Additional Practice Questions Solutions

Hansen 2.21

a) In class, we showed that $\gamma_1 = \beta_1$ under the condition that either $\beta_2 = 0$ or that the included regressor and the omitted variable are uncorrelated. In this case, that would amount to $\mathrm{E}[X \cdot X^2] = 0$. This holds when $\mathrm{E}[X^3] = 0$.

b) Using the same sort of arguments as for part (a), this would holds either if $\theta_2 = 0$ or if $\mathrm{E}[X \cdot X^3] = \mathrm{E}[X^4] = 0$; however, since X^4 is an even power the only case where it can be equal to 0 is if X is always equal to 0 (i.e., X is non-random and always equal to 0) – I think this is the point being made in part (b).

Hansen 3.2

Let's call $\tilde{\beta}$ and $\tilde{\mathbf{e}}$ the OLS estimates and residuals from the regression of $\mathbf{Y}$ on $\mathbf{Z}$. Notice that

$$\begin{aligned}
\tilde{\beta} &= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y} \\
&= \left((\mathbf{XC})'\mathbf{XC}\right)^{-1}(\mathbf{XC})'\mathbf{Y} \\
&= (\mathbf{C}'\mathbf{X}'\mathbf{XC})^{-1}\mathbf{C}'\mathbf{X}'\mathbf{Y} \\
&= \mathbf{C}^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'^{-1}\mathbf{C}'\mathbf{X}'\mathbf{Y} \\
&= \mathbf{C}^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\
&= \mathbf{C}^{-1}\hat{\beta}
\end{aligned}$$

where the second equality holds by plugging in $\mathbf{Z} = \mathbf{XC}$, the third equality holds by taking the transpose of $\mathbf{XC}$, the fourth equality holds because $\mathbf{C}$ and $\mathbf{X}'\mathbf{X}$ are nonsingular, the fifth equality holds by canceling the $\mathbf{C}'^{-1}\mathbf{C}'$, and the last equality holds by the definition of $\hat{\beta}$.

Now, for the residuals, notice that

$$\begin{aligned}
\tilde{\mathbf{e}} &= \mathbf{Y} - \mathbf{Z}\tilde{\beta} \\
&= \mathbf{Y} - \mathbf{XCC}^{-1}\hat{\beta} \\
&= \mathbf{Y} - \mathbf{X}\hat{\beta} \\
&= \hat{\mathbf{e}}
\end{aligned}$$

where this result holds just by plugging in and canceling terms. This says that the residuals from the regression of $\mathbf{Y}$ on $\mathbf{Z}$ are exactly the same as the residuals from the regression of $\mathbf{Y}$ on $\mathbf{X}$.

As a side-comment, a simple example of this problem would be something like scaling all the regressors by, say, 100. If you did this, it would change the value of the estimated coefficients (divide them by 100) but would fit the data equally well.

Hansen 3.3

$$\begin{aligned}\mathbf{X}'\hat{\mathbf{e}} &= \mathbf{X}'(\mathbf{Y} - \mathbf{X}\hat{\beta}) \\ &= \mathbf{X}'(\mathbf{Y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}) \\ &= \mathbf{X}'\mathbf{Y} - \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}) \\ &= \mathbf{X}'\mathbf{Y} - \mathbf{X}'\mathbf{Y} \\ &= \mathbf{0}\end{aligned}$$

which holds by plugging in for $\hat{\mathbf{e}}$ and canceling terms.

Hansen 3.5

The OLS coefficient from a regression of $\hat{\mathbf{e}}$ on $\mathbf{X}$ is given by

$$\begin{aligned}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{e}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{Y} - \mathbf{X}\hat{\beta}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\hat{\beta} \\ &= \hat{\beta} - \hat{\beta} \\ &= 0\end{aligned}$$

where the first part of the third equality holds by the definition of $\hat{\beta}$ and the last part holds by canceling the terms involving $(\mathbf{X}'\mathbf{X})$.

Hansen 3.7

$$\begin{aligned}\mathbf{P}\mathbf{X} &= \mathbf{P}\begin{bmatrix}\mathbf{X}_1 & \mathbf{X}_2\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{P}\mathbf{X}_1 & \mathbf{P}\mathbf{X}_2\end{bmatrix}\end{aligned}$$

Further, since $\mathbf{X} = \begin{bmatrix}\mathbf{X}_1 & \mathbf{X}_2\end{bmatrix}$ and $\mathbf{P}\mathbf{X} = \mathbf{X}$ (from the properties of the projection matrix $\mathbf{P}$), this implies that $\mathbf{P}\mathbf{X}_1 = \mathbf{X}_1$.

Similarly,

$$\begin{aligned}\mathbf{M}\mathbf{X} &= \mathbf{M}\begin{bmatrix}\mathbf{X}_1 & \mathbf{X}_2\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{M}\mathbf{X}_1 & \mathbf{M}\mathbf{X}_2\end{bmatrix}\end{aligned}$$

but we also know that $\mathbf{M}\mathbf{X} = \mathbf{0}_{n\times k} = \begin{bmatrix}\mathbf{0}_{n\times k_1} & \mathbf{0}_{n\times k_2}\end{bmatrix}$ where, for example, $\mathbf{0}_{n\times k_1}$ is an $n \times k_1$ matrix of zeroes. This implies that $\mathbf{M}\mathbf{X}_1 = \mathbf{0}_{n\times k_1}$.

Hansen 3.22

This material will not be on exam, so I'll assign this for a future homework problem.

Extra Question

a)

$$\begin{aligned}
ATE &= \mathrm{E}[Y(1) - Y(0)] \\
&= \mathrm{E}[Y(1)] - \mathrm{E}[Y(0)] \\
&= \mathrm{E}[\mathrm{E}[Y(1)|X]] - \mathrm{E}[\mathrm{E}[Y(0)|X]] \\
&= \mathrm{E}[\mathrm{E}[Y(1)|X, D = 1]] - \mathrm{E}[\mathrm{E}[Y(0)|X, D = 0]] \\
&= \mathrm{E}[\mathrm{E}[Y|X, D = 1]] - \mathrm{E}[\mathrm{E}[Y|X, D = 0]]
\end{aligned}$$

where the first equality is the definition of ATE, the second equality pushes the expectation through the difference, the third equality holds by the law iterated expectations, the fourth equality holds by unconfoundedness, and the last equality holds because $Y = Y(1)$ among the treated group and $Y = Y(0)$ among the untreated group. This shows that ATE is identified.

b) In class, we showed that $ATT = \mathrm{E}[Y|D = 1] - \mathrm{E}[\mathrm{E}[Y|X, D = 0]|D = 1]$. These are notably different. The expression for ATE takes the $\mathrm{E}[Y|X, D = 1]$ (the mean of Y conditional on X among the treated group) and averages it over the distribution of X for the entire population and then subtracts $\mathrm{E}[Y|X, D = 0]$ (the mean of Y conditional on X among the untreated group) averaged over the population distribution of X.

In contrast, ATT compares the mean of Y among the treated group to $\mathrm{E}[Y|X, D = 0]$ where this is averaged over the distribution of X *among the treated group*.

An intuition for why ATE involves averaging over the distribution of X for the entire population is that ATE is the average treatment effect for the entire population.

c) We have that

$$\begin{aligned}
Y_i &= Y_i(0) + D_i(Y_i(1) - Y_i(0)) \\
&= X_i'\beta + e_i + \alpha D_i \\
&= \alpha D_i + X_i'\beta + e_i
\end{aligned}$$

where the first equality holds by writing observed outcomes in terms of potential outcomes, the second equality uses the model for untreated potential outcomes and treatment effect homogeneity, and the last equality rearranges terms.

Furthermore, unconfoundedness implies that $\mathrm{E}[e|X, D] = 0$ which implies that α can be estimated from the regression of Y on D and X.

d) This is exactly the same regressions as we talked about in class after we had identified the ATT. This should not be surprising because, if we restrict the effect of participating in the treatment to be the same across all units, then $ATT = ATE = \alpha$.