 = \sum_{i=1}^n h_i = 20$ have a high-school degree. The average log wage in the subpopulation with $h_i = 0$ is $\frac{1}{n_0} \sum_{i=1}^n (1 - h_i) y_i = 0.2$ and in the subpopulation with $h_i = 1$ is $\frac{1}{n_1} \sum_{i=1}^n h_i y_i = 2$.

(a) Use the model to calculate $\mathbb{E}(y_i|h_i = 0)$ and $\mathbb{E}(y_i|h_i = 1)$. Use the information from the sample to obtain estimators for $\mathbb{E}(y_i|h_i = 0)$ and $\mathbb{E}(y_i|h_i = 1)$. Combine your results from the model and sample to calculate estimators for β_1 and β_2.

(b) Calculate the 2×2 matrix $\sum_{i=1}^n x_i' x_i$, the 2×1 vector $\sum_{i=1}^n x_i' y_i$, and the OLS estimator $\widehat{\beta} = \left(\sum_{i=1}^n x_i' x_i\right)^{-1} \sum_{i=1}^n x_i' y_i$. Compare $\widehat{\beta}$ to your results in (a).

(c) We now also assume homoscedasticity and $\mathbb{E}(u_i^2|h_i) = 15$. Use this information to calculate OLS standard errors for $\widehat{\beta}_1$ and $\widehat{\beta}_2$.

For the following subquestion assume that you calculated $\widehat{\beta}_2 = 0.9$ and $\widehat{\mathrm{std}}(\widehat{\beta}_2) = 0.5$ (these are not the numbers that you should have actually obtained).

(d) Using a large sample t-test, can you reject the null hypothesis $H_0 : \beta_2 \leq 0$ against the alternative hypothesis $H_0 : \beta_2 > 0$ at 95% confidence level? (Hint: The 97.5% quantile of $\mathcal{N}(0,1)$ is 1.96, the 95% quantile of $\mathcal{N}(0,1)$ is 1.64, the 90% quantile of $\mathcal{N}(0,1)$ is 1.28.)

(e) The effect of h_i on y_i is given by β_2 in the above model. Assume now that you also observe the intelligence quotient IQ_i for each individual i. How could you use $(y_i, h_i, \mathrm{IQ}_i)$, $i = 1, \ldots, n$, to obtain a better estimate for the effect of h_i on y_i? Explain why you expect this estimate to be better than the estimates for β_2 above.

Question 2 (20 points)

The data generating process for the three scalar variables x_i, z_i and u_i is given by

$$\begin{pmatrix} x_i \\ z_i \\ u_i \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{xz} & \rho_{xu} \\ \rho_{xz} & 1 & \rho_{zu} \\ \rho_{xu} & \rho_{zu} & 1 \end{pmatrix}\right].$$

where $|\rho_{xz}| < 1$, $|\rho_{xu}| < 1$, $|\rho_{zu}| < 1$, and $\rho_{xz} \neq 0$. The scalar outcome variable y_i is generated from the model $y_i = x_i\,\beta + u_i$, where β is an unknown scalar parameter. For simplicity we do not include a constant into the model. We observe y_i, x_i and z_i for an iid sample of observations $i = 1, \ldots, n$, while u_i is an unobserved error term. The 2SLS (or IV) estimator for β reads

$$\widehat{\beta} = \frac{\sum_{i=1}^n z_i y_i}{\sum_{i=1}^n z_i x_i}.$$

(a) Show that $\widehat{\beta} \to_p \beta^*$ as $n \to \infty$, and find an expression for β^* in terms of β, ρ_{xz}, ρ_{xu} and ρ_{zu}.

(b) Under what condition on ρ_{zu} is $\widehat{\beta}$ consistent for β? What property of the instrumental variable z_i is described by this condition on ρ_{zu}? Also, what property of the instrumental variable z_i is described by the condition $\rho_{xz} \neq 0$?

(c) Assume that we know $\rho_{zu} = 0$. Show that $\sqrt{n}(\widehat{\beta} - \beta) \Rightarrow \mathcal{N}(0, \mathrm{AsyVar}(\sqrt{n}\widehat{\beta}))$, as $n \to \infty$, and find an expression for $\mathrm{AsyVar}(\sqrt{n}\widehat{\beta})$.

(d) Continue to assume that $\rho_{zu} = 0$. For a sample with $n = 125$ observations we calculate $\frac{1}{n}\sum_{i=1}^n z_i y_i = 0.2$, $\frac{1}{n}\sum_{i=1}^n z_i x_i = 0.2$, $\frac{1}{n}\sum_{i=1}^n z_i^2 x_i^2 = 3$, $\frac{1}{n}\sum_{i=1}^n z_i^2 x_i y_i = 0.5$ and $\frac{1}{n}\sum_{i=1}^n z_i^2 y_i^2 = 3$. Use this information and your result in (c) to calculate an asymptotically valid 95% confidence interval for β. (Hint: The 97.5% quantile of $\mathcal{N}(0,1)$ is 1.96, the 95% quantile of $\mathcal{N}(0,1)$ is 1.64, the 90% quantile of $\mathcal{N}(0,1)$ is 1.28.)

Question 3 (20 points)

Consider the linear regression model

$$y_i = \beta_1 w_i + \beta_2 z_i + u_i,$$

where w_i and z_i are two scalar regressors, and u_i is an unobserved error term. We observe an iid sample (y_i, w_i, z_i), $i = 1, \ldots, n$, and we assume that $\mathbb{E}(u_i|w_i, z_i) = 0$. Let $x_i = (w_i, z_i)$ and assume that $\sum_{i=1}^n x_i' x_i$ is invertible. The OLS estimator for $\beta = (\beta_1, \beta_2)'$ reads

$$\widehat{\beta}^{\text{OLS}} = \begin{pmatrix} \widehat{\beta}_1^{\text{OLS}} \\ \widehat{\beta}_2^{\text{OLS}} \end{pmatrix} = \left(\sum_{i=1}^n x_i' x_i \right)^{-1} \sum_{i=1}^n x_i' y_i.$$

Alternatively, we can partition the regression: For given β_1 we can estimate β_2 by $\widehat{\beta}_2 = (\sum_{i=1}^n z_i^2)^{-1} \sum_{i=1}^n z_i (y_i - \beta_1 w_i)$. Plugging this instead of β_2 into the model gives

$$\widetilde{y}_i = \beta_1 \widetilde{w}_i + \widetilde{u}_i,$$

where $\widetilde{y}_i = y_i - z_i (\sum_{i=1}^n z_i^2)^{-1} \sum_{i=1}^n z_i y_i$, and $\widetilde{w}_i = w_i - z_i (\sum_{i=1}^n z_i^2)^{-1} \sum_{i=1}^n z_i w_i$, and $\widetilde{u}_i$ is a new unobserved error term. After this elimination of β_2 the OLS estimator for β_1 reads $\widehat{\beta}_1 = (\sum_{i=1}^n \widetilde{w}_i^2)^{-1} \sum_{i=1}^n \widetilde{w}_i \widetilde{y}_i$. For notational convenience we also define $\Sigma_{ww} = \sum_{i=1}^n w_i^2$, $\Sigma_{zz} = \sum_{i=1}^n z_i^2$, $\Sigma_{wz} = \sum_{i=1}^n w_i z_i$, $\Sigma_{wy} = \sum_{i=1}^n w_i y_i$ and $\Sigma_{zy} = \sum_{i=1}^n z_i y_i$.

(a) Show that $\sum_{i=1}^n \widetilde{w}_i^2 = \sum_{i=1}^n \widetilde{w}_i w_i$ and $\sum_{i=1}^n \widetilde{w}_i \widetilde{y}_i = \sum_{i=1}^n \widetilde{w}_i y_i$. (Hint: You might first want to show that $\sum_{i=1}^n \widetilde{w}_i z_i = 0$. Both of the desired results then follow quickly.)

Even if you cannot show (a) you can still use the results in (a) to solve the following subquestion.

(b) Use the result in (a) to express $\widehat{\beta}_1$ in terms of only Σ_{ww}, Σ_{zz}, Σ_{wz}, Σ_{wy} and Σ_{zy}.

(c) Also express $\widehat{\beta}_1^{\text{OLS}}$ in terms of only Σ_{ww}, Σ_{zz}, Σ_{wz}, Σ_{wy} and Σ_{zy}.

(d) Compare your results in (b) and (c) and calculate the difference $\widehat{\beta}_1 - \widehat{\beta}_1^{\text{OLS}}$. Interpret your result.

(e) Assume that $z_i = 1$ is a constant regressor. Explain the relation between the model $\widetilde{y}_i = \beta_1 \widetilde{w}_i + \widetilde{u}_i$ and the original regression model in that special case.

Section B

Answer **TWO questions from this Section**.

Question 4 (20 points)

Suppose that y_i takes values 0 and 1 and that

$$P(y_i = 0 \,|\, \theta) = \frac{1}{1+e^{\theta}}, \qquad\qquad P(y_i = 1 \,|\, \theta) = \frac{1}{1+e^{-\theta}},$$

where $\theta \in \mathbb{R}$ is a scalar parameter. Note that $P(y_i = 0 \,|\, \theta) + P(y_i = 1 \,|\, \theta) = 1$, because $\frac{1}{1+e^{-\theta}} = \frac{e^{\theta}}{1+e^{\theta}}$ (just multiply numerator and denominator by e^{θ}). We want to do inference on θ from an iid sample $y_1, y_2, \ldots, y_n$ with sample size n.

(a) Write down the log-likelihood function $Q_n(\theta) = \frac{1}{n} \log \prod_{i=1}^{n} f(y_i|\theta)$ for this model, where $f(0|\theta) = P(y_i = 0 \,|\, \theta)$ and $f(1|\theta) = P(y_i = 1 \,|\, \theta)$. Show that

$$\frac{dQ_n(\theta)}{d\theta} = a(\theta) + \frac{1}{n} \sum_{i=1}^{n} y_i,$$

and find the function $a(\theta)$, which only depends on θ and not on $y = (y_1, \ldots, y_n)$.

(b) Calculate the maximum likelihood estimator (MLE) $\widehat{\theta}$ by solving the corresponding first order condition. We do not need to check the second order condition, because $Q_n(\theta)$ is strictly concave in θ.

(c) Calculate the expected Hessian $\mathbb{E}\left[\frac{d^2 \log f(y_i|\theta)}{d\theta^2}\right]$, evaluated at the true θ. From our general theory of the MLE we know that as $n \to \infty$ we have $\sqrt{n}(\widehat{\theta} - \theta) \Rightarrow \mathcal{N}(0, \mathrm{AsyVar}(\sqrt{n}\widehat{\theta}))$. Use your result on the expected Hessian to calculate the asymptotic variance $\mathrm{AsyVar}(\sqrt{n}\widehat{\theta})$, relying on general results from the lecture.

(d) We have $\mathbb{E}(y_i) = b(\theta)$. Calculate the function $b(\theta)$. Write down the method of moments (MM) estimator for θ that is obtained from the moment condition $\mathbb{E}(y_i) = b(\theta)$. This estimator also satisfies $\sqrt{n}(\widehat{\theta}_{\mathrm{MM}} - \theta) \Rightarrow \mathcal{N}(0, \mathrm{AsyVar}(\sqrt{n}\widehat{\theta}_{\mathrm{MM}}))$, as $n \to \infty$. Is the asymptotic variance $\mathrm{AsyVar}(\sqrt{n}\widehat{\theta}_{\mathrm{MM}})$ smaller, larger or equal to the asymptotic variance of the MLE here?

Question 5 (20 points)

Suppose that $y_i \sim \mathcal{N}(0, \sigma^2)$. We observe an iid sample $y_1, y_2, \ldots, y_n$. We want to do inference on the parameter $\sigma > 0$. It is useful to know that $\mathbb{E}(y_i^4) = 3\sigma^4$, $\mathbb{E}\left[\left(y_i^2 - \sigma^2\right)^2\right] = 2\sigma^4$, $\mathbb{E}\left[\left(y_i^2 - \sigma^2\right)\left(y_i^4 - 3\sigma^4\right)\right] = 12\sigma^6$, and $\mathbb{E}\left[\left(y_i^4 - 3\sigma^4\right)^2\right] = 96\sigma^8$.

From our general GMM theory we know that for a vector of moment functions $g(y_i, \sigma)$ that satisfies $\mathbb{E}[g(y_i, \sigma)] = 0$ at the true parameter σ, the corresponding GMM estimator with optimal weight matrix has asymptotic variance-covariance matrix

$$\text{AsyVar}(\sqrt{n}\,\widehat{\sigma}_{\text{GMM}}) = \left\{G'(\text{Var}[g(y_i, \sigma)])^{-1}G\right\}^{-1},$$

where $G = \mathbb{E}[\frac{dg(y_i,\sigma)}{d\sigma}]$. In general G is a matrix, but in our special case here where σ is only a scalar parameter G is simply a vector. You can use this general result on the asymptotic variance without proof, and you can assume that all required regularity conditions are satisfied.

(a) Consider the moment condition $\mathbb{E}(y_i^2) = \sigma^2$. Write down $g(y_i, \sigma)$ that corresponds to that moment condition. Provide a formula for the corresponding GMM estimator as a function of the observed sample $y_1, y_2, \ldots, y_n$. Calculate the asymptotic variance-covariance matrix of this GMM estimator.

(b) Is the choice of weight matrix important for the GMM estimator in (a)? Explain your answer.

(c) Consider the moment conditions $\mathbb{E}(y_i^2) = \sigma^2$ and $\mathbb{E}(y_i^4) = 3\sigma^4$. Write down $g(y_i, \sigma)$ that corresponds to these moment conditions. Write down the corresponding GMM objective function and explain how the GMM estimator can be obtained from that objective function. You do not need to solve for the GMM estimator.

(d) Is the choice of weight matrix important for the GMM estimator in (c)? What is the optimal choice of weight matrix that minimizes the asymptotic variance-covariance matrix? You can rely on general results from the lecture.

(e) Calculate the asymptotic variance-covariance matrix of the GMM estimator in (c) that is obtained when using the optimal weight matrix. Compare to the result in (a). For large samples, is there an efficiency gain from using the additional moment condition $\mathbb{E}(y_i^4) = 3\sigma^4$?

Question 6 (20 points)

We observe data on a scalar outcome variable y_t that is generated from an MA(2) model

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2},$$

where θ_1 and θ_2 are unknown scalar parameters and the ε_t are independently distributed over $t = -2, -1, 0, 1, 2, \ldots, T$, with $\mathbb{E}\varepsilon_t = 0$ and $\mathbb{E}\varepsilon_t^2 = \sigma^2$. For some reason it is decided to incorrectly estimate the model as an AR(1) model. The OLS estimator of the AR(1) coefficient reads

$$\widehat{\rho} = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2}.$$

(a) Calculate the autocovariances $\gamma_j = \mathrm{Cov}(y_t, y_{t-j})$ for $j = 0, 1, 2, 3$ and autocorrelations $\rho_j = \gamma_j / \gamma_0$ for $j = 1, 2, 3$ for the above MA(2) model.

(b) How is the probability limit of $\widehat{\rho}$ related to the autocovariances and autocorrelations in (a)?

In all the following subquestions we assume for simplicity that $\theta_1 = \theta_2 \in [0, 1]$, and we denote this parameter as $\theta = \theta_1 = \theta_2$.

(c) For our sample we find $\widehat{\rho} = 0.5$. Use this information to calculate a consistent estimator for θ.

(d) The OLS regression gives an estimated standard deviation for $\widehat{\rho}$ of $\widehat{\mathrm{std}}(\widehat{\rho}) = 0.02$. Use this information and the delta method to calculate the standard error for the estimator for θ in (c).

G020: Examination in Econometrics

2015-2016

There are two sections to the exam. Answer **ALL questions from Section A** and **TWO questions from Section B**. The total number of points is 100. Section A contributes 60 points and Section B contributes 40 points.

Fully justified answers are required to obtain high marks. However, answers are not expected to exceed 10-15 lines. In cases where a student answers more questions than requested by the examination rubric, the policy of the Economics Department is that the students first set of answers up to the required number will be the ones that count (not the best answers). All remaining answers will be ignored. Calculators are permitted.

Time allowed: 3:00 hours

Section A

Answer **ALL questions from this Section**.

Question A1 (20 points)

Consider the linear regression model with one constant regressor and two additional regressors

$$y_i = \beta_1 + w_i\,\beta_2 + z_i\,\beta_3 + u_i.$$

We assume $\mathbb{E}(u_i|w_i, z_i) = 0$ and $\mathbb{E}(u_i^2|w_i, z_i) = 36$. We observe an iid sample (y_i, w_i, z_i), $i = 1, \ldots, n$, of $n = 25$ observations. For this sample we calculate

$$\frac{1}{n}\sum_{i=1}^{n} y_i = 10, \qquad \frac{1}{n}\sum_{i=1}^{n} w_i = 0, \qquad \frac{1}{n}\sum_{i=1}^{n} z_i = 0, \qquad \frac{1}{n}\sum_{i=1}^{n} w_i z_i = 0,$$
$$\frac{1}{n}\sum_{i=1}^{n} y_i w_i = 9, \qquad \frac{1}{n}\sum_{i=1}^{n} y_i z_i = -8, \qquad \frac{1}{n}\sum_{i=1}^{n} w_i^2 = 9, \qquad \frac{1}{n}\sum_{i=1}^{n} z_i^2 = 16\ . \tag{1}$$

(a) Use the information in (1) to calculate the OLS estimates $\hat\beta_1$, $\hat\beta_2$ and $\hat\beta_3$.

(b) Now, define the alternative regressors $v_i = w_i + z_i$ and $q_i = w_i - z_i$, and consider the linear regression model

$$y_i = \gamma_1 + v_i\,\gamma_2 + q_i\,\gamma_3 + u_i.$$

Use your results in (a) to calculate the OLS estimates for γ_1, γ_2, and γ_3.

(c) Calculate OLS standard errors for $\hat\beta_2$, $\hat\beta_3$, and for $\hat\beta_2 - \hat\beta_3$.

For the following subquestion, please assume that you calculated $\hat\beta_2 = 4$, $\hat\beta_3 = -2$, and $\widehat{\mathrm{std}}(\hat\beta_2 - \hat\beta_3) = 2$ in question (a) and (c). (these are not the numbers that you should have actually obtained)

(d) Test the null hypothesis $H_0 : \beta_2 = \beta_3$ using a large sample two-sided t-test. Can you reject H_0 at 95% confidence level? (Hint: the 97.5% quantile of $\mathcal{N}(0, 1)$ is 1.96. The 95% quantile of $\mathcal{N}(0, 1)$ is 1.64)

(e) Now, assume that $\frac{1}{n}\sum_{i=1}^{n} w_i z_i = -12$, but otherwise the information in (1) is unchanged. Explain what happens to the OLS estimates $\hat\beta_1$, $\hat\beta_2$ and $\hat\beta_3$ in this case.

Question A2 (20 points)

Consider two regressors z_i and w_i with mean and variance-covariance matrix given by

$$\mathbb{E}\left[\begin{pmatrix} z_i \\ w_i \end{pmatrix}\right] = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad \mathrm{Var}\left[\begin{pmatrix} z_i \\ w_i \end{pmatrix}\right] = \begin{pmatrix} 1 & \rho_{ZW} \\ \rho_{ZW} & 1 \end{pmatrix},$$

where $|\rho_{ZW}| < 1$. A linear regression model reads

$$y_i = z_i\beta + w_i\gamma + u_i,$$

where β and γ are unknown scalar parameters, and u_i is an unobserved scalar error term. We define $x_i = (z_i, w_i)$, and we assume that $\mathbb{E}(u_i|x_i) = 0$, $\mathbb{E}(u_i^2|x_i) = \sigma^2$, and $\mathbb{E}(w_i^2|z_i) = 1$. From an iid sample of (y_i, z_i, w_i), $i = 1, \ldots, n$, one can construct the estimators

$$\hat{\beta}^* = \frac{\sum_{i=1}^n z_i y_i}{\sum_{i=1}^n z_i^2}, \qquad \begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = \left(\sum_{i=1}^n x_i' x_i\right)^{-1} \sum_{i=1}^n x_i' y_i .$$

(a) Show that there exists a constant $\beta^* \in \mathbb{R}$ such that $\hat{\beta}^* \to_p \beta^*$ as $n \to \infty$, and find an expression for β^* in terms of β, γ and ρ_{ZW}.

(b) Assume that $\gamma \neq 0$ and $\rho_{ZW} \neq 0$. In that case, would you prefer $\hat{\beta}^*$ or $\hat{\beta}$ to estimate the parameter β? Explain your answer, but no proof required.

We have derived the asymptotic variance of a general OLS estimator under appropriate conditions in the lecture, and you can use those results in the following without proof.

(c) Assume that $\gamma \neq 0$ and $\rho_{ZW} = 0$. Calculate and compare the asymptotic variances of $\hat{\beta}^*$ and $\hat{\beta}$. Which estimator would your prefer in that case?

(d) Assume that $\gamma = 0$ and $\rho_{ZW} \neq 0$. Calculate and compare the asymptotic variances of $\hat{\beta}^*$ and $\hat{\beta}$. Which estimator would your prefer in that case?

(e) Provide an intuitive explanation for the results in (c) and (d) regarding the effect of the inclusion or omission of a regressor on the variance of the estimates for the remaining parameters. A short one sentence explanation for each case is sufficient.

Question A3 (20 points)

Consider the linear regression model

$$y_i = x_i\beta + u_i,$$

where x_i is a single regressor, $\beta \in \mathbb{R}$ is the parameter of interest, and u_i is an unobserved error term. The regressor x_i may be endogenous, but we observe two instrumental variables $z_i = (z_{i1}, z_{i2})$ that satisfy $\mathbb{E}(u_i|z_i) = 0$. For simplicity we assume that all variables y_i, x_i, u_i, z_i have zero mean, and that $\mathbb{E}(z_{i1}z_{i2}) = 0$, implying that the matrix $\mathbb{E}(z_i'z_i)$ is diagonal. In addition, we assume that $\mathbb{E}(u_i^2|z_i) = \sigma^2$, $\mathbb{E}(z_{i1}^2) > 0$, $\mathbb{E}(z_{i2}^2) > 0$, and $\mathbb{E}(x_i z_{i1}) \neq 0$. We observe an iid sample of (y_i, x_i, z_i), $i = 1, \ldots, n$. We consider the IV and 2SLS estimators given by

$$\hat{\beta}_{\text{IV}} = \frac{\sum_{i=1}^n z_{i1}y_i}{\sum_{i=1}^n z_{i1}x_i}, \qquad \hat{\beta}_{\text{2SLS}} = \frac{\left(\sum_{i=1}^n x_i z_i\right)\left(\sum_{i=1}^n z_i'z_i\right)^{-1}\left(\sum_{i=1}^n z_i'y_i\right)}{\left(\sum_{i=1}^n x_i z_i\right)\left(\sum_{i=1}^n z_i'z_i\right)^{-1}\left(\sum_{i=1}^n z_i'x_i\right)}.$$

Notice that $\hat{\beta}_{\text{IV}}$ only uses the first instrument.

(a) Show that $\hat{\beta}_{\text{IV}}$ and $\hat{\beta}_{\text{2SLS}}$ are asymptotically normally distributed as $n \to \infty$, and derive their asymptotic variances. Make sure to use to the assumptions $\mathbb{E}(z_{i1}z_{i2}) = 0$ and $\mathbb{E}(u_i^2|z_i) = \sigma^2$ to simplify the resulting expressions.

(b) Consider $\mathbb{E}(z_{i1}x_i) \neq 0$ and $\mathbb{E}(z_{i2}x_i) = 0$. In this case, is the asymptotic variance of $\hat{\beta}_{\text{2SLS}}$ larger, smaller or equal to the asymptotic variance of $\hat{\beta}_{\text{IV}}$? Which estimator would you recommend to use in this case in both large and small samples?

(c) Explain how you would test whether both instruments are truly exogenous instruments, that is, whether the null hypothesis $H_0 : \mathbb{E}(u_i z_i) = 0$ is satisfied. Explain how to construct a test statistic for this and what its asymptotic distribution should be. No proof required.

(d) Consider again $\mathbb{E}(z_{i1}x_i) \neq 0$ and $\mathbb{E}(z_{i2}x_i) = 0$. Is it still possible to perform your test in part (c)? Does this test still have any power?

Section B

Answer **TWO questions from this Section**.

Question B1 (20 points)

We observe an iid sample (y_i, x_i), $i = 1, \ldots, n$. We assume that the scalar outcome y_i and the single regressor x_i have zero mean, and consider the model:

$$y_i = x_i \beta + u_i, \qquad\qquad u_i \,|\, x_i \sim \mathcal{N}(0, h(x_i)),$$

meaning the unobserved error u_i is normally distributed, conditional on x_i, with mean zero and variance $h(x_i) > 0$. The parameter of interest is $\beta \in \mathbb{R}$. Assume that x_i, $h(x_i)$ and $1/h(x_i)$ have bounded range, to guarantee that all expectations in the following exist.

(a) Assume that $h(x_i)$ is known. In that case we know that the asymptotically efficient estimator for β (i.e. the estimator with smallest asymptotic variance within a certain class of estimators) is the weighted least squares estimator $\hat{\beta}_{\mathrm{WLS}}$, using the inverse of $h(x_i)$ as weights. Write down $\hat{\beta}_{\mathrm{WLS}}$, show that it is asymptotically normally distributed as $n \to \infty$, and derive its asymptotic variance for the current model.

(b) Let $f(x_i)$ be a scalar function of x_i. Show that $\mathbb{E}[(y_i - x_i\beta) f(x_i)] = 0$ is a valid moment condition for β under the above assumptions. Derive the method of moments estimator $\hat{\beta}_{\mathrm{MM}}$ obtained from this moment condition. Under what additional condition is $\hat{\beta}_{\mathrm{MM}}$ consistent for β? No consistency proof required.

(c) Assume that $h(x_i)$ is known. What choice of $f(x_i)$ in (b) guarantees that the resulting estimator $\hat{\beta}_{\mathrm{MM}}$ is asymptotically efficient?

(d) Assume that $h(x_i) = (a + b\,x_i^2 + c\,x_i^4)^{-1}$, where $a, b, c > 0$ are unknown scalar constants. In this case, show that L scalar functions $f_\ell(x_i)$, $\ell \in \{1, \ldots, L\}$, can be chosen such that the two-step efficient GMM method applied to the L moment conditions $\mathbb{E}[(y_i - x_i\beta) f_\ell(x_i)] = 0$, $\ell \in \{1, \ldots, L\}$, gives an asymptotically efficient estimator for β. Explain your answer, but no proof required. The chosen functions $f_\ell(x_i)$ should not depend on any unknown parameters.

(e) Assume that we know $u_i \,|\, x_i \sim t(5)$, that is, conditional on x_i, u_i follows a t-distribution with five degrees of freedom. In this case, is it still true that the WLS estimator in (a) is efficient? Explain your answer, no proof required.

Question B2 (20 points)

We observe an iid sample of discrete outcomes $y_i \in \{0, 1, 2\}$ from individuals $i = 1, \ldots, n$. The probabilites of observing a particular outcome y_i are given by

$$P(y_i = 0) = \theta, \qquad P(y_i = 1) = \theta, \qquad P(y_i = 2) = 1 - 2\theta,$$

where $\theta \in [0, 1/2]$ is an unknown parameter. One can show that this implies that

$$\mathrm{Var}[1(y_i = 0)] = \mathrm{Var}[1(y_i = 1)] = \theta(1-\theta), \qquad \mathrm{Cov}[1(y_i = 0), 1(y_i = 1)] = -\theta^2,$$

where $1(.)$ denotes the indicator function. We assume that the model is correctly specified, we denote the true parameter value by θ_0, and we assume that $\theta_0 \in (0, 1/2)$. We consider the estimators

$$\hat\theta_{\mathrm{MLE}} = \underset{\theta\in[0,1/2]}{\mathrm{argmax}} \sum_{i=1}^{n} f(y_i|\theta), \quad f(y_i|\theta) = [\log\theta]^{1(y_i=0)}\ [\log\theta]^{1(y_i=1)}\ [\log(1-2\theta)]^{1(y_i=2)},$$

$$\hat\theta_{\mathrm{GMM}}(W) = \underset{\theta\in[0,1/2]}{\mathrm{argmin}} \left[\sum_{i=1}^{n} g(y_i,\theta)\right]' W \left[\sum_{i=1}^{n} g(y_i,\theta)\right], \qquad g(y_i,\theta) = \begin{pmatrix} 1(y_i = 0) - \theta \\ 1(y_i = 1) - \theta \end{pmatrix},$$

where W is a 2×2 symmetric positive definite weight matrix. This model satisfies all regularity conditions from the lecture, and therefore we know that as $n \to \infty$ we have $\sqrt{n}\left(\hat\theta_{\mathrm{MLE}} - \theta_0\right) \Rightarrow \mathcal{N}(0, V_{\mathrm{MLE}})$ and $\sqrt{n}\left[\hat\theta_{\mathrm{GMM}}(W) - \theta_0\right] \Rightarrow \mathcal{N}[0, V_{\mathrm{GMM}}(W)]$.

(a) Calculate V_{MLE}. Your result should only depend on θ_0. You can use the general results from the lecture.

(b) We now choose $W = \mathbb{1}_2$, the 2×2 identity matrix. Solve for $\hat\theta_{\mathrm{GMM}}(\mathbb{1}_2)$, and use your explicit formula for $\hat\theta_{\mathrm{GMM}}(\mathbb{1}_2)$ to calculate $V_{\mathrm{GMM}}(\mathbb{1}_2)$. Your result for $V_{\mathrm{GMM}}(\mathbb{1}_2)$ should only depend on θ_0. Is $V_{\mathrm{GMM}}(\mathbb{1}_2)$ larger, smaller, or equal to V_{MLE}?

(c) We know that the asymptotic variance of $\hat\theta_{\mathrm{GMM}}(W)$ can be minimized by choosing W equal to $W^* = \{\mathrm{Var}[g(y_i, \theta_0)]\}^{-1}$. Calculate W^*. Without actually calculating the asymptotic variance, what would you expect for $V_{\mathrm{GMM}}(W^*)$ in the current model? Is $W = W^*$ (or rather $W = cW^*$ for some constant $c > 0$) a sufficient or a necessary condition for minimizing $V_{\mathrm{GMM}}(W^*)$?

(d) The GMM estimator above makes use of the moment conditions $\mathbb{E}[1(y_i = 0) - \theta_0] = 0$ and $\mathbb{E}[1(y_i = 1) - \theta_0] = 0$. Could we improve the efficiency of the GMM estimator by also using the moment condition $\mathbb{E}[1(y_i = 2) - (1 - 2\theta_0)] = 0$? Explain your answer.

Question B3 (20 points)

Consider the Gaussian white noise process $\varepsilon_t \sim \; iid\, \mathcal{N}(0,\sigma^2)$. Thus, ε_t is independently distributed over t, and the pdf of ε_t reads $f(\varepsilon|\sigma^2) = (2\pi\sigma^2)^{-1/2}\exp[-\varepsilon^2/(2\sigma^2)]$. For subquestions (a) and (b) we assume that we observe a sample $(\varepsilon_t \;:\; t = 1,2,\ldots,T)$. We want to estimate the unknown parameter σ^2 from this sample.

(a) Derive an explicit formula for the maximum likelihood estimator $\hat{\sigma}^2_{\text{MLE}}$ that can be obtained from this sample. Is $\hat{\sigma}^2_{\text{MLE}}$ an unbiased estimator for σ^2?

(b) From the general asymptotic theory of the MLE we know that as $T \to \infty$ we have $\sqrt{T}\,(\hat{\sigma}^2_{\text{MLE}} - \sigma^2) \Rightarrow \mathcal{N}(0, V_\infty)$. Calculate the asymptotic variance V_∞. Your result should only depend on σ^2 (or more precisely on the true value of σ^2). You can use the general results from the lecture.

Now, instead of observing ε_t we observe y_t, which is generated from the AR(1) model

$$y_t = \rho\, y_{t-1} + \varepsilon_t \;,$$

with $|\rho| < 1$. We assume that y_0 is also observed, and is distributed such that the process y_t is stationary and ergodic (or equivalently, the model holds for an infinite past history). The goal in the following is to estimate ρ and σ from the sample $(y_t \;:\; t = 0,1,\ldots,T)$. Let $\hat{\rho}$ be the OLS estimator obtained by regressing y_t on y_{t-1}. Let $\hat{\sigma}^2$ be obtained from $\hat{\sigma}^2_{\text{MLE}}$ by replacing ε_t with $y_t - \hat{\rho}\, y_{t-1}$.

(c) Provide explicit formulas for $\hat{\rho}$ and $\hat{\sigma}^2$, and prove that $\hat{\rho}$ is a consistent estimator for ρ.

(d) Derive the asymptotic distribution of $\hat{\sigma}^2$. For this, you can use the result on $\hat{\sigma}^2_{\text{MLE}}$ in part (b), and also the result $\sqrt{T}(\hat{\rho} - \rho) \Rightarrow \mathcal{N}(0, 1-\rho^2)$ as $T \to \infty$, which was derived in the lecture.
(Hint: Start by showing that the expressions for $\hat{\sigma}^2$ and $\hat{\sigma}^2_{\text{MLE}}$ differ only by terms that contain the factor $(\hat{\rho} - \rho)$, and then apply the known results for $\hat{\sigma}^2_{\text{MLE}}$ and $\hat{\rho}$. You can also use that $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \varepsilon_t y_{t-1} \Rightarrow \mathcal{N}(0, \text{Var}(\varepsilon_t y_{t-1}))$, as $T \to \infty$)

G020: Examination in Econometrics

2016-2017

There are two sections to the exam. Answer **ALL questions from Section A** and **TWO questions from Section B**. The total number of points is 100. Section A contributes 60 points and Section B contributes 40 points.

Fully justified answers are required to obtain high marks. However, answers are not expected to exceed 10-15 lines. In cases where a student answers more questions than requested by the examination rubric, the policy of the Economics Department is that the students first set of answers up to the required number will be the ones that count (not the best answers). All remaining answers will be ignored. Calculators are permitted.

Time allowed: 3:00 hours

Section A

Answer **ALL questions from this Section**.

Question A1 (20 points)

Consider the linear regression model

$$y_i = \beta_1 + \beta_2\, d_i + u_i,$$

where y_i is the subjective reported happiness of individual i (measured on some linear scale), d_i is a dummy variable which indicates whether individual i owns a dog ($d_i = 1$) or not ($d_i = 0$), and u_i is an unobserved error term. We assume that $\mathbb{E}(u_i) = 0$ and $\mathbb{E}(u_i d_i) = 0$. We observe an i.i.d. sample of (y_i, d_i), $i = 1, \dots, n$. In total we observe $n = 100$ individuals, of which $n_0 = \sum_{i=1}^n (1 - d_i) = 80$ do not own a dog, and $n_1 = \sum_{i=1}^n d_i = 20$ do own a dog. The average reported happiness in the subpopulation with $d_i = 0$ is $n_0^{-1} \sum_{i=1}^n (1-d_i) y_i = 3$, and in the subpopulation with $d_i = 1$ is $n_1^{-1} \sum_{i=1}^n d_i y_i = 8$. We also calculate the sample average of squared happiness to be $n^{-1} \sum_{i=1}^n y_i^2 = 24$.

(a) Let $x_i = (1, d_i)$ and $\beta = (\beta_1, \beta_2)'$. Calculate $\sum_{i=1}^n x_i' x_i$, and $\sum_{i=1}^n x_i' y_i$, and the OLS estimator

$$\widehat{\beta} = \left(\sum_{i=1}^n x_i' x_i \right)^{-1} \sum_{i=1}^n x_i' y_i.$$

(b) Calculate the OLS estimator obtained from regressing d_i on a constant and y_i, that is, the OLS estimator $\widehat{\gamma} = (\widehat{\gamma}_1, \widehat{\gamma}_2)'$ for the regression model $d_i = \gamma_1 + \gamma_2\, y_i + \epsilon_i$.

(c) Under the unrealistic assumption that dog ownership is assigned randomly to individuals, which of the two estimators $\widehat{\beta}_2$ and $\widehat{\gamma}_2$ estimates a true causal effect? Explain what we mean by causal effect here.

(d) We already made the assumptions $\mathbb{E}(u_i) = 0$ and $\mathbb{E}(u_i d_i) = 0$ above. Are those assumptions also sufficient to guarantee that either $\widehat{\beta}_2$ or $\widehat{\gamma}_2$ estimates a true causal effect? Explain your answer.

(e) We now also assume homoscedasticity, with $\mathbb{E}(u_i^2 | d_i) = 4$. Use this information to calculate the estimator for the OLS standard error of $\widehat{\beta}_2$.

Question A2 (20 points)

The data generating process for the three variables w_i, z_i and u_i is given by

$$\begin{pmatrix} w_i \\ z_i \\ u_i \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{WZ} & 0 \\ \rho_{WZ} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right],$$

where $|\rho_{WZ}| < 1$. The outcome variable y_i is generated from the model

$$y_i = w_i\,\beta_1 + z_i\,\beta_2 + u_i,$$

where β_1 and β_2 are two unknown parameters. For simplicity we do not include a constant into the model. We observe y_i, w_i and z_i for an i.i.d. sample of observations $i = 1, \ldots, n$, while u_i is an unobserved error term. Define $\beta = (\beta_1, \beta_2)'$ and $x_i = (w_i, z_i)$. We consider the estimators

$$\widehat{\beta} = \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix} = \left(\sum_{i=1}^{n} x_i'\,x_i\right)^{-1} \sum_{i=1}^{n} x_i'\,y_i\,, \qquad \widehat{\beta}_1^* = \frac{\sum_{i=1}^{n} y_i\,w_i}{\sum_{i=1}^{n} w_i^2}\,.$$

(a) Show that $\sqrt{n}(\widehat{\beta} - \beta) \Rightarrow \mathcal{N}(0, V)$, as $n \to \infty$, and provide an expression for the 2×2 asymptotic covariance matrix V as a function of ρ_{WZ} only. (Hint: Notice that the assumptions above guarantee homoscedasticity of the error u_i.)

(b) Under what conditions on β_2 and ρ_{WZ} is the estimator $\widehat{\beta}_1^*$ a consistent for β_1? Show that under those conditions we have $\sqrt{n}(\widehat{\beta}_1^* - \beta_1) \Rightarrow \mathcal{N}(0, V_*)$, as $n \to \infty$, and provide an expression for the asymptotic variance V_* that depends only on model parameters.

(c) Assume that $\beta_2 \neq 0$ and $\rho_{WZ} \neq 0$. In that case, taking into account consistency and asymptotic efficiency, which of the two estimators $\widehat{\beta}_1$ and $\widehat{\beta}_1^*$ would you recommend to use to estimate β_1?

(d) Assume that $\beta_2 = 0$ and $\rho_{WZ} \neq 0$. In that case, taking into account consistency and asymptotic efficiency, which of the two estimators $\widehat{\beta}_1$ and $\widehat{\beta}_1^*$ would you recommend to use to estimate β_1?

(e) Assume that $\beta_2 \neq 0$ and $\rho_{WZ} = 0$. In that case, taking into account consistency and asymptotic efficiency, which of the two estimators $\widehat{\beta}_1$ and $\widehat{\beta}_1^*$ would you recommend to use to estimate β_1?

Question A3 (20 points)

We observe an i.i.d. sample (y_i, x_i, z_i), $i = 1, \ldots, n$, where n is the sample size, and y_i, x_i and z_i are three variables with finite second moments. For simplicity we assume that $\mathbb{E}y_i = \mathbb{E}x_i = \mathbb{E}z_i = 0$. We consider the model

$$y_i = x_i \beta + u_i \ ,$$

where u_i is a mean zero error term, and β is the parameter of interest. We assume that $\mathbb{E}(z_i u_i) = 0$ and $\mathbb{E}(x_i z_i) \neq 0$. We consider the estimator

$$\widehat{\beta} = \frac{\sum_{i=1}^n z_i \, y_i}{\sum_{i=1}^n z_i \, x_i} \ .$$

(a) Show that $\sqrt{n}(\widehat{\beta} - \beta) \Rightarrow \mathcal{N}(0, V)$, as $n \to \infty$, and find an expression for the asymptotic variance V.

(b) What property of the instrumental variable z_i is described by our assumption $\mathbb{E}(z_i u_i) = 0$? What property of the instrumental variable z_i is described by our assumption $\mathbb{E}(x_i z_i) \neq 0$?

(c) For a sample with $n = 100$ observations we calculate

$$\frac{1}{n}\sum_{i=1}^n z_i y_i = 0.2, \qquad \frac{1}{n}\sum_{i=1}^n z_i x_i = 0.4, \qquad \frac{1}{n}\sum_{i=1}^n z_i^2 x_i^2 = 8,$$

$$\frac{1}{n}\sum_{i=1}^n z_i^2 x_i y_i = 1, \qquad \frac{1}{n}\sum_{i=1}^n z_i^2 y_i^2 = 3.$$

Use this information and your result in (a) to calculate an asymptotically valid 95% confidence interval for β. Notice that we do not assume homoscedasticity in this question. (Hint: The 97.5% quantile of $\mathcal{N}(0,1)$ is 1.96, the 95% quantile of $\mathcal{N}(0,1)$ is 1.64, the 90% quantile of $\mathcal{N}(0,1)$ is 1.28.)

(d) Formulate a homoscedasticity assumption on u_i. Show that under your assumption we can simplify the asymptotic variance V in (a) such that the variance of u_i, but not u_i itself, appears in the resulting expression for V.

Section B

Answer **TWO questions from this Section**.

Question B1 (20 points)

Suppose we observe independent random draws y_i, $i = 1, \ldots, n$, from the exponential distribution with parameter $\lambda_0 > 0$, that is, y_i takes values in $[0, \infty)$ and its distribution has the probability density function

$$f(y|\lambda_0) = \lambda_0 \exp(-\lambda_0 \, y).$$

For the exponential distribution we furthermore have $\mathbb{E} y_i = \lambda_0^{-1}$, and $\mathbb{E} y_i^2 = 2\, \lambda_0^{-2}$, $\mathbb{E} y_i^3 = 6\, \lambda_0^{-3}$, and $\mathbb{E} y_i^4 = 24\, \lambda_0^{-4}$.

(a) Write down the log-likelihood function $Q_n(\lambda) = \frac{1}{n} \log \prod_{i=1}^n f(y_i|\lambda)$ for this model, and derive the maximum likelihood estimator $\widehat{\lambda}_{\text{MLE}} = \text{argmax}_{\lambda > 0} \, Q_n(\lambda)$. Show that $\widehat{\lambda}_{\text{MLE}}$ can be expressed as a function of the sample mean $\frac{1}{n} \sum_{i=1}^n y_i$ only.

(b) From general results on the MLE we know that $\sqrt{n} \left(\widehat{\lambda}_{\text{MLE}} - \lambda_0 \right) \Rightarrow \mathcal{N}(0, V_{\text{MLE}})$, as $n \to \infty$ (you do not need to show this). Calculate the asymptotic variance V_{MLE}. You can use known general result on the MLE asymptotic variance.

(c) An alternative estimator for λ_0 can be obtained by using the moment condition

$$\mathbb{E} \left(y_i^2 - 2\, \lambda_0^{-2} \right) = 0.$$

Calculate the corresponding method of moments estimator $\widehat{\lambda}_{\text{MM}}$.

(d) Prove that $\sqrt{n} \left(\widehat{\lambda}_{\text{MM}} - \lambda_0 \right) \Rightarrow \mathcal{N}(0, V_{\text{MM}})$, as $n \to \infty$, and calculate the asymptotic variance V_{MM}. Compare V_{MM} to V_{MLE} obtained in (b).
(Hint: One proof method is to apply the CLT to $\frac{1}{\sqrt{n}} \sum_{i=1}^n \left(y_i^2 - 2\lambda_0^{-2} \right)$, and afterwards employ the Delta method.)

Question B2 (20 points)

Suppose that we observe n independent random draws y_i, $i = 1, \dots, n$, of a two component vector $y_i = (y_{1i}, y_{2i})' \in \mathbb{R}^2$, whose mean and variance-covariance matrix are given by

$$\mathbb{E}\, y_i = \begin{pmatrix} \theta_0 \\ 2\theta_0 \end{pmatrix}, \qquad \mathrm{Var}(y_i) = \begin{pmatrix} 1 & 4/5 \\ 4/5 & 1 \end{pmatrix},$$

where $\theta_0 \in \mathbb{R}$ is an unknown parameter of interest.

(a) Let $\widehat{\theta}_1$ be the method of moments estimator for θ_0 based on the moment condition $\mathbb{E}(y_{1i} - \theta_0) = 0$. Let $\widehat{\theta}_2$ be the method of moments estimator for θ_0 based on the moment condition $\mathbb{E}(y_{2i} - 2\theta_0) = 0$. Show that $\sqrt{n}\left(\widehat{\theta}_1 - \theta_0\right) \Rightarrow \mathcal{N}(0, V_1)$ and $\sqrt{n}\left(\widehat{\theta}_2 - \theta_0\right) \Rightarrow \mathcal{N}(0, V_2)$, as $n \to \infty$, and calculate V_1 and V_2.

Now, consider the GMM estimator

$$\widehat{\theta}_{\mathrm{GMM}} = \arg\min_{\theta \in \mathbb{R}} \left[\frac{1}{n} \sum_{i=1}^{n} g(y_i, \theta) \right]' W \left[\frac{1}{n} \sum_{i=1}^{n} g(y_i, \theta) \right],$$

where

$$g(y_i, \theta) = y_i - \begin{pmatrix} \theta \\ 2\theta \end{pmatrix}, \qquad W = \begin{pmatrix} W_{11} & W_{12} \\ W_{12} & W_{22} \end{pmatrix}.$$

It can be shown that (and you can take this result as given)

$$\widehat{\theta}_{\mathrm{GMM}} = w \left(\frac{1}{n} \sum_{i=1}^{n} y_{i1} \right) + (1 - w) \left(\frac{1}{2n} \sum_{i=1}^{n} y_{i2} \right), \quad \text{where} \quad w = \frac{W_{11} + 2W_{12}}{W_{11} + 4W_{12} + 4W_{22}}.$$

(b) Find an optimal GMM weight matrix W_{opt}, that is, a weight matrix W that minimizes the asymptotic variance of $\widehat{\theta}_{\mathrm{GMM}}$ (you can use known general GMM results). For that optimal weight matrix, what is the asymptotic variance of $\widehat{\theta}_{\mathrm{GMM}}$? Compare this to the asymptotic variances of $\widehat{\theta}_1$ and $\widehat{\theta}_2$ in (a).

(c) In addition to the two moment conditions $\mathbb{E}(y_{1i} - \theta_0) = 0$ and $\mathbb{E}(y_{2i} - 2\theta_0) = 0$ above, it is suggested to also use the third moment condition $\mathbb{E}(y_{2i} - y_{1i} - \theta_0) = 0$, which is also a consequence of the above assumptions. Would you expect the efficient GMM estimator based on all three moment conditions to be more efficient than any of the three estimators discussed so far? Explain your answer, but no proof required.

(d) Now, assume that $\mathrm{Var}(y_i)$ and thus also the matrix W_{opt} from question (b) are unknown. Provide formulas for consistent estimators of $\mathrm{Var}(y_i)$ and W_{opt} in that case. You do not need to compute numerical values for those estimators.

Question B3 (20 points)

Consider the AR(1) model

$$y_t = \rho_0 \, y_{t-1} + \varepsilon_t, \qquad\qquad \varepsilon_t \sim i.i.d.\ \mathcal{N}(0, \sigma_0^2),$$

where $\rho_0 \in (-1, 1)$ and $\sigma_0^2 > 0$ are unknown parameters. We observe the initial outcome y_0 at time period zero, and the outcomes $y = (y_1, y_2, \ldots, y_T)$ of T subsequent time periods. The maximum likelihood estimator, conditional on y_0, reads

$$\widehat{\theta} = (\widehat{\rho},\, \widehat{\sigma}^2)' = \arg\max_{\theta} f(y|y_0, \theta),$$

where $\theta = (\rho, \sigma^2)'$ and

$$f(y|y_0, \theta) = \prod_{t=1}^{T} f(y_t|y_{t-1}, \theta), \qquad f(y_t|y_{t-1}, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_t - \rho \, y_{t-1})^2}{2\sigma^2} \right].$$

We assume that the process is stationary (infinite past history).

(a) Derive an explicit formula for $\widehat{\rho}$.

(b) Show that, as $T \to \infty$, we have $\sqrt{T}\,(\widehat{\rho} - \rho_0) \Rightarrow \mathcal{N}(0, V)$, and derive a formula for the asymptotic variance V that only depends on ρ_0.

(c) Is $\widehat{\rho}$ an unbiased estimator for ρ_0? Explain your answer, but no proof required.

Because the process is stationary one can show that $y_0 \sim \mathcal{N}\left(0, \sigma_0^2/(1 - \rho_0^2)\right)$, that is, the density function for the initial y_0 reads

$$f(y_0|\theta) = \sqrt{\frac{1 - \rho^2}{2\pi\sigma^2}} \exp\left[-\frac{y_0^2\,(1 - \rho^2)}{2\sigma^2} \right].$$

(d) Explain how the assumption of stationary initial conditions and the fact that we observe y_0 can be used to obtain an improved maximum likelihood estimator for the parameters $\theta = (\rho, \sigma^2)'$ that fully uses those information. Write down the likelihood maximization problem that defines this improved estimator $\widehat{\theta}_* = (\widehat{\rho}_*,\, \widehat{\sigma}_*^2)$. You do not need to solve for the estimator explicitly.

(e) One can show that, as $T \to \infty$, we have $\sqrt{T}\,(\widehat{\rho}_* - \rho_0) \Rightarrow \mathcal{N}(0, V_*)$. Without actually deriving this result, what do you expect for the asymptotic variance V_* compared to the asymptotic variance V obtained in question (b). Explain your answer, but no proof required.

ECON G020 & GP20: Examination in Econometrics

2017-2018

There are two sections to the exam. Answer **ALL questions from Section A** and **TWO questions from Section B**. The total number of points is 100. Section A contributes 60 points and Section B contributes 40 points.

Fully justified answers are required to obtain high marks. However, answers are not expected to exceed 10-15 lines. In cases where a student answers more questions than requested by the examination rubric, the policy of the Economics Department is that the students first set of answers up to the required number will be the ones that count (not the best answers). All remaining answers will be ignored. Calculators are permitted.

Time allowed: 3:00 hours

Section A

Answer **ALL questions from this Section**.

Question A1 (20 points)

Consider the linear regression model with outcome variable y_i, a constant regressor, and two additional regressors w_i and z_i,

$$y_i = \beta_1 + w_i\beta_2 + z_i\beta_3 + u_i.$$

We assume that the unobserved errors u_i are homoscedastic. We observe a sample (y_i, w_i, z_i), $i = 1, \ldots, n$, of $n = 100$ observations. For this sample we have

$$\begin{aligned} &\sum_{i=1}^n w_i = \sum_{i=1}^n z_i = 0, \qquad \sum_{i=1}^n w_i z_i = 0, \qquad \sum_{i=1}^n y_i w_i = 300, \\ &\sum_{i=1}^n w_i^2 = \sum_{i=1}^n z_i^2 = 100, \qquad \sum_{i=1}^n y_i = 200, \qquad \sum_{i=1}^n y_i z_i = -200, \qquad \sum_{i=1}^n y_i^2 = 5300. \end{aligned} \tag{1}$$

(a) Use the information in (1) to calculate the OLS estimators for β_1, β_2 and β_3.

(b) An estimator for $\sigma^2 = \mathbb{E}(u_i^2|x_i)$ is given by $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \hat{u}_i^2 = \frac{1}{n}\sum_{i=1}^n \hat{u}_i(y_i - \hat{\beta}_1 - w_i\hat{\beta}_2 - z_i\hat{\beta}_3)$, where $\hat{u}_i = y_i - \hat{\beta}_1 - w_i\hat{\beta}_2 - z_i\hat{\beta}_3$ are the regression residuals. Use your result in (a) and the information in (1) to calculate $\hat{\sigma}^2$.
(Hint: you can use $\sum_{i=1}^n \hat{u}_i(\hat{\beta}_1 + w_i\hat{\beta}_2 + z_i\hat{\beta}_3) = 0$ to simplify the calculation, which holds according to the OLS FOC).

(c) Use your result in (b) and the information in (1) to calculate estimators for the standard errors of $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$.

(d) Now, define $v_i = w_i + z_i$ and $s_i = w_i - z_i$, and consider the linear regression model

$$y_i = \gamma_1 + v_i\gamma_2 + s_i\gamma_3 + u_i.$$

Find the OLS estimators for γ_1, γ_2, and γ_3.

Question A2 (20 points)

The data generating process for the three scalar variables x_i, z_i and u_i is given by

$$\begin{pmatrix} x_i \\ z_i \\ u_i \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{xz} & \rho_{xu} \\ \rho_{xz} & 1 & \rho_{zu} \\ \rho_{xu} & \rho_{zu} & 1 \end{pmatrix} \right],$$

where $|\rho_{xz}| < 1$, $|\rho_{xu}| < 1$, $|\rho_{zu}| < 1$, and $\rho_{xz} \neq 0$. Since we assume jointly normally distributed random variables we also know that $\mathbb{E}(z_i^4) = 3$ and $\mathbb{E}(z_i^6) = 15$ and $\mathbb{E}(z_i^2 x_i) = 0$ and $\mathbb{E}(z_i^3 x_i) = 3\rho_{xz}$.

The scalar outcome variable y_i is generated from the model $y_i = x_i \beta + u_i$, where β is an unknown scalar parameter. For simplicity we do not include a constant into the model. We observe y_i, x_i and z_i for an iid sample of observations $i = 1, \ldots, n$, while u_i is an unobserved error term. Consider the estimators

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n z_i y_i}{\sum_{i=1}^n z_i x_i}\,, \qquad \widehat{\beta}_2 = \frac{\sum_{i=1}^n z_i^2 y_i}{\sum_{i=1}^n z_i^2 x_i}\,, \qquad \widehat{\beta}_3 = \frac{\sum_{i=1}^n z_i^3 y_i}{\sum_{i=1}^n z_i^3 x_i}\,.$$

(a) Show that $\widehat{\beta}_1 \to_p \beta_1^*$ as $n \to \infty$, and find an expression for β_1^* in terms of β, ρ_{xz}, ρ_{xu} and ρ_{zu}. Under what condition on ρ_{zu} is $\widehat{\beta}_1$ consistent for β? What property of the instrumental variable z_i is described by this condition on ρ_{zu}?

(b) Assume that we know $\rho_{zu} = 0$. Show that $\sqrt{n}(\widehat{\beta}_1 - \beta) \Rightarrow \mathcal{N}(0, \text{AsyVar}(\sqrt{n}\widehat{\beta}_1))$, as $n \to \infty$, and find an expression for $\text{AsyVar}(\sqrt{n}\widehat{\beta}_1)$ that depends on ρ_{xz} only.

(c) Assume that we know $\rho_{zu} = 0$. Is $\widehat{\beta}_2$ a consistent estimator for β under that assumption? Explain your answer.

(d) Assume that we know $\rho_{zu} = 0$. Show that $\sqrt{n}(\widehat{\beta}_3 - \beta) \Rightarrow \mathcal{N}(0, \text{AsyVar}(\sqrt{n}\widehat{\beta}_3))$, as $n \to \infty$, and find an expression for $\text{AsyVar}(\sqrt{n}\widehat{\beta}_3)$ that depends on ρ_{xz} only. Compare the asymptotic variances of $\widehat{\beta}_1$ and $\widehat{\beta}_3$. Which estimator would you recommend to use here?

Question A3 (20 points)

Consider the linear regression model

$$y_i = \beta_1 w_i + \beta_2 z_i + u_i,$$

where w_i and z_i are two scalar regressors, and u_i is an unobserved error term. We observe an iid sample (y_i, w_i, z_i), $i = 1, \ldots, n$, and we assume that $\mathbb{E}(u_i|w_i, z_i) = 0$, and that $\sum_{i=1}^n w_i z_i \neq 0$ and $\beta_2 \neq 0$. Let $x_i = (w_i, z_i)$ and assume that $\sum_{i=1}^n x_i' x_i$ is invertible. The OLS estimator for $\beta = (\beta_1, \beta_2)'$ reads

$$\widehat{\beta}^{\text{OLS}} = \begin{pmatrix} \widehat{\beta}_1^{\text{OLS}} \\ \widehat{\beta}_2^{\text{OLS}} \end{pmatrix} = \left(\sum_{i=1}^n x_i' x_i \right)^{-1} \sum_{i=1}^n x_i' y_i.$$

Alternatively, we can partition the regression: For given β_1 we can estimate β_2 by $\widehat{\beta}_2 = (\sum_{i=1}^n z_i^2)^{-1} \sum_{i=1}^n z_i (y_i - \beta_1 w_i)$. Plugging this instead of β_2 into the model gives

$$\widetilde{y}_i = \beta_1 \widetilde{w}_i + \widetilde{u}_i,$$

where $\widetilde{y}_i = y_i - z_i (\sum_{i=1}^n z_i^2)^{-1} \sum_{i=1}^n z_i y_i$, and $\widetilde{w}_i = w_i - z_i (\sum_{i=1}^n z_i^2)^{-1} \sum_{i=1}^n z_i w_i$, and $\widetilde{u}_i$ is a new unobserved error term. We consider the following estimators for β_1:

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n \widetilde{w}_i \, \widetilde{y}_i}{\sum_{i=1}^n \widetilde{w}_i^2}, \qquad \widehat{\beta}_1^* = \frac{\sum_{i=1}^n w_i \, \widetilde{y}_i}{\sum_{i=1}^n w_i^2}.$$

One can show that $\sum_{i=1}^n \widetilde{w}_i^2 = \sum_{i=1}^n \widetilde{w}_i w_i$ and $\sum_{i=1}^n \widetilde{w}_i \widetilde{y}_i = \sum_{i=1}^n \widetilde{w}_i y_i$, and you can use those results in the following without proof. For notational convenience we also define $\Sigma_{ww} = \sum_{i=1}^n w_i^2$, $\Sigma_{zz} = \sum_{i=1}^n z_i^2$, $\Sigma_{wz} = \sum_{i=1}^n w_i z_i$, $\Sigma_{wy} = \sum_{i=1}^n w_i y_i$ and $\Sigma_{zy} = \sum_{i=1}^n z_i y_i$.

(a) Express $\widehat{\beta}_1$ in terms of only Σ_{ww}, Σ_{zz}, Σ_{wz}, Σ_{wy} and Σ_{zy}.

(b) Show that $\widehat{\beta}_1 = \widehat{\beta}_1^{\text{OLS}}$.

(c) Do you expect that $\left|\widehat{\beta}_1^*\right| = \left|\widehat{\beta}_1^{\text{OLS}}\right|$ or $\left|\widehat{\beta}_1^*\right| > \left|\widehat{\beta}_1^{\text{OLS}}\right|$ or $\left|\widehat{\beta}_1^*\right| < \left|\widehat{\beta}_1^{\text{OLS}}\right|$. Explain your answer.

(d) Do we obtain a consistent estimator for β_2 by regressing $\widetilde{y}_i$ on z_i? Explain your answer.

Section B

Answer **TWO questions from this Section**.

Question B1 (20 points)

Let $y = (y_1, y_2, \ldots, y_n)$ be a random sample from a uniform distribution on the interval $[\theta_1, \theta_2]$, that is, the $y_i \in \mathbb{R}$ are independent and identically distributed across i, and their pdf reads

$$f(y_i|\theta) = \begin{cases} (\theta_2 - \theta_1)^{-1} & \text{for } y_i \in [\theta_1, \theta_2], \\ 0 & \text{otherwise,} \end{cases}$$

where the parameters of the model are $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$, with $\theta_1 < \theta_2$. One finds

$$\mathbb{E}(y_i) = \frac{\theta_1 + \theta_2}{2}, \qquad \text{Var}(y_i) = \frac{(\theta_2 - \theta_1)^2}{12}.$$

(a) Use the above information on $\mathbb{E}(y_i)$ and $\text{Var}(y_i)$ to write down moment conditions for θ, and provide an explicit formula for the corresponding method of moments estimator $\widehat{\theta}_{\text{MM}} = (\widehat{\theta}_{1,\text{MM}}, \widehat{\theta}_{2,\text{MM}})$. (Hint: It should be easy to express $\widehat{\theta}_{\text{MM}}$ in terms of the sample mean $n^{-1}\sum_{i=1}^n y_i$ and sample variance $n^{-1}\sum_{i=1}^n y_i^2 - \left(n^{-1}\sum_{i=1}^n y_i\right)^2$.)

(b) Provide an explicit formula for the maximum likelihood estimator $\widehat{\theta}_{\text{MLE}} = (\widehat{\theta}_{1,\text{MLE}}, \widehat{\theta}_{2,\text{MLE}})$, which maximizes the likelihood function $f(y|\theta) = \prod_{i=1}^n f(y_i|\theta)$. Is $\widehat{\theta}_{\text{MLE}}$ equal to $\widehat{\theta}_{\text{MM}}$ obtained in part (a)?
(Note that you cannot obtain this MLE by solving the FOC, because the likelihood here is a non-continuous function of θ.)

(c) Let $\widehat{\mu}_{\text{MM}} = \frac{1}{2}\left(\widehat{\theta}_{1,\text{MM}} + \widehat{\theta}_{2,\text{MM}}\right)$ be the estimator for the mean $\mu = \mathbb{E}(y_i)$ obtained from the MM estimator in question (a). Show that $\sqrt{n}(\widehat{\mu}_{\text{MM}} - \mu) \Rightarrow \mathcal{N}(0, V_{\text{MM}})$, as $n \to \infty$, and derive a formula for the asymptotic variance V_{MM}.

One can show that $\widehat{\mu}_{\text{MLE}} = \frac{1}{2}\left(\widehat{\theta}_{1,\text{MLE}} + \widehat{\theta}_{2,\text{MLE}}\right)$ satisfies $\sqrt{n}(\widehat{\mu}_{\text{MLE}} - \mu) \to_p 0$, as $n \to \infty$, implying that $\text{AsyVar}(\sqrt{n}\,\widehat{\mu}_{\text{MLE}}) = 0$. You can use this result in the following without proving it.

(d) Consider $\widehat{\mu} = 2\,\widehat{\mu}_{\text{MM}} - \widehat{\mu}_{\text{MLE}}$. Is $\widehat{\mu}$ a consistent estimator for $\mu = \mathbb{E}(y_i)$? Is $\widehat{\mu}$ more or less efficient than $\widehat{\mu}_{\text{MM}}$? Explain your answer (e.g. by finding the asymptotic distribution of $\widehat{\mu}$).

Question B2 (20 points)

We observe an iid sample of discrete outcomes $y_i \in \{0,1,2\}$ from individuals $i = 1, \dots, n$. The probabilites of observing a particular outcome y_i are given by

$$P(y_i = 0) = 1 - \theta_1 - \theta_2, \qquad P(y_i = 1) = \theta_1, \qquad P(y_i = 2) = \theta_2,$$

where $\theta = (\theta_1, \theta_2) \in (0,1)^2$ are unknown parameters satisfying $\theta_1 + \theta_2 < 1$. This implies that $\mathrm{Var}[1(y_i = 1)] = \theta_1(1-\theta_1)$, $\mathrm{Var}[1(y_i = 2)] = \theta_2(1-\theta_2)$, $\mathrm{Cov}[1(y_i = 1), 1(y_i = 2)] = -\theta_1\theta_2$, where $1(.)$ denotes the indicator function. We assume that the model is correctly specified, and we denote the true parameter values by $\theta_0 = (\theta_{1,0}, \theta_{2,0})$. We consider the estimators

$$\widehat{\theta}_{\mathrm{MLE}} = \left(\widehat{\theta}_{1,\mathrm{MLE}}, \widehat{\theta}_{2,\mathrm{MLE}}\right) = \underset{\theta}{\mathrm{argmax}} \sum_{i=1}^{n} \left[\log(1-\theta_1-\theta_2)\right]^{1(y_i=0)} \left[\log \theta_1\right]^{1(y_i=1)} \left[\log \theta_2\right]^{1(y_i=2)},$$

$$\widehat{\theta}_{\mathrm{GMM}}(W) = \underset{\theta}{\mathrm{argmin}} \left[\sum_{i=1}^{n} g(y_i,\theta)\right]' W \left[\sum_{i=1}^{n} g(y_i,\theta)\right], \qquad g(y_i,\theta) = \begin{pmatrix} 1(y_i = 1) - \theta_1 \\ 1(y_i = 2) - \theta_2 \end{pmatrix},$$

where W is a 2×2 symmetric positive definite weight matrix. This model satisfies all regularity conditions from the lecture, and therefore we know that as $n \to \infty$ we have $\sqrt{n}\left(\widehat{\theta}_{\mathrm{MLE}} - \theta_0\right) \Rightarrow \mathcal{N}(0, V_{\mathrm{MLE}})$ and $\sqrt{n}\left[\widehat{\theta}_{\mathrm{GMM}}(W) - \theta_0\right] \Rightarrow \mathcal{N}[0, V_{\mathrm{GMM}}(W)]$. You can use those results without proof.

(a) Calculate V_{MLE}. Your result should only depend on $\theta_0 = (\theta_{1,0}, \theta_{2,0})$. You can use the general results from the lecture.

(b) Calculate $V_{\mathrm{GMM}}(W)$. Which weight matrix W would you suggest to choose here? Compare V_{MLE} and $V_{\mathrm{GMM}}(W)$.

(c) The GMM estimator above makes use of the moment conditions $\mathbb{E}[1(y_i = 1) - \theta_{1,0}] = 0$ and $\mathbb{E}[1(y_i = 2) - \theta_{2,0}] = 0$. Could we improve the efficiency of the GMM estimator by also using the moment condition $\mathbb{E}[1(y_i = 0) - (1 - \theta_{1,0} - \theta_{2,0})] = 0$? Explain your answer.

(d) Let $\widehat{\beta} = \left(\widehat{\theta}_{1,\mathrm{MLE}}\right)^2 + \left(\widehat{\theta}_{2,\mathrm{MLE}}\right)^2$ be an estimator of $\beta_0 = (\theta_{1,0})^2 + (\theta_{2,0})^2$. Explain why $\sqrt{n}\left(\widehat{\beta} - \beta_0\right) \Rightarrow \mathcal{N}(0, V_\beta)$, as $n \to \infty$. Derive the asymptotic variance V_β under the simplifying assumption that $\theta_{1,0} = \theta_{2,0}$.

Question B3 (20 points)

Consider the MA(1) model

$$y_t = \varepsilon_t + \theta \varepsilon_{t-1}, \qquad \varepsilon_t \sim iid\,\mathcal{N}(0, \sigma^2),$$

where $\theta \in \mathbb{R}$ and $\sigma > 0$ are the parameters of the model. For simplicity, we consider an MA(1) process with zero mean, that is, we have $\mathbb{E}y_t = 0$.

(a) Calculate the variance $\gamma_0 = \mathbb{E}(y_t^2)$, first order autocovariance $\gamma_1 = \mathbb{E}(y_t\, y_{t-1})$, and second order autocovariance $\gamma_2 = \mathbb{E}(y_t\, y_{t-2})$.

(b) For an observed sample $(y_1, y_2, \ldots, y_T)$ of $T = 202$ observations one calculates the sample moments

$$\frac{1}{T}\sum_{t=1}^{T} y_t^2 = 20, \qquad\qquad \frac{1}{T-1}\sum_{t=1}^{T-1} y_t\, y_{t-1} = 8.$$

Use these sample moments and your result in (a) to calculate the method of moments estimators of θ and σ. There may be multiple solutions to this, calculate all of them.

(c) One can show that, as $T \to \infty$, we have

$$\frac{1}{\sqrt{(T-2)/2}} \sum_{t=1}^{(T-2)/2} (y_{2t}\, y_{2t+2} - \gamma_2) \Rightarrow \mathcal{N}(0, V).$$

Find a formula for the asymptotic variance V in terms of the model parameters θ and σ. (Hint: it is useful to first show that $y_{2t}\, y_{2t+2}$ is uncorrelated over t).

(d) For the same sample as in part (b) with $T = 202$ observations for y_t one calculates

$$\frac{1}{(T-2)/2} \sum_{t=1}^{(T-2)/2} y_{2t}\, y_{2t+2} = -3.$$

Based on this information, the information provided in part (b), and using your result in (c), can you reject the null hypothesis that y_t is indeed a mean zero MA(1) process (as described above) at 95% confidence level? (Hint: the 97.5% quantile of $\mathcal{N}(0,1)$ is 1.96. The 95% quantile of $\mathcal{N}(0,1)$ is 1.64)

ECON0064: Examination in Econometrics

2018-2019

There are two sections to the exam. Answer **ALL questions from Section A** and **TWO questions from Section B**. The total number of points is 100. Section A contributes 60 points and Section B contributes 40 points.

Fully justified answers are required to obtain high marks. However, answers are not expected to exceed 10-15 lines. In cases where a student answers more questions than requested by the examination rubric, the policy of the Economics Department is that the students first set of answers up to the required number will be the ones that count (not the best answers). All remaining answers will be ignored. Calculators are permitted.

Time allowed: 3:00 hours

Section A

Answer **ALL questions from this Section**.

Question A1 (20 points)

Consider the linear regression model

$$y_i = \beta_1 + \beta_2 h_i + u_i,$$

where y_i is a measure of health status of individual i, h_i is a dummy variable which indicates whether individual i smokes ($h_i = 1$) or not ($h_i = 0$), and u_i is an unobserved error. Assume that $\mathbb{E}(u_i|h_i) = 0$. Let $x_i = (1, h_i)$ and $\beta = (\beta_1, \beta_2)'$. We observe an iid sample of (y_i, h_i), $i = 1, \ldots, n$. In total we observe $n = 120$ individuals, of which $n_0 = \sum_{i=1}^n (1 - h_i) = 80$ do not smoke and $n_1 = \sum_{i=1}^n h_i = 40$ smoke. The average measure of health in the subpopulation with $h_i = 0$ is $\frac{1}{n_0} \sum_{i=1}^n (1 - h_i) y_i = 4$ and in the subgroup of individuals with $h_i = 1$ is $\frac{1}{n_1} \sum_{i=1}^n h_i y_i = 0.5$.

(a) Calculate the 2×2 matrix $\sum_{i=1}^n x_i' x_i$, the 2×1 vector $\sum_{i=1}^n x_i' y_i$, and the OLS estimator $\widehat{\beta} = \left(\sum_{i=1}^n x_i' x_i\right)^{-1} \sum_{i=1}^n x_i' y_i$.

(b) We now also assume homoscedasticity and $\mathbb{E}(u_i^2|h_i) = 20$. Use this information to calculate OLS standard errors for $\widehat{\beta}_1$ and $\widehat{\beta}_2$.

For the following subquestion assume that you calculated $\widehat{\beta}_2 = -2$ and $\widehat{\text{std}}(\widehat{\beta}_2) = 0.5$ (these are not the numbers that you should have actually obtained).

(c) Using a large sample t-test, can you reject the null hypothesis $H_0 : \beta_2 \geq 0$ against the alternative hypothesis $H_0 : \beta_2 < 0$ at 95% confidence level? (Hint: The 97.5% quantile of $\mathcal{N}(0,1)$ is 1.96, the 95% quantile of $\mathcal{N}(0,1)$ is 1.64, the 90% quantile of $\mathcal{N}(0,1)$ is 1.28.)

(d) Let $\tilde{y}_i = y_i - \frac{1}{n} \sum_{i=1}^n y_i$ and $\tilde{h}_i = h_i - \frac{1}{n} \sum_{i=1}^n h_i$. These are demeaned versions of the variables y_i and h_i. Let $\hat{\delta}$ be the OLS estimator obtained from regressing y_i on $\tilde{h}_i$ only. Let $\hat{\theta}$ be the OLS estimator obtained from regressing $\tilde{y}_i$ on $\tilde{h}_i$ only. What is the relationship between $\hat{\delta}$, $\hat{\theta}$ and $\hat{\beta}_2$?

(e) It is suggested to run a regression of y_i on a constant, h_i and $(1 - h_i)$, that is, to include three regressors $(1, h_i, 1 - h_i)$ in the linear regression. Explain whether this is a good idea, and what you would expect from the resulting OLS estimator.

Question A2 (20 points)

We observe an i.i.d. sample (y_i, x_i, z_i), $i = 1, \ldots, n$, where n is the sample size, and y_i, x_i and z_i are three variables with finite second moments. For simplicity we assume that $\mathbb{E}y_i = \mathbb{E}x_i = \mathbb{E}z_i = 0$ and $\mathbb{E}y_i^2 = \mathbb{E}x_i^2 = \mathbb{E}z_i^2 = 1$. We consider the model

$$y_i = x_i\beta + u_i \ ,$$

where u_i is a mean zero error term, and β is the parameter of interest. We denote $\mathbb{E}(z_i u_i) = \rho_{zu}$, $\mathbb{E}(x_i z_i) = \rho_{xz}$ and $\mathbb{E}(x_i u_i) \neq 0$, where $|\rho_{xz}| < 1$, $|\rho_{xu}| < 1$, $|\rho_{zu}| < 1$, and $\rho_{xz} \neq 0$. We observe y_i, x_i and z_i for an iid sample of observations $i = 1, \ldots, n$, while u_i is an unobserved error term. The 2SLS (or IV) estimator for β reads

$$\widehat{\beta} = \frac{\sum_{i=1}^n z_i y_i}{\sum_{i=1}^n z_i x_i}.$$

(a) Show that $\widehat{\beta} \to_p \beta^*$ as $n \to \infty$, and find an expression for β^* in terms of β, ρ_{xz} and ρ_{zu}.

(b) Under what condition on ρ_{zu} is $\widehat{\beta}$ consistent for β? What property of the instrumental variable z_i is described by this condition on ρ_{zu}? Also, what property of the instrumental variable z_i is described by the condition $\rho_{xz} \neq 0$?

(c) Assume that we know $\rho_{zu} = 0$. Show that $\sqrt{n}(\widehat{\beta} - \beta) \Rightarrow \mathcal{N}(0, \mathrm{AsyVar}(\sqrt{n}\widehat{\beta}))$, as $n \to \infty$, and find an expression for $\mathrm{AsyVar}(\sqrt{n}\widehat{\beta})$.

(d) Continue to assume that $\rho_{zu} = 0$. For a sample with $n = 200$ observations we calculate $\frac{1}{n}\sum_{i=1}^n z_i y_i = 0.4$, $\frac{1}{n}\sum_{i=1}^n z_i x_i = 0.2$, $\frac{1}{n}\sum_{i=1}^n z_i^2 x_i^2 = 3$, $\frac{1}{n}\sum_{i=1}^n z_i^2 x_i y_i = 0.5$ and $\frac{1}{n}\sum_{i=1}^n z_i^2 y_i^2 = 5$. Use this information and your result in (c) to calculate an asymptotically valid 95% confidence interval for β. (Hint: The 97.5% quantile of $\mathcal{N}(0,1)$ is 1.96, the 95% quantile of $\mathcal{N}(0,1)$ is 1.64, the 90% quantile of $\mathcal{N}(0,1)$ is 1.28.)

(e) Now assume that $u_i = w_i\gamma + \epsilon_i$, where $\mathbb{E}(\epsilon_i z_i) = 0$, $\mathbb{E}(w_i z_i) \neq 0$, and (y_i, x_i, z_i, w_i), $i = 1, \ldots, n$, are i.i.d. Is $\widehat{\beta}$ a consistent estimator for β in this case? Show your derivations and explain your answer.

Question A3 (20 points)

We observe an iid sample (y_i, x_i), $i = 1, \ldots, n$. We assume that the scalar outcome y_i and the single regressor x_i have zero mean, and consider the model:

$$y_i = x_i \beta + u_i, \qquad\qquad u_i \,|\, x_i \sim \mathcal{N}(0, h(x_i)),$$

meaning the unobserved error u_i is normally distributed, conditional on x_i, with mean zero and variance $h(x_i) > 0$. The parameter of interest is $\beta \in \mathbb{R}$. Assume that x_i, $h(x_i)$ and $1/h(x_i)$ have bounded range, to guarantee that all expectations in the following exist.

(a) Assume that $h(x_i)$ is known. In that case we know that the asymptotically efficient estimator for β (i.e. the estimator with smallest asymptotic variance within a certain class of estimators) is the weighted least squares estimator $\hat{\beta}_{\text{WLS}}$, using the inverse of $h(x_i)$ as weights. Write down $\hat{\beta}_{\text{WLS}}$ and show that it is a consistent estimator for β.

(b) Show that as $n \to \infty$, $\sqrt{n}(\widehat{\beta}_{\text{WLS}} - \beta) \Rightarrow \mathcal{N}(0, V)$ and provide an expression for V as a function of $h(x_i)$.

(c) We now relax the assumption that $h(x_i)$ is known. Assume instead that $h(x_i) = 3 + \gamma x_i^2$, where $\gamma > 0$ is an unknown scalar constant. Provide a consistent estimator for γ (just write down the estimator, no proof of consistency required). Under this assumption on $h(x_i)$, write down a consistent and efficient estimator for β, denoted by $\hat{\beta}^*$.

(d) Show that as $n \to \infty$, $\sqrt{n}(\widehat{\beta}^* - \beta) \Rightarrow \mathcal{N}(0, V^*)$. Derive an expression for the asymptotic variance V^* as a function of γ. Using the estimator for γ derived in part (c), provide an estimator for V^*.

(e) Assume that we know $u_i \,|\, x_i \sim t(5)$, that is, conditional on x_i, u_i follows a t-distribution with five degrees of freedom. In this case, is it still true that the WLS estimator in (a) is efficient? Explain your answer, no proof required.

Section B

Answer **TWO questions from this Section**.

Question B1 (20 points)

Suppose that y_i takes values 0 and 1 and that

$$P(y_i = 0 \,|\, \theta) = \frac{1}{1+e^{\theta}}, \qquad\qquad P(y_i = 1 \,|\, \theta) = \frac{1}{1+e^{-\theta}},$$

where $\theta \in \mathbb{R}$ is a scalar parameter. Note that $P(y_i = 0 \,|\, \theta) + P(y_i = 1 \,|\, \theta) = 1$, because $\frac{1}{1+e^{-\theta}} = \frac{e^{\theta}}{1+e^{\theta}}$ (just multiply numerator and denominator by e^{θ}). We want to do inference on θ from an iid sample $y_1, y_2, \ldots, y_n$ with sample size n.

(a) Write down the log-likelihood function $Q_n(\theta) = \frac{1}{n} \log \prod_{i=1}^{n} f(y_i|\theta)$ for this model, where $f(0|\theta) = P(y_i = 0 \,|\, \theta)$ and $f(1|\theta) = P(y_i = 1 \,|\, \theta)$. Show that

$$\frac{dQ_n(\theta)}{d\theta} = a(\theta) + \frac{1}{n}\sum_{i=1}^{n} y_i,$$

and find the function $a(\theta)$, which only depends on θ and not on $y = (y_1, \ldots, y_n)$.

(b) Calculate the maximum likelihood estimator (MLE) $\widehat{\theta}$ by solving the corresponding first order condition. We do not need to check the second order condition, because $Q_n(\theta)$ is strictly concave in θ.

(c) Calculate the expected Hessian $\mathbb{E}\left[\frac{d^2 \log f(y_i|\theta)}{d\theta^2}\right]$, evaluated at the true θ. From our general theory of the MLE we know that as $n \to \infty$ we have $\sqrt{n}(\widehat{\theta} - \theta) \Rightarrow \mathcal{N}(0, \mathrm{AsyVar}(\sqrt{n}\widehat{\theta}))$. Use your result on the expected Hessian to calculate the asymptotic variance $\mathrm{AsyVar}(\sqrt{n}\widehat{\theta})$, relying on general results from the lecture.

(d) Suppose you observe a sample with size $n = 20$ and mean $\frac{1}{n}\sum_{i=1}^{n} y_i = 0.4$. Use this information and your result in (b) and (c) to calculate the maximum likelihood estimator $\widehat{\theta}$ and an estimator for the standard error of $\widehat{\theta}$.

(e) We define the odds ratio parameter as $\lambda = \frac{P(y_i=1\,|\,\theta)}{P(y_i=0\,|\,\theta)}$. Show that $\lambda = e^{\theta}$. Given the MLE estimate for $\widehat{\theta}$, provide a consistent estimator $\widehat{\lambda}$ for λ. Show that as $n \to \infty$, $\sqrt{n}(\widehat{\lambda} - \lambda) \Rightarrow \mathcal{N}(0, V)$ and provide an expression for V as a function of θ.

Question B2 (20 points)

Assume that $y_i \sim \mathcal{N}(0, \sigma^2)$. We observe an iid sample $y_1, y_2, \ldots, y_n$. We want to do inference on the parameter $\sigma > 0$.

(a) Use $\mathbb{E}y_i = 0$ and $\mathrm{Var}(y_i) = \sigma^2$ to write down a moment condition for $\sigma > 0$. Provide a formula for the corresponding Method of Moment (MM) estimator as a function of the observed sample $y_1, y_2, \ldots, y_n$. Calculate the asymptotic variance-covariance matrix of this estimator.

From our general GMM theory we know that for a vector of moment functions $g(y_i, \sigma)$ that satisfies $\mathbb{E}[g(y_i, \sigma)] = 0$ at the true parameter σ, the corresponding GMM estimator with optimal weight matrix W^* has asymptotic variance-covariance matrix

$$\mathrm{AsyVar}(\sqrt{n}\,\widehat{\sigma}_{\mathrm{GMM}}) = \left\{G' W^* G\right\}^{-1},$$

where $G = \mathbb{E}[\frac{dg(y_i,\sigma)}{d\sigma}]$. In this special case where σ is only a scalar parameter G is simply a vector. You can use this general result on the asymptotic variance without proof, and you can assume that all required regularity conditions are satisfied.

(b) In addition to the moment condition in (a), use the moment condition $\mathbb{E}(y_i^4) = 3\sigma^4$. Write down $g(y_i, \sigma)$ that corresponds to these moment conditions. Write down the corresponding GMM objective function and explain how the GMM estimator can be obtained from that objective function. You do not need to solve for the GMM estimator.

(c) What is the optimal choice of weight matrix, W^*, that minimizes the asymptotic variance-covariance matrix? Calculate the asymptotic variance-covariance matrix of the GMM estimator in (b) that is obtained when using the optimal weight matrix.

(d) Alternatively, σ can also be estimated by MLE using the assumption that $y_i \sim \mathcal{N}(0, \sigma^2)$. The contribution to the likelihood in this case is given by $f(y_i|\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y_i^2}{2\sigma^2}}$. Using results from the lecture, derive the asymptotic variance for the MLE estimator for $\widehat{\sigma}$.

(e) Compare the asymptotic variance-covariance matrix of the MLE in (d) with that of GMM obtained in part (c) and the one obtained from the method of moments in part (a). Is it possible to add moment conditions to GMM to improve on its efficiency? Explain your answer.

Question B3 (20 points)

Consider the MA(1) model

$$y_t = \varepsilon_t + \theta\varepsilon_{t-1}, \qquad \varepsilon_t \sim iid\,\mathcal{N}(0,\sigma^2),$$

where $\theta \in \mathbb{R}$ and $\sigma > 0$ are the parameters of the model.

(a) Calculate the variance $\gamma_0 = \mathbb{E}(y_t^2)$ and the first and second order autocovariances $\gamma_1 = \mathbb{E}(y_t\, y_{t-1})$ and $\gamma_2 = \mathbb{E}(y_t\, y_{t-2})$.

(b) For an observed sample $y_1, y_2, \ldots, y_T$ one calculates the sample moments

$$\frac{1}{T}\sum_{t=1}^{T} y_t^2 = 45, \qquad\qquad \frac{1}{T-1}\sum_{t=1}^{T-1} y_t y_{t-1} = 18.$$

Use these sample moments and your result in (a) to calculate the method of moments estimators of θ and σ. There may be multiple solutions to this, calculate all of them.

(c) Are the parameters θ and σ uniquely identified from these two moments conditions? Provide a restriction on the support of θ so that the parameters of the model are uniquely identified. Without this restriction, is it possible to use the additional moment conditions provided by higher order autocovariances $\gamma_j = \mathbb{E}(y_t\, y_{t-j})$, $j = 2, 3, 4, \ldots$, to identify the parameters of the model uniquely?

(d) One now suspects that the true model is an AR(1) model

$$y_t = \rho y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim iid\,\mathcal{N}(0,\sigma^2)$$

with $|\rho| < 1$. Calculate $\gamma_2 = \mathbb{E}(y_t\, y_{t-2})$ as a function of $\gamma_0 = \mathbb{E}(y_t^2)$ in this case. Let $\hat{\beta}$ be the OLS estimator obtained from regressing y_t on y_{t-2}. What is the probability limit of $\hat{\beta}$ as $T \to \infty$? Explain how you can use this information to distinguish between the MA(1) and the AR(1) model in the observed sample? (no proof required)

SOLUTIONS

ECON0064: Econometrics

8-HOUR ONLINE EXAMINATION: SUMMER TERM 2021

Section A

Question A1 (20 points)

(a) For $x_i = (1, w_i)$ we have

$$\frac{1}{n}\sum_{i=1}^{n} x_i' x_i = \begin{pmatrix} 1 & \frac{1}{n}\sum_{i=1}^{n} w_i \\ \frac{1}{n}\sum_{i=1}^{n} w_i & \frac{1}{n}\sum_{i=1}^{n} w_i^2 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 3 & 10 \end{pmatrix},$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i' y_i = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} y_i \\ \frac{1}{n}\sum_{i=1}^{n} w_i y_i \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}.$$

We thus find $\det\left(\frac{1}{n}\sum_{i=1}^{n} x_i' x_i\right) = 1$ and

$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i' x_i\right)^{-1} = \begin{pmatrix} 10 & -3 \\ -3 & 1 \end{pmatrix},$$

$$\widehat{\beta} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i' x_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} x_i' y_i\right) = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

i.e. we have $\widehat{\beta}_1 = 1$ and $\widehat{\beta}_2 = 0$.

(b) We have $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \widehat{u}_i^2$, where $\widehat{u}_i = y_i - \widehat{\beta}_1 - w_i\widehat{\beta}_2$. Therefore

$$\begin{aligned}\widehat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{\beta}_1 - w_i\widehat{\beta}_2)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - 1)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(y_i^2 - 2y_i + 1) \\ &= 17 - 2\cdot 1 + 1 = 16.\end{aligned}$$

We found $\widehat{\sigma}^2 = 16$.

(Comment: an alternative way of calculating $\widehat{\sigma}^2$ is to use the formula $\frac{1}{n}\sum_{i=1}^n \widehat{u}_i^2 = \frac{1}{n}\sum_{i=1}^n y_i^2 - \frac{1}{n}\sum_{i=1}^n (\widehat{\beta}_1 + w_i\widehat{\beta}_2)^2$, which is true due to the FOC of the OLS problem: $\sum_{i=1}^n \widehat{u}_i = 0$ and $\sum_{i=1}^n \widehat{u}_i w_i = 0$.)

(c) We have

$$\widehat{\mathrm{Var}}(\widehat{\beta}) = \widehat{\sigma}^2 \left(\sum_{i=1}^n x_i' x_i\right)^{-1} = \frac{\widehat{\sigma}^2}{n}\left(\frac{1}{n}\sum_{i=1}^n x_i' x_i\right)^{-1}$$
$$= \frac{16}{100}\begin{pmatrix} 10 & -3 \\ -3 & 1 \end{pmatrix} = \begin{pmatrix} 8/5 & -12/25 \\ -12/25 & 4/25 \end{pmatrix}.$$

Thus, $\mathrm{se}(\widehat{\beta}_1) = \sqrt{8/5} \approx 1.26$, and $\mathrm{se}(\widehat{\beta}_2) = \sqrt{4/25} = 2/5 = 0.4$.

(d) $t = \frac{\widehat{\beta}_2 - 0}{\mathrm{se}(\widehat{\beta}_2)} = \frac{0.9}{0.5} = 1.8$. Since $|t| > 1.64$ we can reject H_0 at 90% confidence level.

(e) Let $\hat{\theta} = \hat{\beta}_1 + 2\hat{\beta}_2 = 3.0$. First, note that

$$\mathrm{Var}(\hat{\beta}_1 + 2\hat{\beta}_2) = \mathrm{Var}(\hat{\beta}_1) + 4\mathrm{Var}(\hat{\beta}_2) + 4\mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_2).$$

Hence,

$$\widehat{\mathrm{Var}}(\hat{\theta}) = \widehat{\mathrm{Var}}(\hat{\beta}_1 + 2\hat{\beta}_2) = 1.69 + 4 \cdot 0.25 + 4 \cdot (-0.37) = 1.21,$$

so $\mathrm{se}(\hat{\theta}) = \mathrm{se}(\hat{\beta}_1 + 2\hat{\beta}_2) = 1.1$. Hence, we can construct a 95% confidence interval for θ as

$$\left[\hat{\theta} - 1.96\mathrm{se}(\hat{\theta}), \hat{\theta} + 1.96\mathrm{se}(\hat{\theta})\right] = [3.0 - 1.96 \cdot 1.1, 3.0 + 1.96 \cdot 1.1]$$
$$= [0.844, 5.156]\,.$$

Question A2 (20 points)

(a) This is a classical linear regression with an exogenous regressor, so

$$\sqrt{n}(\hat{\beta}^* - \beta) \Rightarrow \mathcal{N}\left(0, \mathbb{E}[x_i^{*\prime} x_i^*]^{-1}\mathbb{E}[u_i^2 x_i^{*\prime} x_i^*]\mathbb{E}[x_i^{*\prime} x_i^*]^{-1}\right).$$

Here, we have $\mathbb{E}[x_i^{*\prime} x_i^*] = \mathbb{E}[x_i^{*2}] = 1$ and $\mathbb{E}[u_i^2 x_i^{*\prime} x_i^*] = \mathbb{E}[u_i^2]\mathbb{E}[x_i^{*2}] = 1$, so

$$\sqrt{n}(\hat{\beta} - \beta) \Rightarrow \mathcal{N}(0, 1).$$

(b) In this case,

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} = \frac{\frac{1}{n}\sum_{i=1}^n x_i y_i}{\frac{1}{n}\sum_{i=1}^n x_i^2} \to_p \frac{\mathbb{E}[x_i y_i]}{\mathbb{E}[x_i^2]}.$$

Here $\mathbb{E}[x_i y_i] = \mathbb{E}[(x_i^* + \varepsilon_i)(x_i^*\beta + u_i)] = \mathbb{E}[x_i^{*2}]\beta = \beta$, and $\mathbb{E}[x_i^2] = \mathbb{E}[(x_i^* + \varepsilon_i)^2] = \mathbb{E}[x_i^{*2}] + \mathbb{E}[\varepsilon_i^2] = 2$, so we conclude

$$\hat{\beta} \to_p \frac{\beta}{2},$$

so $\hat{\beta}$ is not consistent for β.

(c) In this case,

$$\begin{aligned}
\hat{\beta}_{\text{IV}} &= \frac{\sum_{i=1}^n z_i y_i}{\sum_{i=1}^n z_i x_i} \\
&= \frac{\sum_{i=1}^n z_i(x_i\beta + u_i - \varepsilon_i\beta)}{\sum_{i=1}^n z_i x_i} \\
&= \beta + \frac{\sum_{i=1}^n z_i(u_i - \varepsilon_i\beta)}{\sum_{i=1}^n z_i x_i} \\
&= \beta + \frac{\frac{1}{n}\sum_{i=1}^n z_i(u_i - \varepsilon_i\beta)}{\frac{1}{n}\sum_{i=1}^n z_i x_i},
\end{aligned}$$

where the second equality uses $y_i = x_i^*\beta + u_i = x_i\beta + u_i - \varepsilon_i\beta$. By the LLN,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n z_i(u_i - \varepsilon_i\beta) \to_p \mathbb{E}[z_i(u_i - \varepsilon_i\beta)]& \\
&= \mathbb{E}[(x_i^*\rho + \eta_i)(u_i - \varepsilon_i\beta)] = 0,
\end{aligned}$$

where the last equality follows from the mutual independence of $x_i^*, u_i, \varepsilon_i$ and η_i. Similarly, by the LLN, the denominator also converges in probability to a non-stochastic and non-zero limit:

$$\frac{1}{n}\sum_{i=1}^n z_i x_i \to_p \mathbb{E}[z_i x_i] = \mathbb{E}[(x_i^*\rho + \varepsilon_i)(x_i^* + \eta_i)] = \mathbb{E}[x_i^{*2}]\rho = \rho \neq 0.$$

Hence,

$$\hat{\beta} = \beta + \frac{\frac{1}{n}\sum_{i=1}^n z_i(u_i - \varepsilon_i\beta)}{\frac{1}{n}\sum_{i=1}^n z_i x_i} \to_p \beta + \frac{0}{\rho} = \beta,$$

so $\hat{\beta}_{\text{IV}}$ is consistent for β.

(d) Using the previously derived representation, we obtain

$$\sqrt{n}(\hat{\beta}_{\text{IV}} - \beta) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1} z_i(u_i - \varepsilon_i \beta)}{\frac{1}{n} \sum_{i=1}^n z_i x_i}.$$

We have established that $\frac{1}{n} \sum_{i=1}^n z_i x_i \to_p \rho \neq 0$. Since we have also verified that $\mathbb{E}[z_i(u_i - \varepsilon_i \beta)] = 0$, by the CLT, we have

$$\frac{1}{\sqrt{n}} \sum_{i=1} z_i(u_i - \varepsilon_i \beta) \Rightarrow \mathcal{N}\left(0, \operatorname{Var}[z_i(u_i - \varepsilon_i \beta)]\right).$$

Finally,

$$\begin{aligned}
\operatorname{Var}[z_i(u_i - \varepsilon_i \beta)] &= \mathbb{E}[z_i^2(u_i - \varepsilon_i \beta)^2] \\
&= \mathbb{E}[(\rho x_i^* + \eta_i)^2 (u_i - \varepsilon_i \beta)^2] \\
&= \mathbb{E}[(\rho x_i^* + \eta_i)^2] \mathbb{E}[(u_i - \varepsilon_i \beta)^2] \\
&= \left(\rho^2 \mathbb{E}[x_i^{*2}] + \mathbb{E}[\eta_i^2]\right) \left(\mathbb{E}[u_i^2] + \beta^2 \mathbb{E}[\varepsilon_i^2]\right) \\
&= (1 + \rho^2)(1 + \beta^2).
\end{aligned}$$

Hence, applying the Slutsky's theorem, we conclude

$$\begin{aligned}
\sqrt{n}(\hat{\beta}_{\text{IV}} - \beta) &\Rightarrow \frac{\mathcal{N}\left(0, (1 + \rho^2)(1 + \beta^2)\right)}{\rho} \\
&= \mathcal{N}\left(0, \frac{(1 + \rho^2)(1 + \beta^2)}{\rho^2}\right).
\end{aligned}$$

(e) In this case $\mathbb{E}[z_i x_i] = \rho = 0$, so the previously derived result no longer applies. The instrument is no longer relevant and $\hat{\beta}_{\text{IV}}$ is no longer consistent. To derive the asymptotic distribution of $\hat{\beta}_{\text{IV}}$ in this case, note that

$$\begin{aligned}
\hat{\beta}_{\text{IV}} &= \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n z_i y_i}{\frac{1}{\sqrt{n}} \sum_{i=1}^n z_i x_i} \\
&= \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n \eta_i y_i}{\frac{1}{\sqrt{n}} \sum_{i=1}^n \eta_i x_i},
\end{aligned}$$

where $\mathbb{E}[\eta_i y_i] = 0$ and $\mathbb{E}[\eta_i x_i] = 0$. Then, by the CLT,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \begin{pmatrix} \eta_i y_i \\ \eta_i x_i \end{pmatrix} \Rightarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \operatorname{Var}(\eta_i y_i) & \operatorname{Cov}(\eta_i y_i, \eta_i x_i) \\ \operatorname{Cov}(\eta_i y_i, \eta_i x_i) & \operatorname{Var}(\eta_i x_i) \end{pmatrix}\right).$$

By direct inspection, $\mathrm{Var}(\eta_i y_i) = 1 + \beta^2$, $\mathrm{Cov}(\eta_i y_i, \eta_i x_i) = \beta$, $\mathrm{Var}(\eta_i x_i) = 2$, so

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} \eta_i y_i \\ \eta_i x_i \end{pmatrix} \Rightarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1+\beta^2 & \beta \\ \beta & 2 \end{pmatrix} \right).$$

Hence, we conclude that $\hat{\beta}_{\mathrm{IV}}$ converges in distribution to a ratio of two correlated zero-mean normals, i.e., $\hat{\beta}_{\mathrm{IV}} \Rightarrow \frac{\mathcal{Z}_1}{\mathcal{Z}_2}$, where

$$\begin{pmatrix} \mathcal{Z}_1 \\ \mathcal{Z}_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1+\beta^2 & \beta \\ \beta & 2 \end{pmatrix} \right).$$

Since $\mathcal{Z}_1$ can be represented as $\mathcal{Z}_1 = \frac{\beta}{2}\mathcal{Z}_2 + \mathcal{Z}_1^{\perp}$, where

$$\begin{pmatrix} \mathcal{Z}_1^{\perp} \\ \mathcal{Z}_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1+\frac{\beta^2}{2} & 0 \\ 0 & 2 \end{pmatrix} \right),$$

the asymptotic distribution of $\hat{\beta}_{\mathrm{IV}}$ can be simplified as

$$\hat{\beta}_{\mathrm{IV}} \Rightarrow \frac{\beta}{2} + \frac{\mathcal{Z}_1^{\perp}}{\mathcal{Z}_2},$$

or finally

$$\hat{\beta}_{\mathrm{IV}} \Rightarrow \frac{\beta}{2} + \frac{\sqrt{2+\beta^2}}{2} \frac{Z_1}{Z_2},$$

where Z_1 and Z_2 are independent standard normals $\mathcal{N}(0,1)$.

Question A3 (20 points)

(a) By the LLN (across i),

$$\frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{2} x_{ij}^2 \right) \to_p \mathbb{E}[x_{i1}^2 + x_{i2}^2] = 2.$$

Similarly,

$$\frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{2} x_{ij} y_{ij} \right) \to_p \mathbb{E}[x_{i1} y_{i1} + x_{i2} y_{i2}] = 2\beta,$$

where we used $\mathbb{E}[x_{i1} y_{i1}] = \mathbb{E}[x_{i2} y_{i2}] = \mathbb{E}[x_{i1}(x_{i1}\beta + u_{i1})] = \mathbb{E}[x_{i1}^2]\beta = \beta$.

(b) By the CMT,

$$\hat{\beta} = \frac{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{2} x_{ij}y_{ij}}{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{2} x_{ij}^2} \to_p \frac{\mathbb{E}[x_{i1}y_{i1} + x_{i2}y_{i2}]}{\mathbb{E}[x_{i1}^2 + x_{i2}^2]} = \frac{2\beta}{2} = \beta,$$

so $\hat{\beta}$ is consistent for β.

(c) First, note that $\mathbb{E}[x_{i1}u_{i1} + x_{i2}u_{i2}] = 0$, so applying the CLT (across i), we conclude

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\sum_{j=1}^{2} x_{ij}u_{ij}\right) \Rightarrow \mathcal{N}\left(0, \mathrm{Var}(x_{i1}u_{i1} + x_{i2}u_{i2})\right).$$

So, $\Omega = \mathrm{Var}(x_{i1}u_{i1} + x_{i2}u_{i2}) = \mathrm{Var}(x_{i1}u_{i1}) + \mathrm{Var}(x_{i2}u_{i2}) + 2\mathrm{Cov}(x_{i1}u_{i1}, x_{i2}u_{i2}) = 2(1+\kappa\rho)$.

(d) As usual,

$$\hat{\beta} = \beta + \frac{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{2} x_{ij}u_{ij}}{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{2} x_{ij}^2}.$$

Hence, apply the Slutsky's theorem, we conclude

$$\sqrt{n}(\hat{\beta} - \beta) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{2} x_{ij}u_{ij}}{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{2} x_{ij}^2} \Rightarrow \frac{\mathcal{N}(0, 2(1+\kappa\rho))}{2} = \mathcal{N}\left(0, \frac{1+\kappa\rho}{2}\right).$$

(e) Let $u = (u_{11}, u_{12}, u_{21}, u_{22}, \ldots, u_{n1}, u_{n2})'$ denote a $2n \times 1$ vector of the regression errors. The OLS estimator is inefficient because $\Omega = \mathrm{Var}(u|X) = \mathrm{Var}(u) = \mathbb{E}[uu']$ is not a diagonal matrix. Specifically, Ω is a $2n \times 2n$ block diagonal matrix

$$\Omega = \begin{bmatrix} \Omega_u & 0_{2\times 2} & \cdots & 0_{2\times 2} \\ 0_{2\times 2} & \Omega_u & \cdots & 0_{2\times 2} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{2\times 2} & 0_{2\times 2} & \cdots & \Omega_u \end{bmatrix},$$

where

$$\Omega_u = \mathrm{Var}\left[\begin{pmatrix} u_{i1} \\ u_{i2} \end{pmatrix}\right] = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Then Ω^{-1} is a also a block diagonal matrix of the form

$$\Omega^{-1} = \begin{bmatrix} \Omega_u^{-1} & 0_{2\times 2} & \cdots & 0_{2\times 2} \\ 0_{2\times 2} & \Omega_u^{-1} & \cdots & 0_{2\times 2} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{2\times 2} & 0_{2\times 2} & \cdots & \Omega_u^{-1} \end{bmatrix},$$

where

$$\Omega_u^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}.$$

The the GLS estimator takes the form

$$\hat{\beta}_{\text{GLS}} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y = \beta + \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}u,$$

where $X = (x_{11}, x_{12}, x_{21}, x_{22}, \ldots, x_{n1}, x_{n2})'$. As usual,

$$\sqrt{n}(\hat{\beta}_{\text{GLS}} - \beta) = \left(\frac{1}{n}X'\Omega^{-1}X\right)^{-1} \frac{1}{\sqrt{n}}X'\Omega^{-1}u.$$

Since Ω_u^{-1} is block-diagonal, we have

$$\begin{aligned}
\frac{1}{n}X'\Omega^{-1}X &= \frac{1}{n}\sum_{i=1}^{n}(x_{i1}\ x_{i2})\Omega_u^{-1}\begin{pmatrix} x_{i1} \\ x_{i2} \end{pmatrix} \\
&\to_p \mathbb{E}\left[(x_{i1}\ x_{i2})\Omega_u^{-1}\begin{pmatrix} x_{i1} \\ x_{i2} \end{pmatrix}\right] \\
&= \frac{\mathbb{E}[x_{i1}^2 - 2\rho x_{i1}x_{i2} + x_{i2}^2]}{1-\rho^2} \\
&= \frac{2(1-\kappa\rho)}{1-\rho^2},
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{\sqrt{n}}X'\Omega^{-1}u &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(x_{i1}\ x_{i2})\Omega_u^{-1}\begin{pmatrix} u_{i1} \\ u_{i2} \end{pmatrix} \\
&\Rightarrow N\left(0, \operatorname{Var}\left[(x_{i1}\ x_{i2})\Omega_u^{-1}\begin{pmatrix} u_{i1} \\ u_{i2} \end{pmatrix}\right]\right),
\end{aligned}$$

where

$$\begin{aligned}
\operatorname{Var}\left[(x_{i1}\ x_{i2})\Omega_u^{-1}\begin{pmatrix} u_{i1} \\ u_{i2} \end{pmatrix}\right] &= \mathbb{E}\left[(x_{i1}\ x_{i2})\Omega_u^{-1}\begin{pmatrix} u_{i1} \\ u_{i2} \end{pmatrix}(u_{i1}\ u_{i2})\Omega_u^{-1}\begin{pmatrix} x_{i1} \\ x_{i2} \end{pmatrix}\right] \\
&= \mathbb{E}\left[(x_{i1}\ x_{i2})\Omega_u^{-1}\begin{pmatrix} x_{i1} \\ x_{i2} \end{pmatrix}\right] \\
&= \frac{2(1-\kappa\rho)}{1-\rho^2}.
\end{aligned}$$

Hence, we conclude

$$\sqrt{n}(\hat{\beta}_{\text{GLS}} - \beta) \Rightarrow \left(\frac{2(1-\kappa\rho)}{1-\rho^2}\right)^{-1} \mathcal{N}\left(0, \frac{2(1-\kappa\rho)}{1-\rho^2}\right) = \mathcal{N}\left(0, \frac{1-\rho^2}{2(1-\kappa\rho)}\right).$$

If ρ is unknown, it could be consistently estimated by $\hat{\rho} = \frac{1}{n}\sum_{i=1}^{n} \hat{u}_{i1}\hat{u}_{i2}$, where $\hat{u}_{ij} = y_{ij} - x_{ij}\hat{\beta}$ is the OLS residual. In this case, one could employ the FGLS estimator using $\hat{\rho}$ instead of unknown ρ. It would be as asymptotically efficient as the (unfeasible) GLS estimator we discussed first.

Section B

Answer **TWO questions from this Section**.

Question B1 (20 points)

(a) Note that

$$\mathbb{E}[y_i|x_i] = 1 \times P(y_i = 1|x_i) + 0 \times P(y_i = 0|x_i) = x_i\beta.$$

Then $\mathbb{E}[u_i|x_i] = \mathbb{E}[y_i - x_i\beta|x_i] = \mathbb{E}[y_i|x_i] - x_i\beta = 0$, as required.

Also, conditional on x_i, y_i has a Bernoulli distribution with a parameter $p_x = x_i\beta$. Hence, $\text{Var}(u_i|x_i) = \text{Var}(x_i\beta + u_i|x_i) = \text{Var}(y_i|x_i) = p_x(1-p_x) = x_i\beta(1-x_i\beta)$.

(b) Since the standard linear regression assumptions are satisfied, we have

$$\sqrt{n}(\hat{\beta}_{\text{OLS}} - \beta) \Rightarrow \mathcal{N}\left(0, \mathbb{E}[x_i'x_i]^{-1}\mathbb{E}[u_i^2 x_i'x_i]\mathbb{E}[x_i'x_i]^{-1}\right).$$

Using the result of part (a), $\mathbb{E}[u_i^2 x_i'x_i] = \mathbb{E}[\mathbb{E}[u_i^2|x_i]x_i'x_i] = \mathbb{E}[x_i\beta(1-x_i\beta)x_i'x_i]$, so

$$\Sigma_{\text{OLS}} = \mathbb{E}[x_i'x_i]^{-1}\mathbb{E}[x_i\beta(1-x_i\beta)x_i'x_i]\mathbb{E}[x_i'x_i]^{-1}.$$

(c) Since $f(y_i|x_i;b) = (x_i b)^{y_i}(1-x_i b)^{1-y_i}$, we have

$$\begin{aligned}\log f(y|x,b) &= \sum_{i=1}^{n} \log f(y_i|x_i;b) \\ &= \sum_{i=1}^{n} \left\{y_i \log(x_i b) + (1-y_i)\log(1-x_i b)\right\}.\end{aligned}$$

(d) First,

$$\begin{aligned} s(y_i|x_i, b) &= \frac{\partial}{\partial b}\left\{y_i \log(x_i b) + (1-y_i)\log(1-x_i b)\right\} \\ &= \frac{y_i x_i'}{x_i b} - \frac{(1-y_i)x_i'}{1-x_i b} \\ &= \frac{x_i'(y_i - x_i b)}{x_i b(1-x_i b)}. \end{aligned}$$

Note that

$$s(y_i|x_i, \beta) = \frac{x_i' u_i}{x_i\beta(1-x_i\beta)},$$

so

$$\begin{aligned} I(\beta) &= \mathbb{E}[s(y_i|x_i,\beta)s(y_i|x_i,\beta)'] \\ &= \mathbb{E}\left[\frac{x_i' x_i u_i^2}{(x_i\beta)^2(1-x_i\beta)^2}\right] \\ &= \mathbb{E}\left[\frac{x_i' x_i \mathbb{E}[u_i^2|x_i]}{(x_i\beta)^2(1-x_i\beta)^2}\right] \\ &= \mathbb{E}\left[\frac{x_i' x_i}{x_i\beta(1-x_i\beta)}\right], \end{aligned}$$

where we used $\mathbb{E}[u_i^2|x_i] = \text{Var}[u_i|x_i] = x_i\beta(1-x_i\beta)$. Finally,

$$\sqrt{n}(\hat{\beta}_{\text{MLE}} - \beta) \Rightarrow \mathcal{N}(0, \Sigma_{\text{MLE}}),$$

where

$$\Sigma_{\text{MLE}} = I(\beta)^{-1} = \left\{\mathbb{E}\left[\frac{x_i' x_i}{x_i\beta(1-x_i\beta)}\right]\right\}^{-1}$$

The MLE estimator is preferred since it is more efficient, i.e. $\Sigma_{\text{MLE}} < \Sigma_{\text{OLS}}$.

(e) The OLS estimator is not efficient because the model is clearly not homoskedastic since $h(x_i) = \text{Var}[u_i|x_i] = x_i\beta(1-x_i\beta)$. As an alternative to OLS, one could use the FWLS estimator instead defined as

$$\hat{\beta}_{\text{FWLS}} = \arg\min_b \frac{1}{n}\sum_{i=1}^{n} \frac{(y_i - x_i b)^2}{x_i\hat{\beta}_{\text{OLS}}(1-x_i\hat{\beta}_{\text{OLS}})}.$$

Since $\hat{\beta}_{\text{OLS}} \to \beta$, the asymptotic distribution of the FWLS estimator is the same as the asymptotic distribution of the (infeasible) WLS estimator, so we have

$$\hat{\beta}_{\text{FWLS}} \Rightarrow \mathcal{N}\left(0, \Sigma_{\text{FWLS}}\right),$$

where

$$\Sigma_{\text{FWLS}} = \left\{ \mathbb{E}\left[\frac{x_i' x_i}{h(x_i)} \right] \right\}^{-1} = \left\{ \mathbb{E}\left[\frac{x_i' x_i}{x_i\beta(1-x_i\beta)} \right] \right\}^{-1} = \Sigma_{\text{MLE}}.$$

(f) It's not a good idea to use $\tilde{\beta}$ since this estimator is not consistent. Specifically, the population criterion function is clearly not minimized at β because the population FOC is not satisfied at β, i.e. one could verify that

$$\frac{\partial}{\partial b}\tilde{Q}(\beta) \neq 0,$$

where

$$\tilde{Q}(b) = \mathbb{E}\left[\frac{(y_i - x_i b)^2}{x_i b(1-x_i b)} \right].$$

To see that note

$$\begin{aligned} \frac{\partial}{\partial b}\tilde{Q}(\beta) &= \mathbb{E}\left[\frac{-2x_i'(y_i - x_i\beta)}{x_i\beta(1-x_i\beta)} + (y_i - x_i\beta)^2 \frac{x_i'(2x_i\beta - 1)}{(x_i\beta)^2(1-x_i\beta)^2} \right] \\ &= \mathbb{E}\left[\frac{x_i'(2x_i\beta - 1)}{x_i\beta(1-x_i\beta)} \right] \neq 0, \end{aligned}$$

where we used $\mathbb{E}[y_i - x_i\beta | x_i] = \mathbb{E}[u_i|x_i] = 0$ and $\mathbb{E}[(y_i - x_i\beta)^2|x_i] = \mathbb{E}[u_i^2|x_i] = x_i\beta(1-x_i\beta)$.

Question B2 (20 points)

(a) We have $f(x,\theta) = \theta^2 + \theta x$, $f_\theta(x,\theta) = 2\theta + x$, and

$$\sqrt{n}(\hat{\theta}_{\text{NLLS}} - \theta_0) \Rightarrow N(0, \Sigma_{\text{NLLS}}),$$

where

$$\Sigma_{\text{NLLS}} = \mathbb{E}\left[f_\theta(x_i,\theta_0) f_\theta(x_i,\theta_0)'\right]^{-1} \text{Var}\left[u_i f_\theta(x_i,\theta_0)\right] \mathbb{E}\left[f_\theta(x_i,\theta_0) f_\theta(x_i,\theta_0)'\right]^{-1}.$$

Here, $\mathbb{E}\left[f_\theta(x_i,\theta_0) f_\theta(x_i,\theta_0)'\right] = \mathbb{E}[(2\theta_0 + x_i)^2] = 4\theta_0^2 + \sigma_x^2$, and

$$\begin{aligned} \text{Var}\left[u_i f_\theta(x_i,\theta_0)\right] &= \mathbb{E}\left[u_i^2 f_\theta(x_i,\theta_0)^2\right] \\ &= \mathbb{E}[u_i^2(4\theta_0^2 + 4\theta_0 x_i + x_i^2)] \\ &= 4\sigma_u^2\theta_0^2 + \gamma\sigma_u^2\sigma_x^2. \end{aligned}$$

Hence, we conclude

$$\Sigma_{\text{NLLS}} = \frac{\sigma_u^2(4\theta_0^2 + \gamma\sigma_x^2)}{(4\theta_0^2 + \sigma_x^2)^2}.$$

(b) Note that Σ_{NLLS} is still well defined even if we have no variation in x_i, i.e., when $\sigma_x = 0$. In this case, θ_0^2 (and hence also θ_0) is identified as an intercept.

(c) First,

$$\frac{\partial}{\partial\theta} g(y_i, x_i, \theta) = (-2\theta - x_i) \begin{pmatrix} 1 \\ x_i \end{pmatrix} = \begin{pmatrix} -2\theta - x_i \\ -2\theta x_i - x_i^2 \end{pmatrix}.$$

Hence,

$$G = \mathbb{E}\left[\frac{\partial}{\partial\theta} g(y_i, x_i, \theta_0)\right] = \mathbb{E}\left[\begin{pmatrix} -2\theta_0 - x_i \\ -2\theta_0 x_i - x_i^2 \end{pmatrix}\right] = \begin{pmatrix} -2\theta_0 \\ -\sigma_x^2 \end{pmatrix}.$$

Also note that $g(y_i, x_i, \theta_0) = u_i \begin{pmatrix} 1 \\ x_i \end{pmatrix}$, so

$$\Omega = \mathbb{E}\left[g(y_i, x_i, \theta_0) g(y_i, x_i, \theta_0)'\right] = \mathbb{E}\left[\begin{pmatrix} u_i^2 & u_i x_i \\ u_i x_i & u_i^2 x_i^2 \end{pmatrix}\right] = \sigma_u^2 \begin{pmatrix} 1 & 0 \\ 0 & \gamma\sigma_x^2 \end{pmatrix}.$$

Finally, we conclude

$$\sqrt{n}(\hat{\theta}_I - \theta_0) \Rightarrow \mathcal{N}(0, \Sigma_I),$$

$$\Sigma_I = (G'G)^{-1} G'\Omega G (G'G)^{-1} = \frac{\sigma_u^2 (4\theta_0^2 + \gamma\sigma_x^6)}{(4\theta_0^2 + \sigma_x^4)^2}.$$

(d) In this case,

$$W_{opt} = \Omega^{-1} = \sigma_u^{-2} \begin{pmatrix} 1 & 0 \\ 0 & (\gamma\sigma_x^2)^{-1} \end{pmatrix},$$

and

$$\sqrt{n}(\hat{\theta}_{opt} - \theta_0) \Rightarrow N(0, \Sigma_{opt}),$$

where

$$\Sigma_{opt} = (G'\Omega^{-1} G)^{-1} = \frac{\gamma\sigma_u^2}{4\gamma\theta_0^2 + \sigma_x^2}.$$

(e) First, note that the optimal weighting matrix is proportional to

$$\begin{pmatrix} 1 & 0 \\ 0 & (\gamma\sigma_X^2)^{-1} \end{pmatrix},$$

which can be consistently estimated based on the sample variance of x_i. Hence, one can achieve asymptotic efficiency by using

$$\hat{W} = \begin{pmatrix} 1 & 0 \\ 0 & (\gamma\hat{\sigma}_X^2)^{-1} \end{pmatrix},$$

and the first step is not needed.

Second, note that the sample average of the moment function $\frac{1}{n}\sum_{i=1}^{n} g(y_i, x_i, \theta)$ is a quadratic function of θ. Hence, the criterion function

$$\hat{Q}_n(\theta) = -\left[\frac{1}{n}\sum_{i=1}^{n} g(y_i, x_i, \theta)\right]' \hat{W} \left[\frac{1}{n}\sum_{i=1}^{n} g(y_i, x_i, \theta)\right]'$$

is a polynomial of degree 4 with known coefficients, which are determined by the sample moments of (x_i, y_i) and γ only. The criterion function expressed in that form is easy to evaluate (you need to compute the coefficients only once!) and optimize.

Question B3 (20 points)

(a) First, note that since processes x_t and u_t are independent,

$$\gamma_j = \mathrm{Cov}(g_t, g_{t-j}) = \mathbb{E}[g_t g_{t-j}] = \mathbb{E}[x_t u_t x_{t-j} u_{t-j}] = \mathbb{E}[x_t x_{t-j}]\mathbb{E}[u_t u_{t-j}] = \gamma_j^x \gamma_j^u.$$

Since x_t is an MA(1) process, $\gamma_0^x = (1+\theta^2)\sigma_e^2$, $\gamma_1^x = \gamma_{-1}^x = \theta\sigma_e^2$ and $\gamma_j^x = 0$ for $|j| > 1$. Since u_t is an AR(1) process, $\gamma_j^u = \rho^{|j|}\frac{\sigma_\varepsilon^2}{1-\rho^2}$. Hence, we conclude that

$$\gamma_0 = \frac{(1+\theta^2)\sigma_\varepsilon^2\sigma_e^2}{1-\rho^2}, \quad \gamma_{-1} = \gamma_1 = \frac{\rho\theta\sigma_\varepsilon^2\sigma_e^2}{1-\rho^2},$$

and $\gamma_j = 0$ for $|j| > 1$.

(b) Note that since $\gamma_j = 0$ for $|j| > 1$, the g_t has an MA(1) representation and then use the CLT for MA processes. Hence,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t \Rightarrow \mathcal{N}\left(0, \sum_{j=-\infty}^{+\infty} \gamma_j\right).$$

Hence, here

$$\Omega = \gamma_{-1} + \gamma_0 + \gamma_1 = (1+\theta^2+2\rho\theta)\frac{\sigma_\varepsilon^2\sigma_e^2}{1-\rho^2}.$$

(c) By the usual representation,

$$\sqrt{T}(\hat{\beta} - \beta) = \frac{\frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t u_t}{\frac{1}{T} \sum_{t=1}^{T} x_t^2}.$$

Then, by the ergodic LLN,

$$\frac{1}{T} \sum_{t=1}^{T} x_t^2 \to_p \mathbb{E}[x_t^2] = \gamma_0^x = (1 + \theta^2)\sigma_e^2.$$

Moreover, we have established

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t u_t \Rightarrow \mathcal{N}(0, \Omega).$$

By applying the Slutsky's theorem,

$$\sqrt{T}(\hat{\beta} - \beta) \Rightarrow (0, \Sigma),$$

where

$$\Sigma = \frac{\Omega}{(\mathbb{E}[x_t^2])^2} = (1 + \theta^2 + 2\theta\rho)\frac{\sigma_\varepsilon^2}{(1 + \theta^2)^2(1 - \rho^2)\sigma_e^2}.$$

(d) Recall that $\Omega = \gamma_0^x \gamma_0^u + 2\gamma_1^x \gamma_2^u$, so

$$\Sigma = \frac{\gamma_0^x \gamma_0^u + 2\gamma_1^x \gamma_1^u}{(\gamma_0^x)^2}$$

so, Σ can be consistently estimated by

$$\hat{\Sigma} = \frac{\hat{\gamma}_0^x \hat{\gamma}_0^u + 2\hat{\gamma}_1^x \hat{\gamma}_1^u}{(\hat{\gamma}_0^x)^2} = \frac{5 \cdot 4 + 2 \cdot 2.5 \cdot 3.6}{5^2} = \frac{38}{25} = 1.52.$$

Then, we can construct a 95% confidence interval as

$$\left\{ \hat{\beta} - 1.96\sqrt{\frac{\hat{\Sigma}}{T}}, \hat{\beta} + 1.96\sqrt{\frac{\hat{\Sigma}}{T}} \right\} \approx [1.858, 2.342].$$

(e) In this case, $\gamma_j^x = \phi^{|j|}\frac{\sigma_e^2}{1-\phi^2}$, and

$$\Omega = \sum_{j=-\infty}^{\infty} \gamma_j^x \gamma_j^u = \sum_{j=-\infty}^{\infty} (\phi\rho)^{|j|} \frac{\sigma_e^2 \sigma_\varepsilon^2}{(1 - \phi^2)(1 - \rho^2)}.$$

Note that

$$\sum_{j=-\infty}^{\infty} (\phi\rho)^{|j|} = \sum_{j=0}^{\infty} (\phi\rho)^{|j|} + \sum_{j=1}^{\infty} (\phi\rho)^{|j|}$$
$$= (1+\phi\rho) \sum_{j=0}^{\infty} (\phi\rho)^{|j|}$$
$$= \frac{1+\phi\rho}{1-\phi\rho}.$$

Hence,

$$\Omega = \frac{(1+\phi\rho)\sigma_e^2\sigma_\varepsilon^2}{(1-\phi\rho)(1-\phi^2)(1-\rho^2)}.$$

By the ergodic LLN,

$$\frac{1}{T}\sum_{t=1}^{T} x_t^2 \to_p \mathbb{E}[x_t^2] = \gamma_0^x = \frac{\sigma_e^2}{1-\phi^2}.$$

Then,

$$\sqrt{T}(\hat{\beta}-\beta) \Rightarrow \mathcal{N}(0,\Sigma),$$

where

$$\Sigma = \frac{\Omega}{(\mathbb{E}[x_t^2])^2} = \frac{(1+\phi\rho)(1-\phi^2)\sigma_\varepsilon^2}{(1-\phi\rho)(1-\rho^2)\sigma_e^2}.$$

G020: Key Solutions for Econometrics Exam

2011-2012

Question 1

(a) No, incorrect. For perfectly collinear regressors the OLS estimator is not well-defined and thus also not consistent.

(b) Yes, correct.

(c) No, incorrect. Heteroscedasticity robust standard errors also work under homoscedasticity.

(d) No, incorrect. The bias can converge to zero as the sample size becomes large, so that the estimator becomes consistent. This is, for example, the case for 2SLS.

(e) Yes, correct.

(f) No, incorrect. IV needs to be exogenous and relevant, i.e. correlated with the endogenous regressor(s).

(g) No, incorrect. Heteroscedasticity is not a problem for 2SLS.

(h) Yes, correct.

(i) No, incorrect. In the overidentified case the choice of weight matrix matters.

(j) Yes, correct.

Question 2

(a) $X'X = \begin{pmatrix} n_f & 0 \\ 0 & n_m \end{pmatrix} = \begin{pmatrix} 25 & 0 \\ 0 & 100 \end{pmatrix},$

$$(X'X)^{-1} = \begin{pmatrix} 1/25 & 0 \\ 0 & 1/100 \end{pmatrix},$$

$$X'h = \begin{pmatrix} 160\, n_f \\ 169\, n_m \end{pmatrix} = \begin{pmatrix} 4,000 \\ 16,900 \end{pmatrix}.$$

(b) $\hat{\beta} = (X'X)^{-1} X'h = \begin{pmatrix} 160 \\ 169 \end{pmatrix}$,

$\widehat{\text{Var}}(\hat{\beta}) = \sigma_u^2 (X'X)^{-1} = \begin{pmatrix} 20 & 0 \\ 0 & 5 \end{pmatrix}$.

(c) We need $\gamma_1 + \gamma_2 f_i = f_i \beta_1 + m_i \beta_2$. Using $f_i + m_i = 1$ we find $\gamma_1 + \gamma_2 f_i = \beta_2 + f_i(\beta_1 - \beta_2)$. Therefore $\gamma_1 = \beta_2$ and $\gamma_2 = \beta_1 - \beta_2$.

(d) $\hat{\gamma}_2 = \hat{\beta}_1 - \hat{\beta}_2 = -9$,
$\widehat{\text{std}}(\hat{\gamma}_2) = \sqrt{\widehat{\text{Var}}(\hat{\beta}_1) + \widehat{\text{Var}}(\hat{\beta}_2)} = 5$.

(e) $t = \hat{\gamma}_2 / \widehat{\text{std}}(\hat{\gamma}_2) = -18/10 = -9/5 = -1.8$.
Since $|t| < 1.96$ we cannot reject H_0 at 95% confidence level.

(f) $t = -1.8$. Since $t < -1.64$ we can reject H_0 at 95% confidence level.

Question 3

(a) x_i exogenous means $\rho_u = 0$.
$\sqrt{n}(\hat{\beta}_{\text{OLS}} - \beta) = \left(\frac{1}{n}\sum_{i=1}^n x_i^2\right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i u_i \;\Rightarrow\; [\mathbb{E}(x_i^2)]^{-1} \mathcal{N}[0, \mathbb{E}(x_i^2 u_i^2)] = \mathcal{N}(0, \Sigma_{\text{OLS}})$,
where $\Sigma_{\text{OLS}} = \frac{\mathbb{E}(u_i^2)}{\mathbb{E}(x_i^2)} = 1$.

Here we used the WLLN, CLT and Slutsky's theorem.

(b) z_i relevant means $\rho_z \neq 0$.
$\sqrt{n}(\hat{\beta}_{\text{IV}} - \beta) = \left(\frac{1}{n}\sum_{i=1}^n z_i x_i\right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n z_i u_i \;\Rightarrow\; [\mathbb{E}(z_i x_i)]^{-1} \mathcal{N}[0, \mathbb{E}(z_i^2 u_i^2)] = \mathcal{N}(0, \Sigma_{\text{IV}})$,
where $\Sigma_{\text{IV}} = \frac{\mathbb{E}(u_i^2)\mathbb{E}(z_i^2)}{[\mathbb{E}(z_i x_i)]^2} = \frac{1}{\rho_z^2}$.

Here we used the WLLN, CLT and Slutsky's theorem.

(c) Since $|\rho_z| < 1$ we have $\Sigma_{\text{IV}} > \Sigma_{\text{OLS}}$, i.e. if x_i is exogenous one should use the OLS estimator.

Section B

Question 4

(a) Sample analog: $\frac{1}{n}\sum_{i=1}^n y_i = \hat{\theta}$. Thus $\hat{\theta} = 20/100 = 0.2$.

(b) Applying the CLT we find $\sqrt{n}(\hat{\theta} - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (y_i - \theta) \Rightarrow \mathcal{N}(0, \mathbb{E}[(y_i - \theta)^2])$. According to the hint we have $\mathbb{E}[(y_i - \theta)^2] = \theta(1 - \theta)$.

(c) $\widehat{\text{std}}(\hat{\theta}) = \frac{1}{\sqrt{n}} \sqrt{\hat{\theta}(1 - \hat{\theta})} = \frac{1}{10} \sqrt{0.2 \times 0.8} = \frac{1}{25} = 0.04$.

(d) $\hat{\beta} = \hat{\theta}^2 = 0.2^2 = 0.04$.
Since $\frac{d}{d\theta}(\theta^2) = 2\theta$ we have $\widehat{\text{std}}(\hat{\beta}) = (2\hat{\theta})\widehat{\text{std}}(\hat{\theta}) = 0.016$.

Question 5

(a) $Q_n(\lambda) = \frac{1}{n} \log \prod_{i=1}^{n} f(y_i|\lambda) = \frac{1}{n} \sum_{i=1}^{n} [\log \lambda - \lambda y_i]$.
FOC: $0 = \frac{1}{n} \sum_{i=1}^{n} [\hat{\lambda}^{-1} - y_i]$.
Thus $\hat{\lambda} = \left[\frac{1}{n} \sum_{i=1}^{n} y_i\right]^{-1}$.

(b) $\hat{\lambda} = \frac{1}{2}$.

(c) $\log f(y_i|\lambda) = \log \lambda - \lambda y_i$,
$\frac{\partial^2 \log f(y_i|\lambda)}{\partial \lambda^2} = -\frac{1}{\lambda^2}$,
$H(\lambda) = \mathbb{E}\left[\frac{\partial^2 \log f(y_i|\lambda)}{\partial \lambda^2}\right] = -\frac{1}{\lambda^2}$.

(d) $\widehat{\text{std}}(\hat{\lambda}) = \frac{1}{\sqrt{n}}[-H(\hat{\lambda})]^{-1/2} = \frac{1}{\sqrt{n}}\hat{\lambda} = \frac{1}{20} = 0.05$.

Question 6

(a) $\gamma_0 = \sigma^2(1 + \theta^2)$,
$\gamma_1 = \sigma^2\theta$.

(b) The moment conditions are

$$\mathbb{E}(y_t^2) = \sigma^2(1 + \theta^2), \qquad \mathbb{E}(y_t y_{t-1}) = \sigma^2\theta.$$

The MM estimator solves

$$5 = \hat{\sigma}^2(1 + \hat{\theta}^2), \qquad 2 = \hat{\sigma}^2\hat{\theta}.$$

The two solutions to these equations are

$$\hat{\theta}_1 = 2, \qquad \hat{\sigma}_1 = 1,$$

and

$$\hat{\theta}_2 = 1/2, \qquad \hat{\sigma}_2 = 2.$$

(c) There are two solutions in (b), i.e. the parameters are not uniquely identified. Higher order autocovariances don't help, since for the $MA(1)$ model we have $\gamma_j = 0$ for $j > 1$, i.e. those moment conditions don't depend on the parameters.

G020: Examination in Econometrics
SOLUTION

2012-2013

Section A

Question 1 (20 points)

(a) Let $x_i = (1, w_i, z_i)$. From (1) we know that

$$\frac{1}{n}\sum_{i=1}^{n} x_i' x_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/4 \end{pmatrix}, \qquad \frac{1}{n}\sum_{i=1}^{n} x_i' y_i = \begin{pmatrix} 2 \\ 3 \\ -1 \end{pmatrix}.$$

The OLS estimator thus reads

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i' x_i\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} x_i' y_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ -4 \end{pmatrix},$$

i.e. $\hat{\beta}_1 = 2$, $\hat{\beta}_2 = 3$ and $\hat{\beta}_3 = -4$.

(b) We have

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2 = \frac{1}{n}\sum_{i=1}^{n} \left(y_i - \hat{\beta}_1 - w_i\hat{\beta}_2 - z_i\hat{\beta}_3\right)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\sum_{i=1}^{n} \left(\hat{\beta}_1 - w_i\hat{\beta}_2 - z_i\hat{\beta}_3\right)^2, \end{aligned}$$

where in the last step we used the fact that by construction of the OLS estimator the residuals $\hat{u}_i$ are orthogonal to the predicted values $\hat{\beta}_1 - w_i\hat{\beta}_2 - z_i\hat{\beta}_3$. Using (1) and our result for $\hat{\beta}$ we then find

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\sum_{i=1}^{n} \hat{\beta}_1^2 - \frac{1}{n}\sum_{i=1}^{n} w_i^2\hat{\beta}_2^2 - \frac{1}{n}\sum_{i=1}^{n} z_i^2\hat{\beta}_3^2 \\ &= 53 - 2^2 - 1\cdot 3^2 - \frac{1}{4}\cdot(-4)^2 \\ &= 36. \end{aligned}$$

(c) We have

$$\widehat{\mathrm{Var}}(\hat{\beta}) = \hat{\sigma}^2 \left(\sum_{i=1}^{n} x_i' x_i \right)^{-1} = \frac{\hat{\sigma}^2}{n} \left(\frac{1}{n} \sum_{i=1}^{n} x_i' x_i \right)^{-1} = \frac{36}{25} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

$$= \begin{pmatrix} 36/25 & 0 & 0 \\ 0 & 36/25 & 0 \\ 0 & 0 & 144/25 \end{pmatrix}.$$

Thus, $\widehat{\mathrm{std}}(\hat{\beta}_1) = \widehat{\mathrm{std}}(\hat{\beta}_2) = \frac{6}{5} = 1.2$, and $\widehat{\mathrm{std}}(\hat{\beta}_3) = \frac{12}{5} = 2.4$.

(d) We must have $\hat{\gamma}_1 + w_i \hat{\gamma}_2 + v_i \hat{\gamma}_3 = \hat{\beta}_1 + w_i \hat{\beta}_2 + z_i \hat{\beta}_3$ for all possible values of w_i and z_i, where $v_i = 2 + w_i + z_i$, i.e.

$$\hat{\gamma}_1 + w_i \hat{\gamma}_2 + (2 + w_i + z_i)\hat{\gamma}_3 = (\hat{\gamma}_1 + 2\hat{\gamma}_3) + w_i(\hat{\gamma}_2 + \hat{\gamma}_3) + z_i \hat{\gamma}_3 = \hat{\beta}_1 + w_i \hat{\beta}_2 + z_i \hat{\beta}_3.$$

Thus, we need $\hat{\gamma}_1 + 2\hat{\gamma}_3 = \hat{\beta}_1$, $\hat{\gamma}_2 + \hat{\gamma}_3 = \hat{\beta}_2$ and $\hat{\gamma}_3 = \hat{\beta}_3$. From this we find

$$\begin{aligned} \hat{\gamma}_1 &= \hat{\beta}_1 - 2\hat{\beta}_3 = 10, \\ \hat{\gamma}_2 &= \hat{\beta}_2 - \hat{\beta}_3 = 7, \\ \hat{\gamma}_3 &= \hat{\beta}_3 = -4. \end{aligned}$$

(e) We have $t = \frac{\hat{\beta}_3}{\widehat{\mathrm{std}}(\hat{\beta}_3)} = -1.6667$. Since $|t| < 1.96$ we cannot reject H_0 at 95% confidence level.

Question 2 (20 points)

(a) We have

$$\sqrt{n} \left(\hat{\beta}_{\mathrm{WLS}} - \beta \right) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i x_i u_i}{\frac{1}{n} \sum_{i=1}^{n} w_i x_i^2}.$$

For the numerator of this expression we have $\frac{1}{n} \sum_{i=1}^{n} w_i x_i^2 \to_p \mathbb{E}(w_i x_i^2)$ as $n \to \infty$ by the WLLN. For the denominator we have $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i x_i u_i \Rightarrow \mathcal{N}(0, \mathbb{E}[(w_i x_i u_i)^2])$ by the CLT, and since $\mathbb{E}(w_i x_i u_i) = 0$ by assumption. Combining these results and applying Slutsky's theorem thus gives

$$\sqrt{n} \left(\hat{\beta}_{\mathrm{WLS}} - \beta \right) \Rightarrow \mathcal{N} \left(0, \frac{\mathbb{E}[(w_i x_i u_i)^2]}{[\mathbb{E}(w_i x_i^2)]^2} \right),$$

as $n \to \infty$. For the asymptotic variance we thus have

$$\Sigma_{\text{WLS}} = \frac{\mathbb{E}[(w_i x_i u_i)^2]}{[\mathbb{E}(w_i x_i^2)]^2} = \frac{\mathbb{E}[w_i^2 x_i^2 \mathbb{E}(u_i^2|x_i)]}{[\mathbb{E}(w_i x_i^2)]^2} = \frac{\mathbb{E}[w_i^2 x_i^2 (1+x_i^2)]}{[\mathbb{E}(w_i x_i^2)]^2},$$

where we used the law of iterated expectations and $\mathbb{E}(u_i^2|x_i) = 1 + x_i^2$.

(b) For $w_i = 1$ we obtain the OLS estimator, whose asymptotic variance thus reads

$$\Sigma_{\text{OLS}} = \frac{\mathbb{E}[x_i^2(1+x_i^2)]}{[\mathbb{E}(x_i^2)]^2} = \frac{\mathbb{E}(x_i^2) + \mathbb{E}(x_i^4)}{[\mathbb{E}(x_i^2)]^2} = \frac{1+3}{1^2} = 4.$$

(c) The error term u_i is heteroscedastic here, i.e. one needs to use the heterscedasticity robust variance estimator for $\Sigma_{\text{OLS}} = \frac{\mathbb{E}[(x_i u_i)^2]}{[\mathbb{E}(x_i^2)]^2}$. This estimator is obtained from the last formula by replacing expectation by sample means and error u_i by residuals $\hat{u}_i = y_i - x_i \hat{\beta}_{\text{OLS}}$, i.e.

$$\hat{\Sigma}_{\text{OLS}} = \frac{\frac{1}{n}\sum_{i=1}^n x_i^2 \hat{u}_i^2}{\left(\frac{1}{n}\sum_{i=1}^n x_i^2\right)^2}.$$

(d) From the lecture we know that the optimal weight are given by $w_i = \frac{\lambda}{\mathbb{E}(u_i^2|x_i)} = \frac{\lambda}{1+x_i^2}$, where $\lambda > 0$ is an arbitrary scalar factor, which in the following we choose as $\lambda = 1$. We know that for the optimal weights, there should be a cancelation between the denominator and the numerator in Σ_{WLS}, namely we find

$$\Sigma_{\text{WLS}} = \frac{\mathbb{E}[w_i^2 x_i^2(1+x_i^2)]}{[\mathbb{E}(w_i x_i^2)]^2} = \frac{1}{\mathbb{E}(w_i x_i^2)} = \left[\mathbb{E}\left(\frac{x_i^2}{1+x_i^2}\right)\right]^{-1} = \left(\frac{1}{3}\right)^{-1} = 3.$$

Thus, for the optimal weights we have $\Sigma_{\text{WLS}} < \Sigma_{\text{OLS}}$, i.e. the WLS estimator is more efficient than the OLS estimator.

Question 3 (20 points)

(a) We have

$$y_i = p_i \beta + u_i,$$

where $u_i = \varepsilon_i - v_i \beta$. We find

$$\hat{\beta}_{\text{OLS}} - \beta = \frac{\frac{1}{n}\sum_{i=1}^n p_i u_i}{\frac{1}{n}\sum_{i=1}^n p_i^2} \to_p \frac{\mathbb{E} p_i u_i}{\mathbb{E} p_i^2} = \frac{-\beta \mathbb{E} v_i^2}{\mathbb{E}(p_i^*)^2 + \mathbb{E} v_i^2} = \frac{-\beta \sigma_v^2}{1+\sigma_v^2} \neq 0,$$

as $n \to \infty$. Here, we used the WLLN and the continuous mapping theorem. Thus, $\hat{\beta}_{\text{OLS}}$ is not consistent.

(b) As $n \to \infty$ we have

$$\hat{\gamma} = \frac{\sum_{i=1}^n z_i p_i}{\sum_{i=1}^n z_i^2} \to_p \frac{\mathbb{E} z_i p_i}{\mathbb{E} z_i^2} = \frac{\mathbb{E} z_i p_i^*}{\mathbb{E} z_i^2} = \frac{\rho}{1} = \rho,$$
$$\hat{\pi} = \frac{\sum_{i=1}^n z_i y_i}{\sum_{i=1}^n z_i^2} \to_p \frac{\mathbb{E} z_i y_i}{\mathbb{E} z_i^2} = \frac{\beta \mathbb{E} z_i p_i^*}{\mathbb{E} z_i^2} = \frac{\beta\rho}{1} = \beta\rho,$$

where we again used the WLLN and the continuous mapping theorem.

(c) A consistent estimator for β is given by

$$\hat{\beta}_{\mathrm{IV}} = \frac{\hat{\pi}}{\hat{\gamma}}.$$

This is the standard IV (or 2SLS) estimator expressed in terms of the reduced form estimators $\hat{\pi}$ and $\hat{\gamma}$. In the standard 2SLS setting the first stage consists of obtaining $\hat{\gamma}$ and thus $\hat{p}_i = \hat{\gamma} z_i$, followed by the second stage, where one regresses y_i on $\hat{p}_i$, which gives the same $\hat{\beta}_{\mathrm{IV}}$ as above.

(d) z_i is a relevant instrument if $\rho \neq 0$.
The effective error term is $u_i = \varepsilon_i - v_i \beta$, which contains both ε_i and v_i, i.e. exogeneity of z_i requires $\mathbb{E}(z_i \varepsilon_i) = 0$ and $\mathbb{E}(z_i v_i) = 0$.

(e) 2SLS still works, i.e. one regresses y_i on a constant, $\hat{p}_i = \hat{\gamma} z_i$ and w_i. The resulting second stage estimator for β_1, β_2, β_3 is consistent.

Section B

Question 4 (20 points)

(a) We have

$$Q_n(\lambda) = \frac{1}{n} \log \prod_{i=1}^n f(y_i|\lambda) = \frac{1}{n} \sum_{i=1}^n \log f(y_i|\lambda) = \frac{1}{n} \sum_{i=1}^n \log \left[\frac{\lambda^{y_i} \, e^{-\lambda}}{y_i!} \right]$$
$$= \frac{1}{n} \sum_{i=1}^n [y_i \log \lambda - \lambda - \log y_i!] = \left(\frac{1}{n} \sum_{i=1}^n y_i \right) \log \lambda - \lambda - \frac{1}{n} \sum_{i=1}^n \log y_i!.$$

Maximizing this over λ gives the FOC

$$\left(\frac{1}{n} \sum_{i=1}^n y_i \right) \frac{1}{\hat{\lambda}} - 1 = 0,$$

which gives $\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n y_i$.

(b) We have $\log f(y_i|\lambda) = y_i \log \lambda - \lambda - \log y_i!$ and $\frac{\partial^2 \log f(y_i|\lambda)}{\partial \lambda^2} = -\frac{y_i}{\lambda^2}$. Since $\mathbb{E} y_i = \lambda$ we find at the true parameter λ that $H = \mathbb{E}\left[\frac{\partial^2 \log f(y_i|\lambda)}{\partial \lambda^2}\right] = -\frac{1}{\lambda}$.

(c) We have $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^n y_i = 4$,
and $\widehat{\text{std}}(\hat{\lambda}) = \sqrt{-\frac{1}{n\hat{H}}} = \sqrt{\frac{\hat{\lambda}}{n}} = \sqrt{\frac{4}{25}} = \frac{2}{5} = 0.4$.

(d) Yes, the WLLN guarantees consistency of $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^n y_i$ as long as $\mathbb{E} y_i = \lambda$, even if y_i is not Poisson distributed. However, the standard error above assumes that $\text{Var}(y_i) = \lambda$, which is true for the Poisson distribution, but not for other distributions.

Question 5 (20 points)

(a) We have $\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^n y_{i1}$ and $\hat{\theta}_2 = \frac{1}{2n}\sum_{i=1}^n y_{i2}$. Applying the CLT we obtain

$$\sqrt{n}(\hat{\theta}_1 - \theta) \Rightarrow \mathcal{N}(0, \text{Var}(y_{i1})), \qquad \sqrt{n}(\hat{\theta}_2 - \theta) \Rightarrow \mathcal{N}(0, \text{Var}(y_{i2})/4),$$

i.e. $\text{AsyVar}(\hat{\theta}_1) = \text{Var}(y_{i1}) = 1$ and $\text{AsyVar}(\hat{\theta}_2) = \text{Var}(y_{i2})/4 = 1/4$.

(b) The GMM objective reads

$$\left[\frac{1}{n}\sum_{i=1}^n y_i - \binom{1}{2}\theta\right]' W \left[\frac{1}{n}\sum_{i=1}^n y_i - \binom{1}{2}\theta\right].$$

The FOC reads

$$(1,2)\, W \left[\frac{1}{n}\sum_{i=1}^n y_i - \binom{1}{2}\hat{\theta}_{\text{GMM}}\right] = 0.$$

Solving this for $\hat{\theta}_{\text{GMM}}$ gives

$$\begin{aligned}
\hat{\theta}_{\text{GMM}} &= \frac{(1,2)W\frac{1}{n}\sum_{i=1}^n y_i}{(1,2)W\binom{1}{2}} \\
&= \frac{(1,2)W\binom{1}{0}}{(1,2)W\binom{1}{2}}\hat{\theta}_1 + \frac{(1,2)W\binom{0}{2}}{(1,2)W\binom{1}{2}}\hat{\theta}_2 \\
&= w\hat{\theta}_1 + (1-w)\hat{\theta}_2,
\end{aligned}$$

where

$$w = \frac{(1,2)W\binom{1}{0}}{(1,2)W\binom{1}{2}} = \frac{W_{11} + 2W_{12}}{W_{11} + 4W_{12} + 4W_{22}}.$$

Question 6 (20 points)

(a) We find

$$\gamma_0 = \sigma^2(1+\theta^2), \qquad \gamma_1 = \sigma^2\theta, \qquad \gamma_2 = 0,$$

and therefore

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\theta}{1+\theta^2}, \qquad \rho_2 = \frac{\gamma_2}{\gamma_0} = 0.$$

(b) Applying the WLLN (for time series data) and the continuous mapping theorem we find that as $T \to \infty$

$$\hat{\rho} = \frac{\frac{1}{T}\sum_{t=2}^{T} y_t y_{t-1}}{\frac{1}{T}\sum_{t=2}^{T} y_{t-1}^2} \to_p \frac{\mathbb{E} y_t y_{t-1}}{\mathbb{E} y_{t-1}^2} = \frac{\gamma_1}{\gamma_0} = \rho_1.$$

(c) We have to solve

$$\hat{\rho} = \frac{\hat{\theta}}{1+\hat{\theta}^2},$$

which cal also be written as

$$\hat{\theta}^2 - \hat{\rho}^{-1}\,\hat{\theta} + 1 = 0,$$

which is a standard quadratic equation in $\hat{\theta}$ that has two solutions

$$\hat{\theta}_{1/2} = \frac{1}{2\hat{\rho}} \pm \sqrt{\frac{1}{4\hat{\rho}^2} - 1}$$

Plugging in $\hat{\rho} = 0.4$ gives $\hat{\theta}_1 = 0.5$ and $\hat{\theta}_2 = 2$. Since we restrict $\theta \in [0,1]$ the unique solution is

$$\hat{\theta} = 0.5.$$

(d) The relation ship between $\hat{\rho}$ and $\hat{\theta}$ is given by $\hat{\rho} = \frac{\hat{\theta}}{1+\hat{\theta}^2}$. Differentiating this and evaluating at $\hat{\theta} = 0.5$ gives

$$\frac{d\hat{\rho}}{d\hat{\theta}} = \frac{1-\hat{\theta}^2}{(1+\hat{\theta}^2)^2} = 0.48.$$

Thus, according to the delta-method we have

$$\widehat{\text{std}}(\hat{\theta}) = \left(\frac{d\hat{\rho}}{d\hat{\theta}}\right)^{-1} \widehat{\text{std}}(\hat{\rho}) = \frac{0.096}{0.48} = 0.2.$$

SOLUTIONS

G020: Examination in Econometrics

2013-2014

Section A

Question 1 (20 points)

(a) For $x_i = (1, w_i)$ we have

$$\frac{1}{n}\sum_{i=1}^{n} x_i' x_i = \begin{pmatrix} 1 & \frac{1}{n}\sum_{i=1}^{n} w_i \\ \frac{1}{n}\sum_{i=1}^{n} w_i & \frac{1}{n}\sum_{i=1}^{n} w_i^2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix},$$
$$\frac{1}{n}\sum_{i=1}^{n} x_i' y_i = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} y_i \\ \frac{1}{n}\sum_{i=1}^{n} w_i y_i \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \end{pmatrix}.$$

We thus find $\det\left(\frac{1}{n}\sum_{i=1}^{n} x_i' x_i\right) = 1$ and

$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i' x_i\right)^{-1} = \begin{pmatrix} 5 & -2 \\ -2 & 1 \end{pmatrix},$$
$$\widehat{\beta} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i' x_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} x_i' y_i\right) = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

i.e. we have $\widehat{\beta}_1 = 0$ and $\widehat{\beta}_2 = 1$.

(b) We have $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \widehat{u}_i^2$, where $\widehat{u}_i = y_i - \widehat{\beta}_1 - w_i\widehat{\beta}_2$. Therefore

$$\begin{aligned}\widehat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{\beta}_1 - w_i\widehat{\beta}_2)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - w_i)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(y_i^2 - 2y_i w_i + w_i^2) \\ &= 30 - 2\cdot 5 + 5 = 25.\end{aligned}$$

We found $\widehat{\sigma}^2 = 25$.

(Comment: an alternative way of calculating $\widehat{\sigma}^2$ is to use the formula $\frac{1}{n}\sum_{i=1}^n \widehat{u}_i^2 = \frac{1}{n}\sum_{i=1}^n y_i^2 - \frac{1}{n}\sum_{i=1}^n(\widehat{\beta}_1 + w_i\widehat{\beta}_2)^2$, which is true due to the FOC of the OLS problem: $\sum_{i=1}^n \widehat{u}_i = 0$ and $\sum_{i=1}^n \widehat{u}_i w_i = 0$.)

(c) We have

$$\widehat{\mathrm{Var}}(\widehat{\beta}) = \widehat{\sigma}^2\left(\sum_{i=1}^n x_i' x_i\right)^{-1} = \frac{\widehat{\sigma}^2}{n}\left(\frac{1}{n}\sum_{i=1}^n x_i' x_i\right)^{-1}$$
$$= \frac{25}{81}\begin{pmatrix} 5 & -2 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} 125/81 & -50/81 \\ -50/81 & 25/81 \end{pmatrix}.$$

Thus, $\widehat{\mathrm{std}}(\widehat{\beta}_1) = \sqrt{125/81} = 1.242$, and $\widehat{\mathrm{std}}(\widehat{\beta}_2) = \sqrt{25/81} = 5/9 = 0.556$.

(d) $t = \frac{\widehat{\beta}_2 - 0}{\widehat{\mathrm{std}}(\widehat{\beta}_2)} = \frac{9}{5} = 1.8$. Since $t \geq 1.65$ we can reject H_0 at 95% confidence level.

Question 2 (20 points)

(a) Consider $\rho_{zu} = 0$. We first use the model for y_i to obtain

$$\widehat{\beta} - \beta = \left(\frac{1}{n}\sum_{i=1}^n x_i' x_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n x_i' u_i\right).$$

By applying the weak law of large numbers (WLLN) and the continuous mapping theorem (CMT) we obtain as $n \to \infty$

$$\widehat{\beta} - \beta \to_p [\mathbb{E}(x_i' x_i)]^{-1}\,\mathbb{E}(x_i' u_i).$$

The condition $|\rho_{wz}| < 1$ guarantees that $\mathbb{E}(x_i' x_i)$ is invertible. The condition $\rho_{zu} = 0$ guarantees that $\mathbb{E}(x_i' u_i) = 0$. Therefore $[\mathbb{E}(x_i' x_i)]^{-1}\,\mathbb{E}(x_i' u_i) = 0$. We have thus shown that $\widehat{\beta} \to_p \beta$ as $n \to \infty$, i.e. $\widehat{\beta}$ is consistent for β.

(b) The condition $\rho_{zu} = 0$ guarantees that x_i is exogenous. The condition $|\rho_{wz}| < 1$ guarantees that x_i is non-collinear.

If $\rho_{wz} = 1$, then $w_i = z_i$, so that the two regressors in $x_i = (w_i, z_i)$ are identical. The inverse of $\sum_{i=1}^n x_i' x_i$ and thus also $\widehat{\beta}$ are not well-defined in that case.

(c) Consider $\rho_{zu} \neq 0$ and $\rho_{wz} = 0$. We have already derived $\widehat{\beta} - \beta \to_p [\mathbb{E}(x_i' x_i)]^{-1}\,\mathbb{E}(x_i' u_i)$, as $n \to \infty$. This is still valid here. In the current case we have

$$[\mathbb{E}(x_i' x_i)]^{-1}\,\mathbb{E}(x_i' u_i) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{-1}\begin{pmatrix} 0 \\ \rho_{zu} \end{pmatrix} = \begin{pmatrix} 0 \\ \rho_{zu} \end{pmatrix}.$$

This shows that still $\widehat{\beta}_1 - \beta_1 \to_p 0$, i.e. $\widehat{\beta}_1$, i.e. $\underline{\widehat{\beta}_1 \text{ is consistent for } \beta_1}$.

(this was not asked, but we also showed that $\widehat{\beta}_2 - \beta_2 \to_p \rho_{zu} \neq 0$, i.e. $\widehat{\beta}_2$ is not consistent)

(d) We have already derived $\widehat{\beta} - \beta \to_p [\mathbb{E}(x_i' x_i)]^{-1} \mathbb{E}(x_i' u_i)$, as $n \to \infty$. This is still valid here. We have

$$\begin{aligned}
[\mathbb{E}(x_i' x_i)]^{-1} \mathbb{E}(x_i' u_i) &= \begin{pmatrix} 1 & \rho_{wz} \\ \rho_{wz} & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \rho_{zu} \end{pmatrix} \\
&= \frac{1}{1 - \rho_{wz}^2} \begin{pmatrix} 1 & -\rho_{wz} \\ -\rho_{wz} & 1 \end{pmatrix} \begin{pmatrix} 0 \\ \rho_{zu} \end{pmatrix} \\
&= \frac{1}{1 - \rho_{wz}^2} \begin{pmatrix} -\rho_{wz}\rho_{zu} \\ \rho_{zu} \end{pmatrix}.
\end{aligned}$$

Thus, we have shown that $\widehat{\beta}_1 \to_p \beta_1^*$ as $n \to \infty$, where

$$\beta_1^* = \beta_1 - \frac{\rho_{wz}\rho_{zu}}{1 - \rho_{wz}^2}.$$

For $\rho_{zu} \neq 0$ and $\rho_{wz} \neq 0$ we have $\beta_1^* \neq \beta$, i.e. $\underline{\widehat{\beta}_1 \text{ is NOT consistent for } \beta_1.}$

Question 3 (20 points)

(a) Estimators for $\gamma_{xx} = \mathbb{E}(x_i^2)$, $\gamma_{xy} = \mathbb{E}(x_i y_i)$, $\gamma_{xz} = \mathbb{E}(x_i z_i)$ and $\gamma_{yz} = \mathbb{E}(y_i z_i)$ are given by the sample analogs

$$\widehat{\gamma}_{xx} = \frac{1}{n}\sum_{i=1}^n x_i^2, \qquad \widehat{\gamma}_{xy} = \frac{1}{n}\sum_{i=1}^n x_i y_i, \qquad \widehat{\gamma}_{xz} = \frac{1}{n}\sum_{i=1}^n x_i z_i, \qquad \widehat{\gamma}_{yz} = \frac{1}{n}\sum_{i=1}^n y_i z_i.$$

These estimators are consistent, for example, the WLLN guarantees that as $n \to \infty$ we have $\widehat{\gamma}_{xx} = \frac{1}{n}\sum_{i=1}^n x_i^2 \to_p \mathbb{E}(x_i^2) = \gamma_{xx}$.

(b) Using the model we have that $u_i = y_i - x_i\beta$. We thus have

$$\begin{aligned}
0 &= \mathbb{E}(x_i u_i) = \mathbb{E}[x_i(y_i - x_i\beta)] = \mathbb{E}(x_i y_i) - \mathbb{E}(x_i^2)\beta \\
&= \gamma_{xy} - \gamma_{xx}\beta.
\end{aligned}$$

Solving this equation for β gives $\beta = \frac{\gamma_{xy}}{\gamma_{xx}}$. A natural estimator for β is therefore given by

$$\widehat{\beta} = \frac{\widehat{\gamma}_{xy}}{\widehat{\gamma}_{xx}} = \frac{\frac{1}{n}\sum_{i=1}^n x_i y_i}{\frac{1}{n}\sum_{i=1}^n x_i^2} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}.$$

It turns out that this is simply the OLS estimator. We already know consistency of $\widehat{\gamma}_{xy}$ and $\widehat{\gamma}_{xx}$ from (a). Applying this, the CMT, and the assumption that $\gamma_{xx} > 0$ we obtain that as $n \to \infty$ we have

$$\widehat{\beta} = \frac{\widehat{\gamma}_{xy}}{\widehat{\gamma}_{xx}} \to_p \frac{\gamma_{xy}}{\gamma_{xx}} = \beta,$$

i.e. $\widehat{\beta}$ is indeed consistent.

(c) If we want to use z_i as an instrument, then we also need the <u>relevance assumption</u> $\mathbb{E}(x_i z_i) \neq 0$, i.e. $\underline{\gamma_{xz} \neq 0}$. From the exclusion restriction $\mathbb{E}(z_i u_i) = 0$ we find, analogously to the answer in (b), that

$$\begin{aligned} 0 &= \mathbb{E}(z_i u_i) = \mathbb{E}[z_i(y_i - x_i\beta)] = \mathbb{E}(z_i y_i) - \mathbb{E}(z_i x_i)\beta \\ &= \gamma_{yz} - \gamma_{xz}\beta. \end{aligned}$$

Together with $\gamma_{xz} \neq 0$ this implies that $\beta = \frac{\gamma_{yz}}{\gamma_{xz}}$. A consistent estimator for β is then given by

$$\widehat{\beta}_{\mathrm{IV}} = \frac{\gamma_{yz}}{\gamma_{xz}} = \frac{\frac{1}{n}\sum_{i=1}^n z_i y_i}{\frac{1}{n}\sum_{i=1}^n z_i x_i} = \frac{\sum_{i=1}^n z_i y_i}{\sum_{i=1}^n z_i x_i}.$$

(d) As $n \to \infty$ we have

$$\sqrt{n}(\widehat{\beta}_{\mathrm{IV}} - \beta) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n z_i u_i}{\frac{1}{n}\sum_{i=1}^n z_i x_i} \Rightarrow \frac{\mathcal{N}[0, \mathbb{E}(z_i^2 u_i^2)]}{\gamma_{xz}} \sim \mathcal{N}(0, \Sigma_{\mathrm{IV}}),$$

where

$$\Sigma_{\mathrm{IV}} = \frac{\mathbb{E}(z_i^2 u_i^2)}{\gamma_{xz}^2}.$$

Here, we have first used the definition of $\widehat{\beta}_{\mathrm{IV}}$ and the model, and then applied the WLLN, CLT and Slutsky's theorem. A consistent estimator for Σ_{IV} is given by

$$\widehat{\Sigma}_{\mathrm{IV}} = \frac{\frac{1}{n}\sum_{i=1}^n z_i^2\, \widehat{u}_i^2}{\left(\frac{1}{n}\sum_{i=1}^n z_i x_i\right)^2},$$

where $\widehat{u}_i = y_i - x_i\widehat{\beta}_{\mathrm{IV}}$. This is a heteroscedasticity robust variance estimator.

Section B

Question 4 (20 points)

(a) We have

$$\begin{aligned} Q_n(\mu,\sigma) &= \frac{1}{n}\log\prod_{i=1}^{n} f(y_i|\mu,\sigma) = \frac{1}{n}\sum_{i=1}^{n}\log f(y_i|\mu,\sigma) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[-\log\sqrt{2\pi\sigma^2} - \frac{(y_i-\mu)^2}{2\sigma^2}\right] \\ &= -\log\sqrt{2\pi} - \log\sigma - \frac{1}{2\sigma^2}\frac{1}{n}\sum_{i=1}^{n}(y_i-\mu)^2. \end{aligned}$$

The FOC read

$$0 = \frac{\partial Q_n(\widehat{\mu},\widehat{\sigma})}{\partial\mu} = \frac{1}{\widehat{\sigma}^2}\frac{1}{n}\sum_{i=1}^{n}(y_i-\widehat{\mu}),$$

and

$$0 = \frac{\partial Q_n(\widehat{\mu},\widehat{\sigma})}{\partial\sigma} = -\frac{1}{\widehat{\sigma}} + \frac{1}{\widehat{\sigma}^3}\frac{1}{n}\sum_{i=1}^{n}(y_i-\widehat{\mu})^2.$$

(b) Solving the FOC $\frac{1}{\widehat{\sigma}^2}\frac{1}{n}\sum_{i=1}^{n}(y_i-\widehat{\mu}) = 0$ gives

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

(c) Solving the FOC $-\frac{1}{\widehat{\sigma}} + \frac{1}{\widehat{\sigma}^3}\frac{1}{n}\sum_{i=1}^{n}(y_i-\widehat{\mu})^2 = 0$ gives

$$\widehat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\widehat{\mu})^2}.$$

(d) From the lecture we know that for a correctly specified MLE we have $\mathrm{AsyVar}(\sqrt{n}\widehat{\theta}) = [-\mathbb{E}H(y_i,\theta)]^{-1}$, minus the inverse expected Hessian at the true parameter θ. We thus have

$$\begin{aligned} \Sigma_{\mathrm{MLE}} = \mathrm{AsyVar}(\sqrt{n}\widehat{\theta}) &= \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 2/\sigma^2 \end{pmatrix}^{-1} \\ &= \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2/2 \end{pmatrix}, \end{aligned}$$

where here σ refers to the true parameter (which in the lecture we usually denoted as σ_0).

(e) We now only assume that $\mathbb{E} y_i = \mu$ and $\mathrm{Var}(y_i) = \sigma^2$, but the likelihood based on the normal distribution may otherwise be misspecified. Then:

- The estimator $\widehat{\mu}$ is still consistent for μ, because of the WLLN.
- The estimator $\widehat{\sigma}$, which can also be written as $\widehat{\sigma} = \sqrt{\left(\frac{1}{n}\sum_{i=1}^n y_i^2\right) - \widehat{\mu}^2}$, is also still consistent for $\sigma = \sqrt{(\mathbb{E} y_i^2) - \mu^2}$, because of the WLLN and since $\widehat{\mu}$ is still consistent.
- The asymptotic variance of $\widehat{\theta} = (\widehat{\mu}, \widehat{\sigma})'$, in particular the asymptotic variance of $\widehat{\sigma}$, depends on the higher moments of y_i, in particular on $\mathbb{E} y_i^3$ and $\mathbb{E} y_i^4$. For normally distributed y_i those higher moments are determined by μ and σ, but in general this is not true. The above formula for $\widehat{\Sigma}_{\mathrm{MLE}}$ is therefore NOT the correct expression for $\mathrm{AsyVar}(\sqrt{n}\widehat{\theta})$ anymore.

Question 5 (20 points)

(a) The sample analog of the moment condition reads $\frac{1}{n}\sum_{i=1}^n y_i = \widehat{\theta}$, i.e. we have $\widehat{\theta} = \frac{1}{n}\sum_{i=1}^n y_i = 40/400 = 1/10 = 0.1$. Here we used that $n = 400$ and $\sum_{i=1}^n y_i = 40$.

(b) Applying the CLT we find $\sqrt{n}(\widehat{\theta} - \theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (y_i - \theta) \Rightarrow \mathcal{N}(0, \mathbb{E}[(y_i - \theta)^2])$. We thus have $\Sigma_{\mathrm{GMM}} = \mathbb{E}[(y_i - \theta)^2] = \mathrm{Var}(y_i) = \theta(1 - \theta)$.

(c) An estimator for Σ_{GMM} is given by $\widehat{\Sigma}_{\mathrm{GMM}} = \widehat{\theta}(1 - \widehat{\theta})$. Consistency of $\widehat{\Sigma}_{\mathrm{GMM}}$ follows from consistency of $\widehat{\theta}$ and the CMT. For the observed sample we have $\widehat{\Sigma}_{\mathrm{GMM}} = 0.1(1 - 0.1) = 0.09$.

(d) An estimator for the standard error of $\widehat{\theta}$ is given by $\mathrm{std}(\widehat{\theta}) = \sqrt{\widehat{\Sigma}_{\mathrm{GMM}}/n} = 0.015$. The t-test statistics reads

$$t = \frac{\widehat{\theta} - 1/2}{\mathrm{std}(\widehat{\theta})} = \frac{0.1 - 0.5}{0.015} = -26.667.$$

Since $|t| > 1.96$ we reject H_0 at 95% confidence level.

(e) We have $\Sigma_{\mathrm{GMM}} = f(\theta)$, where $f(\theta) = \theta(1 - \theta)$. The first derivative of this function reads $f'(\theta) = 1 - 2\theta$. Since $\widehat{\theta}$ is asymptotically close to θ we have $\widehat{\Sigma}_{\mathrm{GMM}} - \Sigma_{\mathrm{GMM}} =$

$f(\widehat{\theta}) - f(\theta) \approx f'(\theta)\,(\widehat{\theta} - \theta)$. We therefore have $\sqrt{n}(\widehat{\Sigma}_{\text{GMM}} - \Sigma_{\text{GMM}}) \Rightarrow \mathcal{N}(0, \Omega)$, where

$$\begin{aligned}\Omega &= \text{AsyVar}(\sqrt{n}\widehat{\Sigma}_{\text{GMM}}) = [f'(\theta)]^2 \text{AsyVar}(\sqrt{n}\widehat{\theta}) = (1-2\theta)^2 \Sigma_{\text{GMM}} \\ &= \theta(1-\theta)(1-2\theta)^2.\end{aligned}$$

Question 6 (20 points)

(a)
- Since the process is stationary we have $\mathbb{E}y_t = \mathbb{E}y_{t-1}$. Using the model we find $\mathbb{E}y_t = \rho\mathbb{E}y_{t-1} + \mathbb{E}u_t = \rho\mathbb{E}y_{t-1}$, and therefore $\mathbb{E}y_t = \rho\mathbb{E}y_t$ from which we conclude that $\mathbb{E}y_t = 0$.
- Since the process is stationary we have $\mathbb{E}y_t^2 = \mathbb{E}y_{t-1}^2$. Using the model we find $\mathbb{E}y_t^2 = \rho^2\mathbb{E}y_{t-1}^2 + 2\rho\mathbb{E}y_{t-1}u_t + \mathbb{E}u_t^2 = \rho^2\mathbb{E}y_{t-1}^2 + \sigma^2$. Thus, we have $(1-\rho^2)\mathbb{E}y_t^2 = \sigma^2$, i.e. $\gamma_0 = \text{Var}(y_t) = \mathbb{E}y_t^2 = \sigma^2/(1-\rho^2)$.
- We have $\gamma_1 = \text{Cov}(y_t, y_{t-1}) = \mathbb{E}(y_t y_{t-1}) = \rho\mathbb{E}(y_{t-1}^2) = \frac{\rho\sigma^2}{1-\rho^2}$.
- We have $\gamma_2 = \text{Cov}(y_t, y_{t-2}) = \mathbb{E}(y_t y_{t-2}) = \rho\mathbb{E}(y_{t-1}y_{t-2}) = \rho^2\mathbb{E}(y_{t-2}^2) = \frac{\rho^2\sigma^2}{1-\rho^2}$.

(b) As $T \to \infty$ we have

$$\widehat{\rho} = \frac{\frac{1}{T}\sum_{t=2}^T y_t y_{t-1}}{\frac{1}{T}\sum_{t=2}^T y_{t-1}^2} \to_p \frac{\mathbb{E}(y_t y_{t-1})}{\mathbb{E}(y_{t-1}^2)} = \frac{\gamma_1}{\gamma_0} = \rho,$$

where we applied the WLLN (for stationary and ergodic sequences) and the CMT.

(c) No, $\widehat{\rho}$ is NOT unbiased. Normally, we show unbiasedness of the OLS estimator by conditioning on all regressor obervations and evaluating $\mathbb{E}(\widehat{\beta}|X)$, which makes no sense here, since the sequence of regressors is equal to the sequence of outcome variables, i.e. conditioning on X implies conditioning on y and u. When trying to evaluate the unconditional expectation one finds that

$$\mathbb{E}(\widehat{\rho}) = \mathbb{E}\left(\frac{\frac{1}{T}\sum_{t=2}^T y_t y_{t-1}}{\frac{1}{T}\sum_{t=2}^T y_{t-1}^2}\right) \neq \frac{\mathbb{E}(y_t y_{t-1})}{\mathbb{E}(y_{t-1}^2)} = \frac{\gamma_1}{\gamma_0} = \rho,$$

because in general we have $\mathbb{E}(A/B) \neq \mathbb{E}(A)/\mathbb{E}(B)$.

(d) As $T \to \infty$ we have

$$\widehat{\beta} = \frac{\frac{1}{T}\sum_{t=3}^T y_t y_{t-2}}{\frac{1}{T}\sum_{t=3}^T y_{t-2}^2} \to_p \frac{\mathbb{E}(y_t y_{t-2})}{\mathbb{E}(y_{t-2}^2)} = \frac{\gamma_2}{\gamma_0} = \rho^2,$$

where we applied the WLLN (for stationary and ergodic sequences) and the CMT.

(e) The AR(2) model $y_t = \gamma_1 \, y_{t-1} + \gamma_2 \, y_{t-2} + \epsilon_t$ includes the true AR(1) data generating process as a special case, with $\gamma_1 = \rho$, $\gamma_2 = 0$ and $\epsilon_t = u_t$. The OLS estimator is consistent for the true values $\gamma_1 = \rho$ and $\gamma_2 = 0$, i.e. we expect $\widehat{\gamma}_1$ to be close to $\widehat{\rho}$ asymptotically (because both estimators converge to ρ), but we do NOT expect $\widehat{\gamma}_2$ to be close to $\widehat{\beta}$ (because one estimator converges to ρ^2 and one converges to zero).

SOLUTIONS

G020: Examination in Econometrics

2014-2015

Section A

Question 1 (20 points)

(a) We find

$$\mathbb{E}(y_i|h_i = 0) = \beta_1,$$
$$\mathbb{E}(y_i|h_i = 1) = \beta_1 + \beta_2.$$

From the sample we have estimators

$$\widehat{\mathbb{E}}(y_i|h_i = 0) = \frac{1}{n_0}\sum_{i=1}^{n}(1-h_i)y_i = 0.2,$$
$$\widehat{\mathbb{E}}(y_i|h_i = 1) = \frac{1}{n_1}\sum_{i=1}^{n}h_i y_i = 2.$$

Combining this we find $\widehat{\beta}_1 = 0.2$ and $\widehat{\beta}_1 + \widehat{\beta}_2 = 2$, i.e.

$$\widehat{\beta}_1 = 0.2,$$
$$\widehat{\beta}_2 = 2 - 0.2 = 1.8.$$

(b) Using that $h_i^2 = h_i$ we find

$$\sum_{i=1}^{n} x_i' x_i = \sum_{i=1}^{n}\binom{1}{h_i}(1, h_i) = \sum_{i=1}^{n}\begin{pmatrix} 1 & h_i \\ h_i & h_i^2 \end{pmatrix} = \sum_{i=1}^{n}\begin{pmatrix} 1 & h_i \\ h_i & h_i \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^{n} h_i \\ \sum_{i=1}^{n} h_i & \sum_{i=1}^{n} h_i \end{pmatrix}$$
$$= \begin{pmatrix} n & n_1 \\ n_1 & n_1 \end{pmatrix} = \begin{pmatrix} 80 & 20 \\ 20 & 20 \end{pmatrix}.$$

Using that $\sum_{i=1}^n y_i = n_0 \left[\frac{1}{n_0}\sum_{i=1}^n (1-h_i) y_i\right] + n_1 \left[\frac{1}{n_1}\sum_{i=1}^n h_i y_i\right]$ we find

$$\begin{aligned}
\sum_{i=1}^n x_i' y_i &= \sum_{i=1}^n \binom{1}{h_i} y_i = \sum_{i=1}^n \binom{y_i}{h_i y_i} = \binom{\sum_{i=1}^n y_i}{\sum_{i=1}^n h_i y_i} \\
&= \begin{pmatrix} n_0 \left[\frac{1}{n_0}\sum_{i=1}^n (1-h_i) y_i\right] + n_1 \left[\frac{1}{n_1}\sum_{i=1}^n h_i y_i\right] \\ n_1 \left[\frac{1}{n_1}\sum_{i=1}^n h_i y_i\right] \end{pmatrix} = \begin{pmatrix} 60 \times 0.2 + 20 \times 2 \\ 20 \times 2 \end{pmatrix} \\
&= \begin{pmatrix} 52 \\ 40 \end{pmatrix}.
\end{aligned}$$

We thus find

$$\begin{aligned}
\widehat{\beta} = \binom{\widehat{\beta}_1}{\widehat{\beta}_2} &= \left(\sum_{i=1}^n x_i' x_i\right)^{-1} \sum_{i=1}^n x_i' y_i = \begin{pmatrix} 80 & 20 \\ 20 & 20 \end{pmatrix}^{-1} \begin{pmatrix} 52 \\ 40 \end{pmatrix} \\
&= \frac{1}{1.200} \begin{pmatrix} 20 & -20 \\ -20 & 80 \end{pmatrix} \begin{pmatrix} 52 \\ 40 \end{pmatrix} \\
&= \binom{0.2}{1.8}.
\end{aligned}$$

As expected, this is the same result as in (a).

(c) Under homoscedasticity with known $\sigma^2 = \mathbb{E}(u_i^2 | h_i) = 15$ the OLS estimator for the variance of $\widehat{\beta} = \binom{\widehat{\beta}_1}{\widehat{\beta}_2}$ is given by

$$\widehat{\mathrm{Var}}(\widehat{\beta}) = \sigma^2 \left(\sum_{i=1}^n x_i' x_i\right)^{-1} = 15 \, \frac{1}{1.200} \begin{pmatrix} 20 & -20 \\ -20 & 80 \end{pmatrix} = \begin{pmatrix} 1/4 & -1/4 \\ -1/4 & 1 \end{pmatrix}.$$

We thus have

$$\begin{aligned}
\widehat{\mathrm{std}}(\widehat{\beta}_1) &= \sqrt{\widehat{\mathrm{Var}}(\widehat{\beta}_1)} = \frac{1}{2}, \\
\widehat{\mathrm{std}}(\widehat{\beta}_2) &= \sqrt{\widehat{\mathrm{Var}}(\widehat{\beta}_2)} = 1.
\end{aligned}$$

(d) The t-test statistics reads

$$t = \frac{\widehat{\beta}_2}{\widehat{\mathrm{std}}(\widehat{\beta}_2)} = \frac{0.9}{0.5} = 1.8.$$

We can reject $H_0 : \beta_2 \leq 0$ in a one-sided t-test at 95% confidence level if $t > 1.64$. Since indeed $1.8 > 1.64$ we **do reject** H_0.

(e) If we also observe IQ_i, then we could run OLS on the model

$$y_i = \beta_1 + \beta_2 h_i + \beta_3 \text{IQ}_i + u_i,$$

i.e. we include IQ_i as an additional regressor (control variable).

A big concern when estimating the effect of h_i on y_i is unobserved ability, causing an omitted variable bias (endogeneity of h_i due to unobserved ability as a confounding factor). IQ_i is a proxy of unobserved ability, so controlling for IQ_i will mitigate the omitted variable bias (reduce the endogeneity problem), so that the resulting estimate for β_2 is a more plausible estimate of the true causal effect.

Question 2 (20 points)

(a) We have as $n \to \infty$

$$\widehat{\beta} = \frac{\frac{1}{n}\sum_{i=1}^n z_i y_i}{\frac{1}{n}\sum_{i=1}^n z_i x_i} \to_p \frac{\mathbb{E}(z_i y_i)}{\mathbb{E}(z_i x_i)} = \beta^*$$

where we have used the WLLN in both the numerator and the denominator, and we also applied the CMT. By also using the model for y_i and the data generating process for x_i, z_i, u_i we obtain

$$\beta^* = \frac{\mathbb{E}(z_i y_i)}{\mathbb{E}(z_i x_i)} = \frac{\beta \mathbb{E}(z_i x_i) + \mathbb{E}(z_i u_i)}{\mathbb{E}(z_i x_i)} = \beta + \frac{\rho_{zu}}{\rho_{xz}}.$$

(b)
- If $\rho_{zu} = 0$, then $\beta^* = \beta$, so that $\widehat{\beta}$ is consistent for β.
- The condition $\rho_{zu} = 0$ states that z_i is exogenous (uncorrelated with the error u_i), which is also called the exclusion restriction.
- The condition $\rho_{xz} \neq 0$ guarantees that z_i is correlated with x_i, i.e. that the instrument z_i is relevant for the endogenous regressor x_i.

(c) Using the model we find

$$\sqrt{n}\left(\widehat{\beta} - \beta\right) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n z_i u_i}{\frac{1}{n}\sum_{i=1}^n z_i x_i}$$

Since we assume $\rho_{zu} = \mathbb{E}(z_i u_i) = 0$ we can apply the CLT to find that as $n \to \infty$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n z_i u_i \Rightarrow \mathcal{N}(0, \text{Var}(z_i u_i)).$$

As above, by the WLLN we have $\frac{1}{n}\sum_{i=1}^n z_i x_i \to_p \rho_{xz}$. Applying Slutsky's theorem we thus find as $n \to \infty$

$$\sqrt{n}\left(\widehat{\beta} - \beta\right) \Rightarrow \frac{\mathcal{N}(0, \mathrm{Var}(z_i u_i))}{\rho_{xz}} = \mathcal{N}\left(0, \frac{\mathrm{Var}(z_i u_i)}{\rho_{xz}^2}\right),$$

i.e.

$$\mathrm{AsyVar}(\sqrt{n}\widehat{\beta}) = \frac{\mathrm{Var}(z_i u_i)}{\rho_{xz}^2}.$$

Note that $\mathrm{Var}(z_i u_i) = \mathbb{E}[(z_i u_i)^2]$.

It is ok if students stop here. But, since z_i and u_i are normally distributed and uncorrelated they are also independent, so we can further evaluate $\mathrm{Var}(z_i u_i) = \mathbb{E}[(z_i u_i)^2] = \mathbb{E}(z_i^2 u_i^2) = \mathbb{E}(z_i^2)\mathbb{E}(u_i^2) = 1$. We thus find the simplified final result

$$\mathrm{AsyVar}(\sqrt{n}\widehat{\beta}) = \frac{1}{\rho_{xz}^2}.$$

An intermediate level of simplification, which is also correct, would use homoscedasticity and $\sigma_u^2 = 1$ to obtain $\mathrm{AsyVar}(\sqrt{n}\widehat{\beta}) = \frac{\mathbb{E}(z_i^2)}{\rho_{xz}^2}$.

(d) Using the information on the observed sample we calculate

$$\widehat{\beta} = \frac{\frac{1}{n}\sum_{i=1}^n z_i y_i}{\frac{1}{n}\sum_{i=1}^n z_i x_i} = \frac{0.2}{0.2} = 1,$$

$$\widehat{\rho}_{xz} = \frac{1}{n}\sum_{i=1}^n z_i x_i = 0.2.$$

TWO POSSIBLE SOLUTIONS NOW:

(A) If students found the simplified solution in (c) that $\mathrm{AsyVar}(\sqrt{n}\widehat{\beta}) = \frac{1}{\rho_{xz}^2}$, then we have

$$\widehat{\mathrm{AsyVar}}(\sqrt{n}\widehat{\beta}) = \frac{1}{\widehat{\rho}_{xz}^2} = \frac{1}{(0.2)^2} = 25.$$

An estimator for the variance of $\widehat{\beta}$ is thus given by

$$\widehat{\mathrm{Var}}(\widehat{\beta}) = \frac{1}{n}\widehat{\mathrm{AsyVar}}(\sqrt{n}\widehat{\beta}) = \frac{25}{125} = 0.2,$$

implying that $\widehat{\mathrm{std}}(\widehat{\beta}) = \sqrt{0.2} = 0.447$. A 95% confidence interval for β is thus given by

$$\begin{aligned}\mathrm{CI}_{95\%} &= \left[\widehat{\beta} - 1.96 \times \widehat{\mathrm{std}}(\widehat{\beta}), \widehat{\beta} + 1.96 \times \widehat{\mathrm{std}}(\widehat{\beta})\right] = [1 - 0.88, 1 + 0.88] \\ &= [0.12, 1.88].\end{aligned}$$

(B) Alternatively, if students did not find the simplified solution in (c), then we also need to estimate $\mathrm{Var}(z_i u_i)$, which is possible, but more complicated. Define the residuals $\widehat{u}_i = y_i - x_i\widehat{\beta}$. An estimator for $\mathrm{Var}(z_i u_i) = \mathbb{E}[(z_i u_i)^2]$ is then given by

$$\begin{aligned}
\widehat{\mathbb{E}}[(z_i u_i)^2] &= \frac{1}{n}\sum_{i=1}^{n}(z_i\widehat{u}_i)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} z_i^2(y_i - x_i\widehat{\beta})^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} z_i^2 y_i^2 - 2\widehat{\beta}\frac{1}{n}\sum_{i=1}^{n} z_i^2 x_i y_i + \widehat{\beta}^2\frac{1}{n}\sum_{i=1}^{n} z_i^2 y_i^2 \\
&= 3 - 2 \times 0.5 + 3 = 5.
\end{aligned}$$

A consistent estimator for the asymptotic variance is thus given by

$$\widehat{\mathrm{AsyVar}}(\sqrt{n}\widehat{\beta}) = \frac{\widehat{\mathbb{E}}[(z_i u_i)^2]}{\widehat{\rho}_{xz}^2} = \frac{5}{(0.2)^2} = 125.$$

An estimator for the variance of $\widehat{\beta}$ is thus given by

$$\widehat{\mathrm{Var}}(\widehat{\beta}) = \frac{1}{n}\widehat{\mathrm{AsyVar}}(\sqrt{n}\widehat{\beta}) = \frac{125}{125} = 1,$$

implying that $\widehat{\mathrm{std}}(\widehat{\beta}) = 1$. A 95% confidence interval for β is thus given by

$$\begin{aligned}
\mathrm{CI}_{95\%} &= \left[\widehat{\beta} - 1.96 \times \widehat{\mathrm{std}}(\widehat{\beta}), \widehat{\beta} + 1.96 \times \widehat{\mathrm{std}}(\widehat{\beta})\right] = [1 - 1.96, 1 + 1.96] \\
&= [-0.96, 2.96].
\end{aligned}$$

Both solutions (A) and (B) are counted as correct, because both are correctly derived with information given in the question.

Question 3 (20 points)

(a) We have $\widetilde{y}_i = y_i - z_i\Sigma_{zz}^{-1}\Sigma_{zy}$ and $\widetilde{w}_i = w_i - z_i\Sigma_{zz}^{-1}\Sigma_{wz}$. Thus

$$\begin{aligned}
\sum_{i=1}^{n} \widetilde{w}_i^2 &= \sum_{i=1}^{n} \widetilde{w}_i w_i - \sum_{i=1}^{n} \widetilde{w}_i z_i \Sigma_{zz}^{-1}\Sigma_{wz} \\
&= \sum_{i=1}^{n} \widetilde{w}_i w_i - \frac{\Sigma_{wz}}{\Sigma_{zz}} \sum_{i=1}^{n} \widetilde{w}_i z_i, \\
\sum_{i=1}^{n} \widetilde{w}_i \widetilde{y}_i &= \sum_{i=1}^{n} \widetilde{w}_i y_i - \sum_{i=1}^{n} \widetilde{w}_i z_i \Sigma_{zz}^{-1}\Sigma_{zy} \\
&= \sum_{i=1}^{n} \widetilde{w}_i y_i - \frac{\Sigma_{zy}}{\Sigma_{zz}} \sum_{i=1}^{n} \widetilde{w}_i z_i.
\end{aligned}$$

Thus, both of the desired results follow if we can show that $\sum_{i=1}^{n} \widetilde{w}_i z_i = 0$. We find that

$$\begin{aligned}
\sum_{i=1}^{n} \widetilde{w}_i z_i &= \sum_{i=1}^{n} z_i \left(w_i - z_i\Sigma_{zz}^{-1}\Sigma_{wz}\right) \\
&= \Sigma_{wz} - \Sigma_{zz}\Sigma_{zz}^{-1}\Sigma_{wz} = \Sigma_{wz} - \Sigma_{wz} = 0.
\end{aligned}$$

Combining the above we thus indeed find $\sum_{i=1}^{n} \widetilde{w}_i^2 = \sum_{i=1}^{n} \widetilde{w}_i w_i$ and $\sum_{i=1}^{n} \widetilde{w}_i \widetilde{y}_i = \sum_{i=1}^{n} \widetilde{w}_i y_i$.

(b) Using the result in (a) we find

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} \widetilde{w}_i \widetilde{y}_i}{\sum_{i=1}^{n} \widetilde{w}_i^2} = \frac{\sum_{i=1}^{n} \widetilde{w}_i y_i}{\sum_{i=1}^{n} \widetilde{w}_i w_i}.$$

Using the definition of $\widetilde{w}_i$ we thus find

$$\begin{aligned}
\widehat{\beta}_1 &= \frac{\sum_{i=1}^{n} y_i \left(w_i - z_i\Sigma_{zz}^{-1}\Sigma_{wz}\right)}{\sum_{i=1}^{n} w_i \left(w_i - z_i\Sigma_{zz}^{-1}\Sigma_{wz}\right)} = \frac{\Sigma_{wy} - \dfrac{\Sigma_{zy}\Sigma_{wz}}{\Sigma_{zz}}}{\Sigma_{ww} - \dfrac{\Sigma_{wz}^2}{\Sigma_{zz}}} \\
&= \frac{\Sigma_{wy}\Sigma_{zz} - \Sigma_{zy}\Sigma_{wz}}{\Sigma_{ww}\Sigma_{zz} - \Sigma_{wz}^2}.
\end{aligned}$$

(c) Using that $x_i = (w_i, z_i)$ and the inversion formula for 2×2 matrices we find the

OLS estimator that

$$\begin{aligned}
\widehat{\beta}^{\text{OLS}} &= \begin{pmatrix} \widehat{\beta}_1^{\text{OLS}} \\ \widehat{\beta}_2^{\text{OLS}} \end{pmatrix} = \left(\sum_{i=1}^n x_i' x_i \right)^{-1} \sum_{i=1}^n x_i' y_i \\
&= \begin{pmatrix} \Sigma_{ww} & \Sigma_{wz} \\ \Sigma_{wz} & \Sigma_{zz} \end{pmatrix}^{-1} \begin{pmatrix} \Sigma_{wy} \\ \Sigma_{zy} \end{pmatrix} \\
&= \frac{1}{\Sigma_{ww}\Sigma_{zz} - \Sigma_{wz}^2} \begin{pmatrix} \Sigma_{zz} & -\Sigma_{wz} \\ -\Sigma_{wz} & \Sigma_{ww} \end{pmatrix} \begin{pmatrix} \Sigma_{wy} \\ \Sigma_{zy} \end{pmatrix},
\end{aligned}$$

and therefore

$$\widehat{\beta}_1^{\text{OLS}} = \frac{\Sigma_{zz}\Sigma_{wy} - \Sigma_{wz}\Sigma_{zy}}{\Sigma_{ww}\Sigma_{zz} - \Sigma_{wz}^2}.$$

(d) Comparing the above results we find $\widehat{\beta}_1^{\text{OLS}} = \widehat{\beta}_1$, i.e. $\widehat{\beta}_1 - \widehat{\beta}_1^{\text{OLS}} = 0$. Thus, it does not matter whether we regress y_i on x_i and z_i jointly, or whether we partition the regression and first eliminate z_i from the model and afterwards regress $\widetilde{y}_i$ on $\widetilde{x}_i$.

(e) If $z_i = 1$, then $\widetilde{y}_i = y_i - \frac{1}{n}\sum_{i=1}^n y_i$ and $\widetilde{w}_i = w_i - \frac{1}{n}\sum_{i=1}^n w_i$, i.e. $\widetilde{y}_i$ is simply the demeaned y_i, and $\widetilde{w}_i$ the demeaned w_i. Thus $\widetilde{y}_i = \beta_1 \widetilde{w}_i + \widetilde{u}_i$ is obtained from the original model by demeaning, which eliminates the constant z_i.

Section B

Question 4 (20 points)

(a) We find

$$\begin{aligned}
Q_n(\theta) &= \frac{1}{n} \log \prod_{i=1}^n f(y_i|\theta) \\
&= \frac{1}{n} \sum_{i=1}^n \log f(y_i|\theta) \\
&= \frac{1}{n} \sum_{i=1}^n \left[(1-y_i) \log\left(\frac{1}{1+e^{\theta}} \right) + y_i \log\left(\frac{1}{1+e^{-\theta}} \right) \right] \\
&= -\frac{1}{n} \sum_{i=1}^n \left[(1-y_i) \log\left(1+e^{\theta}\right) + y_i \log\left(1+e^{-\theta}\right) \right].
\end{aligned}$$

Taking the derivative wrt θ we obtain

$$\begin{aligned}\frac{dQ_n(\theta)}{d\theta} &= -\frac{1}{n}\sum_{i=1}^{n}\left[(1-y_i)\frac{e^{\theta}}{1+e^{\theta}} - y_i\frac{e^{-\theta}}{1+e^{-\theta}}\right] \\ &= -\frac{1}{n}\sum_{i=1}^{n}\frac{e^{\theta}}{1+e^{\theta}} + \frac{1}{n}\sum_{i=1}^{n} y_i\left(\frac{e^{\theta}}{1+e^{\theta}} + \frac{e^{-\theta}}{1+e^{-\theta}}\right) \\ &= \underbrace{-\frac{e^{\theta}}{1+e^{\theta}}}_{=a(\theta)} + \frac{1}{n}\sum_{i=1}^{n} y_i.\end{aligned}$$

where in the last we used that $\frac{1}{n}\sum_{i=1}^{n} a(\theta) = a(\theta)$ and $\frac{e^{\theta}}{1+e^{\theta}} + \frac{e^{-\theta}}{1+e^{-\theta}} = \frac{e^{\theta}}{1+e^{\theta}} + \frac{1}{e^{\theta}+1} = 1$. Note that $a(\theta)$ can also be rewritten as

$$a(\theta) = -\frac{1}{1+e^{-\theta}}.$$

(b) Using the result in (a) the FOC $\frac{dQ_n(\widehat{\theta})}{d\theta} = 0$ reads

$$\frac{1}{1+\exp(-\widehat{\theta})} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

Thus

$$1+\exp(-\widehat{\theta}) = \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)^{-1},$$

and therefore

$$\widehat{\theta} = -\log\left[\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)^{-1} - 1\right].$$

Alternatively, this can be written as $\widehat{\theta} = \log\left[\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) / \left(1 - \frac{1}{n}\sum_{i=1}^{n} y_i\right)\right]$.

(c) Within the calculation of $\frac{dQ_n(\theta)}{d\theta}$ we have already shown that

$$\frac{d\log f(y_i|\theta)}{d\theta} = -\frac{1}{1+e^{-\theta}} + y_i.$$

Taking one more derivative we obtain

$$\begin{aligned}\frac{d^2\log f(y_i|\theta)}{d\theta^2} &= -\frac{e^{-\theta}}{(1+e^{-\theta})^2} \\ &= -\frac{e^{\theta}}{(1+e^{\theta})^2}.\end{aligned}$$

It is a special feature of this model that this second derivative does not depend on the data anymore. Taking the expectation therefore is trivial, and we obtain

$$\mathbb{E}\left[\frac{d^2 \log f(y_i|\theta)}{d\theta^2}\right] = -\frac{e^\theta}{(1+e^\theta)^2}.$$

From the lecture we know that for a correctly specified likelihood function we can use the information equality to find

$$\text{AsyVar}(\sqrt{n}\widehat{\theta}) = \left\{-\mathbb{E}\left[\frac{d^2 \log f(y_i|\theta)}{d\theta^2}\right]\right\}^{-1},$$

and therefore

$$\text{AsyVar}(\sqrt{n}\widehat{\theta}) = \frac{\left(1+e^\theta\right)^2}{e^\theta} = \left(1+e^\theta\right)\left(1+e^{-\theta}\right).$$

(d) We have

$$\begin{aligned}\mathbb{E}(y_i) &= 0 \times P(y_i = 0 \,|\, \theta) + 1 \times P(y_i = 1 \,|\, \theta) \\ &= P(y_i = 1 \,|\, \theta) = \frac{1}{1+e^{-\theta}},\end{aligned}$$

i.e. we have $b(\theta) = \frac{1}{1+e^{-\theta}}$. The corresponding MM estimator solves the sample analog of this moment condition, i.e.

$$\frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{1+\exp(-\widehat{\theta}_{\text{MM}})}.$$

This is exactly the FOC that was also solved by the MLE $\widehat{\theta}$. We therefore have $\widehat{\theta} = \widehat{\theta}_{\text{MM}}$. Since the two estimators are identical we have $\text{AsyVar}(\sqrt{n}\widehat{\theta}) = \text{AsyVar}(\sqrt{n}\widehat{\theta}_{\text{MM}})$, i.e. the asymptotic variance are also identical.

Question 5 (20 points)

(a)
- A possible moment function reads $g(y_i, \sigma) = y_i^2 - \sigma^2$. However, the moment function is not uniquely determined, it could be multiplied with any non-zero function of σ. The moment function satisfied $\mathbb{E}[g(y_i, \sigma)] = 0$.
- The sample analog of the moment condition reads $\frac{1}{n}\sum_{i=1}^{n} y_i^2 = \widehat{\sigma}^2_{\text{MM}}$. Solving this we find $\widehat{\sigma}_{\text{MM}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^2}$.

– We could apply the CLT and Delta Method to calculate the asymptotic variance directly. But we can also rely on the general MM theory. We have $G = \mathbb{E}[\frac{dg(y_i,\sigma)}{d\sigma}] = -2\sigma$ and $\mathrm{Var}[g(y_i,\sigma)] = \mathbb{E}\left[(y_i^2-\sigma^2)^2\right] = 2\sigma^4$, and applying the general formula thus gives

$$\begin{aligned}\mathrm{AsyVar}(\sqrt{n}\,\widehat{\sigma}_{\mathrm{MM}}) &= \left\{G'(\mathrm{Var}[g(y_i,\sigma)])^{-1}G\right\}^{-1} \\ &= \frac{2\sigma^4}{(-2\sigma)^2} = \frac{\sigma^2}{2}.\end{aligned}$$

(b) The GMM estimator in (a) is actually a MM estimator, because the number of moment conditions is equal to the number of parameters to estimate. For MM estimators, if a solution of the sample analog of the moment conditions exists (as is the case in (a)), then this solution also minimizes the GMM objective function, independent of the choice of moment function. Thus, **the choice of weight function is <u>not</u> important in (a)**. This also explains why in (a) we could use the simplified formula for $\mathrm{AsyVar}(\sqrt{n}\,\widehat{\sigma}_{\mathrm{GMM}})$, which for GMM estimators only holds when an asymptotically optimal weight matrix is chosen.

(c) A possible choice for $g(y_i,\sigma)$ is

$$g(y_i,\sigma) = \begin{pmatrix} y_i^2-\sigma^2 \\ y_i^4-3\sigma^4 \end{pmatrix}.$$

Again, the vector of moment functions is not uniquely determined, since its components could be multiplied by any non-zero function of σ.

The GMM objective function is given by

$$\begin{aligned}Q_n(\sigma) &= \left[\frac{1}{n}\sum_{i=1}^n g(y_i,\sigma)\right]' W \left[\frac{1}{n}\sum_{i=1}^n g(y_i,\sigma)\right] \\ &= \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n y_i^2-\sigma^2 \\ \frac{1}{n}\sum_{i=1}^n y_i^4-3\sigma^4 \end{pmatrix}' W \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n y_i^2-\sigma^2 \\ \frac{1}{n}\sum_{i=1}^n y_i^4-3\sigma^4 \end{pmatrix},\end{aligned}$$

where W is a symmetric positive definite 2×2 weight matrix. The GMM estimator is obtained by minimizing this objective function, i.e.

$$\widehat{\sigma}_{\mathrm{GMM}} = \arg\min_{\sigma>0} Q_n(\sigma).$$

(d) – Yes, **now the choice of weight matrix <u>is</u> important**, because the number of moment conditions is large than the number of parameters to estimate, so that $\widehat{\sigma}_{\mathrm{GMM}}$ depends on W.

– From the lecture we know that the asymptotic variance of the GMM estimator is minimized if we choose the weight matrix (asymptotically equal to)

$$W = (\mathrm{Var}[g(y_i,\sigma)])^{-1}.$$

We calculate

$$\begin{aligned}
\mathrm{Var}[g(y_i,\sigma)] &= \mathbb{E}\left[g(y_i,\sigma)g(y_i,\sigma)'\right] \\
&= \begin{pmatrix} \mathbb{E}\left[\left(y_i^2-\sigma^2\right)^2\right] & \mathbb{E}\left[\left(y_i^2-\sigma^2\right)\left(y_i^4-3\sigma^2\right)\right] \\ \mathbb{E}\left[\left(y_i^2-\sigma^2\right)\left(y_i^4-3\sigma^2\right)\right] & \mathbb{E}\left[\left(y_i^4-3\sigma^2\right)^2\right] \end{pmatrix} \\
&= \begin{pmatrix} 2\sigma^4 & 12\sigma^6 \\ 12\sigma^6 & 96\sigma^8 \end{pmatrix},
\end{aligned}$$

and therefore for the optimal weight matrix

$$W = (\mathrm{Var}[g(y_i,\sigma)])^{-1} = \begin{pmatrix} \frac{2}{\sigma^4} & -\frac{1}{4\sigma^6} \\ -\frac{1}{4\sigma^6} & \frac{1}{24\sigma^8} \end{pmatrix}.$$

However, the weight matrix could be multiplied by an arbitrary positive constant, without changing the GMM estimator.

(e) We calculate

$$G = \mathbb{E}\left[\frac{dg(y_i,\sigma)}{d\sigma}\right] = \begin{pmatrix} -2\sigma \\ -12\sigma^3 \end{pmatrix}.$$

Therefore

$$\begin{aligned}
\mathrm{AsyVar}(\sqrt{n}\,\widehat{\sigma}_{\mathrm{GMM}}) &= \left\{G'(\mathrm{Var}[g(y_i,\sigma)])^{-1}G\right\}^{-1} \\
&= \left\{\begin{pmatrix} -2\sigma \\ -12\sigma^3 \end{pmatrix}\begin{pmatrix} \frac{2}{\sigma^4} & -\frac{1}{4\sigma^6} \\ -\frac{1}{4\sigma^6} & \frac{1}{24\sigma^8} \end{pmatrix}\begin{pmatrix} -2\sigma \\ -12\sigma^3 \end{pmatrix}\right\}^{-1} \\
&= \left\{\frac{2}{\sigma^2}\right\}^{-1} = \frac{\sigma^2}{2}.
\end{aligned}$$

We thus have $\mathrm{AsyVar}(\sqrt{n}\,\widehat{\sigma}_{\mathrm{GMM}}) = \mathrm{AsyVar}(\sqrt{n}\,\widehat{\sigma}_{\mathrm{MM}})$, i.e. using the additional moment condition did not improve the asymptotic variance, so in large samples there is no efficiency gain from using the additional moment condition (there is, in fact, the disadvantage that the additional moment condition only holds under normality of y_i, i.e. is not robust towards violation of the normality assumption).

This result makes sense, because $\widehat{\sigma}_{\mathrm{MM}}$ is in fact equal to the MLE for σ, so it should not be possible to improve on its asymptotic variance.

Question 6 (20 points)

(a) We have $\mathbb{E}y_t = 0$, i.e. the autocovariances can simply be calculated as second moment. We find

$$\begin{aligned}
\gamma_0 &= \mathbb{E}(y_t^2) = \mathbb{E}[(\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2})^2] \\
&= \mathbb{E}(\varepsilon_t^2) + \theta_1^2\mathbb{E}(\varepsilon_{t-1}^2) + \theta_2^2\mathbb{E}(\varepsilon_{t-2}^2) \\
&= \sigma^2\left(1 + \theta_1^2 + \theta_2^2\right), \\
\gamma_1 &= \mathbb{E}(y_t y_{t-1}) = \mathbb{E}[(\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2})(\varepsilon_{t-1} + \theta_1\varepsilon_{t-2} + \theta_2\varepsilon_{t-3})] \\
&= \theta_1\mathbb{E}(\varepsilon_{t-1}^2) + \theta_1\theta_2\mathbb{E}(\varepsilon_{t-2}^2) \\
&= \sigma^2\theta_1(1 + \theta_2), \\
\gamma_2 &= \mathbb{E}(y_t y_{t-2}) = \mathbb{E}[(\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2})(\varepsilon_{t-2} + \theta_1\varepsilon_{t-3} + \theta_2\varepsilon_{t-4})] \\
&= \theta_2\mathbb{E}(\varepsilon_{t-2}^2) \\
&= \sigma^2\theta_2, \\
\gamma_3 &= \mathbb{E}(y_t y_{t-3}) = \mathbb{E}[(\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2})(\varepsilon_{t-3} + \theta_1\varepsilon_{t-4} + \theta_2\varepsilon_{t-5})] \\
&= 0.
\end{aligned}$$

Using this we find for the autocorrelations

$$\begin{aligned}
\rho_1 &= \frac{\gamma_1}{\gamma_0} = \frac{\theta_1(1+\theta_2)}{1+\theta_1^2+\theta_2^2}, \\
\rho_2 &= \frac{\gamma_2}{\gamma_0} = \frac{\theta_2}{1+\theta_1^2+\theta_2^2}, \\
\rho_3 &= \frac{\gamma_3}{\gamma_0} = 0.
\end{aligned}$$

(b) Applying the WLLN (for time series data) and the continuous mapping theorem we find that as $T \to \infty$

$$\widehat{\rho} = \frac{\frac{1}{T}\sum_{t=1}^T y_t y_{t-1}}{\frac{1}{T}\sum_{t=1}^T y_{t-1}^2} \to_p \frac{\mathbb{E}y_t y_{t-1}}{\mathbb{E}y_{t-1}^2} = \frac{\gamma_1}{\gamma_0} = \rho_1.$$

(c) For $\theta = \theta_1 = \theta_2$ we have

$$\rho_1 = \frac{\theta(1+\theta)}{1+2\theta^2},$$

For $\widehat{\rho} = 1/2$ we thus have to solve

$$\frac{1}{2} = \frac{\widehat{\theta}\left(1+\widehat{\theta}\right)}{1+2\widehat{\theta}^2},$$

which cal also be written as

$$1 + 2\widehat{\theta}^2 = 2\widehat{\theta} + 2\widehat{\theta}^2,$$

i.e.

$$1 = 2\widehat{\theta},$$

and therefore

$$\widehat{\theta} = \frac{1}{2}.$$

Since $\widehat{\rho}$ is a consistent estimator for ρ_1 and the relationship between ρ_1 and θ is continuous, we know that $\widehat{\theta}$ is a consistent estimator for θ by the CMT.

(d) The relationship between $\widehat{\rho}$ and $\widehat{\theta}$ is given by $\widehat{\rho} = \frac{\widehat{\theta}\left(1+\widehat{\theta}\right)}{1+2\widehat{\theta}^2}$. Differentiating this and evaluating at $\widehat{\theta} = 0.5$ gives

$$\frac{d\widehat{\rho}}{d\widehat{\theta}} = \frac{1 + 2\widehat{\theta} - 2\widehat{\theta}^2}{\left(1 + 2\widehat{\theta}^2\right)^2} = \frac{2}{3}.$$

Thus, according to the delta-method we have

$$\widehat{\mathrm{std}}(\widehat{\theta}) = \left(\frac{d\widehat{\rho}}{d\widehat{\theta}}\right)^{-1} \widehat{\mathrm{std}}(\widehat{\rho}) = \frac{3}{2}\,0.02 = 0.03.$$

SOLUTIONS

G020: Examination in Econometrics

2015-2016

Section A

Question A1 (20 points)

(a) Let $x_i = (1, w_i, z_i)$. From (1) we know that

$$\frac{1}{n}\sum_{i=1}^{n} x_i' x_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 16 \end{pmatrix}, \qquad \frac{1}{n}\sum_{i=1}^{n} x_i' y_i = \begin{pmatrix} 10 \\ 9 \\ -8 \end{pmatrix}.$$

The OLS estimator thus reads

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i' x_i\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} x_i' y_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/9 & 0 \\ 0 & 0 & 1/16 \end{pmatrix} \begin{pmatrix} 10 \\ 9 \\ -8 \end{pmatrix} = \begin{pmatrix} 10 \\ 1 \\ -1/2 \end{pmatrix},$$

i.e. $\hat{\beta}_1 = 10$, $\hat{\beta}_2 = 1$ and $\hat{\beta}_3 = -1/2$.

(b) Solving the system of two equations

$$v_i = w_i + z_i,$$
$$q_i = w_i - z_i,$$

for w_i and z_i gives

$$w_i = \frac{1}{2}(v_i + q_i),$$
$$z_i = \frac{1}{2}(v_i - q_i)$$

Plugging this into the original regression model gives

$$\begin{aligned} y_i &= \beta_1 + w_i\beta_2 + z_i\beta_3 + u_i \\ &= \beta_1 + \frac{1}{2}(v_i + q_i)\beta_2 + \frac{1}{2}(v_i - q_i)\beta_3 + u_i \\ &= \beta_1 + v_i \left[\frac{1}{2}(\beta_2 + \beta_3)\right] + q_i \left[\frac{1}{2}(\beta_2 - \beta_3)\right] + u_i, \end{aligned}$$

and therefore

$$\gamma_1 = \beta_1,$$
$$\gamma_2 = \frac{1}{2}(\beta_2 + \beta_3),$$
$$\gamma_3 = \frac{1}{2}(\beta_2 - \beta_3).$$

For the estimates we thus find

$$\hat{\gamma}_1 = \hat{\beta}_1 = 10,$$
$$\hat{\gamma}_2 = \frac{1}{2}(\hat{\beta}_2 + \hat{\beta}_3) = 1/4,$$
$$\hat{\gamma}_3 = \frac{1}{2}(\hat{\beta}_2 - \hat{\beta}_3) = 3/4.$$

(c) We have $\mathbb{E}(u_i^2|x_i) = \sigma^2 = 36$ and $n = 25$, and therefore

$$\widehat{\mathrm{Var}}(\hat{\beta}) = \sigma^2 \left(\sum_{i=1}^n x_i' x_i \right)^{-1} = \frac{\sigma^2}{n} \left(\frac{1}{n} \sum_{i=1}^n x_i' x_i \right)^{-1} = \frac{36}{25} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/9 & 0 \\ 0 & 0 & 1/16 \end{pmatrix}$$
$$= \begin{pmatrix} 36/25 & 0 & 0 \\ 0 & 4/25 & 0 \\ 0 & 0 & 9/100 \end{pmatrix}.$$

and therefore

$$\widehat{\mathrm{std}}(\hat{\beta}_2) = \sqrt{4/25} = 2/5 = 0.4,$$
$$\widehat{\mathrm{std}}(\hat{\beta}_3) = \sqrt{9/100} = 3/10 = 0.3,$$
$$\widehat{\mathrm{std}}(\hat{\beta}_2 - \hat{\beta}_3) = \sqrt{9/100 + 4/25 - 2 \times 0} = 1/2 = 0.5,$$

where in the last line we used that $\widehat{\mathrm{Cov}}(\hat{\beta}_2, \hat{\beta}_3) = 0$, because the off-diagonal elements in the estimated variance-covariance matrix $\widehat{\mathrm{Var}}(\hat{\beta})$ are zero.

(d) The null-hypothesis can be rewritten as $H_0 : \beta_2 - \beta_3$. The corresponding $t-$statistic reads

$$t = \frac{\hat{\beta}_2 - \hat{\beta}_3}{\widehat{\mathrm{std}}(\hat{\beta}_2 - \hat{\beta}_3)} = 3.$$

Since $|t| > 1.96$ we reject H_0 at 95% confidence level.

(e) We now have

$$\frac{1}{n}\sum_{i=1}^{n} x_i' x_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 9 & -12 \\ 0 & -12 & 16 \end{pmatrix}.$$

This is a block-diagonal matrix, so we can invert it by inverting each of the diagonal blocks. However, the matrix $\begin{pmatrix} 9 & -12 \\ -12 & 16 \end{pmatrix}$ is not invertible (its determinant equal zero), implying that $\sum_{i=1}^{n} x_i' x_i$ is also not invertible here. This means that there is a perfect collinear relationship between the regressors (we have $4w_i - 3z_i = 0$), and the OLS estimator is undefined in that case.

Question A2 (20 points)

(a) Applying the WLLN and CMT we find that as $n \to \infty$ we have

$$\hat{\beta}^* = \frac{\frac{1}{n}\sum_{i=1}^n z_i y_i}{\frac{1}{n}\sum_{i=1}^n z_i^2} \to_p \frac{\mathbb{E} z_i y_i}{\mathbb{E} z_i^2} = \beta^*.$$

Plugging in the model for y_i we obtain

$$\begin{aligned}
\beta^* = \frac{\mathbb{E} z_i y_i}{\mathbb{E} z_i^2} &= \frac{\mathbb{E} z_i (z_i \beta + w_i \gamma + u_i)}{\mathbb{E} z_i^2} \\
&= \frac{\mathbb{E} z_i^2}{\mathbb{E} z_i^2}\beta + \frac{\mathbb{E} z_i w_i}{\mathbb{E} z_i^2}\gamma + \frac{\mathbb{E} z_i u_i}{\mathbb{E} z_i^2} \\
&= \beta + \rho_{ZW}\gamma.
\end{aligned}$$

(b) In that case $\rho_{ZW}\gamma \neq 0$ and therefore $\beta^* \neq \beta$, that is, $\hat{\beta}^*$ is <u>not</u> consistent for β. In contrast, $\hat{\beta}$ is a consistent estimator for β (all regularity for OLS consistency are imposed in the question). Therefore, we prefer to use $\hat{\beta}$, because it is consistent.

(c) We now have $\rho_{ZW}\gamma = 0$, and therefore $\beta^* = \beta$, so both estimators $\hat{\beta}^*$ and $\hat{\beta}$ are consistent. We now calculate the asymptotic variance of both estimators for the case $\rho_{ZW}\gamma = 0$ (covering both the case in question (c) and (d)):

- From the lecture we know that under homoscedasticity (imposed in the question), as $n \to \infty$, we have

$$\begin{aligned}
\text{AsyVar}\left[\sqrt{n}\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix}\right] &= \sigma^2 \left(\mathbb{E}\, x_i' x_i\right)^{-1} = \sigma^2 \begin{pmatrix} 1 & \rho_{ZW} \\ \rho_{ZW} & 1 \end{pmatrix}^{-1} \\
&= \frac{\sigma^2}{1-\rho_{ZW}^2}\begin{pmatrix} 1 & -\rho_{ZW} \\ -\rho_{ZW} & 1 \end{pmatrix},
\end{aligned}$$

  and therefore

$$\text{AsyVar}\left(\sqrt{n}\hat{\beta}\right) = \frac{\sigma^2}{1-\rho_{ZW}^2}.$$

- For $\hat{\beta}^*$ we note that we can write the model as $y_i = z_i\beta + \epsilon_i$, where $\epsilon_i = w_i\gamma + u_i$ is uncorrelated with z_i and satisfies $\mathbb{E}(\epsilon_i^2|z_i) = \gamma^2\mathbb{E}(w_i^2|z_i) + \mathbb{E}(u_i^2|z_i) = \gamma^2 + \sigma^2$, where we used that $\mathbb{E}(w_i u_i|z_i) = 0$. Using again the general result on the asymptotic variance of an OLS estimator under homoscedasticity we thus find

$$\text{AsyVar}\left(\sqrt{n}\hat{\beta}^*\right) = \frac{\mathbb{E}(\epsilon_i^2)}{\mathbb{E}(z_i^2)} = \gamma^2 + \sigma^2.$$

$\Rightarrow$ for question (c) we have $\rho_{ZW} = 0$, and therefore

$$\text{AsyVar}\left(\sqrt{n}\hat{\beta}\right) = \sigma^2$$
$$\text{AsyVar}\left(\sqrt{n}\hat{\beta}^*\right) = \gamma^2 + \sigma^2.$$

Because $\gamma \neq 0$ this implies

$$\text{AsyVar}\left(\sqrt{n}\hat{\beta}\right) < \text{AsyVar}\left(\sqrt{n}\hat{\beta}^*\right).$$

Thus, we would recommend to use $\hat{\beta}$ in this case, because it is more efficient.

(d) We now have $\gamma = 0$, and therefore

$$\text{AsyVar}\left(\sqrt{n}\hat{\beta}\right) = \frac{\sigma^2}{1 - \rho_{ZW}^2}$$
$$\text{AsyVar}\left(\sqrt{n}\hat{\beta}^*\right) = \sigma^2.$$

Because $\rho_{ZW} \neq 0$ this implies

$$\text{AsyVar}\left(\sqrt{n}\hat{\beta}\right) > \text{AsyVar}\left(\sqrt{n}\hat{\beta}^*\right).$$

Thus, we would recommend to use $\hat{\beta}^*$ in this case, because it is more efficient.

(e) Intuitive Explanation:

$\Rightarrow$ (c): If a relevant omitted regressor is not correlated with remaining regressors, then the OLS estimates for those remaining regressors are still consistent, but are inefficient, **because the omitted regressor could have helped to reduce the variance of the error term.**

$\Rightarrow$ (d): Inclusion of an irrelevant regressor gives larger OLS standard errors, if the irrelevant regressor is correlated with the remaining regressors, **because the irrelevant regressor "absorbs" part of the variation of the other regressors, thus reducing their explanatory power.**

Question A3 (20 points)

(a) – Derivation for $\hat{\beta}_{\text{IV}}$: Using the model for y_i we find

$$\sqrt{n}\left(\hat{\beta}_{\text{IV}}-\beta\right)=\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_{i1}u_i}{\frac{1}{n}\sum_{i=1}^{n} z_{i1}x_i}.$$

Applying the CLT we find, as $n\to\infty$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_{i1}u_i \Rightarrow \mathcal{N}(0,\mathbb{E}(z_{i1}^2u_i^2)),$$

and using the homoscedasticity assumption we can simplify

$$\mathbb{E}(z_{i1}^2u_i^2)=\mathbb{E}(z_{i1}^2\mathbb{E}(u_i^2|z_i))=\sigma^2\mathbb{E}(z_{i1}^2).$$

Applying the WLLN we find, as $n\to\infty$,

$$\frac{1}{n}\sum_{i=1}^{n} z_{i1}x_i \to_p \mathbb{E}(z_{i1}x_i).$$

Combining the above results and applying Slutsky's theorem, we obtain

$$\sqrt{n}\left(\hat{\beta}_{\text{IV}}-\beta\right)\Rightarrow\mathcal{N}(0,V_{\text{IV}}),\qquad V_{\text{IV}}=\frac{\sigma^2\,\mathbb{E}(z_{i1}^2)}{[\mathbb{E}(z_{i1}x_i)]^2}.$$

– Derivation for $\hat{\beta}_{\text{2SLS}}$: Using the model for y_i we find

$$\sqrt{n}\left(\hat{\beta}_{\text{2SLS}}-\beta\right)=\frac{\left(\frac{1}{n}\sum_{i=1}^{n} x_iz_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} z_i'z_i\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i'u_i\right)}{\left(\frac{1}{n}\sum_{i=1}^{n} x_iz_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} z_i'z_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} z_i'x_i\right)}.$$

As above, we apply CLT and WLLN to find, $n\to\infty$,

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i'u_i &\Rightarrow \mathcal{N}(0,\mathbb{E}(u_i^2z_i'z_i)),\\
\frac{1}{n}\sum_{i=1}^{n} x_iz_i &\to_p \mathbb{E}(x_iz_i),\\
\frac{1}{n}\sum_{i=1}^{n} z_i'z_i &\to_p \mathbb{E}(z_i'z_i),\\
\frac{1}{n}\sum_{i=1}^{n} z_i'x_i &\to_p \mathbb{E}(z_i'x_i).
\end{aligned}$$

Applying CMT and Slutsky's theorem, we thus obtain

$$\sqrt{n}\left(\hat{\beta}_{2\text{SLS}} - \beta\right) \Rightarrow \mathcal{N}(0, V_{2\text{SLS}}),$$
$$V_{2\text{SLS}} = \frac{\mathbb{E}(x_i z_i)[\mathbb{E}(z_i' z_i)]^{-1}\mathbb{E}(u_i^2 z_i' z_i)[\mathbb{E}(z_i' z_i)]^{-1}\mathbb{E}(z_i' x_i)}{\{\mathbb{E}(x_i z_i)[\mathbb{E}(z_i' z_i)]^{-1}\mathbb{E}(z_i' x_i)\}^2}.$$

Using the homoscedasticity assumption we can simplify

$$\mathbb{E}(u_i^2 z_i' z_i) = \mathbb{E}(\mathbb{E}(u_i^2|z_i) z_i' z_i) = \sigma^2 \mathbb{E}(z_i' z_i),$$

and using this and the assumption that $\mathbb{E}(z_i' z_i)$ is a diagonal matrix, we can simplify the asymptotic variance as follows

$$\begin{aligned} V_{2\text{SLS}} &= \frac{\sigma^2 \mathbb{E}(x_i z_i)[\mathbb{E}(z_i' z_i)]^{-1}\mathbb{E}(z_i' x_i)}{\{\mathbb{E}(x_i z_i)[\mathbb{E}(z_i' z_i)]^{-1}\mathbb{E}(z_i' x_i)\}^2} \\ &= \frac{\sigma^2}{\mathbb{E}(x_i z_i)[\mathbb{E}(z_i' z_i)]^{-1}\mathbb{E}(z_i' x_i)} \\ &= \sigma^2 \left(\frac{[\mathbb{E}(x_i z_{i1})]^2}{\mathbb{E}(z_{i1}^2)} + \frac{[\mathbb{E}(x_i z_{i2})]^2}{\mathbb{E}(z_{i2}^2)} \right)^{-1}. \end{aligned}$$

(b) For $\mathbb{E}(z_{i1} x_i) \neq 0$ and $\mathbb{E}(z_{i2} x_i) = 0$ we find

$$V_{2\text{SLS}} = V_{\text{IV}} = \frac{\sigma^2\, \mathbb{E}(z_{i1}^2)}{[\mathbb{E}(z_{i1} x_i)]^2},$$

that is, the asymptotic variances are identical in that case. Thus, in very large samples it does not matter which estimator to use. However, in small samples **the estimator $\hat{\beta}_{\text{IV}}$ is preferable, because using the irrelevant instrument z_{i2} introduced additional noise into the estimator $\hat{\beta}_{2\text{SLS}}$, so that we expect the finite sample variance of $\hat{\beta}_{2\text{SLS}}$ to be larger than that of $\hat{\beta}_{\text{IV}}$.**

(c) Overidentification test: Starting from either $\hat{\beta} = \hat{\beta}_{\text{IV}}$ or $\hat{\beta} = \hat{\beta}_{2\text{SLS}}$ (it should not matter which estimator) we construct the residuals $\hat{u}_i = y_i - x_i \hat{\beta}$. We can then construct a test statistic in different ways (students only need to explain one):

- Run a regression of $\hat{u}_i$ on z_i. Use the R^2 of this regression as a test statistic. (standard formulation of Sargan test)
- Run a regression of $\hat{u}_i$ on z_i, call the resulting OLS estimates $\hat{\gamma}$. As test statistic we can use $\hat{\gamma}' \hat{V}^{-1} \hat{\gamma}$, with $\hat{V} = \hat{\sigma}^2 (\sum_i z_i' z_i)^{-1}$, where $\hat{\sigma}^2 = \frac{1}{n} \sum_i \hat{u}_i^2$.
- Directly construct $\frac{n}{\hat{\sigma}^2} (\sum_i \hat{u}_i z_i)(\sum_i z_i' z_i)^{-1}(\sum_i z_i' \hat{u}_i)$ as a test statistic.

$\Rightarrow$ all those test statistics should be distributes as $\chi^2(1)$ under the null hypothesis $H_0 : \mathbb{E}(u_i z_i) = 0$. We reject H_0 if the test statistic is larger than the appropriate quantile of $\chi^2(1)$. Note that this is χ^2 with one degree of freedom, because we have $L = 2$ instruments and $K = 1$ regressors, and $L - K = 1$.

(d) $\mathbb{E}(z_{i1} x_i) \neq 0$ and $\mathbb{E}(z_{i2} x_i) = 0$ is **not a problem** for the overidentification test in (c), it still has power to reject H_0 if it is violated. In this case the second instrument is irrelevant, so we can only use the information from the first instrument to estimate β, but afterwards we still have two instruments that can be used together to perform a nontrivial test of whether $H_0 : \mathbb{E}(u_i z_i) = 0$ is satisfied for both instruments.

Section B

Answer **TWO questions from this Section**.

Question B1 (20 points)

(a) We define the weights

$$w(x_i) = \frac{1}{h(x_i)}.$$

Notice that those weights could be multiplied with any constant, without changing anything in the following. The WLS estimator reads

$$\hat{\beta}_{\text{WLS}} = \frac{\sum_{i=1}^{n} w(x_i)\ x_i y_i}{\sum_{i=1}^{n} w(x_i)\ x_i^2}$$

As usual, using the model, CLT, WLLN and Slutsky's theorem, we find, as $n \to \infty$,

$$\sqrt{n}\left(\hat{\beta}_{\text{WLS}} - \beta\right) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} w(x_i)\ x_i u_i}{\frac{1}{n} \sum_{i=1}^{n} w(x_i)\ x_i^2} \Rightarrow \frac{\mathcal{N}[0, \mathbb{E}(w(x_i)^2\ x_i^2 u_i^2)]}{\mathbb{E}[w(x_i)\ x_i^2]} \sim \mathcal{N}\left(0, V_{\text{WLS}}\right),$$

where we find and simplify the asymptotic variance as follows

$$V_{\text{WLS}} = \frac{\mathbb{E}(w(x_i)^2\ x_i^2 u_i^2)}{\left\{\mathbb{E}[w(x_i)\ x_i^2]\right\}^2} = \frac{\mathbb{E}[w(x_i)^2\ x_i^2 \mathbb{E}(u_i^2|x_i)]}{\left\{\mathbb{E}[w(x_i)\ x_i^2]\right\}^2} = \frac{\mathbb{E}[w(x_i)\ x_i^2]}{\left\{\mathbb{E}[w(x_i)\ x_i^2]\right\}^2} = \left\{\mathbb{E}[w(x_i)\ x_i^2]\right\}^{-1}$$
$$= \left\{\mathbb{E}\left[\frac{x_i^2}{h(x_i)}\right]\right\}^{-1}.$$

(b) Using the model, the law of iterated expectations, and the assumption $\mathbb{E}(u_i|x_i) = 0$, we find

$$\mathbb{E}[(y_i - x_i\beta) f(x_i)] = \mathbb{E}[u_i f(x_i)] = \mathbb{E}[\mathbb{E}(u_i|x_i) f(x_i)] = \mathbb{E}[0 \times f(x_i)]$$
$$= 0.$$

The sample analog of this moment condition reads

$$\frac{1}{n} \sum_{i=1}^{n} [(y_i - x_i \hat{\beta}_{\text{MM}}) f(x_i)] = 0.$$

Solving this for $\hat{\beta}_{\text{MM}}$ gives

$$\hat{\beta}_{\text{MM}} = \frac{\sum_{i=1}^{n} y_i\ f(x_i)}{\sum_{i=1}^{n} x_i\ f(x_i)}.$$

Consistency can be shown in the usual way by applying WLLN and CMT to (the question, however, did not ask to formally show this)

$$\hat{\beta}_{\mathrm{MM}} - \beta = \frac{\frac{1}{n}\sum_{i=1}^{n} u_i \, f(x_i)}{\frac{1}{n}\sum_{i=1}^{n} x_i \, f(x_i)} \to_p \frac{\mathbb{E}[u_i \, f(x_i)]}{\mathbb{E}[x_i \, f(x_i)]} = 0,$$

where in the last step we need the **additional conditions that:**

$$\mathbb{E}[x_i \, f(x_i)] \neq 0.$$

(c) By choosing

$$f(x_i) = \frac{x_i}{h(x_i)},$$

we obtain $\hat{\beta}_{\mathrm{MM}} = \hat{\beta}_{\mathrm{WLS}}$. Because we know that $\hat{\beta}_{\mathrm{WLS}}$ is efficient, the same is true for $\hat{\beta}_{\mathrm{MM}}$ with this choice of $f(x_i)$.

(d) For the $h(x_i) = (a + b\,x_i^2 + c\,x_i^4)^{-1}$ the function $f(x_i)$ that would deliver the efficient MM estimator according to part (c) reads

$$f^*(x_i) = a\,x_i + b\,x_i^3 + c\,x_i^5.$$

However, this MM estimator is not feasible, because a, b, c are unknown. Therefore, we choose $L = 3$ moment conditions corresponding to

$$f_1(x_i) = x_i, \qquad f_2(x_i) = x_i^3, \qquad f_3(x_i) = x_i^5.$$

The moment conditions $\mathbb{E}[(y_i - x_i\beta) f_\ell(x_i)] = 0$, $\ell \in \{1, \ldots, L\}$, can be written as $\mathbb{E} m_i(\beta) = 0$, where

$$m_i(\beta) = \begin{pmatrix} (y_i - x_i\beta)x_i \\ (y_i - x_i\beta)x_i^3 \\ (y_i - x_i\beta)x_i^5 \end{pmatrix},$$

and we have

$$\mathbb{E}[(y_i - x_i\beta) f^*(x_i)] = \mathbb{E}\left[(a, b, c)\; m_i(\beta)\right].$$

Thus, the infeasible optimal moment condition $\mathbb{E}[(y_i - x_i\beta) f^*(x_i)] = 0$ is a linear combination of our feasible vector of moment conditions. The GMM estimator obtained from $\mathbb{E}m_i(\beta) = 0$ is therefore equivalent to the MM estimator obtained from $\mathbb{E}[(y_i - x_i\beta)f^*(x_i)] = 0$ if we choose the GMM weight matrix

$$W = (a, b, c)'(a, b, c).$$

Since two-step efficient GMM minimizes the asymptotic variance of the GMM estimator over all possible choices of weight matrix W, it must be at least as good as the MM estimator obtained from $\mathbb{E}[(y_i - x_i\beta)f^*(x_i)] = 0$. However, since the latter MM estimator is already efficient, it must be the case that two-step efficient GMM is asymptotically efficient as well.

The student answer need not be as detailed as the one provided here, but should contain the key idea, in particular the choice of $f_\ell(x_i)$ and the connection to the MM estimator obtained from $f^*(x_i)$. Notice also that we can add additional moment conditions to $\mathbb{E}m_i(\beta) = 0$, and the two-step efficient GMM estimator will still be efficient.

(e) If $u_i \,|\, x_i \sim t(5)$, then $\hat{\beta}_{\text{WLS}}$ is <u>not</u> efficient anymore, because the MLE has a smaller asymptotic variance. In (a) we assumed $u_i \,|\, x_i \sim \mathcal{N}(0, h(x_i))$, in which case the MLE is equal to the WLS estimator, but for non-normal error this is not the case anymore.

Question B2 (20 points)

(Notice that in the question $f(y_i|\theta)$ refers to the log likelihood, but in the following solution we write $\log f(y_i|\theta)$ for the log likelihood.)

(a) From the log-likelihood

$$\log f(y_i|\theta) = [\log \theta]^{1(y_i=0)} \; [\log \theta]^{1(y_i=1)} \; [\log(1-2\theta)]^{1(y_i=2)},$$

we calculate the score

$$\begin{aligned} s(y_i, \theta) &= \frac{\partial \log f(y_i|\theta)}{\partial \theta} \\ &= \frac{1(y_i=0)}{\theta} + \frac{1(y_i=1)}{\theta} - \frac{2\,1(y_i=2)}{1-2\theta}, \end{aligned}$$

and the Hessian

$$\begin{aligned} H(y_i, \theta) &= \frac{\partial^2 \log f(y_i|\theta)}{\partial \theta^2} \\ &= -\frac{1(y_i=0)}{\theta^2} - \frac{1(y_i=1)}{\theta^2} - \frac{4\,1(y_i=2)}{(1-2\theta)^2}. \end{aligned}$$

Thus, the expected Hessian evaluated at the true parameter reads

$$\begin{aligned} \mathbb{E}[H(y_i, \theta_0)] &= -\frac{1}{\theta_0} - \frac{1}{\theta_0} - \frac{4}{1-2\theta_0} = -\left(\frac{2}{\theta_0} + \frac{4}{1-2\theta_0}\right) \\ &= -\frac{2}{\theta_0(1-2\theta_0)}. \end{aligned}$$

From the lecture we know that for a correctly specified MLE we have

$$\begin{aligned} V_{\text{MLE}} &= -\frac{1}{\mathbb{E}[H(y_i, \theta_0)]} \\ &= \frac{1}{2}\theta_0(1-2\theta_0). \end{aligned}$$

(b) With $W = \mathbb{1}_2$ the GMM objective function reads

$$\sum_{i=1}^{n} [1(y_i=0) - \theta]^2 + \sum_{i=1}^{n} [1(y_i=1) - \theta]^2.$$

Minimizing this gives the following FOC for $\hat{\theta}_{\text{GMM}} = \hat{\theta}_{\text{GMM}}(\mathbb{1}_2)$

$$\sum_{i=1}^{n} \left[1(y_i=0) - \hat{\theta}_{\text{GMM}}\right] + \sum_{i=1}^{n} \left[1(y_i=1) - \hat{\theta}_{\text{GMM}}\right] = 0,$$

which we solve to find

$$\hat{\theta}_{\text{GMM}} = \frac{1}{n}\sum_{i=1}^{n} \frac{1(y_i = 0) + 1(y_i = 1)}{2}.$$

By the CLT we thus have $\sqrt{n}\left(\hat{\theta}_{\text{GMM}} - \theta_0\right) \Rightarrow \mathcal{N}(0, V_{\text{GMM}})$, where according to the CLT we also have

$$\begin{aligned}
V_{\text{GMM}} = V_{\text{GMM}}(\mathbb{1}_2) &= \text{Var}\left(\frac{1(y_i = 0) + 1(y_i = 1)}{2}\right) \\
&= \frac{1}{4}\left\{\text{Var}[1(y_i = 0)] + \text{Var}[1(y_i = 1)] + 2\text{Cov}[1(y_i = 0), 1(y_i = 1)]\right\} \\
&= \frac{1}{4}\left\{2\theta_0(1-\theta_0) - 2\theta^2\right\} \\
&= \frac{1}{2}\theta_0(1 - 2\theta_0)
\end{aligned}$$

Thus, we found $V_{\text{MLE}} = V_{\text{GMM}}$ here. In fact, by solving for $\hat{\theta}_{\text{MLE}}$ explicitly one can show that $\hat{\theta}_{\text{MLE}} = \hat{\theta}_{\text{GMM}}$ for $W = \mathbb{1}_2$.

(c) We find

$$\begin{aligned}
W^* = \{\text{Var}[g(y_i, \theta_0)]\}^{-1} &= \begin{pmatrix} \text{Var}[1(y_i = 0)] & \text{Cov}[1(y_i = 0), 1(y_i = 1)] \\ \text{Cov}[1(y_i = 0), 1(y_i = 1)] & \text{Var}[1(y_i = 1)] \end{pmatrix}^{-1} \\
&= \begin{pmatrix} \theta_0(1-\theta_0) & -\theta_0^2 \\ -\theta_0^2 & \theta_0(1-\theta_0) \end{pmatrix}^{-1} = \frac{1}{\theta_0^2(1-2\theta_0)}\begin{pmatrix} \theta_0(1-\theta_0) & \theta_0^2 \\ \theta_0^2 & \theta_0(1-\theta_0) \end{pmatrix} \\
&= \frac{1}{\theta_0(1-2\theta_0)}\begin{pmatrix} 1-\theta_0 & \theta_0 \\ \theta_0 & 1-\theta_0 \end{pmatrix}.
\end{aligned}$$

Because W^* minimizes the asymptotic variance of GMM we have $V_{\text{GMM}}(W^*) \leq V_{\text{GMM}}(\mathbb{1}_2)$. Efficiency of the MLE guarantees $V_{\text{MLE}} \leq V_{\text{GMM}}(W^*)$. We have already shown that $V_{\text{MLE}} = V_{\text{GMM}}(\mathbb{1}_2)$, so we must have

$$V_{\text{GMM}}(W^*) = V_{\text{GMM}}(\mathbb{1}_2) = V_{\text{MLE}}.$$

Alternatively, one can quite easily show that $\hat{\theta}_{\text{GMM}}(W^*) = \theta_{\text{GMM}}(\mathbb{1}_2)$ by explicitly calculating $\hat{\theta}_{\text{GMM}}(W^*)$, which also gives the same conclusion on the asymptotic variances.

This example shows that $W = W^*$ (or rather $W = cW^*$ for some constant $c > 0$) is a <u>sufficient</u> but <u>not a necessary</u> condition for minimizing the asymptotic variance of the GMM estimator.

(d) The additional moment condition would <u>not</u> improve efficiency of the GMM estimator. Two possible arguments:

- We already found $\hat{\theta}_{\text{MLE}} = \hat{\theta}_{\text{GMM}}$, and the MLE is efficient, so we cannot improve efficiency further.
- $\mathbb{E}[1(y_i = 2) - (1 - 2\theta_0)] = 0$ does not add any new information, because it is simply a linear combination of the two moment conditions we already use, namely we have

$$1(y_i = 2) - (1 - 2\theta) = -\{1(y_i = 0) - \theta\} - \{1(y_i = 1) - \theta\}$$

Question B3 (20 points)

(a) The MLE $\hat{\sigma}^2_{\text{MLE}}$ minimizes the log-likelihood

$$\sum_{t=1}^{T} \log f(\varepsilon_t|\sigma^2) = \sum_{t=1}^{T} \left[-\frac{1}{2}\log(2\pi\sigma^2) - \frac{\varepsilon_t^2}{2\sigma^2} \right].$$

The FOC for this minimization reads

$$\sum_{t=1}^{T} \left[-\frac{1}{2}\frac{1}{\hat{\sigma}^2_{\text{MLE}}} + \frac{\varepsilon_t^2}{2(\hat{\sigma}^2_{\text{MLE}})^2} \right].$$

Solving for $\hat{\sigma}^2_{\text{MLE}}$ gives

$$\hat{\sigma}^2_{\text{MLE}} = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t^2.$$

Calculating the expectation of $\hat{\sigma}^2_{\text{MLE}}$ we find

$$\mathbb{E}(\hat{\sigma}^2_{\text{MLE}}) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}(\varepsilon_t^2) = \frac{1}{T}\sum_{t=1}^{T} \sigma^2 = \sigma^2,$$

so $\hat{\sigma}^2_{\text{MLE}}$ is an unbiased estimator for σ^2.

(b) One way to solve the question is to apply the CLT to $\sqrt{T}\,(\hat{\sigma}^2_{\text{MLE}} - \sigma^2) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\varepsilon_t^2 - \sigma^2)$, which gives $\sqrt{T}\,(\hat{\sigma}^2_{\text{MLE}} - \sigma^2) \Rightarrow \mathcal{N}(0, V_\infty)$, with

$$V_\infty = \mathbb{E}[(\varepsilon_t^2 - \sigma^2)^2].$$

Evaluating this further requires calculating $\mathbb{E}(\varepsilon_t^4) = 3\sigma^4$, which is not given in the question, but could be either be known by the student, or could be calculated by integrating over the pdf.

The above solution would of course be ok, but the more likely solution is the following: For the Hessian we find

$$H(\varepsilon_t, \sigma^2) = \frac{\partial^2 \log f(\varepsilon_t|\sigma^2)}{\partial(\sigma^2)^2} = \frac{1}{2}\frac{1}{(\sigma^2)^2} - \frac{\varepsilon_t^2}{(\sigma^2)^3},$$

and therefore at the true parameter σ^2 (we could write σ_0^2 in the following, but this notation was not introduced in the question) we have

$$\mathbb{E}H(\varepsilon_t, \sigma^2) = \frac{1}{2}\frac{1}{(\sigma^2)^2} - \frac{\sigma^2}{(\sigma^2)^3} = \frac{1}{2\sigma^4}.$$

For the asymptotic variance we thus find

$$V_\infty = \left[\mathbb{E}H(\varepsilon_t, \sigma^2)\right]^{-1} = 2\sigma^4.$$

Notice that the same result could be obtained if not σ^2, but σ is used as the parameter of the distribution of ϵ_t. We then calculate the asymptotic variance of σ first, which afterwards gives V_∞ by the delta method.

(c) We have

$$\hat\rho = \frac{\sum_{t=1}^T y_{t-1}\, y_t}{\sum_{t=1}^T y_{t-1}^2}, \qquad \hat\sigma^2 = \frac{1}{T}\sum_{t=1}^T \left(y_t - \hat\rho\, y_{t-1}\right)^2.$$

Consistency of $\hat\rho$ follows in the standard way: apply the model, WLLN amd CMT to find

$$\hat\rho - \rho = \frac{\frac{1}{T}\sum_{t=1}^T y_{t-1}\,\varepsilon_t}{\frac{1}{T}\sum_{t=1}^T y_{t-1}^2} \to_p \frac{\mathbb{E}y_{t-1}\,\varepsilon_t}{\mathbb{E}y_{t-1}^2} = 0,$$

because $\mathbb{E}y_{t-1}\,\varepsilon_t = 0$ and $\mathbb{E}y_{t-1}^2 > 0$. Notice, however, that we require a time-series WLLN ("Ergodic Theorem") here.

(d) We use the model to obtain

$$\begin{aligned}\hat\sigma^2 &= \frac{1}{T}\sum_{t=1}^T \left[\varepsilon_t - (\hat\rho-\rho)\, y_{t-1}\right]^2 \\ &= \hat\sigma^2_{\mathrm{MLE}} - 2(\hat\rho-\rho)\frac{1}{T}\sum_{t=1}^T \varepsilon_t y_{t-1} + (\hat\rho-\rho)^2\frac{1}{T}\sum_{t=1}^T y_{t-1}^2.\end{aligned}$$

Let $A = \frac{1}{\sqrt{T}}\sum_{t=1}^T \varepsilon_t y_{t-1}$, which according to the hint (and by a time series CLT) satisfies $A \Rightarrow \mathcal{N}(0, \mathrm{Var}(\varepsilon_t y_{t-1}))$. Let $B = \frac{1}{T}\sum_{t=1}^T y_{t-1}^2$, which by a time-series WLLN ("Ergodic Theorem") satisfies $B \to_p \mathbb{E}y_{t-1}^2$. We then have

$$\sqrt{T}\left(\hat\sigma^2 - \sigma^2\right) = \sqrt{T}\left(\hat\sigma^2_{\mathrm{MLE}} - \sigma^2\right) - \underbrace{\frac{2}{\sqrt{T}}A\left[\sqrt{T}(\hat\rho-\rho)\right]}_{\to_p\, 0} + \underbrace{\frac{1}{\sqrt{T}}B\left[\sqrt{T}(\hat\rho-\rho)\right]^2}_{\to_p\, 0},$$

because $\frac{1}{\sqrt{T}} \to 0$ as $T \to \infty$. This shows that $\hat\sigma^2$ and $\hat\sigma^2_{\mathrm{MLE}}$ have the same asymptotic distribution, that is, we have

$$\sqrt{T}\left(\hat\sigma^2 - \sigma^2\right) \Rightarrow \mathcal{N}(0, 2\sigma^4).$$

SOLUTIONS

G020: Examination in Econometrics

2016-2017

Section A

Question A1 (20 points)

(a) Using that $d_i^2 = d_i$ we find

$$\sum_{i=1}^{n} x_i' x_i = \sum_{i=1}^{n} \begin{pmatrix} 1 \\ d_i \end{pmatrix} (1, d_i) = \sum_{i=1}^{n} \begin{pmatrix} 1 & d_i \\ d_i & d_i^2 \end{pmatrix} = \sum_{i=1}^{n} \begin{pmatrix} 1 & d_i \\ d_i & d_i \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^{n} d_i \\ \sum_{i=1}^{n} d_i & \sum_{i=1}^{n} d_i \end{pmatrix}$$
$$= \begin{pmatrix} n & n_1 \\ n_1 & n_1 \end{pmatrix} = \begin{pmatrix} 100 & 20 \\ 20 & 20 \end{pmatrix}.$$

Using that $\sum_{i=1}^{n} y_i = n_0 \left[\frac{1}{n_0}\sum_{i=1}^{n}(1-d_i)y_i\right] + n_1 \left[\frac{1}{n_1}\sum_{i=1}^{n} d_i y_i\right]$ we find

$$\sum_{i=1}^{n} x_i' y_i = \sum_{i=1}^{n} \begin{pmatrix} 1 \\ d_i \end{pmatrix} y_i = \sum_{i=1}^{n} \begin{pmatrix} y_i \\ d_i y_i \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} d_i y_i \end{pmatrix}$$
$$= \begin{pmatrix} n_0 \left[\frac{1}{n_0}\sum_{i=1}^{n}(1-d_i)y_i\right] + n_1 \left[\frac{1}{n_1}\sum_{i=1}^{n} d_i y_i\right] \\ n_1 \left[\frac{1}{n_1}\sum_{i=1}^{n} d_i y_i\right] \end{pmatrix} = \begin{pmatrix} 80 \times 3 + 20 \times 8 \\ 20 \times 8 \end{pmatrix}$$
$$= \begin{pmatrix} 400 \\ 160 \end{pmatrix}.$$

We thus find

$$\widehat{\beta} = \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix} = \left(\sum_{i=1}^{n} x_i' x_i\right)^{-1} \sum_{i=1}^{n} x_i' y_i = \begin{pmatrix} 100 & 20 \\ 20 & 20 \end{pmatrix}^{-1} \begin{pmatrix} 400 \\ 160 \end{pmatrix}$$
$$= \frac{1}{1.600} \begin{pmatrix} 20 & -20 \\ -20 & 100 \end{pmatrix} \begin{pmatrix} 400 \\ 160 \end{pmatrix}$$
$$= \begin{pmatrix} 3 \\ 5 \end{pmatrix}.$$

(b) Define $z_i = (1, y_i)$. We calculate

$$\sum_{i=1}^n z_i' z_i = \sum_{i=1}^n \begin{pmatrix} 1 & y_i \\ y_i & y_i^2 \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i & \sum_{i=1}^n y_i^2 \end{pmatrix} = \begin{pmatrix} 100 & 400 \\ 400 & 2400 \end{pmatrix},$$

and

$$\sum_{i=1}^n z_i' d_i = \begin{pmatrix} \sum_{i=1}^n d_i \\ \sum_{i=1}^n y_i d_i \end{pmatrix} = \begin{pmatrix} 20 \\ 160 \end{pmatrix}.$$

The OLS estimator for the regression model $d_i = \gamma_1 + \gamma_2\, y_i + \epsilon_i$ thus reads

$$\begin{aligned} \widehat{\gamma} = \begin{pmatrix} \widehat{\gamma}_1 \\ \widehat{\gamma}_2 \end{pmatrix} &= \left(\sum_{i=1}^n z_i' z_i \right)^{-1} \sum_{i=1}^n z_i' d_i = \begin{pmatrix} 100 & 400 \\ 400 & 2400 \end{pmatrix}^{-1} \begin{pmatrix} 20 \\ 160 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 4 \\ 4 & 24 \end{pmatrix}^{-1} \begin{pmatrix} 0.2 \\ 1.6 \end{pmatrix} = \frac{1}{8} \begin{pmatrix} 24 & -4 \\ -4 & 1 \end{pmatrix} \begin{pmatrix} 0.2 \\ 1.6 \end{pmatrix} \\ &= \begin{pmatrix} -0.2 \\ 0.1 \end{pmatrix}. \end{aligned}$$

(c)
- Under the assumption that dog ownership is assigned randomly to individuals, the estimator $\widehat{\beta}_2$ estimates the true causal effect of dog ownership on happiness.
- By causal effect we mean that if we change dog ownership from $d_i = 0$ to $d_i = 1$ (giving a dog to somebody who does not own a dog), then happiness changes on average by $\widehat{\beta}_2$.
- The estimator $\widehat{\gamma}_2$ does not describe a causal effect. The value of that estimator also depends a lot on the variance of the unexplained u_i in our sample. Also, the assumption is that we randomly assign dogs, not randomly assign happiness, so we not learn about the causal effect of happiness on dog ownership then.

(d) No, there always exist values of β_1 and β_2 such that simultaneously $y_i = \beta_1 + \beta_2\, d_i + u_i$ and $\mathbb{E}(u_i) = 0$ and $\mathbb{E}(u_i d_i) = 0$ hold. This is a purely statistical result. We know that $\widehat{\beta}_2$ and $\widehat{\gamma}_2$ are measures of correlation between y_i and d_i, but we cannot infer causality.

(e) Under homoscedasticity with known $\sigma^2 = \mathbb{E}(u_i^2 | d_i) = 4$ the OLS estimator for the variance of $\widehat{\beta} = \binom{\widehat{\beta}_1}{\widehat{\beta}_2}$ is given by

$$\widehat{\mathrm{Var}}(\widehat{\beta}) = \sigma^2 \left(\sum_{i=1}^n x_i' x_i \right)^{-1} = 4\, \frac{1}{1.600} \begin{pmatrix} 20 & -20 \\ -20 & 100 \end{pmatrix} = \begin{pmatrix} 0.05 & -0.05 \\ -0.05 & 0.25 \end{pmatrix}.$$

We thus have

$$\widehat{\text{std}}(\widehat{\beta}_2) = \sqrt{\widehat{\text{Var}}(\widehat{\beta}_2)} = \sqrt{0.25} = 0.5.$$

Question A2 (20 points)

(a) We first use the model for y_i to obtain

$$\sqrt{n}\left(\widehat{\beta} - \beta\right) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i' x_i\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i' u_i\right).$$

As $n \to \infty$, we apply the weak law of large numbers (WLLN) to obtain

$$\frac{1}{n}\sum_{i=1}^{n} x_i' x_i \to_p \mathbb{E}(x_i' x_i) = \begin{pmatrix} 1 & \rho_{WZ} \\ \rho_{WZ} & 1 \end{pmatrix},$$

and the central limit theorem (CLT) to obtain

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i' u_i \Rightarrow \mathcal{N}\left(0, \mathbb{E}(u_i^2\, x_i' x_i)\right) \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{WZ} \\ \rho_{WZ} & 1 \end{pmatrix}\right),$$

where in the last step we used that the DGP in the question implies homoscedasticity, $\mathbb{E}(u_i^2|x_i) = 1$, so that $\mathbb{E}(u_i^2\, x_i' x_i) = \mathbb{E}(x_i' x_i)$. Combining the above results and applying Slutsky's theorem we find

$$\sqrt{n}(\widehat{\beta} - \beta) \Rightarrow \mathcal{N}(0, V),$$

with

$$V = \left[\mathbb{E}(x_i' x_i)\right]^{-1} \mathbb{E}(u_i^2\, x_i' x_i) \left[\mathbb{E}(x_i' x_i)\right]^{-1} = \begin{pmatrix} 1 & \rho_{WZ} \\ \rho_{WZ} & 1 \end{pmatrix}^{-1}$$
$$= \frac{1}{1 - \rho_{WZ}^2} \begin{pmatrix} 1 & -\rho_{WZ} \\ -\rho_{WZ} & 1 \end{pmatrix}.$$

(b) By the WLLN and continuous mapping theorem (CMT) we have, as $n \to \infty$,

$$\widehat{\beta}_1^* = \frac{\frac{1}{n}\sum_{i=1}^{n} y_i\, w_i}{\frac{1}{n}\sum_{i=1}^{n} w_i^2} \to_p \frac{\mathbb{E} y_i\, w_i}{\mathbb{E} w_i^2} = \beta_1 + \beta_2 \frac{\mathbb{E} z_i\, w_i}{\mathbb{E} w_i^2} = \beta_1 + \beta_2\, \rho_{WZ},$$

where we also use the model for y_i. Thus, $\widehat{\beta}_1^*$ is consistent iff

$$\beta_2\, \rho_{WZ} = 0,$$

or equivalently iff

$$\beta_2 = 0 \qquad \text{or} \qquad \rho_{WZ} = 0.$$

Under that assumption we can apply the WLLN, CLT and Slutsky's theorem, as in part (a), to find that

$$\sqrt{n}(\widehat{\beta}_1^* - \beta_1) \Rightarrow \mathcal{N}(0, V_*),$$

where using homoscedasticity of the effective error term $\epsilon_i = z_i\,\beta_2 + u_i$ we have

$$V_* = \frac{\mathrm{Var}(z_i\,\beta_2 + u_i)}{\mathbb{E}w_i^2} = (\beta_2)^2 + 1.$$

(c) For $\beta_2 \neq 0$ and $\rho_{WZ} \neq 0$ we have (as shown above) that

- $\widehat{\beta}_1$ is consistent.
- $\widehat{\beta}_1^*$ is inconsistent.

We would therefore recommend to use $\widehat{\beta}_1$.

(d) For $\beta_2 = 0$ and $\rho_{WZ} \neq 0$ both $\widehat{\beta}_1$ and $\widehat{\beta}_1^*$ are consistent, but we find (from part (a) and (b))

$$\mathrm{AsyVar}(\widehat{\beta}_1) = V_{11} = \frac{1}{1 - \rho_{WZ}^2} > 1, \qquad \mathrm{AsyVar}(\widehat{\beta}_1^*) = V_* = (\beta_2)^2 + 1 = 1,$$

and therefore

$$\mathrm{AsyVar}(\widehat{\beta}_1) > \mathrm{AsyVar}(\widehat{\beta}_1^*).$$

We would therefore recommend to use $\widehat{\beta}_1^*$.

(e) For $\beta_2 \neq 0$ and $\rho_{WZ} = 0$ both $\widehat{\beta}_1$ and $\widehat{\beta}_1^*$ are consistent, but we find (from part (a) and (b))

$$\mathrm{AsyVar}(\widehat{\beta}_1) = V_{11} = \frac{1}{1 - \rho_{WZ}^2} = 1, \qquad \mathrm{AsyVar}(\widehat{\beta}_1^*) = V_* = (\beta_2)^2 + 1 > 1,$$

and therefore

$$\mathrm{AsyVar}(\widehat{\beta}_1) < \mathrm{AsyVar}(\widehat{\beta}_1^*).$$

We would therefore recommend to use $\widehat{\beta}_1$.

Question A3 (20 points)

(a) Using the model we find

$$\sqrt{n}\left(\widehat{\beta}-\beta\right)=\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i u_i}{\frac{1}{n}\sum_{i=1}^{n} z_i x_i}.$$

We apply the CLT to find that, as $n\to\infty$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i u_i \Rightarrow \mathcal{N}(0,\mathrm{Var}(z_i u_i)).$$

By the WLLN we have $\frac{1}{n}\sum_{i=1}^{n} z_i x_i \to_p \mathbb{E}(z_i x_i)$. Applying Slutsky's theorem we thus find, as $n\to\infty$,

$$\sqrt{n}\left(\widehat{\beta}-\beta\right)\Rightarrow\frac{\mathcal{N}(0,\mathrm{Var}(z_i u_i))}{\rho_{xz}}=\mathcal{N}\left(0,\frac{\mathrm{Var}(z_i u_i)}{[\mathbb{E}(z_i x_i)]^2}\right).$$

Thus,

$$V=\mathrm{AsyVar}(\sqrt{n}\widehat{\beta})=\frac{\mathrm{Var}(z_i u_i)}{[\mathbb{E}(z_i x_i)]^2}=\frac{\mathbb{E}[(z_i u_i)^2]}{[\mathbb{E}(z_i x_i)]^2}.$$

(b)
- The condition $\mathbb{E}(z_i u_i)=0$ states that z_i is exogenous (uncorrelated with the error u_i), which is also called the exclusion restriction.
- The condition $\mathbb{E}(z_i x_i)\neq 0$ guarantees that z_i is correlated with x_i, i.e. that the instrument z_i is relevant for the endogenous regressor x_i.

(c) Using the information on the observed sample we calculate

$$\widehat{\beta}=\frac{\frac{1}{n}\sum_{i=1}^{n} z_i y_i}{\frac{1}{n}\sum_{i=1}^{n} z_i x_i}=\frac{0.2}{0.4}=0.5,$$

$$\widehat{\mathbb{E}}(z_i x_i)=\frac{1}{n}\sum_{i=1}^{n} z_i x_i=0.4.$$

Define the residuals $\widehat{u}_i=y_i-x_i\widehat{\beta}$. An estimator for $\mathrm{Var}(z_i u_i)=\mathbb{E}[(z_i u_i)^2]$ is then given by

$$\begin{aligned}\widehat{\mathbb{E}}[(z_i u_i)^2]&=\frac{1}{n}\sum_{i=1}^{n}(z_i\widehat{u}_i)^2\\&=\frac{1}{n}\sum_{i=1}^{n} z_i^2(y_i-x_i\widehat{\beta})^2\\&=\frac{1}{n}\sum_{i=1}^{n} z_i^2 y_i^2-2\widehat{\beta}\frac{1}{n}\sum_{i=1}^{n} z_i^2 x_i y_i+\widehat{\beta}^2\frac{1}{n}\sum_{i=1}^{n} z_i^2 x_i^2\\&=3-2\times 0.5\times 1+0.5^2\times 8=4.\end{aligned}$$

A consistent estimator for the asymptotic variance is thus given by

$$\widehat{\text{AsyVar}}(\sqrt{n}\widehat{\beta}) = \frac{\widehat{\mathbb{E}}[(z_i u_i)^2]}{[\widehat{\mathbb{E}}(z_i x_i)]^2} = \frac{4}{(0.4)^2} = 25.$$

An estimator for the variance of $\widehat{\beta}$ is thus given by

$$\widehat{\text{Var}}(\widehat{\beta}) = \frac{1}{n}\widehat{\text{AsyVar}}(\sqrt{n}\widehat{\beta}) = \frac{25}{100} = 0.25,$$

implying that $\widehat{\text{std}}(\widehat{\beta}) = 0.5$. A 95% confidence interval for β is thus given by

$$\begin{aligned}\text{CI}_{95\%} &= \left[\widehat{\beta} - 1.96 \times \widehat{\text{std}}(\widehat{\beta}), \widehat{\beta} + 1.96 \times \widehat{\text{std}}(\widehat{\beta})\right] = [0.5 - 1.96 \times 0.5, 0.5 + 1.96 \times 0.5] \\ &= [-0.48, 1.48].\end{aligned}$$

(d) If we assume that $\mathbb{E}(u_i^2|z_i) = \sigma^2$, then by applying the law of iterated expectations (LIE) we find

$$\text{AsyVar}(\sqrt{n}\widehat{\beta}) = \frac{\mathbb{E}[(z_i u_i)^2]}{[\mathbb{E}(z_i x_i)]^2} = \frac{\sigma^2\,\mathbb{E}[z_i^2]}{[\mathbb{E}(z_i x_i)]^2}.$$

Section B

Answer **TWO questions from this Section**.

Question B1 (20 points)

(a) We have

$$Q_n(\lambda) = \frac{1}{n}\log\prod_{i=1}^{n} f(y_i|\lambda) = \frac{1}{n}\sum_{i=1}^{n}[\log\lambda - \lambda y_i].$$

The MLE thus solves the first order condition

$$0 = \frac{1}{n}\sum_{i=1}^{n}[\widehat{\lambda}_{\text{MLE}}^{-1} - y_i].$$

Solving this we obtain

$$\widehat{\lambda} = \left[\frac{1}{n}\sum_{i=1}^{n} y_i\right]^{-1}.$$

(b) Three possibilities to do this (all correct, of course, only one way needs to be presented):

– By the CLT we know that $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i - \lambda_0^{-1}) \Rightarrow \mathcal{N}(0, \mathrm{Var}(y_i))$, as $n \to \infty$, where $\mathrm{Var}(y_i) = 2\lambda_0^{-2} - \lambda_0^{-2} = \lambda_0^{-2}$. Applying the delta method to

$$\widehat{\lambda}_{\mathrm{MLE}} = \frac{1}{\lambda_0^{-1} + \frac{1}{\sqrt{n}}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i - \lambda_0^{-1})\right]} \approx \lambda_0 - \frac{\lambda_0^2}{\sqrt{n}}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i - \lambda_0^{-1})\right]$$

we have

$$\sqrt{n}\left(\widehat{\lambda}_{\mathrm{MLE}} - \lambda_0\right) \approx \lambda_0^2\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i - \lambda_0^{-1})\right],$$

and therefore

$$V_{\mathrm{MLE}} = \mathrm{AsyVar}(\sqrt{n}\,\widehat{\lambda}_{\mathrm{MLE}}) = \lambda_0^4\,\mathrm{Var}(y_i) = \lambda_0^2.$$

For more details on the delta method see the solution to part (d) below, where the derivation is more formal. Here, we have just used the Taylor expansion that underlies the delta method directly, but we did not justify the approximation $\approx$ above formally, which requires a mean value expansion.

– Alternatively, we can calculate the expected Hessian. We have $\log f(y|\lambda) = \log\lambda - \lambda y_i$, and therefore

$$\frac{\partial^2 \log f(y|\lambda)}{\partial\lambda^2} = -\frac{1}{\lambda^2}.$$

We thus obtain

$$H(\lambda_0) = \mathbb{E}\left[\frac{\partial^2 \log f(y_i|\lambda_0)}{\partial\lambda^2}\right] = -\frac{1}{\lambda_0^2}.$$

Therefore,

$$V_{\mathrm{MLE}} = [-H(\lambda_0)]^{-1} = \lambda_0^2.$$

– Finally, we can also calculate the variance of the score. We have

$$\mathrm{Var}\left(\frac{\partial \log f(y|\lambda_0)}{\partial\lambda}\right) = \mathrm{Var}\left(\frac{1}{\lambda} - y_i\right) = \mathrm{Var}\,(y_i) = \lambda_0^{-2},$$

and therefore

$$V_{\mathrm{MLE}} = \left[\mathrm{Var}\left(\frac{\partial \log f(y|\lambda_0)}{\partial\lambda}\right)\right]^{-1} = \lambda_0^2.$$

Thus, all three methods give the same results for V_{MLE}.

(c) The sample analog of the moment condition $\mathbb{E}\left(y_i^2 - 2\,\lambda_0^{-2}\right) = 0$ reads

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i^2 - 2\,\widehat{\lambda}_{\mathrm{MM}}^{-2}\right) = 0.$$

Solving this we obtain

$$\widehat{\lambda}_{\mathrm{MM}} = \left[\frac{1}{2}\,\frac{1}{n}\sum_{i=1}^{n} y_i^2\right]^{-1/2}.$$

(d) We have $\mathbb{E}\left(y_i^2 - 2\lambda_0^{-2}\right) = 0$ and $\mathrm{Var}\left(y_i^2 - 2\lambda_0^{-2}\right) = \mathrm{Var}(y_i^2) = \mathbb{E}y_i^4 - (\mathbb{E}y_i^2)^2 = 20\lambda_0^{-4}$. Applying the CLT we thus have, as $n \to \infty$,

$$\sqrt{n}\left[\left(\frac{1}{n}\sum_{i=1}^{n} y_i^2\right) - 2\lambda_0^{-2}\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(y_i^2 - 2\lambda_0^{-2}\right) \Rightarrow \mathcal{N}(0, 20\,\lambda_0^{-4}).$$

Applying the delta method we find for differentiable functions $g : \mathbb{R} \to \mathbb{R}$ that

$$\sqrt{n}\left[g\left(\frac{1}{n}\sum_{i=1}^{n} y_i^2\right) - g\left(2\lambda_0^{-2}\right)\right] \Rightarrow \mathcal{N}\left[0, 20\,\lambda_0^{-4}\,g'\left(2\lambda_0^{-2}\right)\right].$$

In particular, choosing $g(x) = (x/2)^{-1/2} = \frac{\sqrt{2}}{\sqrt{x}}$ have $g'(x) = -\frac{1}{\sqrt{2}\,x^{3/2}}$ and

$$g\left(\frac{1}{n}\sum_{i=1}^{n} y_i^2\right) = \widehat{\lambda}_{\mathrm{MM}}, \qquad g\left(2\lambda_0^{-2}\right) = \lambda_0, \qquad g'\left(2\lambda_0^{-2}\right) = -\frac{\lambda_0^3}{4}.$$

We thus obtain

$$\sqrt{n}\left(\widehat{\lambda}_{\mathrm{MM}} - \lambda_0\right) \Rightarrow \mathcal{N}\left(0, V_{\mathrm{MM}}\right), \qquad V_{\mathrm{MM}} = \frac{5}{4}\,\lambda_0^2.$$

We thus have

$$V_{\mathrm{MM}} > V_{\mathrm{MLE}}.$$

As expected, the MLE is more efficient than the MM estimator.

Question B2 (20 points)

(a) We have $\widehat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} y_{i1}$ and $\widehat{\theta}_2 = \frac{1}{2n}\sum_{i=1}^{n} y_{i2}$. Applying the CLT we obtain

$$\sqrt{n}(\widehat{\theta}_1 - \theta_0) \Rightarrow \mathcal{N}(0, \mathrm{Var}(y_{i1})), \qquad \sqrt{n}(\widehat{\theta}_2 - \theta_0) \Rightarrow \mathcal{N}(0, \mathrm{Var}(y_{i2})/4),$$

i.e. $\mathrm{AsyVar}(\widehat{\theta}_1) = \mathrm{Var}(y_{i1}) = 1$ and $\mathrm{AsyVar}(\widehat{\theta}_2) = \mathrm{Var}(y_{i2})/4 = 1/4$.

(b) In general, the optimal weight matrix satisfies (up to an arbitrary scalar factor) $W_{\text{opt}} = \text{Var}(g(y_i,\theta))^{-1}$, evaluated at the true θ. For the current model this gives (again, up to an arbitrary scalar factor)

$$W_{\text{opt}} = \text{Var}(y_i)^{-1} = \begin{pmatrix} 1 & 4/5 \\ 4/5 & 1 \end{pmatrix}^{-1} = \left[\frac{1}{5}\begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}\right]^{-1} = \frac{5}{9}\begin{pmatrix} 5 & -4 \\ -4 & 5 \end{pmatrix}$$
$$= \begin{pmatrix} 25/9 & -20/9 \\ -20/9 & 25/9 \end{pmatrix}.$$

To calculate the asymptotic variance of $\widehat{\theta}_{\text{GMM}}$ there are two possibilities:

– Applying the general formula for the asymptotic variance of the optimal GMM estimator we obtain

$$\begin{aligned}\text{AsyVar}(\widehat{\theta}_{\text{GMM}}) &= \left[\frac{\partial g(y_i,\theta_0)}{\partial \theta'} \text{Var}(g(y_i,\theta_0))^{-1} \frac{\partial g(y_i,\theta_0)}{\partial \theta}\right]^{-1} \\ &= \left[\begin{pmatrix} -1 \\ -2 \end{pmatrix}' W_{\text{opt}} \begin{pmatrix} -1 \\ -2 \end{pmatrix}\right]^{-1} \\ &= \frac{1}{5}.\end{aligned}$$

– Alternatively (e.g. if students do not remember this formula), one can also use the explicit formula for the GMM estimator given in the question:

$$\widehat{\theta}_{\text{GMM}} = w\left(\frac{1}{n}\sum_{i=1}^{n} y_{i1}\right) + (1-w)\left(\frac{1}{2n}\sum_{i=1}^{n} y_{i2}\right), \quad w = \frac{W_{11}+2W_{12}}{W_{11}+4W_{12}+4W_{22}}.$$

The weight w corresponding to $W = W_{\text{opt}}$ is

$$w_{\text{opt}} = -1/3,$$

and the asymptotic variance of the corresponding optimal GMM estimator can then be calculated as

$$\begin{aligned}\text{AsyVar}(\widehat{\theta}_{\text{GMM}}) &= \begin{pmatrix} w_{\text{opt}} \\ (1-w_{\text{opt}})/2 \end{pmatrix}' \text{Var}(y) \begin{pmatrix} w_{\text{opt}} \\ (1-w_{\text{opt}})/2 \end{pmatrix} \\ &= \begin{pmatrix} -1/3 \\ 2/3 \end{pmatrix}' \begin{pmatrix} 1 & 4/5 \\ 4/5 & 1 \end{pmatrix} \begin{pmatrix} -1/3 \\ 2/3 \end{pmatrix} \\ &= \frac{1}{5}.\end{aligned}$$

Thus, both methods to calculate $\text{AsyVar}(\widehat{\theta}_{\text{GMM}})$ give the same result. Of course, students only need to present one calculation.

As expected, $\text{AsyVar}(\widehat{\theta}_{\text{GMM}})$ is smaller than both $\text{AsyVar}(\widehat{\theta}_1)$ and $\text{AsyVar}(\widehat{\theta}_2)$.

(c) No, the new moment condition is just a linear combination of the two existing ones, so we cannot construct a more efficient estimator than $\widehat{\theta}_{\text{GMM}}$ in (b) using this additional moment condition. There is simply no additional information in that moment condition.

(d) To estimate $\text{Var}(y_i)$ we first need some preliminary consistent estimator for θ_0, e.g. $\widehat{\theta}_2$ (but $\widehat{\theta}_1$ or $\widehat{\theta}_{\text{GMM}}$ for an arbitrary weight matrix would also work, of course). Then we can estimate $\text{Var}(y_i)$ via the sample variance, namely

$$\widehat{\text{Var}}(y_i) = \frac{1}{n} \sum_{i=1}^{n} \left[y_i - \begin{pmatrix} \widehat{\theta}_2 \\ 2\widehat{\theta}_2 \end{pmatrix} \right] \left[y_i - \begin{pmatrix} \widehat{\theta}_2 \\ 2\widehat{\theta}_2 \end{pmatrix} \right]'.$$

Having $\widehat{\text{Var}}(y_i)$ we can estimate W_{opt} via $\widehat{W}_{\text{opt}} = \widehat{\text{Var}}(y_i)^{-1}$.

Question B3 (20 points)

(a) Maximizing the likelihood is equivalent to maximizing the log-likelihood

$$\log f(y|y_0, \theta) = \sum_{t=1}^{T} \left[-\frac{1}{2} \log(2\pi\sigma^2) - \frac{(y_t - \rho\, y_{t-1})^2}{2\sigma^2} \right].$$

Since we are only interested in $\widehat{\rho}$, we find that maximzing the log-likelihood over ρ, for any value of σ, is equivalent to minimizing the SSR

$$\sum_{t=1}^{T} (y_t - \rho\, y_{t-1})^2.$$

The FOC of this minimization problem is

$$\sum_{t=1}^{T} y_{t-1}(y_t - \widehat{\rho}\, y_{t-1}) = 0,$$

and solving this gives the OLS estimator

$$\widehat{\rho} = \frac{\sum_{t=1}^{T} y_{t-1}\, y_t}{\sum_{t=1}^{T} y_{t-1}^2}.$$

(b) Using the model for y_t we obtain

$$\sqrt{T}\,(\widehat{\rho} - \rho_0) = \frac{\frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_{t-1}\,\varepsilon_t}{\frac{1}{T} \sum_{t=1}^{T} y_{t-1}^2}.$$

Applying time series CLT and WLLN we find, as $T \to \infty$,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_{t-1}\,\varepsilon_t \Rightarrow \mathcal{N}(0, \mathbb{E}(y_{t-1}^2\,\varepsilon_t^2)), \qquad \frac{1}{T} \sum_{t=1}^{T} y_{t-1}^2 \to_p \mathbb{E}y_{t-1}^2,$$

and by Slutsky's theorem we thus have

$$\sqrt{T}\,(\widehat{\rho} - \rho_0) \Rightarrow \mathcal{N}\,(0, V)\,, \qquad V = \frac{\mathbb{E}(y_{t-1}^2\,\varepsilon_t^2)}{(\mathbb{E}y_{t-1}^2)^2}.$$

By stationary we have $\mathrm{Var}(y_{t-1}) = \mathrm{Var}(y_t)$ and also using the model we thus find that

$$\begin{aligned}\mathrm{Var}(y_t) &= \mathrm{Var}(\rho_0\, y_{t-1} + \varepsilon_t) = \rho_0^2 \mathrm{Var}(y_{t-1}) + \mathrm{Var}(\varepsilon_t)\\ &= \rho_0^2 \mathrm{Var}(y_t) + \sigma_0^2.\end{aligned}$$

Solving this for $\mathrm{Var}(y_t)$ gives

$$\mathrm{Var}(y_t) = \frac{\sigma_0^2}{1 - \rho_0^2}.$$

We have thus shown that $\mathbb{E}y_{t-1}^2 = \frac{\sigma_0^2}{1-\rho_0^2}$. Using independence of ε_t and y_{t-1} we also obtain that

$$\mathbb{E}(y_{t-1}^2\,\varepsilon_t^2) = \mathbb{E}(y_{t-1}^2)\mathbb{E}(\varepsilon_t^2) = \frac{\sigma_0^4}{1 - \rho_0^2}.$$

Plugging those results into the above expression for the asymptotic variance gives

$$V = 1 - \rho_0^2.$$

(c) No, $\widehat{\rho}$ is not unbiased. Our unbiasedness proof for OLS crucially requires that for all observations t we have $E(\varepsilon_t|X) = 0$, where X are all the regressors, that is, $X = (y_0, y_1, \ldots, y_{T-1})$ here. This assumption for unbiasedness cannot be satisfied, because $E(\varepsilon_t|X) = y_t - \rho_0 y_{t-1}$ here. One can show that $\widehat{\rho}$ is biased for any finite T, but that the bias vanishes at a rate of $1/T$ as $T \to \infty$.

(d) The improved estimator solves

$$\widehat{\theta}_* = (\widehat{\rho}_*, \widehat{\sigma}_*^2)' = \arg\max_\theta f(y_0, y|\theta),$$

where

$$f(y_0, y|\theta) = \left[\prod_{t=1}^{T} f(y_t|y_{t-1}, \theta)\right] f(y_0|\theta).$$

This is simply the likelihood over all observations $(y_0, y) = (y_0, y_1, y_2, \ldots, y_T)$. This estimator uses more information than $\widehat{\theta}$, because the information on the parameters contained in y_0 is also used. We expect the finite sample variance of $\widehat{\theta}_*$ to be smaller than of $\widehat{\theta}$.

(e) We expect that

$$V = V_*.$$

The reason is that we have only added one observation y_0 to the likelihood, which is roughly equivalent to going form T to $T+1$. As $T \to \infty$ this one observation contains a negligible fraction of the total information in the data, so we do not expect that the asymptotic variance changes.

SOLUTIONS

ECON G020 & GP20: Examination in Econometrics

2017-2018

Section A

Question A1 (20 points)

(a) Let $x_i = (1, w_i, z_i)$. From (1) we know that

$$\frac{1}{n}\sum_{i=1}^{n} x_i' x_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \frac{1}{n}\sum_{i=1}^{n} x_i' y_i = \begin{pmatrix} 2 \\ 3 \\ -2 \end{pmatrix}.$$

The OLS estimator thus reads

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i' x_i\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} x_i' y_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 2 \\ 3 \\ -2 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ -2 \end{pmatrix},$$

i.e. $\hat{\beta}_1 = 2$, $\hat{\beta}_2 = 3$ and $\hat{\beta}_3 = -2$.

(b) We have

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2 = \frac{1}{n}\sum_{i=1}^{n} \left(y_i - \hat{\beta}_1 - w_i\hat{\beta}_2 - z_i\hat{\beta}_3\right)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\sum_{i=1}^{n} \left(\hat{\beta}_1 + w_i\hat{\beta}_2 + z_i\hat{\beta}_3\right)^2, \end{aligned}$$

where in the last step we used the fact that by construction of the OLS estimator the residuals $\hat{u}_i$ are orthogonal to the predicted values $\hat{\beta}_1 + w_i\hat{\beta}_2 + z_i\hat{\beta}_3$. Using (1) and our result for $\hat{\beta}$ we then find

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\sum_{i=1}^{n} \hat{\beta}_1^2 - \frac{1}{n}\sum_{i=1}^{n} w_i^2\hat{\beta}_2^2 - \frac{1}{n}\sum_{i=1}^{n} z_i^2\hat{\beta}_3^2 \\ &= 53 - 2^2 - 1 \cdot 3^2 - 1 \cdot (-2)^2 \\ &= 36. \end{aligned}$$

(c) We have

$$\widehat{\mathrm{Var}}(\hat{\beta}) = \hat{\sigma}^2 \left(\sum_{i=1}^{n} x_i' x_i \right)^{-1} = \frac{\hat{\sigma}^2}{n} \left(\frac{1}{n} \sum_{i=1}^{n} x_i' x_i \right)^{-1} = \frac{36}{100} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} 0.36 & 0 & 0 \\ 0 & 0.36 & 0 \\ 0 & 0 & 0.36 \end{pmatrix}.$$

Thus, $\widehat{\mathrm{std}}(\hat{\beta}_1) = \widehat{\mathrm{std}}(\hat{\beta}_2) = \widehat{\mathrm{std}}(\hat{\beta}_3) = 0.6$.

(d) We must have $\hat{\beta}_1 + w_i \hat{\beta}_2 + z_i \hat{\beta}_3 = \hat{\gamma}_1 + v_i \hat{\gamma}_2 + s_i \hat{\gamma}_3$ for all possible values of w_i and z_i, where $v_i = w_i + z_i$ and $s_i = w_i - z_i$. We thus have

$$\begin{aligned} \hat{\beta}_1 + w_i \hat{\beta}_2 + z_i \hat{\beta}_3 &= \hat{\gamma}_1 + (w_i + z_i)\hat{\gamma}_2 + (w_i - z_i)\hat{\gamma}_3 \\ &= \hat{\gamma}_1 + w_i(\hat{\gamma}_2 + \hat{\gamma}_3) + z_i(\hat{\gamma}_2 - \hat{\gamma}_3). \end{aligned}$$

Thus, we need $\hat{\beta}_1 = \hat{\gamma}_1$, $\hat{\beta}_2 = \hat{\gamma}_2 + \hat{\gamma}_3$ and $\hat{\beta}_3 = \hat{\gamma}_2 - \hat{\gamma}_3$. From this we find

$$\begin{aligned} \hat{\gamma}_1 &= \hat{\beta}_1 = 2, \\ \hat{\gamma}_2 &= \frac{\hat{\beta}_2 + \hat{\beta}_3}{2} = \frac{1}{2}, \\ \hat{\gamma}_3 &= \frac{\hat{\beta}_2 - \hat{\beta}_3}{2} = \frac{5}{2}. \end{aligned}$$

Question A2 (20 points)

(a) We have as $n \to \infty$

$$\widehat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^n z_i y_i}{\frac{1}{n}\sum_{i=1}^n z_i x_i} \to_p \frac{\mathbb{E}(z_i y_i)}{\mathbb{E}(z_i x_i)} = \beta_1^*$$

where we have used the WLLN in both the numerator and the denominator, and we also applied the CMT. By also using the model for y_i and the data generating process for x_i, z_i, u_i we obtain

$$\beta_1^* = \frac{\mathbb{E}(z_i y_i)}{\mathbb{E}(z_i x_i)} = \frac{\beta\mathbb{E}(z_i x_i) + \mathbb{E}(z_i u_i)}{\mathbb{E}(z_i x_i)} = \beta + \frac{\rho_{zu}}{\rho_{xz}}.$$

If $\rho_{zu} = 0$, then $\beta_1^* = \beta$, so that $\widehat{\beta}_1$ is consistent for β.

The condition $\rho_{zu} = 0$ states that z_i is exogenous (uncorrelated with the error u_i), which is also called the exclusion restriction (or exogeneity condition) for the instrumental variables z_i.

(b) Using the model we find

$$\sqrt{n}\left(\widehat{\beta}_1 - \beta\right) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n z_i u_i}{\frac{1}{n}\sum_{i=1}^n z_i x_i}$$

Since we assume $\rho_{zu} = \mathbb{E}(z_i u_i) = 0$ we can apply the CLT to find that as $n \to \infty$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n z_i u_i \Rightarrow \mathcal{N}(0, \operatorname{Var}(z_i u_i)).$$

As above, by the WLLN we have $\frac{1}{n}\sum_{i=1}^n z_i x_i \to_p \mathbb{E}(z_i x_i)$. Applying Slutsky's theorem we thus find as $n \to \infty$

$$\sqrt{n}\left(\widehat{\beta}_1 - \beta\right) \Rightarrow \frac{\mathcal{N}(0, \operatorname{Var}(z_i u_i))}{\mathbb{E}(z_i x_i)} = \mathcal{N}\left(0, \frac{\operatorname{Var}(z_i u_i)}{[\mathbb{E}(z_i x_i)]^2}\right),$$

i.e.

$$\operatorname{AsyVar}(\sqrt{n}\widehat{\beta}_1) = \frac{\operatorname{Var}(z_i u_i)}{[\mathbb{E}(z_i x_i)]^2} = \frac{\mathbb{E}[(z_i u_i)^2]}{[\mathbb{E}(z_i x_i)]^2} = \frac{\mathbb{E}(z_i^2)\mathbb{E}(u_i^2)}{[\mathbb{E}(z_i x_i)]^2} = \frac{1}{\rho_{xz}^2}.$$

Notice that for $\rho_{zu} = 0$ we know that z_i and u_i are independent (because uncorrelated normally distributed random variables are independent), which implies $\mathbb{E}(z_i^2 u_i^2) = \mathbb{E}(z_i^2)\mathbb{E}(u_i^2)$, which was used in the last display.

(c) Because $\mathbb{E}(z_i^2 x_i) = 0$ the instrument z_i^2 is <u>not relevant</u>, and the IV estimator $\widehat{\beta}_2$ using this instrument is therefore <u>inconsistent</u>.

(d) Independence of z_i and u_i also implies that $\mathbb{E}(z_i^3 u_i) = 0$. Analogous to part (b) we thus find that

$$\sqrt{n}\left(\widehat{\beta}_3 - \beta\right) \Rightarrow \frac{\mathcal{N}(0, \mathrm{Var}(z_i^3 u_i))}{\mathbb{E}(z_i^3 x_i)} = \mathcal{N}\left(0, \frac{\mathrm{Var}(z_i^3 u_i)}{[\mathbb{E}(z_i^3 x_i)]^2}\right),$$

i.e.

$$\mathrm{AsyVar}(\sqrt{n}\widehat{\beta}_3) = \frac{\mathrm{Var}(z_i^3 u_i)}{[\mathbb{E}(z_i^3 x_i)]^2} = \frac{\mathbb{E}(z_i^6 u_i^2)}{[\mathbb{E}(z_i^3 x_i)]^2} = \frac{\mathbb{E}(z_i^6)\mathbb{E}(u_i^2)}{[\mathbb{E}(z_i^3 x_i)]^2} = \frac{15}{9\rho_{xz}^2}.$$

We have $1 < 15/9$, that is $\mathrm{AsyVar}(\sqrt{n}\widehat{\beta}_1) < \mathrm{AsyVar}(\sqrt{n}\widehat{\beta}_3)$. We would therefore recommend to use $\widehat{\beta}_1$.

Question A3 (20 points)

(a) We have $\widetilde{y}_i = y_i - z_i \Sigma_{zz}^{-1} \Sigma_{zy}$ and $\widetilde{w}_i = w_i - z_i \Sigma_{zz}^{-1} \Sigma_{wz}$. We thus find

$$\begin{aligned}\sum_{i=1}^{n} \widetilde{w}_i z_i &= \sum_{i=1}^{n} z_i \left(w_i - z_i \Sigma_{zz}^{-1} \Sigma_{wz} \right) \\ &= \Sigma_{wz} - \Sigma_{zz} \Sigma_{zz}^{-1} \Sigma_{wz} = \Sigma_{wz} - \Sigma_{wz} = 0,\end{aligned}$$

and also

$$\begin{aligned}\sum_{i=1}^{n} \widetilde{w}_i^2 &= \sum_{i=1}^{n} \widetilde{w}_i w_i - \sum_{i=1}^{n} \widetilde{w}_i z_i \Sigma_{zz}^{-1} \Sigma_{wz} \\ &= \sum_{i=1}^{n} \widetilde{w}_i w_i - \underbrace{\frac{\Sigma_{wz}}{\Sigma_{zz}} \sum_{i=1}^{n} \widetilde{w}_i z_i}_{=0}, \\ \sum_{i=1}^{n} \widetilde{w}_i \widetilde{y}_i &= \sum_{i=1}^{n} \widetilde{w}_i y_i - \sum_{i=1}^{n} \widetilde{w}_i z_i \Sigma_{zz}^{-1} \Sigma_{zy} \\ &= \sum_{i=1}^{n} \widetilde{w}_i y_i - \underbrace{\frac{\Sigma_{zy}}{\Sigma_{zz}} \sum_{i=1}^{n} \widetilde{w}_i z_i}_{=0}.\end{aligned}$$

which was already stated in the question itself, and thus need not actually be shown here by the students, but is included in this solutions for completeness.

Using the above we obtain

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} \widetilde{w}_i \widetilde{y}_i}{\sum_{i=1}^{n} \widetilde{w}_i^2} = \frac{\sum_{i=1}^{n} \widetilde{w}_i y_i}{\sum_{i=1}^{n} \widetilde{w}_i w_i}.$$

Using the definition of $\widetilde{w}_i$ we thus find

$$\begin{aligned}\widehat{\beta}_1 &= \frac{\sum_{i=1}^{n} y_i \left(w_i - z_i \Sigma_{zz}^{-1} \Sigma_{wz} \right)}{\sum_{i=1}^{n} w_i \left(w_i - z_i \Sigma_{zz}^{-1} \Sigma_{wz} \right)} = \frac{\Sigma_{wy} - \dfrac{\Sigma_{zy} \Sigma_{wz}}{\Sigma_{zz}}}{\Sigma_{ww} - \dfrac{\Sigma_{wz}^2}{\Sigma_{zz}}} \\ &= \frac{\Sigma_{wy} \Sigma_{zz} - \Sigma_{zy} \Sigma_{wz}}{\Sigma_{ww} \Sigma_{zz} - \Sigma_{wz}^2}.\end{aligned}$$

(b) Using that $x_i = (w_i, z_i)$ and the inversion formula for 2×2 matrices we find the

OLS estimator that

$$
\begin{aligned}
\widehat{\beta}^{\mathrm{OLS}} = \begin{pmatrix} \widehat{\beta}_1^{\mathrm{OLS}} \\ \widehat{\beta}_2^{\mathrm{OLS}} \end{pmatrix} &= \left(\sum_{i=1}^{n} x_i' x_i \right)^{-1} \sum_{i=1}^{n} x_i' y_i \\
&= \begin{pmatrix} \Sigma_{ww} & \Sigma_{wz} \\ \Sigma_{wz} & \Sigma_{zz} \end{pmatrix}^{-1} \begin{pmatrix} \Sigma_{wy} \\ \Sigma_{zy} \end{pmatrix} \\
&= \frac{1}{\Sigma_{ww}\Sigma_{zz} - \Sigma_{wz}^2} \begin{pmatrix} \Sigma_{zz} & -\Sigma_{wz} \\ -\Sigma_{wz} & \Sigma_{ww} \end{pmatrix} \begin{pmatrix} \Sigma_{wy} \\ \Sigma_{zy} \end{pmatrix},
\end{aligned}
$$

and therefore

$$
\widehat{\beta}_1^{\mathrm{OLS}} = \frac{\Sigma_{zz}\Sigma_{wy} - \Sigma_{wz}\Sigma_{zy}}{\Sigma_{ww}\Sigma_{zz} - \Sigma_{wz}^2}.
$$

Comparing the above results we find $\widehat{\beta}_1^{\mathrm{OLS}} = \widehat{\beta}_1.$. Thus, it does not matter whether we regress y_i on x_i and z_i jointly, or whether we partition the regression and first eliminate z_i from the model and afterwards regress $\widetilde{y}_i$ on $\widetilde{x}_i$.

(c) We already know that $\widehat{\beta}_1^{\mathrm{OLS}} = \widehat{\beta}_1$, and we have

$$
\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} w_i \, \widetilde{y}_i}{\sum_{i=1}^{n} \widetilde{w}_i^2} = \widehat{\beta}_1^* \, \frac{\sum_{i=1}^{n} w_i^2}{\sum_{i=1}^{n} \widetilde{w}_i^2},
$$

and therefore

$$
\frac{\widehat{\beta}_1^{\mathrm{OLS}}}{\widehat{\beta}_1^*} = \frac{\sum_{i=1}^{n} w_i^2}{\sum_{i=1}^{n} \widetilde{w}_i^2} > 1,
$$

because projecting our z_i reduces the total variation in w_i. We therefore have

$$
\left| \widehat{\beta}_1^* \right| < \left| \widehat{\beta}_1^{\mathrm{OLS}} \right|.
$$

(d) Regressing $\widetilde{y}_i$ on z_i does <u>not</u> give a consistent estimator for β_2, because we have $\sum_{i=1}^{n} \widetilde{y}_i z_i = 0$, so we always obtain an estimator equal to zero from this regression, while we have $\beta_2 \neq 0$.

Section B

Answer **TWO questions from this Section**.

Question B1 (20 points)

(a) Moment conditions are given by

$$\mathbb{E}\left[y_i - \frac{\theta_1+\theta_2}{2}\right] = 0,$$
$$\mathbb{E}\left[y_i^2 - \frac{(\theta_2-\theta_1)^2}{12} - \left(\frac{\theta_1+\theta_2}{2}\right)^2\right] = 0.$$

These are two moment conditions in two unknown parameters. The corresponding MM estimator solves the sample analog of the moment conditions

$$\frac{1}{n}\sum_{i=1}^{n}\left[y_i - \frac{\widehat{\theta}_{1,\mathrm{MM}}+\widehat{\theta}_{2,\mathrm{MM}}}{2}\right] = 0,$$
$$\frac{1}{n}\sum_{i=1}^{n}\left[y_i^2 - \frac{(\widehat{\theta}_{2,\mathrm{MM}}-\widehat{\theta}_{1,\mathrm{MM}})^2}{12} - \left(\frac{\widehat{\theta}_{1,\mathrm{MM}}+\widehat{\theta}_{2,\mathrm{MM}}}{2}\right)^2\right] = 0.$$

Thus, $\frac{\widehat{\theta}_{1,\mathrm{MM}}+\widehat{\theta}_{2,\mathrm{MM}}}{2}$ equals the sample mean, and $\frac{(\widehat{\theta}_{2,\mathrm{MM}}-\widehat{\theta}_{1,\mathrm{MM}})^2}{12}$ equals the sample variance, that is, we have

$$\frac{\widehat{\theta}_{1,\mathrm{MM}}+\widehat{\theta}_{2,\mathrm{MM}}}{2} = \frac{1}{n}\sum_{i=1}^{n} y_i,$$
$$\frac{\widehat{\theta}_{2,\mathrm{MM}}-\widehat{\theta}_{1,\mathrm{MM}}}{\sqrt{12}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)^2}.$$

Solving this system of linear equations for $\widehat{\theta}_{1,\mathrm{MM}}$ and $\widehat{\theta}_{1,\mathrm{MM}}$ gives

$$\widehat{\theta}_{1,\mathrm{MM}} = \frac{1}{n}\sum_{i=1}^{n} y_i - \sqrt{3\left[\frac{1}{n}\sum_{i=1}^{n} y_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)^2\right]},$$
$$\widehat{\theta}_{2,\mathrm{MM}} = \frac{1}{n}\sum_{i=1}^{n} y_i + \sqrt{3\left[\frac{1}{n}\sum_{i=1}^{n} y_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)^2\right]}.$$

(b) We can write

$$f(y_i|\theta) = \frac{1}{\theta_2 - \theta_1} 1(y_i \in [\theta_1, \theta_2]),$$

where 1(.) is the indicator function. The likelihood function for the whole sample thus reads

$$\begin{aligned} f(y|\theta) &= \prod_{i=1}^{n} f(y_i|\theta) \\ &= \frac{1}{(\theta_2 - \theta_1)^n} \prod_{i=1}^{n} 1(y_i \in [\theta_1, \theta_2]) \\ &= \begin{cases} \frac{1}{(\theta_2-\theta_1)^n} & \text{if } y_i \in [\theta_1, \theta_2] \text{ for all } i = 1, \dots, n, \\ 0 & \text{otherwise} \end{cases} \\ &= \frac{1}{(\theta_2 - \theta_1)^n} 1 \left[\left(\min_{i=1,\dots,n} y_i \right) \leq \theta_1 \leq \theta_2 \leq \left(\max_{i=1,\dots,n} y_i \right) \right]. \end{aligned}$$

Thus, maximizing $f(y_i|\theta)$ over $\theta = (\theta_1, \theta_2)$ is equivalent to solving

$$\min_{\theta} \ (\theta_2 - \theta_1) \qquad \text{subject to} \qquad \left(\min_{i=1,\dots,n} y_i \right) \leq \theta_1 \leq \theta_2 \leq \left(\max_{i=1,\dots,n} y_i \right).$$

The solution to this is

$$\widehat{\theta}_{\text{MLE}} = (\widehat{\theta}_{1,\text{MLE}}, \widehat{\theta}_{2,\text{MLE}}) = \left(\min_{i=1,\dots,n} y_i \, , \, \max_{i=1,\dots,n} y_i \right).$$

Thus, we have $\widehat{\theta}_{\text{MLE}} \neq \widehat{\theta}_{\text{MM}}$.

(c) We have

$$\widehat{\mu}_{\text{MM}} = \frac{1}{n} \sum_{i=1}^{n} y_i,$$

and by applying the CLT for iid processes we therefore find as $n \to \infty$ that

$$\sqrt{n}(\widehat{\mu}_{\text{MM}} - \mu) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (y_i - \mathbb{E} y_i) \Rightarrow \mathcal{N}\left(0, V_{\text{MM}}\right), \quad V_{\text{MM}} = \text{Var}(y_i) = \frac{(\theta_2 - \theta_1)^2}{12}.$$

(d) Because both $\widehat{\mu}_{\text{MM}}$ and $\widehat{\mu}_{\text{MLE}}$ are consistent for μ, we find that $\widehat{\mu} = 2\widehat{\mu}_{\text{MM}} - \widehat{\mu}_{\text{MLE}}$ is also <u>consistent for μ</u>.

For the asymptotic distribution of $\widehat{\mu}$ we find

$$\begin{aligned}
\sqrt{n}(\widehat{\mu} - \mu) &= 2\sqrt{n}(\widehat{\mu}_{\mathrm{MM}} - \mu) - \sqrt{n}(\widehat{\mu}_{\mathrm{MLE}} - \mu) \\
&\to_p 2\sqrt{n}(\widehat{\mu}_{\mathrm{MM}} - \mu) \\
&\Rightarrow \mathcal{N}\left(0, 4\, V_{\mathrm{MM}}\right).
\end{aligned}$$

Thus, we have $\mathrm{AsyVar}(\widehat{\mu}) > \mathrm{AsyVar}(\widehat{\mu}_{\mathrm{MM}})$, that is, $\widehat{\mu}$ is the <u>less efficient</u> estimator.

Question B2 (20 points)

(a) ONE POSSIBLE SOLUTION: From the log-likelihood

$$\log f(y_i|\theta) = [\log(1-\theta_1-\theta_2)]^{1(y_i=0)}\ [\log\theta_1]^{1(y_i=1)}\ [\log\theta_2]^{1(y_i=2)},$$

we calculate the score

$$\begin{aligned} s(y_i,\theta) &= \frac{\partial \log f(y_i|\theta)}{\partial\theta} \\ &= \begin{pmatrix} -\frac{1(y_i=0)}{1-\theta_1-\theta_2} + \frac{1(y_i=1)}{\theta_1} \\ -\frac{1(y_i=0)}{1-\theta_1-\theta_2} + \frac{1(y_i=2)}{\theta_2} \end{pmatrix}, \end{aligned}$$

and the Hessian

$$\begin{aligned} H(y_i,\theta) &= \frac{\partial^2 \log f(y_i|\theta)}{\partial\theta\partial\theta'} \\ &= \begin{pmatrix} -\frac{1(y_i=0)}{(1-\theta_1-\theta_2)^2} - \frac{1(y_i=1)}{\theta_1^2} & -\frac{1(y_i=0)}{(1-\theta_1-\theta_2)^2} \\ -\frac{1(y_i=0)}{(1-\theta_1-\theta_2)^2} & -\frac{1(y_i=0)}{(1-\theta_1-\theta_2)^2} - \frac{1(y_i=2)}{\theta_2^2} \end{pmatrix}. \end{aligned}$$

Thus, the expected Hessian evaluated at the true parameter reads

$$\mathbb{E}[H(y_i,\theta_0)] = \begin{pmatrix} -\frac{1}{1-\theta_{1,0}-\theta_{2,0}} - \frac{1}{\theta_{1,0}} & -\frac{1}{1-\theta_{1,0}-\theta_{2,0}} \\ -\frac{1}{1-\theta_{1,0}-\theta_{2,0}} & -\frac{1}{1-\theta_{1,0}-\theta_{2,0}} - \frac{1}{\theta_{2,0}} \end{pmatrix}.$$

From the lecture we know that for a correctly specified MLE we have

$$\begin{aligned} V_{\text{MLE}} &= -\left\{\mathbb{E}[H(y_i,\theta_0)]\right\}^{-1} \\ &= \begin{pmatrix} \theta_{1,0}(1-\theta_{1,0}) & -\theta_{1,0}\,\theta_{2,0} \\ -\theta_{1,0}\,\theta_{2,0} & \theta_{2,0}(1-\theta_{2,0}) \end{pmatrix}. \end{aligned}$$

SECOND POSSIBLE SOLUTION: Alternatively, one can calculate the MLE explicitly. Solving the FOC $\sum_{i=1}^n s(y_i,\theta)$ gives

$$\widehat{\theta}_{1,\text{MLE}} = \frac{1}{n}\sum_{i=1}^n 1(y_i=1), \qquad \widehat{\theta}_{2,\text{MLE}} = \frac{1}{n}\sum_{i=1}^n 1(y_i=2).$$

Calculating V_{MLE} is then just an application of the CLT, see solution for (b) below. Of course, one finds the same result for V_{MLE}.

(b) The number of moment conditions equals the number of parameters here, that is, the GMM estimator is simply the MM estimator, which solves

$$\sum_{i=1}^{n} g(y_i, \widehat{\theta}_{\text{GMM}}) = 0.$$

Thus, we already know that <u>the choice of weight matrix W does not matter here,</u> we can choose any positive definite matrix W and obtain the same estimator $\widehat{\theta}_{\text{GMM}} = \widehat{\theta}_{\text{GMM}}(W)$, and the asymptotic variance $V_{\text{GMM}} = V_{\text{GMM}}(W)$ also does not depend on W either.

Solving

$$\sum_{i=1}^{n} \begin{pmatrix} 1(y_i = 1) - \widehat{\theta}_{1,\text{GMM}} \\ 1(y_i = 2) - \widehat{\theta}_{2,\text{GMM}} \end{pmatrix} = 0$$

we find

$$\widehat{\theta}_{1,\text{GMM}} = \frac{1}{n} \sum_{i=1}^{n} 1(y_i = 1), \qquad \widehat{\theta}_{2,\text{GMM}} = \frac{1}{n} \sum_{i=1}^{n} 1(y_i = 2).$$

Applying the CLT we thus find that

$$\sqrt{n} \left(\widehat{\theta}_{\text{GMM}} - \theta_0 \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} 1(y_i = 1) - \theta_{1,0} \\ 1(y_i = 2) - \theta_{2,0} \end{pmatrix} \Rightarrow \mathcal{N}(0, V_{\text{GMM}}),$$

where

$$\begin{aligned} V_{\text{GMM}} = \operatorname{Var} \begin{pmatrix} 1(y_i = 1) \\ 1(y_i = 2) \end{pmatrix} &= \begin{pmatrix} \operatorname{Var} 1(y_i = 1) & \operatorname{Cov}[1(y_i = 1), 1(y_i = 2)] \\ \operatorname{Cov}[1(y_i = 1), 1(y_i = 2)] & \operatorname{Var} 1(y_i = 2) \end{pmatrix} \\ &= \begin{pmatrix} \theta_{1,0}(1 - \theta_{1,0}) & -\theta_{1,0}\, \theta_{2,0} \\ -\theta_{1,0}\, \theta_{2,0} & \theta_{2,0}(1 - \theta_{2,0}) \end{pmatrix}. \end{aligned}$$

We thus have $V_{\text{GMM}} = V_{\text{MLE}}$. Alternatively, one could have arrived by that result by noticing that $\widehat{\theta}_{\text{GMM}} = \widehat{\theta}_{\text{MLE}}$ — that would also be a correct solution, of course.

(c) The additional moment condition would <u>not</u> improve efficiency of the GMM estimator. Two possible arguments:

- We already found $\hat{\theta}_{\text{MLE}} = \hat{\theta}_{\text{GMM}}$, and the MLE is efficient, so we cannot improve efficiency further.

- The moment condition $\mathbb{E}[1(y_i = 0) - (1 - \theta_{1,0} - \theta_{2,0})] = 0$ does not add any new information, because it is simply a linear combination of the two moment conditions we already use, namely we have

$$1(y_i = 0) - (1 - \theta_1 - \theta_2) = -\{1(y_i = 1) - \theta_1\} - \{1(y_i = 1) - \theta_2\}.$$

(d) The result $\sqrt{n}\left(\widehat{\beta} - \beta_0\right) \Rightarrow \mathcal{N}(0, V_\beta)$ follows from the delta method, because we already know that $\widehat{\theta}_{\mathrm{MLE}}$ is asymptotically normally distributed, and the function $f(\theta) = \theta_1^2 + \theta_2^2$ is differentiable.

The gradient of the function reads

$$\nabla f(\theta) = \begin{pmatrix} 2\theta_1 \\ 2\theta_2 \end{pmatrix},$$

and we thus find that

$$\begin{aligned} V_\beta &= [\nabla f(\theta_0)]' \, V_{\mathrm{MLE}} \, [\nabla f(\theta_0)] \\ &= \begin{pmatrix} 2\theta_{1,0} \\ 2\theta_{2,0} \end{pmatrix}' \begin{pmatrix} \theta_{1,0}(1-\theta_{1,0}) & -\theta_{1,0}\,\theta_{2,0} \\ -\theta_{1,0}\,\theta_{2,0} & \theta_{2,0}(1-\theta_{2,0}) \end{pmatrix} \begin{pmatrix} 2\theta_{1,0} \\ 2\theta_{2,0} \end{pmatrix} \\ &= 8\theta_{1,0}^3 - 16\theta_{1,0}^4 = 8(1-2\theta_{1,0})\theta_{1,0}^3, \end{aligned}$$

where in the last line we used the simplifying assumption that $\theta_{1,0} = \theta_{2,0}$.

Question B3 (20 points)

(a) We calculate

$$\gamma_0 = \sigma^2(1+\theta^2),$$
$$\gamma_1 = \sigma^2\theta,$$
$$\gamma_2 = 0.$$

(b) The moment conditions are

$$\mathbb{E}\left\{y_t^2 - \left[\sigma^2(1+\theta^2)\right]\right\} = 0, \qquad \mathbb{E}\left(y_t y_{t-1} - \sigma^2\theta\right) = 0.$$

The corresponding MM estimator solves

$$20 = \hat{\sigma}^2(1+\hat{\theta}^2), \qquad 8 = \hat{\sigma}^2\hat{\theta}.$$

The two solutions to these equations are

$$\hat{\theta}_1 = 2, \qquad \hat{\sigma}_1 = 2,$$

and

$$\hat{\theta}_2 = 1/2, \qquad \hat{\sigma}_2 = 4.$$

(c) We have $\gamma_2 = 0$. Define $z_t = y_{2t}\, y_{2t+2}$. The question already states that

$$\frac{1}{\sqrt{(T-2)/2}} \sum_{t=1}^{(T-2)/2} z_t \Rightarrow \mathcal{N}(0, V),$$

which follows by an application of an appropriate time series CLT.

It is easy to verify that the process z_t has mean zero and is uncorrelated over $t = 1, \ldots, (T-2)/2$,[1] and we therefore simply have

$$V = \mathrm{Var}(z_t) = \mathbb{E}z_t^2 = \mathbb{E}(y_{2t}^2\, y_{2t+2}^2) = \mathbb{E}(y_{2t}^2)\mathbb{E}(y_{2t+2}^2) = \gamma_0^2 = \sigma^4(1+\theta^2)^2.$$

[1] Because we wanted z_t to be uncorrelated we defined $z_t = y_{2t}\, y_{2t+2}$, and not $z_t = y_t\, y_{t+2}$. With the latter definition we would have that z_t and z_{t+1} are correlated, which would complicate the computation of the asymptotic variance in this question.

(d) In (a) we found that $\gamma_2 = 0$, that is, we can test whether y_t is an MA(1) process by testing the null hypothesis that $H_0 : \mathbb{E}(y_t\, y_{t+2}) = 0$. To test this null hypothesis we use the test statistics

$$t = \frac{1}{\sqrt{\widehat{V}}} \frac{1}{\sqrt{(T-2)/2}} \sum_{t=1}^{(T-2)/2} y_{2t}\, y_{2t+2},$$

where $\widehat{V} = \widehat{\gamma}_0^2 = \left(\frac{1}{T}\sum_{t=1}^{T} y_t^2\right)^2 = 20^2 = 400$ is a consistent estimator for the asymptotic variance V. From the result in part (c) and Slutsky's theorem we then know that $t \Rightarrow \mathcal{N}(0,1)$ as $T \to \infty$. Thus, we would reject H_0 in a two-sides asymptotic test if we would find that $|t| > 1.96$.

From the information provided we calcuate

$$\begin{aligned} t &= \left(\frac{1}{\sqrt{\widehat{V}}}\right) \sqrt{(T-2)/2} \left(\frac{1}{(T-2)/2} \sum_{t=1}^{(T-2)/2} y_{2t}\, y_{2t+2}\right) \\ &= \frac{1}{20}\sqrt{100}\,(-2) = -1.5. \end{aligned}$$

Thus, because $|t| < 1.96$ we cannot reject the null hypotheses that the observed process is an MA(1) process.

SOLUTIONS

ECON0064: Examination in Econometrics

2018-2019

Section A

Question A1 (20 points)

(a) Using that $h_i^2 = h_i$ we find

$$\sum_{i=1}^{n} x_i' x_i = \sum_{i=1}^{n} \binom{1}{h_i} (1, h_i) = \sum_{i=1}^{n} \begin{pmatrix} 1 & h_i \\ h_i & h_i^2 \end{pmatrix} = \sum_{i=1}^{n} \begin{pmatrix} 1 & h_i \\ h_i & h_i \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n h_i \\ \sum_{i=1}^n h_i & \sum_{i=1}^n h_i \end{pmatrix}$$
$$= \begin{pmatrix} n & n_1 \\ n_1 & n_1 \end{pmatrix} = \begin{pmatrix} 120 & 40 \\ 40 & 40 \end{pmatrix}.$$

Using that $\sum_{i=1}^n y_i = n_0 \left[\frac{1}{n_0} \sum_{i=1}^n (1-h_i) y_i \right] + n_1 \left[\frac{1}{n_1} \sum_{i=1}^n h_i y_i \right]$ we find

$$\sum_{i=1}^{n} x_i' y_i = \sum_{i=1}^{n} \binom{1}{h_i} y_i = \sum_{i=1}^{n} \binom{y_i}{h_i y_i} = \binom{\sum_{i=1}^n y_i}{\sum_{i=1}^n h_i y_i}$$
$$= \begin{pmatrix} n_0 \left[\frac{1}{n_0} \sum_{i=1}^n (1-h_i) y_i \right] + n_1 \left[\frac{1}{n_1} \sum_{i=1}^n h_i y_i \right] \\ n_1 \left[\frac{1}{n_1} \sum_{i=1}^n h_i y_i \right] \end{pmatrix} = \begin{pmatrix} 80 \times 4 + 40 \times 0.5 \\ 40 \times 0.5 \end{pmatrix}$$
$$= \begin{pmatrix} 340 \\ 20 \end{pmatrix}.$$

We thus find

$$\widehat{\beta} = \binom{\widehat{\beta}_1}{\widehat{\beta}_2} = \left(\sum_{i=1}^{n} x_i' x_i \right)^{-1} \sum_{i=1}^{n} x_i' y_i = \begin{pmatrix} 120 & 40 \\ 40 & 40 \end{pmatrix}^{-1} \begin{pmatrix} 340 \\ 20 \end{pmatrix}$$
$$= \frac{1}{80} \begin{pmatrix} 1 & -1 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} 340 \\ 20 \end{pmatrix}$$
$$= \binom{4}{-3.5}.$$

(b) Under homoscedasticity with known $\sigma^2 = \mathbb{E}(u_i^2|h_i) = 20$ the OLS estimator for the variance of $\widehat{\beta} = \binom{\widehat{\beta}_1}{\widehat{\beta}_2}$ is given by

$$\widehat{\mathrm{Var}}(\widehat{\beta}) = \sigma^2 \left(\sum_{i=1}^{n} x_i' x_i \right)^{-1} = 20 \, \frac{1}{80} \begin{pmatrix} 1 & -1 \\ -1 & 3 \end{pmatrix} = \begin{pmatrix} 1/4 & -1/4 \\ -1/4 & 3/4 \end{pmatrix}.$$

We thus have

$$\widehat{\mathrm{std}}(\widehat{\beta}_1) = \sqrt{\widehat{\mathrm{Var}}(\widehat{\beta}_1)} = \frac{1}{2},$$
$$\widehat{\mathrm{std}}(\widehat{\beta}_2) = \sqrt{\widehat{\mathrm{Var}}(\widehat{\beta}_2)} = \frac{\sqrt{3}}{2}.$$

(c) The t-test statistics reads

$$t = \frac{\widehat{\beta}_2}{\widehat{\mathrm{std}}(\widehat{\beta}_2)} = \frac{-2}{0.5} = -4.$$

We can reject $H_0 : \beta_2 \geq 0$ in a one-sided t-test at 95% confidence level if $t < -1.64$. Since indeed $-4 < -1.64$ we **do reject** H_0.

(d) We have $\hat{\delta} = \hat{\theta} = \hat{\beta}_2$. This is an example of a partitioned regression: $\tilde{y}$ is the part of y that is not explained by the constant, and $\tilde{h}$ is the part of h that is not explained by the constant. Thus, regressing y on $\tilde{h}$ gives the same result that we get by regressing y on h and a constant. We get the same result from regressing y on $\tilde{h}$ as we get from regressing $\tilde{y}$ on $\tilde{h}$, because $\tilde{h}$ is orthogonal to the constant, i.e. $\sum_{i=1}^{n} \tilde{h}_i = 0$.

(e) Yes, the assumption $\mathbb{E}(u_i|h_i) = 0$ is sufficient to ensure that we estimate true causal effects. To guarantee that $\widehat{\beta}_2$ estimates the true causal effect of smoking on health we would need that the regressor is as good as randomly assigned to individuals. The question specifies that.

Question A2 (20 points)

(a) We have as $n \to \infty$

$$\widehat{\beta} = \frac{\frac{1}{n}\sum_{i=1}^{n} z_i y_i}{\frac{1}{n}\sum_{i=1}^{n} z_i x_i} \to_p \frac{\mathbb{E}(z_i y_i)}{\mathbb{E}(z_i x_i)} = \beta^*$$

where we have used the WLLN in both the numerator and the denominator, and we also applied the CMT. By also using the model for y_i and the information on x_i, z_i, u_i we obtain

$$\beta^* = \frac{\mathbb{E}(z_i y_i)}{\mathbb{E}(z_i x_i)} = \frac{\beta\mathbb{E}(z_i x_i) + \mathbb{E}(z_i u_i)}{\mathbb{E}(z_i x_i)} = \beta + \frac{\rho_{zu}}{\rho_{xz}}.$$

(b)
- If $\rho_{zu} = 0$, then $\beta^* = \beta$, so that $\widehat{\beta}$ is consistent for β.
- The condition $\rho_{zu} = 0$ states that z_i is exogenous (uncorrelated with the error u_i), which is also called the exclusion restriction.
- The condition $\rho_{xz} \neq 0$ guarantees that z_i is correlated with x_i, i.e. that the instrument z_i is relevant for the endogenous regressor x_i.

(c) Using the model we find

$$\sqrt{n}\left(\widehat{\beta} - \beta\right) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i u_i}{\frac{1}{n}\sum_{i=1}^{n} z_i x_i}$$

Since we assume $\rho_{zu} = \mathbb{E}(z_i u_i) = 0$ we can apply the CLT to find that as $n \to \infty$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i u_i \Rightarrow \mathcal{N}(0, \mathrm{Var}(z_i u_i)).$$

As above, by the WLLN we have $\frac{1}{n}\sum_{i=1}^{n} z_i x_i \to_p \rho_{xz}$. Applying Slutsky's theorem we thus find as $n \to \infty$

$$\sqrt{n}\left(\widehat{\beta} - \beta\right) \Rightarrow \frac{\mathcal{N}(0, \mathrm{Var}(z_i u_i))}{\rho_{xz}} = \mathcal{N}\left(0, \frac{\mathrm{Var}(z_i u_i)}{\rho_{xz}^2}\right),$$

i.e.

$$\mathrm{AsyVar}(\sqrt{n}\widehat{\beta}) = \frac{\mathrm{Var}(z_i u_i)}{\rho_{xz}^2}.$$

Note that $\mathrm{Var}(z_i u_i) = \mathbb{E}[(z_i u_i)^2]$.

(d) Using the information on the observed sample we calculate

$$\widehat{\beta} = \frac{\frac{1}{n}\sum_{i=1}^{n} z_i y_i}{\frac{1}{n}\sum_{i=1}^{n} z_i x_i} = \frac{0.4}{0.2} = 2,$$

$$\widehat{\rho}_{xz} = \frac{1}{n}\sum_{i=1}^{n} z_i x_i = 0.2.$$

We need to estimate $\mathrm{Var}(z_i u_i)$.

Define the residuals $\widehat{u}_i = y_i - x_i\widehat{\beta}$. An estimator for $\mathrm{Var}(z_i u_i) = \mathbb{E}[(z_i u_i)^2]$ is then given by

$$\begin{aligned}
\widehat{\mathbb{E}}[(z_i u_i)^2] &= \frac{1}{n}\sum_{i=1}^{n}(z_i \widehat{u}_i)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} z_i^2 (y_i - x_i\widehat{\beta})^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} z_i^2 y_i^2 - 2\widehat{\beta}\frac{1}{n}\sum_{i=1}^{n} z_i^2 x_i y_i + \widehat{\beta}^2\frac{1}{n}\sum_{i=1}^{n} z_i^2 x_i^2 \\
&= 5 - 2 \times 2 \times 0.5 + 2^2 \times 3 = 15.
\end{aligned}$$

A consistent estimator for the asymptotic variance is thus given by

$$\widehat{\mathrm{AsyVar}}(\sqrt{n}\widehat{\beta}) = \frac{\widehat{\mathbb{E}}[(z_i u_i)^2]}{\widehat{\rho}_{xz}^2} = \frac{15}{(0.2)^2} = 375.$$

An estimator for the variance of $\widehat{\beta}$ is thus given by

$$\widehat{\mathrm{Var}}(\widehat{\beta}) = \frac{1}{n}\widehat{\mathrm{AsyVar}}(\sqrt{n}\widehat{\beta}) = \frac{375}{200} = 1.875,$$

implying that $\widehat{\mathrm{std}}(\widehat{\beta}) = 1.369$. A 95% confidence interval for β is thus given by

$$\begin{aligned}
\mathrm{CI}_{95\%} &= \left[\widehat{\beta} - 1.96 \times \widehat{\mathrm{std}}(\widehat{\beta}), \widehat{\beta} + 1.96 \times \widehat{\mathrm{std}}(\widehat{\beta})\right] = [2 - 1.96 \times 1.369, 2 + 1.96 \times 1.369] \\
&= [-0.683, 4.683].
\end{aligned}$$

(e) We have:

$$y_i = x_i\beta + u_i = x_i\beta + w_i\gamma + \epsilon_i \ ,$$

By using the model for y_i and the information on x_i, z_i, w_i, ϵ_i we obtain:

$$\widehat{\beta} = \frac{\frac{1}{n}\sum_{i=1}^{n} z_i y_i}{\frac{1}{n}\sum_{i=1}^{n} z_i x_i} = \frac{\beta\frac{1}{n}\sum_{i=1}^{n}(z_i x_i) + \gamma\frac{1}{n}\sum_{i=1}^{n}(z_i w_i) + \frac{1}{n}\sum_{i=1}^{n}(z_i\epsilon_i)}{\frac{1}{n}\sum_{i=1}^{n} z_i x_i}$$

As $n \to \infty$

$$\widehat{\beta} \to_p \frac{\beta \mathbb{E}(z_i x_i) + \gamma \mathbb{E}(z_i w_i) + \mathbb{E}(z_i \epsilon_i)}{\mathbb{E}(z_i x_i)} = \beta + \gamma \frac{\mathbb{E}(z_i w_i)}{\mathbb{E}(z_i x_i)}.$$

where we have used the WLLN in both the numerator and the denominator, and we also applied the CMT.

We conclude that $\widehat{\beta}$ is not a consistent estimator for β in this case since w_i are omitted variables in the IV regression.

Question A3 (20 points)

(a) We define the weights

$$w(x_i) = \frac{1}{h(x_i)}.$$

Notice that those weights could be multiplied with any constant, without changing anything in the following. The WLS estimator reads

$$\hat{\beta}_{\text{WLS}} = \frac{\sum_{i=1}^n w(x_i)\ x_i y_i}{\sum_{i=1}^n w(x_i)\ x_i^2}$$

Using the model, CLT, WLLN and Slutsky's theorem, we find, as $n \to \infty$,

$$\hat{\beta}_{\text{WLS}} = \frac{\sum_{i=1}^n w(x_i)\ x_i y_i}{\sum_{i=1}^n w(x_i)\ x_i^2} \to_p \frac{\beta \mathbb{E}(w(x_i)\ x_i y_i)}{\mathbb{E} w(x_i)\ x_i^2} = \beta + \frac{\mathbb{E}(w(x_i)\ x_i u_i)}{\mathbb{E}[w(x_i)\ x_i^2]} = \beta$$

(b) As usual, using the model, CLT, WLLN and Slutsky's theorem, we find, as $n \to \infty$,

$$\sqrt{n}\left(\hat{\beta}_{\text{WLS}} - \beta\right) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n w(x_i)\ x_i u_i}{\frac{1}{n} \sum_{i=1}^n w(x_i)\ x_i^2} \Rightarrow \frac{\mathcal{N}[0, \mathbb{E}(w(x_i)^2\ x_i^2 u_i^2)]}{\mathbb{E}[w(x_i)\ x_i^2]} \sim \mathcal{N}\left(0, V\right),$$

where we find and simplify the asymptotic variance as follows

$$V = \frac{\mathbb{E}(w(x_i)^2\ x_i^2 u_i^2)}{\{\mathbb{E}[w(x_i)\ x_i^2]\}^2} = \frac{\mathbb{E}[w(x_i)^2\ x_i^2 \mathbb{E}(u_i^2|x_i)]}{\{\mathbb{E}[w(x_i)\ x_i^2]\}^2} = \frac{\mathbb{E}[w(x_i)\ x_i^2]}{\{\mathbb{E}[w(x_i)\ x_i^2]\}^2} = \left\{\mathbb{E}[w(x_i)\ x_i^2]\right\}^{-1}$$
$$= \left\{\mathbb{E}\left[\frac{x_i^2}{h(x_i)}\right]\right\}^{-1}.$$

(c) Define the residuals $\widehat{u}_i = y_i - x_i \widehat{\beta}$. We can estimate γ by applying OLS to the following equation:

$$\widehat{u}_i^2 - 3 = \gamma x_i^2 + \epsilon_i$$

The resulting estimator reads:

$$\widehat{\gamma} = \frac{\sum_{i=1}^n x_i^2 (\widehat{u}_i^2 - 3)}{\sum_{i=1}^n x_i^4}$$

We can still estimate β by the weighted least squares (WLS) estimator in a consistent and efficient way. The optimal weights are given by $w_i = \frac{1}{3+\gamma x_i^2}$ (but they could also be multiplied with an arbitrary constant, which would have no effect on $\hat{\beta}_{\text{WLS}}$). To make the WLS estimator feasible, we have to use the feasible weights $\widehat{w}_i = \frac{1}{3+\widehat{\gamma} x_i^2}$ where $\widehat{\gamma}$ is the above OLS estimator.

(d) The derivation of asymptotic normality in part (b) is still valid, since the feasible weights are consistent estimator for the true weights. As usual, using the model, CLT, WLLN and Slutsky's theorem, we find, as $n \to \infty$,

$$\sqrt{n}\left(\hat{\beta}_{\text{FWLS}} - \beta\right) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widehat{w}(x_i)\ x_i u_i}{\frac{1}{n} \sum_{i=1}^{n} \widehat{w}(x_i)\ x_i^2} \Rightarrow \frac{\mathcal{N}[0, \mathbb{E}(w(x_i)^2\ x_i^2 u_i^2)]}{\mathbb{E}[w(x_i)\ x_i^2]} \sim \mathcal{N}\left(0, V^*\right),$$

where the asymptotic variance is

$$V^* = \left\{\mathbb{E}\left[x_i^2 w(x_i)\right]\right\}^{-1} = \left\{\mathbb{E}\left[\frac{x_i^2}{3 + \gamma x_i^2}\right]\right\}^{-1}$$

An estimator for V^* is obtained by sample analog and by CMT as follows:

$$\widehat{V}^* = \left\{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{x_i^2}{3 + \widehat{\gamma} x_i^2}\right]\right\}^{-1}$$

where $\widehat{\gamma}$ is the OLS estimator from part (c).

(e) If $u_i \,|\, x_i \sim t(5)$, then $\hat{\beta}_{\text{WLS}}$ is <u>not</u> efficient anymore, because the MLE has a smaller asymptotic variance. In (a) we assumed $u_i \,|\, x_i \sim \mathcal{N}(0, h(x_i))$, in which case the MLE is equal to the WLS estimator, but for non-normal error this is not the case anymore.

Section B

Answer **TWO questions from this Section**.

Question B1 (20 points)

(a) We find

$$
\begin{aligned}
Q_n(\theta) &= \frac{1}{n}\log\prod_{i=1}^{n} f(y_i|\theta) \\
&= \frac{1}{n}\sum_{i=1}^{n}\log f(y_i|\theta) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[(1-y_i)\log\left(\frac{1}{1+e^{\theta}}\right) + y_i\log\left(\frac{1}{1+e^{-\theta}}\right)\right] \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left[(1-y_i)\log\left(1+e^{\theta}\right) + y_i\log\left(1+e^{-\theta}\right)\right].
\end{aligned}
$$

Taking the derivative wrt θ we obtain

$$
\begin{aligned}
\frac{dQ_n(\theta)}{d\theta} &= -\frac{1}{n}\sum_{i=1}^{n}\left[(1-y_i)\frac{e^{\theta}}{1+e^{\theta}} - y_i\frac{e^{-\theta}}{1+e^{-\theta}}\right] \\
&= -\frac{1}{n}\sum_{i=1}^{n}\frac{e^{\theta}}{1+e^{\theta}} + \frac{1}{n}\sum_{i=1}^{n}y_i\left(\frac{e^{\theta}}{1+e^{\theta}} + \frac{e^{-\theta}}{1+e^{-\theta}}\right) \\
&= \underbrace{-\frac{e^{\theta}}{1+e^{\theta}}}_{=a(\theta)} + \frac{1}{n}\sum_{i=1}^{n}y_i.
\end{aligned}
$$

where in the last we used that $\frac{1}{n}\sum_{i=1}^{n}a(\theta) = a(\theta)$ and $\frac{e^{\theta}}{1+e^{\theta}} + \frac{e^{-\theta}}{1+e^{-\theta}} = \frac{e^{\theta}}{1+e^{\theta}} + \frac{1}{e^{\theta}+1} = 1$. Note that $a(\theta)$ can also be rewritten as

$$
a(\theta) = -\frac{1}{1+e^{-\theta}}.
$$

(b) Using the result in (a) the FOC $\frac{dQ_n(\widehat{\theta})}{d\theta} = 0$ reads

$$
\frac{1}{1+\exp(-\widehat{\theta})} = \frac{1}{n}\sum_{i=1}^{n}y_i.
$$

Thus

$$1+\exp(-\widehat{\theta}) = \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)^{-1},$$

and therefore

$$\widehat{\theta} = -\log\left[\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)^{-1} - 1\right].$$

Alternatively, this can be written as $\widehat{\theta} = \log\left[\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) / \left(1-\frac{1}{n}\sum_{i=1}^{n} y_i\right)\right]$.

(c) Within the calculation of $\frac{dQ_n(\theta)}{d\theta}$ we have already shown that

$$\frac{d\log f(y_i|\theta)}{d\theta} = -\frac{1}{1+e^{-\theta}} + y_i.$$

Taking one more derivative we obtain

$$\begin{aligned}\frac{d^2\log f(y_i|\theta)}{d\theta^2} &= -\frac{e^{-\theta}}{\left(1+e^{-\theta}\right)^2}\\ &= -\frac{e^{\theta}}{\left(1+e^{\theta}\right)^2}.\end{aligned}$$

It is a special feature of this model that this second derivative does not depend on the data anymore. Taking the expectation therefore is trivial, and we obtain

$$\mathbb{E}\left[\frac{d^2\log f(y_i|\theta)}{d\theta^2}\right] = -\frac{e^{\theta}}{\left(1+e^{\theta}\right)^2}.$$

From the lecture we know that for a correctly specified likelihood function we can use the information equality to find

$$\text{AsyVar}(\sqrt{n}\widehat{\theta}) = \left\{-\mathbb{E}\left[\frac{d^2\log f(y_i|\theta)}{d\theta^2}\right]\right\}^{-1},$$

and therefore

$$\text{AsyVar}(\sqrt{n}\widehat{\theta}) = \frac{\left(1+e^{\theta}\right)^2}{e^{\theta}} = \left(1+e^{\theta}\right)\left(1+e^{-\theta}\right).$$

(d) We have that $\widehat{\theta} = -\log\left[\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)^{-1} - 1\right] = -\log\left[(0.4)^{-1}-1\right] = -0.405$ and $\widehat{\text{std}}(\widehat{\theta}) = \sqrt{\frac{1}{n}\text{AsyVar}(\sqrt{n}\widehat{\theta})} = \sqrt{\frac{1}{n}\left(1+e^{\widehat{\theta}}\right)\left(1+e^{-\widehat{\theta}}\right)} = \sqrt{\frac{1}{20}\left(1+e^{-0.405}\right)\left(1+e^{0.405}\right)} = 0.46$.

(e) Using the definition for λ, we find

$$\lambda = \frac{P(y_i = 1 \mid \theta)}{P(y_i = 0 \mid \theta)} = \frac{1}{1+e^{-\theta}} \Big/ \frac{1}{1+e^{\theta}} = \frac{e^{\theta}}{1+e^{\theta}} \Big/ \frac{1}{1+e^{\theta}} = e^{\theta}.$$

A direct application of the Continuous Mapping Theorem (CMT) gives $\widehat{\lambda} = e^{\widehat{\theta}}$, where $\widehat{\theta}$ is the MLE estimator from part (b). The CMT applies in this case because the exponential function is continuous.

The asymptotic distribution of $\widehat{\lambda}$ follows from applying the Delta Method. Since the exponential function is differentiable, we get

$$\mathrm{AsyVar}(\sqrt{n}\widehat{\lambda}) = \left(\frac{d\lambda}{d\theta}\right)^2 \mathrm{AsyVar}(\sqrt{n}\widehat{\theta}) = e^{2\theta}\left(1+e^{\theta}\right)\left(1+e^{-\theta}\right).$$

Question B2 (20 points)

(a)
- The moment function reads $g(y_i, \sigma) = y_i^2 - \sigma^2$. However, the moment function is not uniquely determined, it could be multiplied with any non-zero function of σ. The moment function satisfies $\mathbb{E}[g(y_i, \sigma)] = 0$.
- The sample analog of the moment condition reads $\frac{1}{n}\sum_{i=1}^n y_i^2 = \widehat{\sigma}^2_{\text{MM}}$. Solving this we find $\widehat{\sigma}_{\text{MM}} = \sqrt{\frac{1}{n}\sum_{i=1}^n y_i^2}$.
- We could apply the CLT and Delta Method to calculate the asymptotic variance directly. But we can also rely on the general MM theory. We have $G = \mathbb{E}[\frac{dg(y_i,\sigma)}{d\sigma}] = -2\sigma$ and $\text{Var}[g(y_i, \sigma)] = \mathbb{E}\left[(y_i^2 - \sigma^2)^2\right] = 2\sigma^4$, and applying the general formula thus gives

$$\begin{aligned}
\text{AsyVar}(\sqrt{n}\,\widehat{\sigma}_{\text{MM}}) &= \left\{G'(\text{Var}[g(y_i,\sigma)])^{-1}G\right\}^{-1} \\
&= \frac{2\sigma^4}{(-2\sigma)^2} = \frac{\sigma^2}{2}.
\end{aligned}$$

(b) $g(y_i, \sigma)$ is given by

$$g(y_i,\sigma) = \begin{pmatrix} y_i^2 - \sigma^2 \\ y_i^4 - 3\sigma^4 \end{pmatrix}.$$

Again, the vector of moment functions is not uniquely determined, since its components could be multiplied by any non-zero function of σ.

The GMM objective function is given by

$$\begin{aligned}
Q_n(\sigma) &= \left[\frac{1}{n}\sum_{i=1}^n g(y_i,\sigma)\right]' W \left[\frac{1}{n}\sum_{i=1}^n g(y_i,\sigma)\right] \\
&= \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n y_i^2 - \sigma^2 \\ \frac{1}{n}\sum_{i=1}^n y_i^4 - 3\sigma^4 \end{pmatrix}' W \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n y_i^2 - \sigma^2 \\ \frac{1}{n}\sum_{i=1}^n y_i^4 - 3\sigma^4 \end{pmatrix},
\end{aligned}$$

where W is a symmetric positive definite 2×2 weight matrix. The GMM estimator is obtained by minimizing this objective function, i.e.

$$\widehat{\sigma}_{\text{GMM}} = \arg\min_{\sigma > 0} Q_n(\sigma).$$

(c)
- From the lecture we know that the asymptotic variance of the GMM estimator is minimized if we choose the weight matrix (asymptotically equal to)

$$W = (\text{Var}[g(y_i,\sigma)])^{-1}.$$

We calculate

$$\begin{aligned}
\operatorname{Var}[g(y_i,\sigma)] &= \mathbb{E}\left[g(y_i,\sigma)g(y_i,\sigma)'\right] \\
&= \begin{pmatrix} \mathbb{E}\left[\left(y_i^2-\sigma^2\right)^2\right] & \mathbb{E}\left[\left(y_i^2-\sigma^2\right)\left(y_i^4-3\sigma^2\right)\right] \\ \mathbb{E}\left[\left(y_i^2-\sigma^2\right)\left(y_i^4-3\sigma^2\right)\right] & \mathbb{E}\left[\left(y_i^4-3\sigma^2\right)^2\right] \end{pmatrix} \\
&= \begin{pmatrix} 2\sigma^4 & 12\sigma^6 \\ 12\sigma^6 & 96\sigma^8 \end{pmatrix},
\end{aligned}$$

and therefore for the optimal weight matrix

$$W = \left(\operatorname{Var}[g(y_i,\sigma)]\right)^{-1} = \begin{pmatrix} \frac{2}{\sigma^4} & -\frac{1}{4\sigma^6} \\ -\frac{1}{4\sigma^6} & \frac{1}{24\sigma^8} \end{pmatrix}.$$

However, the weight matrix could be multiplied by an arbitrary positive constant, without changing the GMM estimator.

– We calculate

$$G = \mathbb{E}\left[\frac{dg(y_i,\sigma)}{d\sigma}\right] = \begin{pmatrix} -2\sigma \\ -12\sigma^3 \end{pmatrix}.$$

Therefore

$$\begin{aligned}
\operatorname{AsyVar}(\sqrt{n}\,\widehat{\sigma}_{\text{GMM}}) &= \left\{G'\left(\operatorname{Var}[g(y_i,\sigma)]\right)^{-1}G\right\}^{-1} \\
&= \left\{\begin{pmatrix} -2\sigma \\ -12\sigma^3 \end{pmatrix}\begin{pmatrix} \frac{2}{\sigma^4} & -\frac{1}{4\sigma^6} \\ -\frac{1}{4\sigma^6} & \frac{1}{24\sigma^8} \end{pmatrix}\begin{pmatrix} -2\sigma \\ -12\sigma^3 \end{pmatrix}\right\}^{-1} \\
&= \left\{\frac{2}{\sigma^2}\right\}^{-1} = \frac{\sigma^2}{2}.
\end{aligned}$$

(d) The logarithm of the pdf is $\ln(f(y_i|\sigma)) = -\ln(\sqrt{2\pi}) - \ln(\sigma) - \frac{y_i^2}{2\sigma^2}$. It follows that the score function is $s(y_i,\sigma) = \frac{d\ln f}{d\sigma} = -\frac{1}{\sigma} + \frac{y_i^2}{\sigma^3}$. The corresponding Hessian is $H(y_i,\sigma) = \frac{1}{\sigma^2} - 3y_i^2\frac{1}{\sigma^4}$. From the lecture, we know that $\sqrt{n}(\widehat{\sigma}_{\text{MLE}} - \sigma) \Rightarrow \mathcal{N}(0, -\mathbb{E}\left[H(y_i;\sigma)\right]^{-1})$. Applying this result to the Hessian function, we obtain

$$\mathbb{E}\left[H(y_i,\sigma)\right] = \frac{1}{\sigma}^2 - 3\mathbb{E}(y_i^2)\frac{1}{\sigma}^4 = -\frac{2}{\sigma}^2 \implies -\mathbb{E}\left[H(y_i,\sigma)\right]^{-1} = \frac{\sigma^2}{2}.$$

(e) Putting together the results on the asymptotic variance for each estimator, we get

$$\operatorname{AsyVar}(\sqrt{n}\,\widehat{\sigma}_{\text{GMM}}) = \operatorname{AsyVar}(\sqrt{n}\,\widehat{\sigma}_{\text{MM}}) = \operatorname{AsyVar}(\sqrt{n}\,\widehat{\sigma}_{\text{MLE}}) = \frac{\sigma^2}{2}.$$

In this case, all estimators have the same asymptotic variance. We can conclude that adding extra moments to GMM should not improve efficiency.

The asymptotic variance for the MLE estimator is a useful benchmark in this context, because we showed in the lecture that it is a lower bound (the Cramer-Rao lower bound) on the asymptotic variance of all unbiased estimators. So it should not be possible to improve on its asymptotic variance.

Question B3 (20 points)

(a) $\gamma_0 = \mathbb{E}(y_t^2) = \mathbb{E}(\varepsilon_t^2) + 2\theta\mathbb{E}(\varepsilon_t\varepsilon_{t-1}) + \theta^2\mathbb{E}(\varepsilon_{t-1}^2) = \sigma^2(1+\theta^2)$,
$\gamma_1 = \mathbb{E}(y_t\, y_{t-1}) = \mathbb{E}(\varepsilon_t\varepsilon_{t-1}) + \theta\mathbb{E}(\varepsilon_{t-1}^2) + \theta\mathbb{E}(\varepsilon_t\varepsilon_{t-2}) + \theta^2\mathbb{E}(\varepsilon_{t-1}\varepsilon_{t-2}) = \sigma^2\theta$,
$\gamma_2 = \mathbb{E}(y_t\, y_{t-2}) = \mathbb{E}(\varepsilon_t\varepsilon_{t-2}) + \theta\mathbb{E}(\varepsilon_{t-1}\varepsilon_{t-2}) + \theta\mathbb{E}(\varepsilon_t\varepsilon_{t-3}) + \theta^2\mathbb{E}(\varepsilon_{t-1}\varepsilon_{t-3}) = 0$.

(b) The moment conditions are

$$\mathbb{E}(y_t^2) = \sigma^2(1+\theta^2), \qquad \mathbb{E}(y_t y_{t-1}) = \sigma^2\theta.$$

The MM estimator solves

$$45 = \hat{\sigma}^2(1+\hat{\theta}^2), \qquad 18 = \hat{\sigma}^2\hat{\theta}.$$

Solving these two equations for $\hat{\sigma}$ and $\hat{\theta}$ one finds that there are two solutions, given by

$$\hat{\theta}_1 = 2, \qquad \hat{\sigma}_1 = 3,$$

and

$$\hat{\theta}_2 = 1/2, \qquad \hat{\sigma}_2 = 6.$$

(c) There are two solutions in (b), i.e. the parameters are not uniquely identified.

Any restriction on θ that excludes one of the solutions above is enough to guarantee that the model is uniquely identified. For example, we can assume that $\theta = 1/2$.

Higher order autocovariances do not help, since for the $MA(1)$ model we have $\gamma_j = 0$ for $j > 1$, i.e. those moment conditions don't depend on the parameters, i.e. also deliver no information about the parameters.

(d) We know from the course that $|\rho| < 1$ guarantees that the AR(1) process is stationary and ergodic. Substituting the value for y_t recursively, we get

$$\gamma_2 = \mathbb{E}(y_t\, y_{t-2}) = \rho\mathbb{E}((y_{t-1}\, y_{t-2})) = \rho^2\mathbb{E}(y_{t-2}^2) = \rho^2\gamma_0$$

since $\gamma_0 = \mathbb{E}(y_t^2) = \mathbb{E}(y_{t-2}^2)$ by stationarity. The difference between the MA(1) and AR(1) model is that autocovariances slowly die out at rate ρ as the process moves forward.

The OLS estimator obtained from regressing y_t on y_{t-2} satisfies as $T \to \infty$

$$\hat{\beta} \to_p \frac{\mathbb{E}(y_t\, y_{t-2})}{\mathbb{E}(y_{t-2}^2)} = \frac{\gamma_2}{\gamma_0}.$$

For the MA(1) model we have $\hat{\beta} \to_p 0$, while for the AR(1) model $\hat{\beta} \to_p \rho^2$. We can therefore distinguish the MA(1) and AR(1) by testing whether $\hat{\beta}$ is close to zero or not (i.e. testing the hypothesis $H_0 : \beta = 0$).