PhD Econometrics 1: Revision

Imperial College London

Hormoz Ramian

Question 1: Consider the model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$, where $\boldsymbol{X}$ is of rank k, and $\boldsymbol{u}$ has mean $\boldsymbol{0}_n$ and a positive definite variance $\boldsymbol{\Sigma}$ that is known. Denote the GLS estimator $\widehat{\boldsymbol{\beta}} := (\boldsymbol{X}'\Sigma^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$ with $\boldsymbol{\Omega} := \text{var}(\widehat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\Sigma^{-1}\boldsymbol{X})^{-1}$. Suppose we know that $\boldsymbol{R}'\boldsymbol{\beta} = \boldsymbol{c}$ where $\boldsymbol{R}$ is a nonrandom $k \times r$ matrix of rank $r < k$ and $\boldsymbol{c}$ is a nonrandom vector.

(1.1) Obtain the constrained least squares (CLS) estimator $\widetilde{\beta}$ by minimizing $\frac{1}{2}\boldsymbol{u}'\boldsymbol{\Sigma}^{-1}\boldsymbol{u}$ subject to $\boldsymbol{R}'\boldsymbol{\beta} = \boldsymbol{c}$.

(1.2) Assuming $\boldsymbol{X}$ is nonrandom, derive the mean and variance of $\widetilde{\boldsymbol{\beta}}$ in this model.

(1.3) Show that $\text{var}(\widehat{\beta}) - \text{var}(\widetilde{\beta})$ is positive semidefinite.

(1.4) Show that:

$$\widetilde{\boldsymbol{u}}'\boldsymbol{\Sigma}^{-1}\widetilde{\boldsymbol{u}} - \widehat{\boldsymbol{u}}'\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{u}} \;=\; (\widetilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}})'\boldsymbol{\Omega}^{-1}(\widetilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}) \;=\; (\boldsymbol{R}'\widehat{\boldsymbol{\beta}} - \boldsymbol{c})'(\boldsymbol{R}'\boldsymbol{\Omega}\boldsymbol{R})^{-1}(\boldsymbol{R}'\widehat{\boldsymbol{\beta}} - \boldsymbol{c})'$$

where $\widetilde{\boldsymbol{u}}$ and $\widehat{\boldsymbol{u}}$ are the CLS and GLS residuals, respectively. How do you reconcile the first equality with part (1.3) when $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_n$?

(1.5) Suppose that $\boldsymbol{u}$ is spherically distributed (hence $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_n$), independently of $\boldsymbol{X}$. Prove that:

$$\frac{n-k}{r} \times \frac{\widetilde{\boldsymbol{u}}'\widetilde{\boldsymbol{u}} - \widehat{\boldsymbol{u}}'\widehat{\boldsymbol{u}}}{\widehat{\boldsymbol{u}}'\widehat{\boldsymbol{u}}} \sim \mathrm{F}(r, n-k).$$

Question 2: Consider estimating the elasticity of a consumption commodity demand denoted by $q_i^d = \alpha_0 + \alpha_1 p_i + u_i$ where $p_i = \ln P_i$, P_i is the actual price, and $q_i^d = \ln Q_i^d$ with Q_i^d denoting the actual quantity demanded. The term u_i represents other factors besides price that affect demand, such as income and consumer taste. The supply equation is in the same form, and is given by $q_i^s = \beta_0 + \beta_1 p_i + v_i$ where the term v_i represents the factors that affect supply, such as weather conditions production factor prices, and union status. Assuming first, that the two error terms are uncorrelated, and second that the equilibrium condition $q_i^d = q_i^s = q_i$ holds, then:

(2.1) Derive the the reduce form system in terms of p_i and q_i and provide the terms inside the 2×2 symmetric variance-covariance matrix $\text{cov}(p_i, q_i)$.

(2.2) Show that the ordinary least squares estimator resulting from the regression of q_i on p_i is biased for both structural parameters α_1 and β_1.

(2.3) Now suppose that variable z_i (e.g. weather condition) is available which is uncorrelated with u_i

such that $\text{cov}(z_i, u_i) = 0 \ \forall_i$ but $\text{cov}(z_i, p_i) \neq 0$. Show that using z_i, an instrumental variable approach is able to estimate α_1.

Question 3: Consider the normal linear model with $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$, and that $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$ where n is the sample size, $\boldsymbol{y}$ and $\boldsymbol{\epsilon}$ are each an $n \times 1$ random vectors, $\boldsymbol{X}$ is an $n \times k$ matrix of k nonrandom regressors, $n > k$ and that $\sigma^2 > 0$. Suppose that the rank of $\boldsymbol{X}$is k, so that $\boldsymbol{X}'\boldsymbol{X}$ is positive definite.

(3.1) Derive the log-likelihood ℓ function, score vector, and Hessian matrix for the $m = k+1$ parameters $\boldsymbol{\theta}' := (\boldsymbol{\beta}', \sigma^2)$ of this model.

(3.2) Derive the maximum likelihood estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}^2$ of the parameters, and calculate $\ell(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2)$.

Suggested Answers

Question 1

(1.1) Note that $r < k$ and then let's setup the Lagrangian problem[1]:

$$\begin{aligned} S(\boldsymbol{\beta}, \boldsymbol{\lambda}) &:= \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}) + \boldsymbol{\lambda}'(\boldsymbol{c} - \boldsymbol{R}'\boldsymbol{\beta}) \\ &= \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}) + (\boldsymbol{c} - \boldsymbol{R}'\boldsymbol{\beta})'\boldsymbol{\lambda} \end{aligned}$$

which contains quadratic and linear forms in $\boldsymbol{\beta}$, and $\boldsymbol{\lambda}$ whic is an r-dimensional vector of Lagrange multipliers. Differentiating,

$$\frac{\partial S(\boldsymbol{\beta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta}} = -\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}) - \boldsymbol{R\lambda}$$

and the first-order conditions are

$$\begin{aligned} \boldsymbol{0}_k &= -\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}) - \boldsymbol{R}\widetilde{\boldsymbol{\lambda}} \\ \boldsymbol{0}_r &= \boldsymbol{R}'\widetilde{\boldsymbol{\beta}} - \boldsymbol{c} \end{aligned}$$

re-write the $(k + r)$-system above as:

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X} & \boldsymbol{R} \\ \boldsymbol{R}' & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \widetilde{\boldsymbol{\beta}} \\ \widetilde{\boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{\Sigma}^{-1} \\ \boldsymbol{c} \end{bmatrix}$$

where the first matrix is symmetric thus we can denote $\boldsymbol{\Omega} := (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}$ for the variance of the GLS estimator $\widehat{\boldsymbol{\beta}} = \boldsymbol{\Omega X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$. Using the formula for a partitioned inverse:

$$\begin{bmatrix} \boldsymbol{\Omega}^{-1} & \boldsymbol{R} \\ \boldsymbol{R}' & \boldsymbol{0} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{\Omega}^{1/2}(\boldsymbol{I}_k - \boldsymbol{P})\boldsymbol{\Omega}^{1/2} & \boldsymbol{\Omega R}(\boldsymbol{R}'\boldsymbol{\Omega R})^{-1} \\ (\boldsymbol{\Omega R}(\boldsymbol{R}'\boldsymbol{\Omega R})^{-1})' & -(\boldsymbol{R}'\boldsymbol{\Omega R})^{-1} \end{bmatrix} \tag{1}$$

[1] Note that if we had $r = k$ then $\boldsymbol{\beta} = \boldsymbol{R}'^{-1}\boldsymbol{c}$ and there would be no need to go any further.

note that the $r \times r$ matrix $\boldsymbol{R}'\boldsymbol{\Omega}\boldsymbol{R}$ is full rank and we have the symmetric idempotent form:

$$\boldsymbol{P} := \boldsymbol{\Omega}^{1/2}\boldsymbol{R}(\boldsymbol{R}'\boldsymbol{\Omega}\boldsymbol{R})^{-1}\boldsymbol{R}'\boldsymbol{\Omega}^{1/2}$$

satisfying $\operatorname{tr}(\boldsymbol{P}) = r$, then write,

$$\begin{bmatrix} \widetilde{\boldsymbol{\beta}} \\ \widetilde{\boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\Omega}\boldsymbol{R}(\boldsymbol{R}'\boldsymbol{\Omega}\boldsymbol{R})^{-1}(\boldsymbol{R}'\widehat{\beta} - \boldsymbol{c}) \\ (\boldsymbol{R}'\boldsymbol{\Omega}\boldsymbol{R})^{-1}(\boldsymbol{R}'\widehat{\beta} - \boldsymbol{c}) \end{bmatrix}$$

this the minimizer to the Lagrangian problem since the objective function is convex and the constraint is linear in parameters.

(1.2) Since $\mathbb{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ we have,

$$\begin{aligned} \mathbb{E}(\widehat{\boldsymbol{\beta}}) &= \boldsymbol{\beta} - \boldsymbol{\Omega}\boldsymbol{R}(\boldsymbol{R}'\boldsymbol{\Omega}\boldsymbol{R})^{-1}(\boldsymbol{R}'\widehat{\beta} - \boldsymbol{c}) \\ &= \boldsymbol{\beta} \end{aligned}$$

by $\boldsymbol{R}'\boldsymbol{\beta} = \boldsymbol{c}$, and for the variance, we can collect the terms in $\widehat{\boldsymbol{\beta}}$ and write:

$$\widetilde{\beta} = \boldsymbol{\Omega}^{1/2}(\boldsymbol{I}_k - \boldsymbol{P})\boldsymbol{\Omega}^{-1/2}\widehat{\beta} + \boldsymbol{b}$$

where $\boldsymbol{b} := \boldsymbol{\Omega}\boldsymbol{R}(\boldsymbol{R}'\boldsymbol{\Omega}\boldsymbol{R})^{-1}\boldsymbol{c}$, is a nonrandom term. Then useing $\operatorname{var}(\widehat{\beta}) = \boldsymbol{\Omega}$ we obtain:

$$\begin{aligned} \operatorname{var}(\widetilde{\boldsymbol{\beta}}) &= \boldsymbol{\Omega}^{1/2}(\boldsymbol{I}_k - \boldsymbol{P})\boldsymbol{\Omega}^{-1/2}\operatorname{var}(\widehat{\boldsymbol{\beta}})\boldsymbol{\Omega}^{-1/2}(\boldsymbol{I}_k - \boldsymbol{P})\boldsymbol{\Omega}^{1/2} \\ &= \boldsymbol{\Omega}^{1/2}(\boldsymbol{I}_k - \boldsymbol{P})(\boldsymbol{I}_k - \boldsymbol{P})\boldsymbol{\Omega}^{1/2} \\ &= \boldsymbol{\Omega}^{1/2}(\boldsymbol{I}_k - \boldsymbol{P})\boldsymbol{\Omega}^{1/2} \\ &= \boldsymbol{\Omega} - \boldsymbol{\Omega}\boldsymbol{R}(\boldsymbol{R}'\boldsymbol{\Omega}\boldsymbol{R})^{-1}\boldsymbol{R}'\boldsymbol{\Omega} \end{aligned}$$

since $\boldsymbol{I}_k - \boldsymbol{P}$ is idempotent. Note that this the term that we derived in equation (1) in the block of the partitioned inverse (first row, first column).

(1.3) For any vector $\boldsymbol{a}$, write

$$\boldsymbol{a}'(\operatorname{var}(\widehat{\boldsymbol{\beta}}) - \operatorname{var}(\widetilde{\boldsymbol{\beta}}))\boldsymbol{a} = \boldsymbol{a}'\boldsymbol{\Omega}^{1/2}\boldsymbol{P}\boldsymbol{\Omega}^{1/2}\boldsymbol{a} = \boldsymbol{\alpha}'\boldsymbol{\alpha} \tag{2}$$

where $\boldsymbol{\alpha} := \boldsymbol{P}\boldsymbol{\Omega}^{1/2}\boldsymbol{a}$ and we can use $\boldsymbol{P} = \boldsymbol{P}^2$ (idempotent $\boldsymbol{P}$). Therefore, this quadratic form is nonnegative and the difference of variances is positive semidefinite[2].

(1.4) Express $\widetilde{\boldsymbol{u}}$ in terms of $\widehat{\boldsymbol{u}}$ to arrange:

$$\begin{aligned} \widetilde{\boldsymbol{u}} &= \boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}} \\ &= \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} + (\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}) \\ &= \widehat{\boldsymbol{u}} - \boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}) \end{aligned}$$

[2] Incorporating the restriction $\boldsymbol{R}'\boldsymbol{\beta} = \boldsymbol{c}$ into the estimation of $\boldsymbol{\beta}$ has improved the efficiency (precision) of out estimator. As a special case, we get the implication that including irrelevant variables (those whose true parameter value is zero) in the regression reduces the precision of the estimates.

then

$$\begin{aligned}
\widetilde{\boldsymbol{u}}'\boldsymbol{\Sigma}^{-1}\widetilde{\boldsymbol{u}} &= \left(\widehat{\boldsymbol{u}} - \boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})\right)' \boldsymbol{\Sigma}^{-1} \left(\widehat{\boldsymbol{u}} - \boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})\right) \\
&= \widehat{\boldsymbol{u}}'\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{u}} + 2(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{u}} + (\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}) \\
&= \widehat{\boldsymbol{u}}'\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{u}} + (\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})
\end{aligned}$$

because $\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{u}} = \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) = \boldsymbol{0}_k$ from the definition of $\widehat{\boldsymbol{\beta}}$ in the first part. Then,

$$\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}} = -\boldsymbol{\Omega}\boldsymbol{R}(\boldsymbol{R}'\boldsymbol{\Omega}\boldsymbol{R})^{-1}(\boldsymbol{R}'\widehat{\boldsymbol{\beta}} - \boldsymbol{c}) \tag{3}$$

therefore,

$$\begin{aligned}
\widetilde{\boldsymbol{u}}'\boldsymbol{\Sigma}^{-1}\widetilde{\boldsymbol{u}} - \widehat{\boldsymbol{u}}'\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{u}} &= \underbrace{(\boldsymbol{R}'\widehat{\boldsymbol{\beta}} - \boldsymbol{c})'(\boldsymbol{R}'\boldsymbol{\Omega}\boldsymbol{R})^{-1}\boldsymbol{R}'\boldsymbol{\Omega}}_{\text{pre-multiplied}}\,\boldsymbol{\Omega}^{-1}\,\underbrace{\boldsymbol{\Omega}\boldsymbol{R}(\boldsymbol{R}'\boldsymbol{\Omega}\boldsymbol{R})^{-1}(\boldsymbol{R}'\widehat{\boldsymbol{\beta}} - \boldsymbol{c})}_{\text{post-multiplied}} \\
&= (\boldsymbol{R}'\widehat{\boldsymbol{\beta}} - \boldsymbol{c})'(\boldsymbol{R}'\boldsymbol{\Omega}\boldsymbol{R})^{-1}(\boldsymbol{R}'\widehat{\boldsymbol{\beta}} - \boldsymbol{c})
\end{aligned}$$

To reconcile with part (1.3) when $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_n$ then,

$$\widetilde{\boldsymbol{u}}'\widetilde{\boldsymbol{u}} - \widehat{\boldsymbol{u}}'\widehat{\boldsymbol{u}} = (\boldsymbol{R}'\widehat{\boldsymbol{\beta}} - \boldsymbol{c})'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{R}'\widehat{\boldsymbol{\beta}} - \boldsymbol{c}) \geq 0$$

Note that when estimating the model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}$, imposing the constraint on $\widetilde{\beta}$ reduces the variability of the estimator at the cost of a larger sum of squared residuals $\widetilde{\boldsymbol{u}}'\widetilde{\boldsymbol{u}}$. Given $\boldsymbol{X}$, the difference in the sums of squared residuals is small whenever $\widetilde{\boldsymbol{\beta}}$ is close to the unconstrained $\widehat{\boldsymbol{\beta}}$, that is when $\boldsymbol{R}'\widehat{\boldsymbol{\beta}}$ is close to $\boldsymbol{c}$.

(1.5) Let's begin by mentioning the generality of this result first, which holds even when $\boldsymbol{u}$ is not normally distributed. Define the idempotent matrix:

$$\boldsymbol{C} := \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' \tag{4}$$

where $tr(\boldsymbol{C}) = n - k$ and that

$$\begin{aligned}
\widetilde{\boldsymbol{u}} &= \boldsymbol{y} - \widehat{\boldsymbol{y}} \\
&= \widehat{\boldsymbol{u}} - \boldsymbol{X}(\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}) \\
&= \widehat{\boldsymbol{u}} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u} \\
&= \boldsymbol{C}\boldsymbol{u} \qquad (5)
\end{aligned}$$

and then that $\boldsymbol{\Omega} = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$ and that:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}\left[\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}\right]^{-1}(\boldsymbol{R}'\widehat{\boldsymbol{\beta}} - \boldsymbol{c}) \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}\left[\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}\right]^{-1}(\boldsymbol{R}'\boldsymbol{\beta} + \boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u} - \boldsymbol{c}) \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}\left[\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}\right]^{-1}(\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u})
\end{aligned}$$

since $\boldsymbol{R}'\boldsymbol{\beta} = \boldsymbol{c}$. Now we can re-arrange the expression above as:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}\left[\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}\right]^{-1}\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1/2}\underbrace{(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{R}\left[\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}\right]^{-1}\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{X})^{-1/2}}_{\text{denote } \boldsymbol{P} \text{ as } \boldsymbol{P}=\boldsymbol{P}^2}(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{X}'\boldsymbol{u}
\end{aligned}$$

and then

$$\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}} = \left\{(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{P}(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\right\}\boldsymbol{X}'\boldsymbol{u} \tag{6}$$

Consider the ratio (substitude in the denominator from equation (5) and in the numerator from equation (6)):

$$\begin{aligned}
\frac{\widetilde{\boldsymbol{u}}'\widetilde{\boldsymbol{u}} - \widehat{\boldsymbol{u}}'\widehat{\boldsymbol{u}}}{\widehat{\boldsymbol{u}}'\widehat{\boldsymbol{u}}} &= \frac{(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})}{\widehat{\boldsymbol{u}}'\widehat{\boldsymbol{u}}} \\
&= \frac{\boldsymbol{u}'\boldsymbol{X}\left\{(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{P}(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\right\}\boldsymbol{X}'\boldsymbol{X}\left\{(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{P}(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\right\}\boldsymbol{X}'\boldsymbol{u}}{\boldsymbol{u}'\boldsymbol{C}\boldsymbol{u}} \\
&= \frac{\boldsymbol{u}'\boldsymbol{B}\boldsymbol{u}}{\boldsymbol{u}'\boldsymbol{C}\boldsymbol{u}}
\end{aligned}$$

where $\boldsymbol{B} := \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{P}(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{X}'$ which is idempotent since:

$$\boldsymbol{B} = \boldsymbol{B}^2 = \underbrace{\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{P}(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{X}'}\cdot\underbrace{\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{P}(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{X}'}$$

and that $tr(\boldsymbol{X}) = tr(\boldsymbol{P}) = r$. Because $\boldsymbol{X}'\boldsymbol{C} = \boldsymbol{0}$ and $\sigma^2 > 0$ and that $\boldsymbol{X}$ does not affect the conditional distribution, the results hold unconditionally too.

Question 2

(2.1) A regressor is endogenous if it is not predetermined (i.e., not orthogonal to the error term), that is, if it does not satisfy the orthogonality condition. When the equation includes the intercept, the orthogonality condition is violated and hence the regressor is endogenous, if and only if the regressor is correlated with the error term. In the present example, the regressor p_i is necessarily endogenous in both equations. To see why, treat the system of two simultaneous equations and solve for (p_i, q_i) as,

$$p_i = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{v_i - u_i}{\alpha_1 - \beta_1} \tag{7}$$

$$q_i = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} + \frac{\alpha_1 v_i - \beta_1 u_i}{\alpha_1 - \beta_1} \tag{8}$$

such that the price is a function of the two error terms. We can calculate the covariance of the regressor p_i with the demand shifter u_i and the supply shifter v_i:

$$\operatorname{cov}(p_i, u_i) = \operatorname{cov}\left[\frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{v_i - u_i}{\alpha_1 - \beta_1}, u_i\right] = -\frac{1}{\alpha_1 - \beta_1}\operatorname{var}(u_i) \quad (9)$$

$$\operatorname{cov}(p_i, v_i) = \operatorname{cov}\left[\frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{v_i - u_i}{\alpha_1 - \beta_1}, v_i\right] = \frac{1}{\alpha_1 - \beta_1}\operatorname{var}(v_i) \quad (10)$$

which are not zero (unless $\operatorname{var}(ui) = 0$ and $\operatorname{var}(vi) = 0$). Therefore, price is correlated positively with the demand shifter and negatively with the supply shifter, if the demand curve is downward-sloping ($\alpha_1 < 0$) and the supply curve upwardsloping ($\beta_1 > 0$). In this example, endogeneity is a result of market equilibrium.

(2.2) When quantity is regressed on a constant and price, it estimates neither the demand curve or the supply curve, because price is endogenous in both the demand and supply equations. The OLS estimator is consistent for the least squares projection coefficients but in this setting projection of q_i on a constant and p_i, the coefficient of p_i is $\operatorname{cov}(p_i, q_i)/\operatorname{var}(p_i)$:

$$\begin{aligned}\operatorname{cov}(p_i, q_i) &= \alpha_1 + \frac{\operatorname{cov}(p_i, u_i)}{\operatorname{var}(p_i)} = \beta_1 + \frac{\operatorname{cov}(p_i, v_i)}{\operatorname{var}(p_i)} \\ &= \frac{\alpha_1 \operatorname{var}(v_i) + \beta_1 \operatorname{var}(u_i)}{\operatorname{var}(v_i)\operatorname{var}(u_i)} \\ &\in (\alpha_1, \beta_1)\end{aligned}$$

therefore the OLS estimator for the slope measures a weighted average of α_1 and β_1. Such a bias is called the simultaneous equations bias. The OLS cannot consistently estimate the true values because both curves are shifted by other factors besides price, and we cannot tell from data whether the change in price and quantity is due to a demand shift or a supply shift. If, however, $u_i = 0$ (that is, the demand curve stays still), then the equilibrium prices and quantities will trace out the demand curve and the OLS is consistent to obtain α_1.

(2.3) From the results in previous parts, p_i has one part correlated with u_i which is $-\frac{u_i}{\alpha_1-\beta_1}$ and one part uncorrelated with u_i which is $\frac{v_i}{\alpha_1-\beta_1}$. If we can isolate the second part, then we can focus on those variations in p_i that are uncorrelated with u_i and disregard the variations in p_i that biases the OLS estimator. Take a supply shifter z_i (e.g., weather), which can be considered to be uncorrelated with the demand shifter u_i then:

$$\operatorname{cov}(z_i, q_i) = \alpha_1 . \operatorname{cov}(z_i, p_i)$$

which yields $\widehat{\alpha} = \widehat{\operatorname{cov}}(z_i, q_i)/\widehat{\operatorname{cov}}(z_i, p_i)$ which is the IV estimator. Another method to estimate α_1 as suggested above is to run regression:

$$q_i = \alpha_0 + \alpha_1 \widehat{p}_i + \widetilde{u}_i$$

where $\widehat{p}_i$ is the predicted value from the following regression:

$$p_i = \gamma_0 + \gamma_1 z_i + \eta_i$$

and that $\widetilde{u}_i = \alpha_1(p_i - \widehat{p}_i) + u_i$, and that $\mathrm{cov}(\widehat{p}_i, \widehat{u}_i) = 0$ thus the estimation is consistent (two-stage least squares).

Question 3

(3.1) The likelihood function is:

$$\mathcal{L}(\boldsymbol{\beta}, \sigma^2) \;=\; \prod_{i=1}^{n} \left[\frac{\exp\left\{-\frac{1}{2\sigma^2}\epsilon_i^2\right\}}{\sigma\sqrt{2\pi}} \right] \;=\; \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{\epsilon}'\boldsymbol{\epsilon}\right\}$$

which is an expression in terms of unobservable $\boldsymbol{\epsilon}$ but it can be re-arranged in terms of the data $\boldsymbol{y}$ and $\boldsymbol{X}$ and the parameters to-be-estimated, using the linear model $\boldsymbol{\epsilon} = \boldsymbol{y} - \boldsymbol{X\beta}$, then log-likelihood is:

$$\begin{aligned}
\ell(\boldsymbol{\beta}, \sigma^2) &= \log \prod_{i=1}^{n} \left[\frac{\exp\left\{-\frac{1}{2\sigma^2}\epsilon_i^2\right\}}{\sigma\sqrt{2\pi}} \right] \;=\; \log \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{\epsilon}'\boldsymbol{\epsilon}\right\} \\
&= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta})
\end{aligned}$$

Differentiating this quadratic function of $\boldsymbol{\beta}$ gives:

$$\frac{\partial \ell(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} \;=\; \underbrace{\frac{1}{\sigma^2}\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X\beta})}_{k\text{-equations}}$$

the score is obtained by stacking k-vector above atop 1×1 derivative w.r.t. σ^2:

$$\frac{\partial \ell(\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} \;=\; \underbrace{-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta})}_{\text{1-equation}}$$

Then the Hessian matrix is a $(k+1) \times (k+1)$ partitioned array below:

$$H \;=\; \begin{bmatrix} \frac{\partial^2 \ell(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} & \frac{\partial^2 \ell(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta} \partial \sigma^2} \\ \left(\frac{\partial^2 \ell(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta} \partial \sigma^2}\right)' & \frac{\partial^2 \ell(\boldsymbol{\beta}, \sigma^2)}{(\partial \sigma^2)^2} \end{bmatrix}$$

Let's denote the block elements inside the Hessian matrix with:

$$\begin{aligned}
H_{11} &= -\frac{1}{\sigma^2}\boldsymbol{X}'\boldsymbol{X} \\
H_{12} &= -\frac{1}{\sigma^4}\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X\beta}) \;=\; H_{21}' \\
H_{22} &= \frac{n}{2\sigma^2} - \frac{1}{\sigma^6}(\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta})
\end{aligned}$$

(3.2) The ML estimators are given by solving the score vector when it's equal to zeros, particularly:

$$\begin{bmatrix} \frac{1}{\widehat{\sigma}^2}\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \\ -\frac{n}{2\widehat{\sigma}^2} + \frac{1}{2\widehat{\sigma}^4}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \end{bmatrix} \;=\; \begin{bmatrix} \mathbf{0}_k \\ 0 \end{bmatrix}$$

resulting in:

$$\begin{aligned}
\boldsymbol{X}'\boldsymbol{y} &= \boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} \\
\widehat{\sigma}^2 &= \frac{1}{n}\widehat{\epsilon}'\widehat{\epsilon}
\end{aligned}$$

where $\widehat{\epsilon} := \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ and that $\widehat{\sigma}^2$ is the sample average of the square residuals $\widehat{\epsilon}_i^2$, $\forall_i$. The matrix $\boldsymbol{X}$ is not invertible but is full columns rank and that $\boldsymbol{X}'\boldsymbol{X}$ is square and invertible to give

$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$, the expanding $\widehat{\epsilon}'\widehat{\epsilon} = (\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) = \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}\widehat{\beta} - \widehat{\beta}'\boldsymbol{X}'\boldsymbol{y} + \widehat{\beta}'\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}}$ noting that this simplifies to $\widehat{\epsilon}'\widehat{\epsilon} = \boldsymbol{y}'\boldsymbol{y} + \widehat{\beta}'\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}}$, and that $\widehat{H}$ is negative definite:

$$\widehat{H} = \begin{bmatrix} -\frac{1}{\widehat{\sigma}^2}\boldsymbol{X}'\boldsymbol{X} & -\frac{1}{\widehat{\sigma}^4}\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \\ . & \frac{n}{2\widehat{\sigma}^2} - \frac{1}{\widehat{\sigma}^6}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \end{bmatrix} = \begin{bmatrix} -\frac{1}{\widehat{\sigma}^2}\boldsymbol{X}'\boldsymbol{X} & \mathbf{0}_k \\ \mathbf{0}_k' & -\frac{n}{2\widehat{\sigma}^4} \end{bmatrix}$$

since $\widehat{\epsilon}'\widehat{\epsilon} = n\widehat{\sigma}^2$ and that $\boldsymbol{X}'\widehat{\boldsymbol{\epsilon}} = \boldsymbol{X}'\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \mathbf{0}_k$

(3.3)

$$\ell(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\widehat{\sigma}^2) - \frac{1}{2\widehat{\sigma}^2}\widehat{\epsilon}'\widehat{\epsilon} = -\frac{n}{2}(\log(2\pi) + 1) - \frac{n}{2}\log(\widehat{\sigma}^2)$$