Econ 7010 - Midterm - 2023

Ryan T. Godwin

The exam is 70 minutes long, and has 70 marks (1 mark per minute). You may not use any outside materials, only writing implements. Write your answers in the booklet provided.

Short Answer - 6 marks each

1. Use the M and P matrices to show that, if the variables in the X matrix change units, the residuals and predicted values remain the same.

A change of units in X can be effected using a matrix A. Denote the transformed matrix of regressors as $X^\star$, where $X^\star = XA$. The predicted values from the transformed model are:

$$\begin{aligned} P_{X^\star} = P_{XA} &= XA\left(A'X'XA\right)^{-1}A'X' \\ &= XAA^{-1}\left(X'X\right)^{-1}A'^{-1}A'X' \\ &= X\left(X'X\right)^{-1}X' \\ &= P_X \end{aligned}$$

The predicted values are the same from both models. Also:

$$M_{XA} = I - P_{XA} = I - P_X = M_X,$$

so the residuals are the same as well.

2. What is the role of "A2: No Perfect Multicollinearity" in the derivation of the Least Squares estimator?

(i) It's needed for $(X'X)$ to be invertible. That is, $(X'X)^{-1}$ doesn't exist, and neither does the LS estimator, unless we have have A2.

(ii) It ensures that $\frac{\partial^2 \boldsymbol{e}'\boldsymbol{e}}{\partial \boldsymbol{b} \partial \boldsymbol{b}'}$ is positive definite, so that we have minimized $\boldsymbol{e}'\boldsymbol{e}$, not maximized it.

3. Explain why R^2 increases when a variable is added to the model. What is the maximum value for R^2, and why?

$$R^2 = \left[1 - \frac{\boldsymbol{e}'\boldsymbol{e}}{\boldsymbol{y}'M_i\boldsymbol{y}}\right]$$

The LS estimate $\boldsymbol{b}$ solve the minimization problem:

$$\min \left(\boldsymbol{e}'\boldsymbol{e}\right)$$

When a variable is added to the model, the minimization problem gets easier, $\boldsymbol{e}'\boldsymbol{e}$ gets smaller, and R^2 increases.

The best that the model can fit is if the regression "line" passes through each data point, all residuals are 0, and then R^2 would attain the maximum of 1 (this would not be a good or interesting model, however).

4. Prove that the LS estimator is unbiased, stating any assumptions that you use.

5. Explain how to interpet a p-value OR explain how to interpret a confidence interval.

A p-value is the probability of obtaining an estimate (or test statistic) that is more adverse to the null hypothesis, than the estimate just obtained.

A (1-α)% confidence interval is a random interval, where (1-α)% of such intervals contain the true parameter value. Alternatively, a (1-α)% confidence interval contains all values for the null hypothesis that will not be rejected at the *alpha*% level.

6. Explain why an intercept is not needed in a model where all data has been "centred" (all data are mean zero).

It helps to understand that the intercept is the $\hat{\boldsymbol{y}}$ for when all the x variables are evaluated at 0.

Data can be centred by $M_{\boldsymbol{i}}\boldsymbol{y}$ and $M_{\boldsymbol{i}}X$. By FWL, $\boldsymbol{y}$ and X are then orthogonal to $\boldsymbol{i}$ (the intercept), and $\boldsymbol{i}$ can be dropped without affecting the estimates for $\boldsymbol{\beta}$.

Alternatively:

Without an intercept, the LS "line" is forced to pass through the origin ($\hat{\boldsymbol{y}} = 0$ when $X = \mathbf{0}$). But the LS line passes through the sample means of the data, so it passes through the origin anyway, and the intercept isn't needed.

$$\bar{X}\boldsymbol{b} = \bar{y} = 0$$

For this answer, it helps to understand that the intercept is the $\hat{y}$ for when all the x variables are evaluated at 0.

7. When estimating σ^2, why is s^2 typically used, and not $\hat{\sigma}^2$?

   s^2 is an unbiased estimator (under assumptions). $\hat{sigma}^2$ is a biased estimator.

8. Show that **either** the fitted regression line passes through the sample means of the data **or** that the sample mean of the fitted values $\hat{y}_i$ equals the sample mean of the actual values y_i.

From the first row of

$$X'\boldsymbol{y} = X'X\boldsymbol{b}$$

we get:

$$\sum_i y_i = nb_1 + b_2 \sum_i x_{i2} + \cdots + b_k \sum_i x_{ik}$$

and dividing by n we get:

$$\bar{y} = b_1 + b_2\overline{x}_2 + \cdots + b_k\overline{x}_k$$

---

or:

$$y_i = \hat{y}_i + e_i$$

taking the average of both sides:

$$\bar{y} = \overline{\hat{y}} + \bar{e}$$

but since the residuals sum to 0 (if there's an intercept in the model):

$$\bar{y} = \overline{\hat{y}} + 0 = \overline{\hat{y}}$$

Long Answer - 11 marks each

9. The true population model is:

$$\boldsymbol{y} = X_1\boldsymbol{\beta}_1 + X_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon} \tag{1}$$

but the model that is estimated is:

$$\boldsymbol{y} = X_1\boldsymbol{\beta}_1 + \boldsymbol{u} \tag{2}$$

However, X_1 and X_2 are **expected to be orthogonal**. Is b_1 from equation (2) biased? Prove.

$\boldsymbol{b}_1$ from model (2) is unbiased.

$$\begin{aligned}
\boldsymbol{b}_1 &= \left(X_1'X\right)^{-1} X_1'\boldsymbol{y} \\
&= \left(X_1'X\right)^{-1} X_1'\left(X_1\boldsymbol{\beta}_1 + X_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}\right) \\
&= \boldsymbol{\beta}_1 + \left(X_1'X_1\right)^{-1} X_1'X_2\boldsymbol{\beta}_2 + \left(X_1'X_1\right)^{-1} X_1'\boldsymbol{\epsilon}
\end{aligned}$$

Using A.5 and taking expectations we get:

$$\mathbb{E}\left[\boldsymbol{b}_1\right] = \boldsymbol{\beta}_1 + \left(X_1'X_1\right)^{-1}\mathbb{E}\left[X_1'X_2\right]\boldsymbol{\beta}_2 + \left(X_1'X_1\right)^{-1}\mathbb{E}\left[X_1'\boldsymbol{\epsilon}\right]$$

If X_1 and X_2 are expected to be orthogonal then $\mathbb{E}\left[X_1'X_2\right] = \mathbf{0}$ and by A.5 $\mathbb{E}\left[X_1\boldsymbol{\epsilon}\right] = \mathbf{0}$, so $\mathbb{E}\left[\boldsymbol{b}_1\right] = \boldsymbol{\beta}_1$.

Alternatively: using FWL you can show that the partial estimator for $\boldsymbol{b}_1$ under model 1 is the same as that for $\boldsymbol{b}_1$ under model 2 (due to the orthogonality), so that if it's unbiased in one model it's unbiased in both. Many students opted for this answer.

10. Derive the variance-covariance matrix for $\boldsymbol{b}$, stating any assumptions that you use. What does the Gauss-Markov theorem have to say about this matrix? How is this matrix used in hypothesis testing?

$$
\begin{aligned}
\mathbb{V}(\boldsymbol{b}) &= \mathbb{V}\left[\boldsymbol{\beta} + \left(X'X\right)^{-1} X'\boldsymbol{\epsilon}\right] \\
&= \mathbb{V}\left[\left(X'X\right)^{-1} X'\boldsymbol{\epsilon}\right] \\
&= \left(X'X\right)^{-1} X'\mathbb{V}(\boldsymbol{\epsilon}) X \left(X'X\right)^{-1} \quad ; \quad \text{by A.5} \\
&= \left(X'X\right)^{-1} X'\sigma^2 I_n X \left(X'X\right)^{-1} \quad ; \quad \text{by A.4} \\
&= \sigma^2 \left(X'X\right)^{-1}
\end{aligned}
$$

The GM theorem says that $\mathbb{V}(\hat{\boldsymbol{\beta}}) - \mathbb{V}(\boldsymbol{b})$ is positive definite for any linear and unbiased $\hat{\boldsymbol{\beta}}$, making LS the estimator with minimum variance.

The square roots of the diagonal elements of the matrix $\sigma^2 \left(X'X\right)^{-1}$ are the *standard errors* of the $\boldsymbol{b}_i$. These are needed for hypothesis testing, and for constructing confidence intervals. For example, the z-test statistic is:

$$
z = \frac{b_i - \beta_{i,0}}{\sqrt{\sigma^2 \left(X'X\right)^{-1}}_{i,i}}
$$

and replacing σ^2 with s^2 gives the t-statistic.

END.