Winter Vacation Mock Exam

Part A

Answer all of the following questions.

1

Consider the variables y_i, x_i, z_i for $i = 1, ..., n$ and the two regression equations

$$y_i = \alpha + \beta x_i + u_i, \tag{1}$$
$$x_i = \gamma + \phi y_i + v_i. \tag{2}$$

1.1

What are the least squares estimators for β and ϕ? Is the product of those estimators equal to unity? Interpret the result.

By the FWL theorem we have

$$\begin{aligned} \mathbf{M}_{\boldsymbol{\iota}}\mathbf{y} &= \mathbf{M}_{\boldsymbol{\iota}}\mathbf{x}\beta + \mathbf{M}_{\boldsymbol{\iota}}\mathbf{u}, \\ \implies \hat{\beta} &= (\mathbf{x}^\top \mathbf{M}_{\boldsymbol{\iota}}\mathbf{x})^{-1}\mathbf{x}^\top \mathbf{M}_{\boldsymbol{\iota}}\mathbf{y} \\ &= \frac{\mathbf{x}^\top \mathbf{M}_{\boldsymbol{\iota}}\mathbf{y}}{\mathbf{x}^\top \mathbf{M}_{\boldsymbol{\iota}}\mathbf{x}}, \end{aligned}$$

and

$$\begin{aligned} \mathbf{M}_{\boldsymbol{\iota}}\mathbf{x} &= \mathbf{M}_{\boldsymbol{\iota}}\mathbf{y}\phi + \mathbf{M}_{\boldsymbol{\iota}}\mathbf{v}, \\ \implies \hat{\phi} &= (\mathbf{y}^\top \mathbf{M}_{\boldsymbol{\iota}}\mathbf{y})^{-1}\mathbf{y}^\top \mathbf{M}_{\boldsymbol{\iota}}\mathbf{x} \\ &= \frac{\mathbf{y}^\top \mathbf{M}_{\boldsymbol{\iota}}\mathbf{x}}{\mathbf{y}^\top \mathbf{M}_{\boldsymbol{\iota}}\mathbf{y}}. \end{aligned}$$

Their product is given by

$$\begin{aligned} \hat{\beta}\hat{\phi} &= \frac{\mathbf{x}^\top \mathbf{M}_{\boldsymbol{\iota}}\mathbf{y}}{\mathbf{x}^\top \mathbf{M}_{\boldsymbol{\iota}}\mathbf{x}} \frac{\mathbf{y}^\top \mathbf{M}_{\boldsymbol{\iota}}\mathbf{x}}{\mathbf{y}^\top \mathbf{M}_{\boldsymbol{\iota}}\mathbf{y}} \\ &= \frac{\mathbf{x}^\top (\mathbf{I} - \mathbf{P}_{\boldsymbol{\iota}})\mathbf{y}}{\mathbf{x}^\top (\mathbf{I} - \mathbf{P}_{\boldsymbol{\iota}})\mathbf{x}} \frac{\mathbf{y}^\top (\mathbf{I} - \mathbf{P}_{\boldsymbol{\iota}})\mathbf{x}}{\mathbf{y}^\top (\mathbf{I} - \mathbf{P}_{\boldsymbol{\iota}})\mathbf{y}} \\ &= \frac{\mathbf{x}^\top (\mathbf{y} - \bar{y}\boldsymbol{\iota})}{\mathbf{x}^\top (\mathbf{x} - \bar{x}\boldsymbol{\iota})} \frac{\mathbf{y}^\top (\mathbf{x} - \bar{x}\boldsymbol{\iota})}{\mathbf{y}^\top (\mathbf{y} - \bar{y}\boldsymbol{\iota})}, \end{aligned} \tag{3}$$

where we use the facts that $\boldsymbol{\iota}^\top \boldsymbol{\iota} = n$ and $\boldsymbol{\iota}^\top \mathbf{x} = n\bar{x}$. Further, notice that 3 is analogous to

$$\frac{\widehat{\text{Cov}}(x_i, y_i)^2}{\widehat{\text{Var}}(x_i)\widehat{\text{Var}}(y_i)} = \frac{ESS}{TSS}$$

We know that typically for any vector $\mathbf{y}$ and matrix of regressors $\mathbf{X}$

$$TSS = ||\mathbf{y}||^2 = ||\mathbf{P_X y}||^2 + ||\mathbf{M_X y}||^2 = ESS + SSR,$$

and the goodness of fit or coefficient of determination is given by

$$R^2 = \frac{ESS}{TSS}.$$

Therefore (3) is the goodness of fit, $0 \leq R^2 \leq 1$.

1.2

What are the instrumental variable estimators for β and ϕ when using z_i as the instrument? Is the product of those estimators equal to unity? Interpret the result.

Suppose we estimate x_i by z_i in (1) and y_i by z_i in (2). We then have the following first stage regressions

$$\begin{aligned}\mathbf{x} &= \mathbf{z}\pi_1 + \mathbf{e}_1,\\ \mathbf{y} &= \mathbf{z}\pi_2 + \mathbf{e}_2,\end{aligned}$$

which we use to get

$$y_i = \alpha + \beta\hat{x}_i + u_i, \tag{4}$$

$$x_i = \gamma + \phi\hat{y}_i + v_i, \tag{5}$$

where

$$\begin{aligned}\hat{\mathbf{x}} &= \mathbf{P_z x} = \mathbf{z}\underbrace{(\mathbf{z}^\top\mathbf{z})^{-1}\mathbf{z}^\top\mathbf{x}}_{\hat{\pi}_1},\\ \hat{\mathbf{y}} &= \mathbf{P_z x} = \mathbf{z}\underbrace{(\mathbf{z}^\top\mathbf{z})^{-1}\mathbf{z}^\top\mathbf{x}}_{\hat{\pi}_1},\end{aligned}$$

and so, by the FWL theorem we can get 2SLS estimates for the coefficients

$$\begin{aligned}\mathbf{M}_{\boldsymbol{\iota}}\mathbf{y} &= \mathbf{M}_{\boldsymbol{\iota}}\mathbf{P_z x}\beta + \mathbf{M}_{\boldsymbol{\iota}}\mathbf{u},\\ \implies \hat{\beta} &= (\mathbf{x}^\top\mathbf{M}_{\boldsymbol{\iota}}\mathbf{P_z x})^{-1}\mathbf{x}^\top\mathbf{M}_{\boldsymbol{\iota}}\mathbf{P_z y}\\ &= \frac{\mathbf{x}^\top\mathbf{M}_{\boldsymbol{\iota}}\mathbf{P_z y}}{\mathbf{x}^\top\mathbf{M}_{\boldsymbol{\iota}}\mathbf{P_z x}},\end{aligned}$$

and

$$\begin{aligned}\mathbf{M}_{\boldsymbol{\iota}}\mathbf{x} &= \mathbf{M}_{\boldsymbol{\iota}}\mathbf{P_z y}\phi + \mathbf{M}_{\boldsymbol{\iota}}\mathbf{v},\\ \implies \hat{\phi} &= (\mathbf{y}^\top\mathbf{M}_{\boldsymbol{\iota}}\mathbf{P_z y})^{-1}\mathbf{y}^\top\mathbf{M}_{\boldsymbol{\iota}}\mathbf{P_z x}\\ &= \frac{\mathbf{y}^\top\mathbf{M}_{\boldsymbol{\iota}}\mathbf{P_z x}}{\mathbf{y}^\top\mathbf{M}_{\boldsymbol{\iota}}\mathbf{P_z y}}.\end{aligned}$$

The product of $\hat{\beta}$ and $\hat{\phi}$ is now

$$\begin{aligned}\hat{\beta}\hat{\phi} &= \frac{\mathbf{x}^\top\mathbf{M}_{\boldsymbol{\iota}}\mathbf{P_z y}}{\mathbf{x}^\top\mathbf{M}_{\boldsymbol{\iota}}\mathbf{P_z x}}\frac{\mathbf{y}^\top\mathbf{M}_{\boldsymbol{\iota}}\mathbf{P_z x}}{\mathbf{y}^\top\mathbf{M}_{\boldsymbol{\iota}}\mathbf{P_z y}}\\ \Leftrightarrow &\frac{\widehat{\mathrm{Cov}}(z_i y_i)}{\widehat{\mathrm{Cov}}(z_i, x_i)}\frac{\widehat{\mathrm{Cov}}(z_i x_i)}{\widehat{\mathrm{Cov}}(z_i, y_i)} = 1.\end{aligned}$$

Part B

2

The random variables (x_i, u_i, z_i) are generated by the process

$$x_i = w_{1i} + aw_{2i} + bw_{3i}$$
$$u_i = w_{3i} + cw_{4i}$$
$$z_i = w_{2i} + dw_{5i}$$

with

$$w_i \sim N(0, I_5)$$

where $w_i = (w_{1i}, w_{2i}, w_{3i}, w_{4i}, w_{5i})^\top$ and I_5 is a 5×5 identity matrix. The parameters a, b, c, d are scalar constants.
The random variable y_i is generated as

$$y_i = \beta x_i + u_i$$

where the parameter β is a scalar constant.

2.1

First consider the case in which $b = 0$. Derive $\mathbb{E}[x_i]$, $\mathbb{E}[x_i^2]$, and $\mathbb{E}[x_i u_i]$. What is the joint distribution of the random vector $(x_i, u_i)^\top$?

Given what we know,

$$\begin{aligned} \mathbb{E}[x_i] &= \mathbb{E}[w_{1i}] + a\mathbb{E}[w_{2i}] + b\mathbb{E}[w_{3i}] \\ &= 0, \end{aligned}$$

$$\begin{aligned} \mathbb{E}[x_i^2] &= \mathbb{E}[(w_{1i} + aw_{2i} + bw_{3i})(w_{1i} + aw_{2i} + bw_{3i})] \\ &= \mathbb{E}[w_{1i}^2 + 2w_{1i}aw_{2i} + 2w_{1i}bw_{3i} + a^2 w_{2i}^2 + 2aw_{2i}bw_{3i} + b^2 w_{3i}^2] \\ &= 1 + a^2 + \underbrace{b^2}_{=0} \\ \therefore \mathbb{E}[x_i^2] &= 1 + a^2, \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}[x_i u_i] &= \mathbb{E}[(w_{1i} + aw_{2i} + bw_{3i})(w_{3i} + cw_{4i})] \\ &= \mathbb{E}[w_{1i}w_{3i} + w_{1i}cw_{4i} + aw_{2i}w_{3i} + aw_{2i}cw_{4i} + bw_{3i}^2 + bw_{3i}cw_{4i}] \\ &= \underbrace{b}_{=0} \\ \therefore \mathbb{E}[x_i u_i] &= 0, \end{aligned}$$

which implies that x_i and u_i are independent (diagonal variance-covariance matrix). Then we can write the joint distribution as

$$\begin{pmatrix} x_i \\ u_i \end{pmatrix} \overset{d}{\sim} N\left(\begin{pmatrix} \mu_x \\ \mu_u \end{pmatrix}, \begin{pmatrix} \Omega_{xx} & \Omega_{xu} \\ \Omega_{ux} & \Omega_{uu} \end{pmatrix} \right),$$

where $\mu_x = 0$, $\mu_u = 0$, $\Omega_{xx} = 1 + a^2$, $\Omega_{xu} = \Omega_{ux} = 0$, and $\Omega_{uu} = 1 + c^2$. Alternatively, we could write the bivariate distribution as

$$F(x, u) = \Pr\left((X \leq x) \cap (U \leq u)\right),$$

but since x and u are independent, the joint CDF of (x, u) is

$$F(x, u) = F(x, \infty)F(\infty, u),$$

where the first factor on the RHS is the joint probability that $X \leq x$ and $U \leq u$. The function $F(x, \infty)$ is called the marginal CDF of X and is thus just the CDF of X considered by itself since the second inequality imposes no constraint. We can also express independence in terms of marginal density by taking the derivative of the marginal CDF:

$$\begin{aligned} f(x) &\equiv F_x(x, \infty) \\ f(u) &\equiv F_u(\infty, u). \end{aligned}$$

2.2

Again consider the case in which $b = 0$.

2.2.1

What is the joint distribution of the random vector $(x_i, u_i, y_i)^\top$?

The joint distribution can be written as:

$$\begin{pmatrix} x_i \\ u_i \\ y_i \end{pmatrix} \overset{d}{\sim} N\left(\begin{pmatrix} \mu_x \\ \mu_u \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Omega_{xx} & \Omega_{xu} & \Omega_{xy} \\ \Omega_{ux} & \Omega_{uu} & \Omega_{uy} \\ \Omega_{yx} & \Omega_{yu} & \Omega_{yy} \end{pmatrix} \right),$$

where we have $\mu_x = 0$, $\mu_u = 0$, $\Omega_{xx} = 1 + a^2$, $\Omega_{uu} = 1 + c^2$, and $\Omega_{xu} = \Omega_{ux} = 0$ from before. In addition we have

$$\mathbb{E}[y_i] = \mu_y = 0,$$

and

$$\begin{aligned} \operatorname{Var}(y_i) &= \beta^2 \operatorname{Var}(x_i) + \operatorname{Var}(u_i) + 2\operatorname{Cov}(x_i, u_i) \\ \therefore \Omega_{yy} &= \beta^2 + a^2 + 1 + c^2, \\ \mathbb{E}[y_i x_i] &= \mathbb{E}[(\beta x_i + u_i)x_i] \\ &= \beta \mathbb{E}[x_i^2] \\ \therefore \Omega_{xy} = \Omega_{yx} &= \beta(1 + a^2), \\ \mathbb{E}[y_i u_i] &= \mathbb{E}[(\beta x_i + u_i)u_i] \\ &= \mathbb{E}[u_i^2] \\ \therefore \Omega_{yu} = \Omega_{uy} &= 1 + c^2. \end{aligned}$$

2.2.2

What is the marginal distribution of y_i? What is the conditional distribution of $y_i|x_i$?

The unconditional expectation of y_i is:

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

with an unconditional variance of Ω_{yy}. Since x_i and u_i are an additive product of normally distributed random variables, it follows that y_i must also be normally distributed:

$$y_i \overset{d}{\sim} N(0, \Omega_{yy}).$$

The conditional expectation of y_i is:

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

and the conditional variance is:

$$\begin{aligned}
\mathrm{Var}(y_i|x_i) &= \mathrm{Var}(\beta x_i|x_i) + \mathrm{Var}(u_i|x_i) \\
&= 0 + \mathrm{Var}(u_i) \\
&= 1 + c^2,
\end{aligned}$$

since there is no covariance between u_i and x_i. Thus we have

$$y_i|x_i \overset{d}{\sim} N(\beta x_i, 1 + c^2).$$

2.2.3

Comment briefly on the implications of your answer in part 2.2.2 for the efficient estimation of the parameter β from a sample of independent and identically distributed observations on (y_i, x_i) generated by this process.

With independent observations, the model $y_i = \beta x_i + u_i$ is a classical linear regression model with a stochastic regressor (x_i) and a homoskedastic and normally distributed error term (u_i). Using vector notation, independence implies:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u},$$

with

$$\mathbb{E}[\mathbf{y}|\mathbf{X}] = \mathbf{X}\beta, \ \mathrm{Var}(\mathbf{y}|\mathbf{X}) = (1 + c^2)\mathbf{I}_n,$$

where n is the sample size. This model satisfies the assumptions of the Gauss-Markov theorem, so that OLS is the BLUE of β, conditional on the realised data. The OLS estimator $\hat{\beta}_{\mathrm{OLS}}$ is efficient in the sense that:

$$\mathrm{Var}(\hat{\beta}_{\mathrm{OLS}}|\mathbf{X}) \leq \mathrm{Var}(\tilde{\beta}|\mathbf{X}),$$

for any other unbiased estimator $\tilde{\beta}$ constructed as a linear function of the random vector $\mathbf{y}|\mathbf{X}$.

2.3

Now consider the case in which $b \neq 0$.

2.3.1

What is the joint distribution of the random vector $(x_i, u_i, z_i)^\top$?

The joint distribution can be written as:

$$\begin{pmatrix} x_i \\ u_i \\ z_i \end{pmatrix} \overset{d}{\sim} N\left(\begin{pmatrix} \mu_x \\ \mu_u \\ \mu_z \end{pmatrix}, \begin{pmatrix} \Omega_{xx} & \Omega_{xu} & \Omega_{xz} \\ \Omega_{ux} & \Omega_{uu} & \Omega_{uz} \\ \Omega_{zx} & \Omega_{zu} & \Omega_{zz} \end{pmatrix} \right),$$

where we now have,

$$\begin{aligned} \mu_x &= \mathbb{E}[x_i] = 0, \\ \mu_u &= \mathbb{E}[u_i] = 0, \\ \mu_z &= \mathbb{E}[z_i] = 0, \end{aligned}$$

with the following on the diagonal of the variance-covariance matrix:

$$\begin{aligned} \Omega_{xx} &= \mathbb{E}[x_i^2] = 1 + a^2 + b^2, \\ \Omega_{uu} &= \mathbb{E}[u_i^2] = 1 + c^2, \\ \Omega_{zz} &= \mathbb{E}[z_i^2] = 1 + d^2, \end{aligned}$$

and the off-diagonal elements:

$$\begin{aligned} \Omega_{xu} = \Omega_{ux} = \mathbb{E}[x_i u_i] &= \mathbb{E}[(w_{1,i} + a w_{2,i} + b w_{3,i})(w_{3,i} + c w_{4,i})] \\ &= b, \\ \Omega_{xz} = \Omega_{zx} = \mathbb{E}[x_i z_i] &= \mathbb{E}[(w_{1,i} + a w_{2,i} + b w_{3,i})(w_{2,i} + d w_{5,i})] \\ &= a, \\ \Omega_{zu} = \Omega_{uz} = \mathbb{E}[z_i u_i] &= \mathbb{E}[(w_{2,i} + d w_{5,i})(w_{3,i} + c w_{4,i})] \\ &= 0. \end{aligned}$$

2.3.2

What is the correlation between x_i and u_i, $Corr(x_i, u_i)$?

Plugging in the formula for correlation, we have:

$$\begin{aligned} \operatorname{Corr}(x_i, u_i) &= \frac{\operatorname{Cov}(x_i, u_i)}{\sqrt{\operatorname{Var}(x_i)\operatorname{Var}(u_i)}} = \frac{\Omega_{xu}}{\sqrt{\Omega_{xx}\Omega_{uu}}} \\ &= \frac{b}{\sqrt{(1 + a^2 + b^2)(1 + c^2)}}. \end{aligned}$$

2.3.3

What is the conditional distribution of $u_i|z_i$?

Start by finding the conditional expectation:

$$\mathbb{E}[u_i|z_i] = \mathbb{E}[w_{3,i} + cw_{4,i}|z_i] = 0,$$

which is the same as the unconditional expectation. We also have

$$\text{Var}(u_i|z_i) = 1 + c^2,$$

which is also the same as the unconditional variance. So:

$$u_i|z_i \overset{d}{\sim} N(0, 1 + c^2),$$

where the variance of u_i is clearly homoskedastic.

2.3.4

Derive expressions for the scalar parameter π and the conditional variance of $r_i|z_i$ in the linear regression

$$x_i = \pi z_i + r_i$$

We know that

$$\hat{\pi} = \frac{\text{Cov}(x_i, z_i)}{\text{Var}(z_i)} = \frac{a}{1 + d^2}.$$

Then, derive the conditional distribution $x_i|z_i$. Start with the conditional expectation:

$$\mathbb{E}[x_i|z_i] = \mathbb{E}[\pi z_i + r_i|z_i] = \pi z_i,$$

and the conditional variance:

$$\begin{aligned}
\text{Var}(x_i|z_i) &= \text{Var}(\pi z_i + r_i|z_i) \\
&= 0 + \text{Var}(r_i|z_i) \\
&= \text{Var}(r_i),
\end{aligned}$$

by construction. So

$$x_i|z_i \overset{d}{\sim} N\left(\frac{a}{1 + d^2} z_i, \text{Var}(r_i)\right),$$

where

$$\begin{aligned}
\text{Var}(r_i) &= \text{Var}(x_i + \pi z_i) \\
&= \mathbb{E}[x_i^2] + \pi^2 \mathbb{E}[z_i] + 2\text{Cov}(x_i, z_i) \\
&= 1 + a^2 + b^2 + \pi^2(1 + d^2) \\
&= 1 + a^2 + b^2 + \frac{a^2}{1 + d^2}.
\end{aligned}$$

2.3.5

Comment briefly on the implications of your answers to parts 2.3.2 - 2.3.4 for the efficient estimation of the parameter β from a sample of independent and identically distributed observations on (y_i, x_i, z_i) generated by this process.

With independent observations, the model $y_i = \beta x_i + u_i$ is a linear model with a single endogenous explanatory variable (x_i), a single instrumental variable (z_i) that is valid and informative (so long as $a \neq 0$), and an error term that is conditional homoskedastic and hence IID conditional on z_i. Two stage least squares provides an asymptotically efficient estimator of β in this case.

3

The data generation process for (y_i, x_i) for $i = 1, 2, ..., n$ is of the form

$$\begin{aligned} y_i &= \beta x_i + u_i, \\ x_i &= v_i. \end{aligned}$$

3.1

First suppose it is known that (u_i, v_i) is independent and identically distributed over $i = 1, 2, ..., n$ with

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{pmatrix} \right]$$

and $\sigma_u^2 > 0$ and $\sigma_v^2 > 0$. Define the $n \times 1$ vectors $y = (y_1, y_2, ..., y_n)^\top$ and $X = (x_1, x_2, ..., x_n)^\top$.

3.1.1

Derive $\mathbb{E}[\mathbf{y}|\mathbf{X}]$ and $Var(\mathbf{y}|\mathbf{X})$.

Begin with

$$\mathbb{E}[\mathbf{y}|\mathbf{X}] = \mathbf{X}\beta + \mathbb{E}[\mathbf{u}|\mathbf{X}],$$

and by the law of iterated expectations

$$\begin{aligned} \mathbb{E}_X[\mathbb{E}_{u|X}[\mathbf{u}|\mathbf{X}]] &= \mathbb{E}[\mathbf{u}] = 0, \\ \therefore \mathbb{E}[\mathbf{y}|\mathbf{X}] &= \mathbf{X}\beta. \end{aligned}$$

For the conditional variance, begin by noting that $\text{Var}(\mathbf{y}|\mathbf{X}) = \text{Var}((\mathbf{X}\beta + \mathbf{u})|\mathbf{X})$, and so

$$\begin{aligned} \text{Var}((\mathbf{X}\boldsymbol{\beta} + \mathbf{u})|\mathbf{X}) &= \beta^2 \text{Var}(\mathbf{X}|\mathbf{X}) + \text{Var}(\mathbf{u}|\mathbf{X}) + 2\beta \text{Cov}(\mathbf{X}, \mathbf{u}) \\ &= \text{Var}(\mathbf{u}|\mathbf{X}) \\ \therefore \text{Var}(\mathbf{y}|\mathbf{X}) &= \sigma_u^2, \end{aligned}$$

where we assume by the law of total variance

$$\begin{aligned} \text{Var}(\mathbf{u}) &= \text{Var}(\mathbb{E}[\mathbf{u}|\mathbf{X}]) + \mathbb{E}[\text{Var}(\mathbf{u}|\mathbf{X})] \\ \text{Var}(\mathbf{u}) &= \text{Var}(\mathbf{u}|\mathbf{X}) = \sigma_u^2. \end{aligned}$$

3.1.2

What is the distribution of y conditional on X?

Given the information above

$$\mathbf{y}|\mathbf{X} \overset{d}{\sim} N(\mathbf{X}\beta, \sigma_u^2 \mathbf{I})$$

3.1.3

Explain how data on (y_i, x_i) for $i = 1, 2, ..., n$ could be used to construct an exact 95% confidence interval for the parameter β.

Since y is normally distributed (or rather, u is normally distributed) and we have a finite sample, we can use an exact distribution for our test statistic to test the null hypothesis $H_0 : \beta = \beta_0$. We also assume that we use a two-tailed test, and so our confidence interval is symmetrical. Our test statistic is

$$t_{\beta_0} = \frac{\hat{\beta} - \beta_0}{SE(\hat{\beta})} \overset{d}{\sim} t(n-k),$$

and thus our 95% confidence interval is given by

$$\begin{aligned}
\Pr\left(t_{\frac{\alpha}{2}} \leq \frac{\hat{\beta} - \beta_0}{SE(\hat{\beta})} \leq t_{1-\frac{\alpha}{2}}\right) &= 0.95 \\
\Pr\left(t_{\frac{\alpha}{2}} SE(\hat{\beta}) \leq \hat{\beta} - \beta_0 \leq t_{1-\frac{\alpha}{2}} SE(\hat{\beta})\right) &= 0.95 \\
\Pr\left(\hat{\beta} - t_{1-\frac{\alpha}{2}} SE(\hat{\beta}) \leq \beta_0 \leq \hat{\beta} - t_{\frac{\alpha}{2}} SE(\hat{\beta})\right) &= 0.95,
\end{aligned}$$

and so

$$\text{CI} = \left(\hat{\beta} - t_{1-\frac{\alpha}{2}} SE(\hat{\beta}), \hat{\beta} - t_{\frac{\alpha}{2}} SE(\hat{\beta})\right).$$

3.2

Now suppose it is known that (u_i, v_i) is independently and identically distributed over $i = 1, 2, ..., n$ with $\mathbb{E}[u_i] = \mathbb{E}[v_i] = 0$, $Var(u_i) = \sigma_u^2 > 0$, $Var(v_i) = \sigma_v^2 > 0$, and that u_i and v_i are independent of each other.

3.2.1

Derive $\mathbb{E}[x_i u_i]$ and $\mathbb{E}[u_i^2 | x_i]$.

The key point here is that u_i and v_i are no longer normally distributed. They're merely IID. Begin with the covariance of x_i and u_i:

$$\mathbb{E}[x_i u_i] = \mathbb{E}[x_i]\mathbb{E}[u_i] = 0,$$

and by the law of total variance we get the conditional variance of u_i:

$$\begin{aligned}
\mathbb{E}[u_i^2] = \sigma_u^2 &= \mathbb{E}[\text{Var}(u_i|x_i)] + \text{Var}(\mathbb{E}[u_i|x_i]) \\
\implies \sigma_u^2 &= \text{Var}(u_i|x_i).
\end{aligned}$$

3.2.2

Explain how data on (y_i, x_i) for $i = 1, 2, ..., n$ could be used to construct an approximate 95% confidence interval for the parameter β, assuming that the sample size n is large.

Now that the conditional distribution of y (or rather, u) is no longer NID, and given that we have a sufficiently large n, we will have to rely on asymptotic theory to test our null hypothesis and construct an approximate confidence interval. Our test statistic to test the null hypothesis $H_0 : \beta = \beta_0$ is

$$t_{\beta_0} = \frac{\hat{\beta} - \beta_0}{SE(\hat{\beta})} \overset{a}{\sim} N(0, 1),$$

and our 95% confidence interval is given by

$$\begin{aligned}
\Pr\left(z_{\frac{\alpha}{2}} \leq \frac{\hat{\beta} - \beta_0}{SE(\hat{\beta})} \leq z_{1-\frac{\alpha}{2}} \right) &\approx 0.95 \\
\Pr\left(z_{\frac{\alpha}{2}} SE(\hat{\beta}) \leq \hat{\beta} - \beta_0 \leq z_{1-\frac{\alpha}{2}} SE(\hat{\beta}) \right) &\approx 0.95 \\
\Pr\left(\hat{\beta} - z_{1-\frac{\alpha}{2}} SE(\hat{\beta}) \leq \beta_0 \leq \hat{\beta} - z_{\frac{\alpha}{2}} SE(\hat{\beta}) \right) &\approx 0.95,
\end{aligned}$$

and so

$$\text{CI} \approx \left(\hat{\beta} - z_{1-\frac{\alpha}{2}} SE(\hat{\beta}), \hat{\beta} - z_{\frac{\alpha}{2}} SE(\hat{\beta}) \right).$$

3.3

Now suppose it is known that

$$\begin{aligned}
u_i &= \gamma w_i + e_i, \\
v_i &= \theta_1 z_{1i} + \theta_2 z_{2i} + \phi w_i + \epsilon_i.
\end{aligned}$$

The vector $(w_i, e_i, z_{1i}, z_{2i}, \epsilon_i)$ is independently and identically distributed over $i = 1, 2, ..., n$; each element has expectation zero and a strictly positive variance, and the individual elements are mutually uncorrelated (unless stated otherwise below).

3.3.1

Derive $\mathbb{E}[x_i u_i]$.

Start with

$$\begin{aligned}
\mathbb{E}[x_i u_i] &= \mathbb{E}[v_i(\gamma w_i + e_i)] \\
&= \mathbb{E}[(\theta_1 z_{1i} + \theta_2 z_{2i} + \phi w_i + \epsilon_i)(\gamma w_i + e_i)] \\
&= \mathbb{E}[\phi \gamma w_i^2] \\
\therefore \mathbb{E}[x_i u_i] &= \phi \gamma \text{Var}(w_i),
\end{aligned}$$

and so x_i and u_i are now correlated if $\phi \neq 0$ and $\gamma \neq 0$.

3.3.2

Explain how data on $(y_i, x_i, z_{1i}, z_{2i})$ for $i = 1, 2, ..., n$ could be used to estimate the parameter β consistently in the case where both $\gamma \neq 0$ and $\phi \neq 0$.

Two stage least squares (2SLS) could be used used to estimate β, using z_1 and z_2 as instruments to estimate x if we assume that $(z_1, z_2) \perp u$. The first stage regression could be

$$\mathbf{X} = \mathbf{Z}\boldsymbol{\Theta} + \mathbf{r},$$

where $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2]$, $\boldsymbol{\Theta} = [\theta_1, \theta_2]^\top$, and $\mathbf{r}$ is an error vector which is orthogonal to $\mathbf{X}$ by construction. The OLS estimator for $\boldsymbol{\Theta}$ is

$$\hat{\boldsymbol{\Theta}} = (\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top \mathbf{X},$$

and this parameter estimate would be used to attain $\hat{X}$, which would then be used for the second stage regression

$$\mathbf{y} = \hat{\mathbf{X}}\beta + \mathbf{u},$$

where we attain the 2SLS estimate of β,

$$\begin{aligned} \hat{\beta}_{2SLS} &= (\hat{\mathbf{X}}^\top \hat{\mathbf{X}})^{-1} \hat{\mathbf{X}}^\top \mathbf{y} \\ \Leftrightarrow \hat{\beta}_{2SLS} &= (\mathbf{X}^\top \mathbf{P_Z} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{P_Z} \mathbf{y}, \end{aligned}$$

where $\mathbf{P_Z} = \mathbf{Z}(\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top$.

3.3.3

Discuss estimation problems that could arise, and how the presence of these problems could be investigated, in settings where: i) z_{1i} is correlated with e_i; ii) the parameters θ_1 and θ_2 are both close to zero.

If z_1 is correlate with e then z_1 would fail to be valid instrument, as it too would be correlated with the error term u, leading to inconsistent estimates of β. One test for instrument validity would be the Sargan test, which essentially tests the following:

$$\begin{aligned} H_0 &: \mathbb{E}[Z^\top u] = 0, \\ H_1 &: \mathbb{E}[Z^\top u] \neq 0, \end{aligned}$$

and is based on the following test statistic

$$J = \frac{(\mathbf{y} - \mathbf{X}\hat{\beta}_{2SLS})^\top \mathbf{P_Z} (\mathbf{y} - \mathbf{X}\hat{\beta}_{2SLS})}{\hat{\sigma}^2} \overset{a}{\sim} \chi^2(l-k),$$

where we would reject the null hypothesis for significantly high values. Note, however, that even if we were to reject the null hypothesis (and thus concluding that our instruments are invalid), we would not know whether or not z_1 or z_2 are valid. If we were absolutely certain that z_2 is a valid instrument, and the Sargan test leads us to reject the null hypothesis, then we would drop z_1 as an instrument. Note that by dropping z_1 we would be moving from an over-identifying case to a just-identified case, so there would be no concerns over dimensionality or indeterminacy arising from failing to meet the moment conditions.

If the parameters θ_1 and θ_2 are close to zero, then we would call the z_1 and z_2 weak or ineffective instruments. That is, they have little explanatory power with respect to x. We could run a test of weak instruments by stating

$$\begin{aligned} H_0 &: \Theta = 0 = \Theta_0, \\ H_1 &: \Theta \neq 0, \end{aligned}$$

where we would test the null hypothesis using a Wald test:

$$W = (\hat{\Theta} - \Theta_0)^\top \left(\hat{\text{Var}}(\hat{\Theta}) \right)^{-1} (\hat{\Theta} - \Theta_0) \overset{a}{\sim} \frac{\chi^2(p)}{p},$$

where $p = 2$. As a general rule of thumb, Staiger and Stock and Stock and Yogo suggest that the value of the Wald test statistic should be at least 10 here for the standard asymptotic approximation of the distribution of the 2SLS estimator to be considered reliable.