Exercise 1

We are using the model $y_i = \Phi(x_i'\theta_0) + u_i$ where $\dim(x_i) = \dim(\theta) = k$.

Part A: Estimation with Exogenous Regressors

Assume that $E[u_i|x_i] = 0$. This means that all of the regressors are exogenous.

(i) The nonlinear least squares estimator is

$$\hat{\theta} = \arg\max_{\theta\in\Theta} -\frac{1}{2n}\sum_{i=1}^{n}\big[y_i - \Phi(x_i'\theta)\big]^2.$$

That division by 2 doesn't affect the maximizer since it's monotonic, but it will make things nicer once we take derivatives.

From the law of iterated expectations, the exogeneity assumption implies that $E[x_i u_i] = 0$. Since $u_i = y_i - \Phi(x_i'\theta_0)$, we can define the method of moments estimator to be the one that satisfies

$$\frac{1}{n}\sum_{i=1}^{n} x_i\big[y_i - \Phi(x_i'\hat{\theta}_{MOM})\big] = 0.$$

(ii) For the nonlinear least squares version, we need the following assumptions for consistency:

- **Existence** of $\hat{\theta}$
- **Identification** of θ_0
- **ULLN**: i.i.d.; compactness of Θ; measureability in (y_i, x_i) for all θ; continuity in θ; existence of a dominating function

For the method of moments, we need roughly the same things. Details aren't worth getting into as far as I'm concerned.

For asymptotic normality, we need

- **Consistency** of $\hat{\theta}$
- **Mean value expansion validity**: a twice continuously differentiable function and interior θ_0
- **Consistency** of the inverted part
- Normality of the other part

This is worth going into detail over. Let's use M-estimation for the NLS version. Then we have

$$
\begin{aligned}
& m_i = -\frac{1}{2}\left[y_i - \Phi(x_i'\theta_0)\right]^2 \\
\implies \quad & s_i = x_i'\left[y_i - \Phi(x_i'\hat{\theta})\right]\Phi'(x_i'\hat{\theta}) \\
\implies \quad & H_i = x_i x_i'\left[y_i - \Phi(x_i'\hat{\theta})\right]\Phi''(x_i'\hat{\theta}) - x_i x_i'\Phi'(x_i'\hat{\theta})^2
\end{aligned}
$$

Let $Q_n(\theta) = \sum_{i=1}^n -\frac{1}{2}\left[y_i - \Phi(x_i'\theta_0)\right]^2$. Then use the mean value expansion

$$
\begin{aligned}
\frac{\partial Q_n(\hat{\theta})}{\partial \theta} &= \frac{\partial Q_n(\hat{\theta}_0)}{\partial \theta_0} + \frac{\partial^2 Q_n(\theta^*)}{\partial \theta \partial \theta'}(\hat{\theta} - \theta_0) \\
\implies \quad \frac{1}{n}\sum_{i=1}^n s_i(\hat{\theta}) &= \frac{1}{n}\sum_{i=1}^n s_i(\theta_0) + \frac{1}{n}\sum_{i=1}^n H(\theta^*)(\hat{\theta} - \theta_0).
\end{aligned}
$$

The first term is zero just by the definition of $\hat{\theta}$. So then we can write

$$
\begin{aligned}
(\hat{\theta} - \theta_0) &= -\left[\frac{1}{n}\sum_{i=1}^n H(\theta^*)\right]^{-1} \frac{1}{n}\sum_{i=1}^n s_i(\theta_0) \\
&= -\left[\frac{1}{n}\sum_{i=1}^n x_i x_i'\left[y_i - \Phi(x_i'\theta^*)\right]\Phi''(x_i'\theta^*) - x_i x_i'\Phi'(x_i'\theta^*)^2\right]^{-1} \frac{1}{n}\sum_{i=1}^n x_i'\left[y_i - \Phi(x_i'\theta_0)\right]\Phi'(x_i'\theta_0)
\end{aligned}
$$

The asymptotic normality follows from

$$
\begin{aligned}
\sqrt{n}(\hat{\theta} - \theta_0) &= -\left[\frac{1}{n}\sum_{i=1}^n H(\theta^*)\right]^{-1} \frac{1}{n}\sum_{i=1}^n s(\theta_0) \\
&= -\left[\frac{1}{n}\sum_{i=1}^n x_i x_i'\left[y_i - \Phi(x_i'\theta^*)\right]\Phi''(x_i'\theta^*) - x_i x_i'\Phi'(x_i'\theta^*)^2\right]^{-1} \frac{\sqrt{n}}{n}\sum_{i=1}^n x_i'\left[y_i - \Phi(x_i'\theta_0)\right]\Phi'(x_i'\theta_0)
\end{aligned}
$$

The Hessian needs to be consistent and positive definite. The score needs to have zero mean—which it does—and finite second moments so that we can apply the Lindeberg-Levy central limit theorem. It will be normally distributed with mean zero and the rather horrifying asymptotic variance

$$\begin{aligned}\text{Avar}\left(\sqrt{n}[\hat{\theta}-\theta_0]\right) = & E\left[\sum_{i=1}^{n} x_i x_i' \left[y_i - \Phi(x_i'\theta^*)\right]\Phi''(x_i'\theta^*) - x_i x_i' \Phi'(x_i'\theta^*)^2\right]^{-1} \\ & E\left[x_i x_i' \left[y_i - \Phi(x_i'\theta_0)\right]^2 \Phi'(x_i'\theta_0)^2\right] \\ & E\left[\sum_{i=1}^{n} x_i x_i' \left[y_i - \Phi(x_i'\theta^*)\right]\Phi''(x_i'\theta^*) - x_i x_i' \Phi'(x_i'\theta^*)^2\right]^{-1}\end{aligned}$$

Now let's think of this in terms of the method of moments. In this case, we'll have h be the "moment function," that is, $h_i = x_i u_i = x_i[y_i - \Phi(x_i'\theta)]$. So $\hat{\theta}_M$ is defined by

$$\frac{1}{n}\sum_{i=1}^{n} h_i = \frac{1}{n}\sum_{i=1}^{n} x_i[y_i - \Phi(x_i'\hat{\theta}_M)] = 0.$$

Now let's do an MVE around it to get

$$\frac{1}{n}\sum_{i=1}^{n} x_i[y_i - \Phi(x_i'\hat{\theta}_M)] = \frac{1}{n}\sum_{i=1}^{n} x_i[y_i - \Phi(x_i'\theta_0)] - \frac{1}{n}\sum_{i=1}^{n} x_i x_i' \Phi'(x_i'\theta^*)(\hat{\theta}_M - \theta_0),$$

which we can turn into

$$\sqrt{n}(\hat{\theta}_M - \theta_0) = \left[\frac{1}{n}\sum_{i=1}^{n} x_i x_i' \Phi'(x_i'\theta^*)\right]^{-1} \frac{\sqrt{n}}{n}\sum_{i=1}^{n} x_i[y_i - \Phi(x_i'\theta_0)].$$

Great. For asymptotic normality, we need consistency and positive definiteness of the inverted term; and the other term needs to have zero mean and finite second moment. The zero mean is satisfied since that term is just the moment condition. It will be normally distributed with zero mean and asymptotic normality of

$$\begin{aligned}\text{Avar}\left(\sqrt{n}[\hat{\theta}_M-\theta_0]\right) = & E\left[\sum_{i=1}^{n} x_i x_i' \Phi'(x_i'\theta^*)\right]^{-1} E\left[x_i x_i' \left(y_i - \Phi(x_i'\theta_0)\right)^2\right] \\ & E\left[\sum_{i=1}^{n} x_i x_i' \Phi'(x_i'\theta^*)\right]^{-1}.\end{aligned}$$

Part B: Estimation with Instrumental Variables

Now let's assume that $E[u_i|x_i] \neq 0$ so that there is some endogeneity in the regressors. We have instruments z_i where $\dim(z_i) = \ell$ and $\ell > k$. In other words, we have more instruments than parameters. The fact that z_i are instruments implies that $E[z_i u_i] = 0$. We can still do estimates, but we we won't have square matrices and so we won't be able to invert things. Another approach is needed.

(i) Some gobbledygook about linear projections. Don't care.

(ii) Okay, so we can't invert things the way we'd like. Let's try something else. Oh hey, we have the moment condition

$$E[z_i u_i] = E\left[z_i\big(y_i - \Phi(x_i'\theta_{IV})\big)\right] = 0.$$

Consistency is pretty much the same rigmarole. But now we have to get a bit tricky with the mean value expansion. In particular, we'll need some weighting matrix $\hat{W}$ that converges to W_0. Furthermore, we'll do the mean value expansion on

$$\left[\frac{1}{n}\sum_{i=1}^{n} h_i(\hat{\theta}_{IV})\right]' \hat{W}\frac{1}{n}\sum_{i=1}^{n} h_i(\hat{\theta}_{IV}),$$

which is nice and symmetric. Our new estimator in this setup is

$$\hat{\theta}_{IV} = \arg\max_{\theta\in\Theta} -\frac{1}{n}\sum_{i=1}^{n} h_i'(\theta)\hat{W}\frac{1}{n}\sum_{i=1}^{n} h_i(\theta).$$

Based on the moment condition, we'll have $h_i = z_i\big(y_i - \Phi(x_i'\hat{\theta}_{IV})\big)$. So then we'd like to do the mean value expansion

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} z_i\big(y_i - \Phi(x_i'\hat{\theta}_{IV})\big) &= \frac{1}{n}\sum_{i=1}^{n} z_i\big(y_i - \Phi(x_i'\hat{\theta}_0)\big)\\ &\quad -\frac{1}{n}\sum_{i=1}^{n} z_i x_i'\Phi'(x_i'\theta^*)\big(\hat{\theta}_{IV} - \theta_0\big).\end{aligned}$$

The problem is, the LHS is not zero like we would want. What we can do is pre-multiply both sides by $\hat{\mathcal{H}}_n'\hat{W}$, where

$$\hat{\mathcal{H}}_n = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial h_i(\hat{\theta})}{\partial\theta'} = -\frac{1}{n}\sum_{i=1}^{n} z_i x_i'\Phi'(x_i'\theta),$$

which gives

$$\left[\frac{1}{n}\sum_{i=1}^{n} -z_i x_i' \Phi'(x_i'\hat{\theta}_{IV})\right]' \hat{W} \sum_{i=1}^{n} z_i\big(y_i - \Phi(x_i'\hat{\theta}_{IV})\big) =$$

$$\left[-\frac{1}{n}\sum_{i=1}^{n} z_i x_i' \Phi'(x_i'\hat{\theta}_{IV})\right]' \hat{W} \frac{1}{n}\sum_{i=1}^{n} z_i\big(y_i - \Phi(x_i'\hat{\theta}_0)\big)$$

$$+ \left[\frac{1}{n}\sum_{i=1}^{n} z_i x_i' \Phi'(x_i'\hat{\theta}_{IV})\right]' \hat{W} \frac{1}{n}\sum_{i=1}^{n} z_i x_i' \Phi'(x_i'\theta^*)\big(\hat{\theta}_{IV} - \theta_0\big).$$

The reason this is useful is because the LHS is actually zero now. This is because we know that

$$\frac{1}{n}\sum_{i=1}^{n} h_i'(\hat{\theta}_{IV})\hat{W}\frac{1}{n}\sum_{i=1}^{n} h_i(\hat{\theta}_{IV}) = 0,$$

and since $\hat{W}$ is symmetric, matrix differentiation gives

$$2\hat{W}\frac{1}{n}\sum_{i=1}^{n} h_i(\hat{\theta}_{IV}) = 0.$$

So divide out the 2 and multiply both sides by $\hat{\mathcal{H}}_n'$ and we have the zero. Now we can actually continue with the mean value expansion. In particular,

$$\sqrt{n}\big(\hat{\theta}_{IV} - \theta_0\big) = \left(\left[\frac{1}{n}\sum_{i=1}^{n} z_i x_i' \Phi'(x_i'\hat{\theta}_{IV})\right]' \hat{W} \frac{1}{n}\sum_{i=1}^{n} z_i x_i' \Phi'(x_i'\theta^*)\right)^{-1}$$

$$\left[\frac{1}{n}\sum_{i=1}^{n} z_i x_i' \Phi'(x_i'\hat{\theta}_{IV})\right]' \hat{W} \frac{\sqrt{n}}{n}\sum_{i=1}^{n} z_i\big(y_i - \Phi(x_i'\hat{\theta}_0)\big).$$

Yeah this is a damn mess, but you get the picture. So the inverted chunk has to converge to something positive semidefinite; the inverted term within the inverted chunk has to have full column rank k. The other term has to have zero mean and finite second moment so we can invoke Lindeberg-Levy. The asymptotic variance is exactly what you'd expect.

Exercise 2

We'll use the model $y_i = x_i'\theta_0 + u_i$ where $\dim(x_i) = \dim(\theta) = k$. x_i is not exogenous, so $E[u_i x_i] \neq 0$. But we have some z_i that satisfies $E[u_i z_i] = 0$ and $x_i = \gamma' z_i + \epsilon_i$. The dimension of z_i is $\ell \times 1$ and the dimension of γ is $\ell \times k$. Keep in mind that z contains all of the exogenous regressors as well. In order to make any progress, we need to make sure that there are at least as many instruments as there are endogenous variables.

I'm going to break this down a bit further. Partition x into endognous x_1 and exogenous x_2. Then define $z = [x_2 \quad \tilde{z}]$, where $\tilde{z}$ are the instruments. If there are exactly as many instruments as endogenous regressors, i.e. $\dim(x_1) = \dim(\tilde{z})$, then $\dim(z) = \ell = k$. This is the just-identified case. If there are more instruments than endogenous regressors, then $\dim(x_1) < \dim(\tilde{z})$, so $\dim(z) = \ell > k$. This is the over-identified case. The just-identification case is nice because we'll have a nice square matrix where things can be inverted. If we are over-identified, then we have to be trickier.

Part A: Just-Identification

Let's first suppose that $\ell = k$ so that we have as many instruments as endogenous variables.

(i) The population moment that defines θ_0 is $E[z_i(y_i - x_i'\theta_0)] = 0$. There are k moment conditions (based on the dimension of the thing in the expectation) and k estimands/parameters we'd like to estimate because of $\dim(\theta) = k$.

(ii) The primitive conditions for identification of θ_0 are $E[z_i u_i] = 0$ and $y_i = x_i'\theta_0 + u_i$ for a unique θ_0.

(iii) From the moment condition, we can derive an estimator

$$
\begin{aligned}
0 &= \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\hat{\theta})z_i \\
&= \frac{1}{n}\sum_{i=1}^{n} z_i y_i - \frac{1}{n}\sum_{i=1}^{n} z_i x_i'\hat{\theta} \\
\implies \quad \hat{\theta} &= \left[\frac{1}{n}\sum_{i=1}^{n} z_i x_i'\right]^{-1} \frac{1}{n}\sum_{i=1}^{n} z_i y_i.
\end{aligned}
$$

(iv) Let's examine consistency. Since $y = x_i'\theta_0 + u_i$, we can plug that in and get

$$
\begin{aligned}
\hat{\theta} &= \left[\frac{1}{n}\sum_{i=1}^{n} z_i x_i'\right]^{-1} \frac{1}{n}\sum_{i=1}^{n} z_i(x_i'\theta_0 + u_i) \\
&= \left[\frac{1}{n}\sum_{i=1}^{n} z_i x_i'\right]^{-1} \frac{1}{n}\sum_{i=1}^{n} z_i x_i'\theta_0 + \left[\frac{1}{n}\sum_{i=1}^{n} z_i x_i'\right]^{-1} \frac{1}{n}\sum_{i=1}^{n} u_i \\
&= \theta_0 + \left[\frac{1}{n}\sum_{i=1}^{n} z_i x_i'\right]^{-1} \frac{1}{n}\sum_{i=1}^{n} z_i u_i.
\end{aligned}
$$

So we need $E[z_i x_i']$ to have full rank. We need $E[z_i u_i] = 0$. We need i.i.d. data. Then we know that the inverted term converges to something finite and the second term converges to zero. By Slutsky's theorem, the two then multiply to zero in the plim. And thus we have consistency.

(v) For asymptotic normality, we can go with

$$\sqrt{n}(\hat{\theta} - \theta_0) = \left[\frac{1}{n}\sum_{i=1}^{n} z_i x_i'\right]^{-1} \frac{\sqrt{n}}{n}\sum_{i=1}^{n} z_i u_i.$$

We need finite $Var[z_i u_i]$, in which case we can use Lindeberg-Levy to establish that

$$\sqrt{n}(\hat{\theta} - \theta_0) \sim \mathcal{N}\left(0, E[z_i x_i']^{-1} E[u_i^2 z_i z_i'] E[z_i x_i']^{-1}\right).$$

(vi) If errors are homoskedastic, then $E[u_i^2 z_i z_i'] = \sigma^2 E[z_i z_i']$. So you could just write the asymptotic variance as $\sigma^2 E[z_i x_i']^{-1} E[z_i z_i'] E[z_i x_i']^{-1}$. Not sure why that's useful.

Part B: Overidentification and GMM

Now let's assume that $\ell > k$, so we have more instruments than endogenous regressors.

(i) We'll just use the same population moment, $E[z_i(y_i - x_i'\theta_0)] = 0$. There are still k parameters that we want to estimate; and now the moment "matrix" is $(\ell \times 1)(1 \times k) = \ell \times k$ in dimension, so there are ℓ moment conditions.

(ii) Identification requires the same things as before: the primitive conditions for identification of θ_0 are $E[z_i u_i] = 0$ and $y_i = x_i'\theta_0 + u_i$ for a unique θ_0.

(iii) Define $h_i(\theta) = z_i(y_i - x_i'\theta)$. Then the GMM estimator satisfies

$$\hat{\theta} = \arg\max_{\theta \in \Theta} -\frac{1}{n}\sum_{i=1}^{n} z_i(y_i - x_i'\theta)' \, \hat{W} \, \frac{1}{n}\sum_{i=1}^{n} z_i(y_i - x_i'\theta).$$

(iv) For ULLN, we write

$$\begin{aligned}
&\sup_{\theta}\left|\frac{1}{n}\sum_{i=1}^{n} z_i(y_i - x_i'\theta) - E[z_i(y_i - x_i'\theta)]\right| \\
=&\sup_{\theta}\left|\frac{1}{n}\sum_{i=1}^{n} z_i y_i - \frac{1}{n}\sum_{i=1}^{n} z_i x_i'\theta - E[z_i y_i] + E[z_i x_i'\theta]\right| \\
\leq&\sup_{\theta}\left|\frac{1}{n}\sum_{i=1}^{n} z_i y_i - E[z_i y_i]\right| + \left|\frac{1}{n}\sum_{i=1}^{n} z_i x_i' - E[z_i x_i']\right| \sup_{\theta}|\theta|\,.
\end{aligned}$$

We need $\sup_\theta |\theta|$ to be finite. Furthermore we need i.i.d. data, always with the i.i.d. data. And finally we need finite $E[z_i y_i]$ and $E[z_i x_i']$. Then we can invoke Khinchine to show that this whole things converges in probability to zero.

(v) To find the sampling error, first do the mean value expansion with $h_i(\theta)$ to get

$$\frac{1}{n}\sum_{i=1}^{n} h_i(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n} h_i(\theta_0) + \frac{1}{n}\sum_{i=1}^{n} \frac{\partial h_i(\theta^*)}{\partial \theta}(\hat{\theta} - \theta_0).$$

Define

$$\hat{\mathcal{H}}_n = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial h_i(\theta)}{\partial \theta'}.$$

Premultiply both sides by $\hat{\mathcal{H}}_n' \hat{W}$ and the result is

$$\hat{\mathcal{H}}_n' \hat{W} \frac{1}{n}\sum_{i=1}^{n} h_i(\hat{\theta}) = \hat{\mathcal{H}}_n' \hat{W} \frac{1}{n}\sum_{i=1}^{n} h_i(\theta_0) + \hat{\mathcal{H}}_n' \hat{W} \frac{1}{n}\sum_{i=1}^{n} \frac{\partial h_i(\theta^*)}{\partial \theta}(\hat{\theta} - \theta_0).$$

The LHS is zero from the moment condition. Then we can solve for

$$\sqrt{n}(\hat{\theta} - \theta_0) = -\left[\hat{\mathcal{H}}_n' \hat{W} \frac{1}{n}\sum_{i=1}^{n} \frac{\partial h_i(\theta^*)}{\partial \theta}\right]^{-1} \hat{\mathcal{H}}_n' \; \hat{W} \; \frac{\sqrt{n}}{n} \sum_{i=1}^{n} h_i(\theta_0).$$

We have $h_i(\theta) = z_i(y_i - x_i'\theta) = z_i u_i$. Therefore

$$\frac{\partial h_i(\theta)}{\partial \theta} = -z_i x_i',$$

$$\mathcal{H}_0 = -E[z_i x_i].$$

I guess we can just plug in $\mathcal{H}_0$ and W for some reason. Plugging in the rest gives

$$\sqrt{n}(\hat{\theta} - \theta_0) = \left[E[z_i x_i]' W \frac{1}{n}\sum_{i=1}^{n} z_i x_i'\right]^{-1} E[z_i x_i]' \; W \; \frac{\sqrt{n}}{n} \sum_{i=1}^{n} z_i u_i.$$

For asymptotic normality, we need finite $E[z_i x_i']$, full rank of $E[z_i x_i]$, positive semidefiniteness of the inverted term. Then the inverted term with Slutsky's theorem gives $(E[z_i x_i']' W E[z_i x_i'])^{-1}$.

For the other term, we just need finite $Var(z_i u_i)$. Then we can go Lindeberg-Levy all over it for normality with zero mean. The asymptotic variance will end up being

$$\text{Avar}\left(\sqrt{n}(\hat{\theta} - \theta_0)\right) = (E[z_i x_i']' W E[z_i x_i'])^{-1} E[z_i x_i]' W E[u^2 z_i z_i'] W E[z_i x_i] (E[z_i x_i']' W E[z_i x_i'])^{-1}.$$

Christ.

(vi) We haven't said anything about W yet other than it's symmetric. The efficient weighting matrix is actually

$$W = E[h_i(\theta_0)h_i(\theta_0)']^{-1} = E[z_i(y_i - x_i'\theta)z_i(y_i - x_i'\theta)']^{-1} = E[u_i^2 z_i z_i']^{-1}.$$

This is going to nicely cancel out some stuff in the asymptotic variance, giving

$$\text{Avar}\left(\sqrt{n}(\hat{\theta} - \theta_0)\right) = (E[z_i x_i']' E[u_i^2 z_i z_i']^{-1} E[z_i x_i'])^{-1}.$$

(vii) Under homoskedasticity, we can write $E[u_i^2 z_i z_i'] = \sigma^2 E[z_i z_i']$. Therefore

$$\text{Avar}\left(\sqrt{n}(\hat{\theta} - \theta_0)\right) = \sigma^2 (E[z_i x_i']' E[z_i z_i']^{-1} E[z_i x_i'])^{-1}.$$

Homoskedasticity is stupid.

Part C: Overidentification and Two-Stage Least Squares

(i) Now let's do a two-stage least squares estimate. The idea is that since x_i is correlated with z_i, we can have $x_i = \Gamma' z_i + \epsilon_i$. So regress x_i on z_i. Doing the typical OLS rigmarole gives $\hat{\Gamma} = (\sum_{i=1}^n z_i z_i')^{-1} \sum_{i=1}^n z_i x_i'$. It follows that

$$\hat{x}_i = \left[\left(\sum_{i=1}^n z_i z_i'\right)^{-1} \sum_{i=1}^n z_i x_i'\right]' z_i.$$

The OLS estimator is

$$\begin{aligned}\hat{\theta} &= \left(\sum_{i=1}^n \hat{x}_i x_i'\right)^{-1} \sum_{i=1}^n \hat{x}_i y_i \\ &= \left(\sum_{i=1}^n \hat{\Gamma}' z_i x_i'\right)^{-1} \sum_{i=1}^n \hat{\Gamma}' z_i y_i.\end{aligned}$$

Okay, so what's the transpose of Γ? The first term is symmetric, and the second we can just exchange x_i and z_i. So when we take the transpose, we get

$$\hat{\Gamma}' = \sum_{i=1}^n x_i z_i' \left(\sum_{i=1}^n z_i z_i'\right)^{-1}$$

We can continue with the OLS estimator

$$
\begin{aligned}
\hat{\theta} &= \left(\sum_{i=1}^{n} \hat{\Gamma}' z_i x_i'\right)^{-1} \sum_{i=1}^{n} \hat{\Gamma}' z_i y_i \\
&= \left[\sum_{i=1}^{n} x_i z_i' \left(\sum_{i=1}^{n} z_i z_i'\right)^{-1} \sum_{i=1}^{n} z_i x_i'\right]^{-1} \sum_{i=1}^{n} x_i z_i' \left(\sum_{i=1}^{n} z_i z_i'\right)^{-1} \sum_{i=1}^{n} z_i y_i.
\end{aligned}
$$

(ii) We want the sampling error to be asymptotically normal and centered around zero.

$$
\begin{aligned}
\hat{\theta} &= \left[\frac{1}{n}\sum_{i=1}^{n} x_i z_i' \left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} z_i x_i'\right]^{-1} \frac{1}{n}\sum_{i=1}^{n} x_i z_i' \left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} z_i (x_i'\theta_0 + u_i) \\
&= \theta_0 + \left[\frac{1}{n}\sum_{i=1}^{n} x_i z_i' \left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} z_i x_i'\right]^{-1} \frac{1}{n}\sum_{i=1}^{n} x_i z_i' \left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} z_i u_i.
\end{aligned}
$$

So right away we need i.i.d. data, we need for $E[z_i z_i']$ have full row rank and be finite, and we need $E[z_i x_i']$ to be finite. Then the entire thing in the inverse needs to be positive semidefinite. We need $E[z_i u_i] = 0$. Then by Slutsky's theorem, we have consistency.

So for normality, we can then take

$$
\sqrt{n}\left(\hat{\theta} - \theta_0\right) = \left[\frac{1}{n}\sum_{i=1}^{n} x_i z_i' \left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} z_i x_i'\right]^{-1} \frac{1}{n}\sum_{i=1}^{n} x_i z_i' \left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} z_i u_i.
$$

We need $\mathrm{Var}(z_i u_i)$ to have finite second moment so that we can get our Lindeberg-Levy on. Then we have normality with mean zero and the horrendous asymptotic variance

$$
\begin{aligned}
\mathrm{Avar}\left(\sqrt{n}\left(\hat{\theta} - \theta_0\right)\right) &= \left(E[x_i z_i'] E[z_i z_i']^{-1} E[z_i x_i']\right)^{-1} \\
&\quad E[x_i z_i'] E[z_i z_i']^{-1} E[u^2 z_i z_i'] E[z_i z_i']^{-1} E[z_i x_i'] \\
&\quad \left(E[x_i z_i'] E[z_i z_i']^{-1} E[z_i x_i']\right)^{-1}.
\end{aligned}
$$

(iii) Under homoskedasticity we can have $E[u^2 z_i z_i] = \sigma^2 E[z_i z_i']$. This will make the asymptotic variance less horrific:

$$
\mathrm{Avar}\left(\sqrt{n}\left(\hat{\theta} - \theta_0\right)\right) = \sigma^2 \left(E[x_i z_i'] E[z_i z_i']^{-1} E[z_i x_i']\right)^{-1}.
$$

This is the same asymptotic variance as found earlier when we used the efficient weighting matrix W. So yes, this is efficient.

(iv) Without homoskedasticity, we have a different asymptotic variance, and therefore it is not efficient.

Exercise 3

We're using the model $y_i = x_{1i}'\theta_1 + x_{2i}'\theta_2 + u_i$ where

- $\dim(x_{1i}) = k_1$
- $\dim(x_{2i}) = 1$
- $E[u_i|x_{1i}] = 0$
- $E[u_i|x_{2i}] \neq 0$
- $E[u_i|z_{2i}] = 0$
- $\dim(z_{2i}) = 2$
- $E[z_i x_i] \neq 0$.

(i) There are $k+2$ exogenous variables from x_{1i} and z_{2i}, and 1 endogenous regressor from x_{2i}. The model is overidentified because we have one more instrument than endogenous variable.

(ii) We can define $x_i = [x_{1i}' \;\; x_{2i}']'$ and $z_i = [x_{1i}' \;\; z_{2i}']'$.

(iii) The estimator is the same as that which came before, in particular,

$$\hat{\theta} = \left[\sum_{i=1}^{n} x_i z_i' \left(\sum_{i=1}^{n} z_i z_i'\right)^{-1} \sum_{i=1}^{n} z_i x_i'\right]^{-1} \sum_{i=1}^{n} x_i z_i' \left(\sum_{i=1}^{n} z_i z_i'\right)^{-1} \sum_{i=1}^{n} z_i y_i.$$

It's just the same two-stage thing I just did. This is a strange question to have.