ECMT 676: Econometrics II

Homework II - Due date 02/27/2025

Sebastian Sanin-Restrepo*

Exercise 1

$$\hat{\beta} = \left(\mathbb{E}_G[X_g Z_g']\right)^{-1} \left(\mathbb{E}_G[Y_g Z_g]\right)$$

where $\mathbb{E}_G$ represents the average over $g = 1, ..., G$ i.e. $\mathbb{E}_G[X_g Z_g'] = \frac{1}{G}\sum_{g=1}^{G} X_g Z_g'$ and $\mathbb{E}_G[Y_g Z_g'] = \frac{1}{G}\sum_{g=1}^{G} Y_g Z_g'$.

Prove that this estimator is numerically equivalent to the IV estimator considered in the last section.

Exercise 2

Let $X_n = O_p(r_n)$ for a sequence $r_n \to 0$, i.e., that $X_n/r_n = O_p(1)$. Demonstrate that $X_n = o_p(1)$. Using this, show that asymptotic normality of $\sqrt{G}(\hat{\beta} - \beta)$ implies that $\hat{\beta} \to_p \beta$.

First part: Let $X_n = O_p(r_n)$ for a sequence $r_n \to 0$, i.e., that $X_n/r_n = O_p(1)$. I want to show that $X_n = o_p(1)$.

$X_n = O_p(r_n)$ means that for any $\epsilon > 0$, there exists a constant $M > 0$ and an integer N such that for all $n \geq N$,

$$P\left(\left|\frac{X_n}{r_n}\right| > M\right) < \epsilon$$

We want to show that, $X_n = o_p(1)$ means that X_n converges in probability to zero, that is, for every $\epsilon > 0$

$$P(|X_n| > \epsilon) \to 0, \quad n \to \infty$$

Let's start by noticing that the event $\{|X_n| > \epsilon\}$ is equivalent to $\left\{\frac{|X_n|}{r_n} > \frac{\epsilon}{r_n}\right\}$. Using the assumption that $X_n = O_p(r_n)$, then for any $\delta > 0$ there exists a constant $M > 0$ and N such that $\forall n \geq N$,

$$P\left(\frac{|X_n|}{r_n} > M\right) < \delta$$

Since $r_n \to 0$, for large enough n, we have,

$$\frac{\epsilon}{r_n} > M$$

*Economics, Ph.D. Student. E-mail: sr.sebastian@tamu.edu

Therfore, for these large n we obtain,

$$P\left(\frac{|X_n|}{r_n} > \frac{\epsilon}{r_n}\right) \leq P\left(\frac{|X_n|}{r_n} > M\right) < \delta$$

Since, $\delta > 0$ was arbitrarily chose,

$$P(|X_n| > \epsilon) = P\left(\frac{|X_n|}{r_n} > \frac{\epsilon}{r_n}\right) < \delta$$

which is the definition of $X_n = o_p(1)$.

Second part: Using the previous result, I want to show that asymptotic normality of $\sqrt{G}(\hat{\beta} - \beta)$ implies that $\hat{\beta} \to_p \beta$.

By Theorem 1.1. in Nonstandard inference lecture notes we know that $\sqrt{G}\Sigma^{-1/2}(\hat{\beta} - \beta) \to_d N(0, I_{d_x})$, meaning that its probability does not escape to infinity, therefore,

$$\sqrt{G}\Sigma^{-1/2}(\hat{\beta} - \beta) = O_p(1)$$

Since $\Sigma^{-1/2}$ is non-random invertible matrix (by assumption 2), this implies that the scaled difference is also $O_p(1)$, that is

$$\sqrt{G}(\hat{\beta} - \beta) = O_p(1)$$

Now, note that $\sqrt{G}(\hat{\beta} - \beta) = O_p(1)$ implies,

$$\hat{\beta} - \beta = O_p\left(\frac{1}{\sqrt{G}}\right)$$

Since $\frac{1}{\sqrt{G}} \to 0$, as $G \to \infty$, we have a similar result as in $X_n = O_p(r_n)$ with $r_n \to 0$ then $X_n = o_p(1)$. Therefore,

$$\hat{\beta} - \beta = o_p(1)$$

Thus,

$$\hat{\beta} \to_p \beta$$

Exercise 3

The requirement that the number of units per cluster, n_g, is the same across clusters $g = 1, ..., G$ is unnecessarily restrictive. Explain why this (type of) requirement is needed, i.e., what data-generating processes are being ruled out, and suggest a relaxation of the assumption that could still generate the asymptotic normality result.

The assumption of $n_g = n_{g'}\ \forall g, g' \in G$ is a technical convenience. It ensures that when we aggregate the data within each cluster (i.e. when we compute $Z_g = \sum_{i=1}^{n_g} Z_{ig}$), each cluster contributes equally (in terms of scale) to the sample averages. This homogeneity simplifies the derivations for the asymptotic distribution of the estimator.

By using this assumption, we rule out the cases in which we have heterogeneous cluster sizes, which is a common setup when working with real data. This is the case of working with geographic clusters, where we can find very different numbers of individuals in various regions. We also rule out cases in which there are clusters with a dominant influence, that is the case when there are some clusters with very large n_g. In this case, those dominant clusters may affect the estimators, affecting the moment conditions needed for standard asymptotic results.

A possible relaxation is to allow n_g to vary between clusters but impose conditions that control this variation. For example, we would need to assume a constant C exists such that $n_g \leq C \ \forall \ g \in G$. This ensures no cluster is too large, keeping the relative contribution to the estimator comparable. Alternatively, one can assume that the ratio of the sizes between the largest and smallest clusters remains bounded as $G \to \infty$. That is, if:

$$\frac{\max_{1 \leq g \leq G} n_g}{\min_{1 \leq g \leq G} n_g} \leq C$$

Then the variability in cluster sizes does not distort the asymptotic behavior.

Exercise 4

Examining the form of the asymptotic variance in Theorem 1.1. we can see that it looks similar to the variance of the standard IV estimator under non-clustering asymptotics that we derived in the last lecture note. However, "cluster-robust" confidence intervals are typically larger than their non-cluster-robust counterparts. Why is this the case?

Cluster-robust confidence intervals are typically larger because they consider the within-cluster correlation that is ignored under standard non-cluster robust (iid) assumptions. In a non-clustered sample that is *iid*, each observation is independent. In contrast, when data are clustered, observations within a cluster are correlated. Even in cases where we have a large number of observations per cluster $n_g \to \infty$, they do not provide as much independent information as if they were truly independent. As a result, the estimator's variance increases when clustering is taken into account.

Suppose the model is given by

$$Y_{ig} = X'_{ig}\beta + \epsilon_{ig}, \quad i = 1, ..., n_g. \quad g = 1, ..., G.$$

where clusters g are independent, observations within a cluster may be correlated. For simplicity suppose $n_g = n_0 \ \forall \ g \in G$. Assume that the error process within each cluster has the following structure:

$$Var(\epsilon_{ig}) = \sigma^2, \quad Cov(\epsilon_{ig}, \epsilon_{jg}) = \rho\sigma^2 \quad for\ i \neq j$$

with $0 \leq \rho \leq 1$. When $\rho = 0$, the errors are independent; in another case, they are positively or negatively correlated.

Let's consider first the case of nonclustered (iid) data. Then, the variance of the sum of errors in one cluster would be:

$$\text{Var}\left(\sum_{i=1}^{n_0} \epsilon_{ig}\right) = \sum_{i=1}^{n_0} \text{Var}(\epsilon_{ig}) = n_0\sigma^2$$

So, the variance for a sample average across all $N = n_0 \cdot G$ observations will be:

$$\text{Var}\left(\frac{1}{N}\sum_{g=1}^{G}\sum_{i=1}^{n_0} \epsilon_{ig}\right) = \frac{Gn_0\sigma^2}{n_0^2G^2} = \frac{\sigma^2}{n_0G}$$

Now, with clustering, we first aggregate the observations within the cluster $\epsilon_g = \sum_{i=1}^{n_0} \epsilon_{ig}$. Since we have within-cluster correlation, the variance of ϵ_g is given by:

$$\text{Var}(\epsilon_g) = \sum_{i=1}^{n_0} Var(\epsilon_{ig}) + \sum_{i \neq j} Cov(\epsilon_{ig}, \epsilon_{jg}) = n_0\sigma^2 + n_0(n_0-1)\rho\sigma^2 = n_0\sigma^2(1 + \rho(n_0 - 1))$$

If we then average over clusters, the variance of the cluster-level average becomes

$$\text{Var}\left(\frac{1}{G}\sum_{g=1}^{G} \epsilon_g\right) = \frac{1}{G^2}\sum_{g=1}^{G} \text{Var}(\epsilon_g) = \frac{1}{G}\frac{n_0\sigma^2[1 + \rho(n_0 - 1)]}{G}$$

The variability of the cluster-level quantities drives the estimator's asymptotic variance. Comparing this with the non-clustered case where the variance shrinks at the rate $1/N = 1/(n_0G)$, where the variance is **inflated** by a factor $[1 + \rho(n_0 - 1)]$. Even if G is large, if $\rho > 0$, this factor could be substantially more significant than 1.

Now, since $\widehat{\text{Var}}(\hat{\beta}) \propto \widehat{\text{Var}}(\epsilon_g)$, then the confidence intervals computed for $\hat{\beta}$ will be affected by the inflated variance in the clustered case, which will widen them.

Exercise 5

Show that Assumption 1.1. and 1.2. imply the following conditions on the variables (Y_g, X_g, Z_g), there exists a constant $C_g < \infty$ such that $\mathbb{E}[||X_g Z_g^{'}||] \leq C_g$ for all $g = 1, ..., G$ and the eigenvalues of $\mathbb{E}[\epsilon_g^2 Z_g Z_g^{'}]$ are bounded by C_g for all $g = 1, ..., G$. (Hint: recall that $||X + Y|| \leq ||X|| + ||Y||$). Explain how this helps us establish Theorem 1.1.

We want to show that Assumptions 1.1. and 1.2. imply that the aggregated variables

$$\{(Y_g, X_g, Z_g)\}_{g=1}^{G}, \quad \text{with} \quad Y_g = \sum_{i=1}^{n_g} Y_{ig}, \quad X_g = \sum_{i=1}^{n_g} X_{ig}, \quad Z_g = \sum_{i=1}^{n_g} Z_{ig}$$

satisfy appropriate moment conditions, such that there exists a constant $C_g < \infty$ such that,

$$\mathbb{E}[||X_g Z_g'||] \leq C_g \quad \forall \ g = 1, ..., G$$

and, the eigenvalues of $\mathbb{E}[\epsilon_g^2 Z_g Z_g']$ are bounded by C_g $\forall$ $g = 1, ..., G$.

Under Assumption 1.2(ii) we have that for each individual observation $\mathbb{E}[||X_i Z_i'|| \leq C$, for some $C < \infty$. Now, consider the product $X_g Z_g'$. By the triangle inequality, we have,

$$||X_g Z_g'|| \leq \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} ||X_{ig} Z_{jg}'||$$

Taking expectations, we obtain,

$$\mathbb{E}[||X_g Z_g'||] \leq \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \mathbb{E}[||X_{ig} Z_{jg}'||]$$

Since by Assumption 1.2(ii) each $\mathbb{E}[||X_{ig} Z_{jg}'||] < C$, then there exists a constant $C_g < \infty$ such that $\mathbb{E}[||X_g Z_g'||] < C_g$.

Now, by Assumption 1.2(iii), the eigenvalues of $\mathbb{E}[\epsilon_i^2 Z_i Z_i']$ are bounded by some constant C for each observation i. When we have $\epsilon_g = \sum_{i=1}^{n_g} \epsilon_{ig}$ and $Z_g = \sum_{i=1}^{n_g} Z_{ig}$, we have:

$$\mathbb{E}[\epsilon_g^2 Z_g Z_g] = \mathbb{E}\left[\left(\sum_{i=1}^{n_g} \epsilon_{ig}^2\right)\left(\sum_{i=1}^{n_g} Z_{ig}\right)\left(\sum_{i=1}^{n_g} Z_{ig}\right)'\right]$$

Expanding the square gives

$$\left(\sum_{i=1}^{n_0} \epsilon_{ig}\right)^2 = \sum_{i=1}^{n_0} \epsilon_{ig}^2 + 2 \sum_{1 \leq i < j \leq n_0} \epsilon_{ig}\, \epsilon_{jg}$$

Thus, the expectation involves a finite sum (with n_0 fixed) of terms of the form $\epsilon_{ig}\epsilon_{jg}\, Z_{ig} Z_{jg}'$. We use the fact that for any matrix A, the largest eigenvalue is bounded by its operator norm $\|A\|$. Hence,

$$\lambda_{\max}\Big(\mathbb{E}[\epsilon_g^2\, Z_g Z_g']\Big) \leq \Big\|\mathbb{E}[\epsilon_g^2\, Z_g Z_g']\Big\|.$$

By using Jensen's inequality and the triangle inequality, we have

$$\left\|\mathbb{E}[\epsilon_g^2\, Z_g Z_g']\right\| \leq \mathbb{E}\Big[\left\|\epsilon_g^2\, Z_g Z_g'\right\|\Big]$$

Since ϵ_g^2 is a scalar and using the property that for any vector v, $\|vv'\| = \|v\|^2$, we can write

$$\left\|\epsilon_g^2\, Z_g Z_g'\right\| = \epsilon_g^2\, \|Z_g Z_g'\| = \epsilon_g^2\, \|Z_g\|^2.$$

Thus,

$$\left\|\mathbb{E}[\epsilon_g^2\, Z_g Z_g']\right\| \leq \mathbb{E}\left[\epsilon_g^2\, \|Z_g\|^2\right].$$

Since $Z_g = \sum_{i=1}^{n_0} Z_{ig}$ by the triangle inequality we have

$$\|Z_g\| \leq \sum_{i=1}^{n_0} \|Z_{ig}\| \implies \|Z_g\|^2 \leq \left(\sum_{i=1}^{n_0} \|Z_{ig}\|\right)^2.$$

Similarly,

$$\epsilon_g = \sum_{i=1}^{n_0} \epsilon_{ig} \implies |\epsilon_g| \leq \sum_{i=1}^{n_0} |\epsilon_{ig}| \implies \epsilon_g^2 \leq \left(\sum_{i=1}^{n_0} |\epsilon_{ig}|\right)^2.$$

Putting these together,

$$\epsilon_g^2\, \|Z_g\|^2 \leq \left(\sum_{i=1}^{n_0} |\epsilon_{ig}|\right)^2 \left(\sum_{i=1}^{n_0} \|Z_{ig}\|\right)^2$$

Since n_0 is fixed, expanding these sums yields only finitely many terms. By Assumption 1.2(iii), the individual moments $\mathbb{E}\big[\epsilon_{ig}^2\, Z_{ig} Z_{ig}'\big]$ have eigenvalues bounded by some constant, and by a similar moment condition we can assume that

$$\mathbb{E}[\|Z_{ig}\|^2] \leq C_Z \quad \text{and} \quad \mathbb{E}[|\epsilon_{ig}|^2] \leq C_\epsilon,$$

for each i and for some finite constants C_Z, C_ϵ. Thus, by applying Cauchy–Schwarz and the linearity of expectation, we obtain

$$\mathbb{E}\left[\epsilon_g^2\, \|Z_g\|^2\right] \leq C_g,$$

for some constant C_g that depends on n_0, C_Z, and C_ϵ. This later implies that all eigenvalues of $\mathbb{E}[\epsilon_g^2\, Z_g Z_g']$ are bounded by C_g.

Exercise 6

Using what you have established in Exercise 5, propose a consistent estimator for the asymptotic variance σ, $\hat{\Sigma}$. Formally show that $\hat{\Sigma} \to_p \Sigma$. What Theorem do we apply to then show that $\hat{\Sigma}^{-1} \to_p \Sigma^{-1}$ and why does it apply in this setting? How does your variance estimator differ from the variance estimator for the standard iid setting? (Hint: you will need to use an estimator $\hat{\beta}$ of β in place of β. For simplicity, you can ignore estimation error in $\hat{beta}$ when establishing consistency of the variance estimate).

Recall from Theorem 1.1. that the asymptotic variance is given by:

$$\Sigma = (\mathbb{E}[X_g Z_g'])^{-1}\mathbb{E}[\epsilon_g^2 Z_g Z_g'](\mathbb{E}[Z_g X_g'])^{-1}$$

A natural estimator will replace ϵ_g by its sample analog by $\hat{\epsilon}_g = Y_g - X_g'\hat{\beta}$, where $\hat{\beta}$ is a consistent estimator for β (for simplycity we ignore estimation error in $\hat{\beta}$ when establishing consistency of the variance estimate as suggested). Then, we define the estimator $\hat{\Sigma}$ for Σ as follows,

$$\hat{\Sigma} = \left(\frac{1}{G}\sum_{g=1}^{G} X_g Z_g'\right)^{-1} \left(\frac{1}{G}\sum_{g=1}^{G} \hat{\epsilon}_g^2 Z_g Z_g'\right)\left(\frac{1}{G}\sum_{g=1}^{G} Z_g X_g'\right)^{-1}$$

By Assumptions 1.1 and 1.2, we see that the cluster-level moments are well-behaved. Thus, by the Law of Large Numbers, we have,

$$\frac{1}{G}\sum_{g=1}^{G} X_g Z_g' \to_p \mathbb{E}[X_g Z_g']$$

Since data is independent across clusters, we can apply the Continuous Mapping Theorem (CMT) over the previous result using the matrix inversion function (that is, continuous function); we get,

$$\left(\frac{1}{G}\sum_{g=1}^{G} X_g Z_g'\right)^{-1} \to_p (\mathbb{E}[X_g Z_g'])^{-1}$$

Now, by assuming consistency of $\hat{\beta}$, we obtain $\hat{\epsilon}_g \to_p \epsilon_g$, and hence

$$\frac{1}{G}\sum_{g=1}^{G} \hat{\epsilon}_g^2 Z_g Z_g' \to_p \mathbb{E}[\epsilon_g^2 Z_g Z_g']$$

Combining both results, using again the CMT using as a function the matrix multiplication, we obtain,

$$\hat{\Sigma} \to_p \Sigma$$

As mentioned before, this estimator is different from the one in the *iid* case. This is because it uses clustered data, which might include correlation within clustered observations. This potential correlation makes this variance estimator higher than the *iid* case.

Exercise 7

Using the result from above (and in the case where $d_x = d_z = 1$), formally show that $\hat{\beta}$ is inconsistent for β. In particular, for an $\epsilon > 0$ what is $\lim_{n\to\infty} Pr(|\hat{\beta} - \beta| > \epsilon)$?

Recall from the weak instrument asymptotics we obtained,

$$\hat{\beta} - \beta \to_d \frac{\mathbb{Z}_1}{\mathbb{Z}_2}$$

where

$$\mathbb{Z} = \begin{pmatrix} \mathbb{Z}_1 \\ \mathbb{Z}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \Pi \end{pmatrix}, \begin{pmatrix} Var(\epsilon Z) & \mathbb{E}[\epsilon X Z^2] \\ \mathbb{E}[\epsilon X Z^2] & Var(XZ) \end{pmatrix} \right),$$

and $\Pi \neq 0$. To formally show inconsistency, note that an estimator $\hat{\beta}$ is consistent for β if $\forall\ \epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(|\hat{\beta} - \beta| > \epsilon \right) = 0$$

However, using the weak instrument limit we have,

$$\lim_{n\to\infty} \mathbb{P}\left(|\hat{\beta} - \beta| > \epsilon \right) = \mathbb{P}\left(\left| \frac{\mathbb{Z}_1}{\mathbb{Z}_2} \right| > \epsilon \right) > 0$$

This last result is because the distribution of $\frac{\mathbb{Z}_1}{\mathbb{Z}_2}$ is non-degenerate (because the denominator is normally distributed with mean Π and finite variance, hence it does not converge to a constant). So, it does not matter how large the sample size n is; there is always a nonzero probability that $\hat{\beta}$ deviates from β by more than ϵ. So $\hat{\beta}$ is inconsistent for β.

Exercise 8

Using the analysis above, show that for any $M > 0$ and $\epsilon > 0$ we have

$$\lim_{n\to\infty} Pr(|\hat{\beta} - \beta| > \epsilon) \leq Pr(|\mathbb{Z}_1| > \epsilon M) + Pr(|\mathbb{Z}_2| \leq M)$$

For any fixed $M > 0$ and $\epsilon > 0$, consider the event $A = \left\{ \left| \frac{\mathbb{Z}_1}{\mathbb{Z}_2} \right| > \epsilon \right\}$. Notice that this event can be decomposed into two mutually exclusive cases. If $|\mathbb{Z}_2| \leq M$, the event A will occur regardless of $\mathbb{Z}_1$, since the ratio may exceed ϵ regardless of the size of $\mathbb{Z}_1$. So, $\{\mathbb{Z}_2 \leq M\} \subset \left\{ \left| \frac{\mathbb{Z}_1}{\mathbb{Z}_2} \right| > \epsilon \right\}$. In contrast, if $|\mathbb{Z}_2| > M$, then it must be the case that

$$|\mathbb{Z}_1| > \epsilon |Z_2| > \epsilon M$$

Therefore,

$$\left\{ \left| \frac{\mathbb{Z}_1}{\mathbb{Z}_2} \right| > \epsilon \quad \text{and} \quad |\mathbb{Z}_2| > M \right\} \subset \{ |\mathbb{Z}_1| > \epsilon M \}$$

Combining these two cases, we have that the event $\left\{\left|\frac{\mathbb{Z}_1}{\mathbb{Z}_2}\right| > \epsilon\right\}$ is contained in the union $\{|\mathbb{Z}_2| \leq M\} \cup \{|\mathbb{Z}_1| > \epsilon M\}$. Taking probabilities yields,

$$\mathbb{P}\left(\left|\frac{\mathbb{Z}_1}{\mathbb{Z}_2}\right| > \epsilon\right) \leq \mathbb{P}\left(|\mathbb{Z}_2| \leq M\right) + \mathbb{P}\left(|\mathbb{Z}_1| > \epsilon M\right)$$

From Exercise 7, we know that

$$\lim_{n\to\infty} \mathbb{P}\left(|\hat{\beta} - \beta| > \epsilon\right) = \mathbb{P}\left(\left|\frac{\mathbb{Z}_1}{\mathbb{Z}_2}\right| > \epsilon\right)$$

so,

$$\lim_{n\to\infty} \mathbb{P}\left(|\hat{\beta} - \beta| > \epsilon\right) = \mathbb{P}\left(\left|\frac{\mathbb{Z}_1}{\mathbb{Z}_2}\right| > \epsilon\right) \leq \mathbb{P}\left(|\mathbb{Z}_2| \leq M\right) + \mathbb{P}\left(|\mathbb{Z}_1| > \epsilon M\right)$$

Exercise 9

Assume that $\Pi > 0$ and $Var(\epsilon Z) = Var(ZX) = 1$. Show that

$$\lim_{n\to\infty} Pr(|\hat{\beta} - \beta| > \epsilon) \leq 3\Phi\left(-\frac{\epsilon\pi}{1+\epsilon}\right)$$

where $\Phi(\cdot)$ is the CDF of a standard normal random variable. (Hint: use the result in Exercise 8 and set $M = \pi/(1+\epsilon)$.

Recall from Exercise 8 that,

$$\lim_{n\to\infty} \mathbb{P}\left(|\hat{\beta} - \beta| > \epsilon\right) \leq \mathbb{P}\left(|\mathbb{Z}_2| \leq M\right) + \mathbb{P}\left(|\mathbb{Z}_1| > \epsilon M\right)$$

Here $\mathbb{Z}_1 \sim N(0,1)$ since $\text{Var}(\epsilon Z) = 1$ and $\mathbb{Z}_2 \sim N(\Pi, 1)$ since $\text{Var}(ZX) = 1$ (by assumption). Choosing the suggested $M = \frac{\Pi}{1+\epsilon}$. Since $\mathbb{Z}_1 \sim N(0,1)$,

$$\mathbb{P}\left(|\mathbb{Z}_1| > \epsilon M\right) = 2\mathbb{P}\left(|\mathbb{Z}_1| > \epsilon M\right) = 2\Phi(-\epsilon M) = 2\Phi\left(-\frac{\epsilon\Pi}{1+\epsilon}\right)$$

Now, consider $\mathbb{Z}_2 \sim N(\Pi, 1)$, then,

$$\begin{aligned}
\mathbb{P}\left(|\mathbb{Z}_2| \leq M\right) =& \mathbb{P}\left(-M \leq \mathbb{Z}_2 \leq M\right) \\
=& \mathbb{P}\left(\mathbb{Z}_2 \geq -M\right) + \mathbb{P}\left(\mathbb{Z}_2 \leq M\right) \\
=& \Phi\left(M - \Pi\right) - \Phi\left(-M - \Pi\right) \\
=& \Phi\left(\frac{\Pi}{1+\epsilon} - \Pi\right) - \Phi\left(-\frac{\Pi}{1+\epsilon} - \Pi\right) \\
=& \Phi\left(-\frac{\epsilon\Pi}{1+\epsilon}\right) - \Phi\left(-\frac{\Pi(2+\epsilon)}{1+\epsilon}\right)
\end{aligned}$$

Notice that the second term is very small since $-M - \Pi < -\Pi$. So, we can define a convenient bound as,

$$\mathbb{P}\left(|\mathbb{Z}_2| \leq M\right) \leq \Phi\left(-\frac{\epsilon\Pi}{1+\epsilon}\right)$$

Plugging in the previous results for $\mathbb{P}\left(|\mathbb{Z}_2| \leq M\right)$ and $\mathbb{P}\left(|\mathbb{Z}_1| > \epsilon M\right)$ into

$$\lim_{n\to\infty} \mathbb{P}\left(|\hat{\beta} - \beta| > \epsilon\right) \leq \mathbb{P}\left(|\mathbb{Z}_2| \leq M\right) + \mathbb{P}\left(|\mathbb{Z}_1| > \epsilon M\right)$$

we obtain,

$$\lim_{n\to\infty} \mathbb{P}\left(|\hat{\beta} - \beta| > \epsilon\right) \leq \Phi\left(-\frac{\epsilon\Pi}{1+\epsilon}\right) + 2\Phi\left(-\frac{\epsilon\Pi}{1+\epsilon}\right) = 3\Phi\left(-\frac{\epsilon\Pi}{1+\epsilon}\right)$$

Exercise 10

Suppose we want to make sure that the limiting probability that $|\hat{\beta} - \beta|$ exceeds 0.1 is at most 0.1. How large would Π need to be according to the upper bound you derived in the last exercise? Propose a test of whether Π satisfies this requirement.

Using the upper bound found in Exercise 9, we have,

$$\lim_{n\to\infty} \mathbb{P}\left(|\hat{\beta} - \beta| > \epsilon\right) \leq 3\Phi\left(-\frac{\epsilon\Pi}{1+\epsilon}\right)$$

Setting $\epsilon = 0.1$ and requiring that this probability is at most 0.1, we need,

$$\lim_{n\to\infty} \mathbb{P}\left(|\hat{\beta} - \beta| > 0.1\right) \leq 3\Phi\left(-\frac{0.1\Pi}{1+0.1}\right) \leq 0.1$$

We proceed to solve for Π,

$$\begin{aligned} 3\Phi\left(-\frac{0.1\Pi}{1+0.1}\right) \leq 0.1 &\implies 3\left[1 - \Phi\left(\frac{0.1}{1.1}\Pi\right)\right] \leq 0.1 \implies 1 - \Phi\left(\frac{0.1}{1.1}\Pi\right) \leq \frac{0.1}{3} \\ &\implies 1 - \frac{0.1}{3} \leq \Phi\left(\frac{0.1}{1.1}\Pi\right) \implies 0.9666667 \leq \Phi\left(\frac{0.1}{1.1}\Pi\right) \\ &\implies 1.833915 \leq \frac{0.1}{1.1}\Pi \implies 20.17306 \leq \Pi \end{aligned}$$

Therefore, $\Pi \geq 20.17$ is the requirement. So, the proposed test for the instrument to be strong enough (so that the limiting probability of $|\hat{\beta} - \beta| > 0.1$ is lower than 10%, is that $\Pi > 20.17$. The test can be done as follows:

1. In the first-stage regression (for the scalar case), estimate the coefficient on the instrument. Note that in the weak-identification asymptotics, we consider the scaling $\Pi = \sqrt{n}\pi$

2. Construct a confidence interval for Π using the standard error from the first stage.

3. Reject the null hypothesis that $\Pi \geq 20.17$, concluding that the instrument is too weak if the lower bound of the confidence interval is below 20.17.

Exercise 11

Intuitively, why might we expect inconsistency of $\hat{\beta}$ under weak identification? Try to use the language from the top of this section when we are discussing the instrument as exogenously inducing some agents to change their treatment decisions.

Under strong identification, the instrument has a significant enough effect on the treatment, so many individuals in the sample are shifted by the instrument. This movement creates enough variation in the treatment variable X_i to accurately and precisely estimate the causal effect β. In other words, the instrument can shift a substantial portion of the population, and thus, our sample provides rich information about how changes in X_i relate to changes in the outcome Y_i.

By contrast, under weak identification, the instrument only affects the treatment decision for a few individuals—even as the sample size grows. The instrument does not change the treatment status of most of the population. Thus, even in large samples, the variation in X_i induced by the instrument is very small, meaning that the instrument does not shift enough individuals to reliably identify the causal effect.

Intuitively, we can see it as trying to measure the effect of a new policy by looking at only a few people who change their behavior in response to the policy. Even if it is applied exogenously, the fact that only a part of individuals change their behavior means that random fluctuations heavily influence our estimator in those few cases. Thus, the estimator $\hat{\beta}$ does not accommodate to the true value of β as the sample size increases.

Exercise 12

Prove Theorem 2.1. (Hint: recall that if $Z \sim N(0, I_d)$ then $Z'Z \sim \chi^2_d$).

Theorem 2.1: *Suppose that the eigenvalues of $\mathbb{E}[\epsilon_i^2 Z_i Z_i']$ are finite. Then, under H_0, $AR(\beta_0) \to_d \chi^2_{d_z}$.*

$$
\begin{aligned}
AR(\beta_0) &:= n\mathbb{E}_n[\epsilon_i(\beta_0)Z_i'](\mathbb{E}_n[\epsilon_i(\beta_0)Z_iZ_i'])^{-1}\mathbb{E}_n[\epsilon_i(\beta_0)Z_i'] \\
&= \left(\sum_{i=1}^n \epsilon_i(\beta_0)Z_i'\right)\left(\sum_{i=1}^n \epsilon_i(\beta_0)Z_iZ_i'\right)^{-1}\left(\sum_{i=1}^n \epsilon_i(\beta_0)Z_i'\right)
\end{aligned}
$$

where $\epsilon_i(\beta_0) = Y_i - X_i'\beta_0$ are the implied errors under the null hypothesis of $\beta = \beta_0$.

Notice that by the Central Limit Theorem,

$$\sqrt{n}\ \mathbb{E}_n[\epsilon_i(\beta_0)Z_i] \to_d N(0, \Omega), \quad \Omega = \mathbb{E}[\epsilon_i(\beta_0)Z_iZ_i']$$

By the Law of Large Numbers,

$$\mathbb{E}_n[\epsilon_i(\beta_0)Z_iZ_i^{'}] \to_p \mathbb{E}[\epsilon_i(\beta_0)Z_iZ_i^{'}]$$

and by the Continuous Mapping Theorem,

$$(\mathbb{E}_n[\epsilon_i(\beta_0)Z_iZ_i^{'}])^{-1} \to_p (\mathbb{E}[\epsilon_i(\beta_0)Z_iZ_i^{'}])^{-1}$$

So, combining these results and using Slutsky's theorem to replace $(\mathbb{E}_n[\epsilon_i(\beta_0)Z_iZ_i^{'}])^{-1}$ by Ω^{-1}, we obtain,

$$AR(\beta_0) = \left(\sqrt{n}\mathbb{E}_n[\epsilon_i(\beta_0)Z_i]\right)' \Omega^{-1} \left(\sqrt{n}\mathbb{E}_n[\epsilon_i(\beta_0)Z_i]\right)$$

Notice that since the eigen values of $\mathbb{E}[\epsilon_i^2 Z_i Z_i']$ are finite, and that Ω is positive definite, $\Omega^{-1} = \Omega^{-1/2}\Omega^{-1/2}$, we can define,

$$W := \Omega^{-1/2}\sqrt{n}\mathbb{E}_n[\epsilon_i(\beta_0)Z_i]$$

Therefore, we can write the Anderson-Rubin (AR) statistic can be written as,

$$AR(\beta_0) = W'W$$

Since we know that $W \to_d N(0, I_{d_z})$, and the square of a normal distribution follows a chi-square distribution, we have:

$$AR(\beta_0) = W'W \to_d \chi^2_{d_z}$$

Which is the result of Theorem 2.1.

Exercise 13

Show that the restriction on ϵ_i in (3.1) ($\mathbb{E}[\epsilon_i|Z_i] = 0$) implies the restriction $\mathbb{E}[\epsilon_i Z_i] = 0$ from before. In the context of linear regression, where $Z_i = X_i$, how might we interpret the assumption that $\mathbb{E}[\epsilon_i|X_i] = 0$?

We assume that $\mathbb{E}[\epsilon_i|Z_i] = 0$. Consider the expectation of the product $\epsilon_i Z_i$ as $\mathbb{E}[\epsilon_i Z_i]$, and by Law of Iterated Expectations we obtain,

$$\mathbb{E}[\epsilon_i Z_i] = \mathbb{E}[\mathbb{E}[\epsilon_i Z_i|Z_i]] = \mathbb{E}[Z_i\mathbb{E}[\epsilon_i|Z_i]] = \mathbb{E}[Z_i \cdot 0] = 0$$

Therefore, $\mathbb{E}[\epsilon_i|Z_i] = 0 \implies \mathbb{E}[\epsilon_i Z_i] = 0$. In the context of linear regression (when $Z_i = X_i$), we may interpret this assumption $\mathbb{E}[\epsilon_i|Z_i] = 0$, as exogeneity of regressors since it means that the error term ϵ_i has a zero conditional mean given the regressors X_i. This also guarantees the unbiasedness of OLS. Finally, it guarantees that there is no omitted variable bias.

Exercise 14

Show that the above moment condition is implied by the restriction $\mathbb{E}[\epsilon_i|Z_i] = 0$. The moment condition is $\mathbb{E}[(Y_i - X_i'\beta)\tilde{Z}_i] = 0$, where $\tilde{Z}_i = \frac{\mathbb{E}[X_i|Z_i]}{\sigma^2(z_i)}$, and $\sigma^2(z_i) = \mathbb{E}[\epsilon_i^2|Z_i]$.

Starting with the model $Y_i = X_i'\beta + \epsilon_i$, $\mathbb{E}[\epsilon_i|Z_i] = 0$. We want to show that $\mathbb{E}[(Y_i - X_i'\beta)\tilde{Z}_i] = 0$, where $\tilde{Z}_i = \frac{\mathbb{E}[X_i|Z_i]}{\sigma^2(z_i)}$, and $\sigma^2(z_i) = \mathbb{E}[\epsilon_i^2|Z_i]$.

Notice that the moment condition can be written as,

$$\mathbb{E}[(Y_i - X_i'\beta)\tilde{Z}_i] = \mathbb{E}[\epsilon_i \tilde{Z}_i]$$

Under the assumption that $\mathbb{E}[\epsilon_i|Z_i] = 0$, and noting that $\tilde{Z}_i$ is a function of Z_i since both $\mathbb{E}[X|Z]$ and $\sigma^2(Z_i)$ are functions of Z_i, we can use the Law of Iterated Expectations to obtain,

$$\mathbb{E}[\epsilon_i \tilde{Z}_i] = \mathbb{E}[\mathbb{E}[\epsilon_i \tilde{Z}_i|Z_i]] = \mathbb{E}[\tilde{Z}_i \mathbb{E}[\epsilon_i|Z_i]] = \mathbb{E}[\tilde{Z}_i \cdot 0] = 0$$

Therefore, the moment condition $\mathbb{E}[(Y_i - X_i'\beta)\tilde{Z}_i] = 0$ is implied by $\mathbb{E}[\epsilon_i|Z_i] = 0$.

Exercise 15

Verify that the estimator in (3.5) is equivalent to the two-stage least squares estimator. (Hint: you can use the first order conditions of OLS to show that $\mathbb{E}_n[X_i\hat{Z}_i'] = \mathbb{E}_n[\hat{Z}_i\hat{Z}_i']$). The estimator in (3.5) is $\hat{\beta} = (\mathbb{E}_n[X_i\hat{Z}_i'])^{-1}(\mathbb{E}_n[Y_i\hat{Z}_i'])$, where $\hat{Z}_i = Z_i\hat{\Pi}$, coming from the OLS estimation of $\mathbb{E}[X_i|Z_i] = Z_i'\Pi$.

The estimator from (3.5) is given by,

$$\hat{\beta} = (\mathbb{E}_n[X_i\hat{Z}_i'])^{-1}(\mathbb{E}_n[Y_i\hat{Z}_i'])$$

for which the estimated optimal instrument is,

$$\hat{Z}_i = Z_i\hat{\Pi}$$

obtained by regressing $\mathbb{E}[X_i|Z_i] = Z_i'\Pi$. In the classical 2SLS framework, the firs stage provides the fitted values $\hat{X} = Z_i\hat{\Pi}$, which are then used in the regression of Y_i on $\hat{X}_i$.

The first-order condition for the OLS estimation of Π implies that the residuals are,

$$\hat{u}_i = X_i - Z_i'\hat{\Pi},$$

then the FOC give

$$\mathbb{E}_n[Z_i \hat{u}_i'] = \mathbb{E}_n \left[Z_i \left(X_i - Z_i' \hat{\Pi} \right)' \right] = 0$$

Because $\hat{Z}_i = Z_i \hat{\Pi}$ is a linear function of Z_i, we have,

$$\mathbb{E}_n \left[Z_i \left(X_i - Z_i' \hat{\Pi} \right)' \right] = 0 \implies \mathbb{E}_n \left[Z_i \left(X_i - \hat{Z}_i \right)' \right] = 0 \implies \mathbb{E}_n \left[Z_i X_i' - Z_i \hat{Z}_i{}' \right] = 0$$

Therefore,

$$\mathbb{E}_n \left[Z_i X_i' \right] = \mathbb{E}_n \left[Z_i \hat{Z}_i{}' \right]$$

Premultiplying by $\hat{\Pi}$ both sides,

$$\hat{\Pi} \; \mathbb{E}_n \left[Z_i X_i' \right] = \mathbb{E}_n \left[Z_i \hat{Z}_i{}' \right] \; \hat{\Pi} \implies \mathbb{E}_n \left[X_i \hat{Z}_i' \right] = \mathbb{E}_n \left[\hat{Z}_i \hat{Z}_i{}' \right]$$

Substituting the above result into the expression for $\hat{\beta}$,

$$\hat{\beta} = \left(\mathbb{E}_n [X_i \hat{Z}_i'] \right)^{-1} \mathbb{E}_n [Y_i \hat{Z}_i'] = \left(\mathbb{E}_n [\hat{Z}_i \hat{Z}_i{}'] \right)^{-1} \mathbb{E}_n [Y_i \hat{Z}_i']$$

This is the standard form of the 2SLS estimator, where the second stage uses the fitted values $\hat{X}_i = \hat{Z}_i$. In 2SLS, one typically writes the estimator as

$$\hat{\beta}_{2SLS} = (X P_Z X')^{-1} (X' P_Z Y),$$

where P_Z is a matrix such that $P_Z = P_Z P_Z$, then $X_i P_Z X_i' = X_i P_Z P_Z X_i' = \hat{Z}_i \hat{Z}_i'$, and $X_i' P_Z Y_i = Y_i \hat{Z}_i'$, so,

$$\hat{\beta}_{2SLS} = (X P_Z X')^{-1} (X' P_Z Y) = \left(\hat{Z}_i \hat{Z}_i' \right)^{-1} (Y_i \hat{Z}_i')$$

Exercise 16

In the previous lecture note we established that two stage least squares was efficient under homoskedasticity. Here, two stage least squares is efficient under homoskedasticity and one extra assumption. What is the extra assumption and what makes this setting different than the setting we considered in the previous note?

In the previous lecture note, the efficiency of two-stage least squares (2SLS) was derived under the assumption that the instruments are uncorrelated with the error, i.e., $\mathbb{E}[\epsilon_i Z_i] = 0$. Now, we impose a stronger assumption, $\mathbb{E}[\epsilon_i | Z_i] = 0$. The stronger conditional moment restriction allows the construction of the optimal instrument, which leads to an efficient estimator under homoskedasticity. Under homoskedasticity $\sigma^2(Z_i)$ is constant, and the optimal instrument simplifies to $\mathbb{E}[X_i | Z_i]$. This is different from the earlier setting where only the unconditional moment condition was used, and the optimal instrument was not explicitly constructed. This setting is used because we may have many instruments or we seek to incorporate machine learning techniques for the first-stage estimation. The focus is on exploiting a more precise structure of the relationship between X_i and Z_i to achieve efficiency. In contrast, the previous note typically considered a simpler framework with a less restrictive error assumption.

Exercise 17

Suppose that $s^2 log^M(d_z n)/n \to 0$. Show that the result in Theorem 3.1 implies that $||\hat{\Pi} - \Pi||_1 = o_p(1)$

By Theorem 3.1, we know that with probability approaching one there exists a constant $C > 0$ such that,

$$\|\hat{\Pi} - \Pi\|_1 \leq C \frac{s^2 log^M(d_z n)}{n}$$

Now, assuming that $s^2 log^M(d_z n)/n \to 0$ as $n \to \infty$, meaning that the upper bound on the ℓ_1-norm error shrinks to zero. This implies that for large n, the quantity $C \frac{s^2 log^M(d_z n)}{n}$ becomes arbitrarily small.

Recall that the definition of $o_p(1)$ is that X_n is $o_p(1)$ if for every $\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}(|X_n| > \epsilon) = 0$. In this context, $X_n = \|\hat{\Pi} - \Pi\|_1$, we want to show that for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(\|\hat{\Pi} - \Pi\|_1 > \epsilon) = 0$$

Given that with probability approaching one,

$$\|\hat{\Pi} - \Pi\|_1 \leq C \frac{s^2 log^M(d_z n)}{n}$$

and since by assumption $s^2 log^M(d_z n)/n \to 0$, then for any $\epsilon > 0$ there exists an N such that for all $n > N$,

$$C \frac{s^2 log^M(d_z n)}{n} < \epsilon$$

So, we know that with probability approaching one,

$$\|\hat{\Pi} - \Pi\|_1 \leq C \frac{s^2 log^M(d_z n)}{n} < \epsilon$$

This directly implies that,

$$\mathbb{P}\left(\|\hat{\Pi} - \Pi\|_1 \right\rangle \epsilon) \to 0$$

Therefore, $\|\hat{\Pi} - \Pi\|_1 = o_p(1)$

Exercise 18

Derive the limiting distribution of $\sqrt{n}(\tilde{\beta}-\beta)$. What assumptions do you need? Use this, along with the fact that the limiting distribution of $\sqrt{n}(\tilde{\beta}-\beta)$ is the same as that of $\sqrt{n}(\hat{\beta}-\beta)$ to propose a feasible for the null hypothesis $H_0 : \beta = 5$ against an alternative $H_1 : \beta \neq 5$. When proposing your feasible testing procedure keep in mind that $\tilde{Z}_i$ is not known.

We begin with the "infeasible" estimator constructed using the optimal instrument, given by,

$$\tilde{\beta} = (\mathbb{E}_n[X_i\tilde{Z}_i'])^{-1}(\mathbb{E}_n[Y_i\tilde{Z}_i])$$

where $\tilde{Z}_i = \mathbb{E}[X_i|Z_i]$. The structural equation was given by,

$$Y_i = X_i\beta + \epsilon_i, \quad \mathbb{E}[\epsilon_i|Z_i] = 0$$

Recall that we showed before that since $\tilde{Z}_i$ is a function of Z_i, it follows that $\mathbb{E}[\epsilon_i\tilde{Z}_i] = 0$.

We can rewrite $\tilde{\beta}$ as follows,

$$\begin{aligned}
\tilde{\beta} =& (\mathbb{E}_n[X_i\tilde{Z}_i'])^{-1}(\mathbb{E}_n[X_i\beta\tilde{Z}_i] + \mathbb{E}_n[\epsilon\tilde{Z}_i])\\
\tilde{\beta} =& \beta + (\mathbb{E}_n[X_i\tilde{Z}_i'])^{-1}\mathbb{E}_n[\epsilon\tilde{Z}_i]\\
\sqrt{n}(\tilde{\beta}-\beta) =& (\mathbb{E}_n[X_i\tilde{Z}_i'])^{-1}\sqrt{n}\mathbb{E}_n[\epsilon\tilde{Z}_i]
\end{aligned}$$

Using the standard assumptions such as $\{(Y_i, X_i, Z_i)\}$ are i.i.d, finite second moments of ϵ_i and $\tilde{Z}_i$, and $\mathbb{E}[X_i\tilde{Z}_i']$ is invertible, then the sample average

$$\sqrt{n}\mathbb{E}_n[\epsilon\tilde{Z}_i] \to_d N(0,\Omega) \quad \Omega = \mathbb{E}[\epsilon_i^2\tilde{Z}_i\tilde{Z}_i']$$

by the Central Limit Theorem. Since $\mathbb{E}_n[X_i\tilde{Z}_i'] \to_p \mathbb{E}[X_i\tilde{Z}_i']$, by Slutksy's theorem we have,

$$\sqrt{n}(\tilde{\beta}-\beta) \to_d N(0,\Sigma) \quad \Sigma = (\mathbb{E}[X_i\tilde{Z}_i'])^{-1}\mathbb{E}[\epsilon_i^2\tilde{Z}_i\tilde{Z}_i'](\mathbb{E}[\tilde{Z}_iX_i'])^{-1}$$

To construct a feasible test for $H_0 : \beta = 5$, we need the established distribution for $\tilde{\beta}$. Since $\sqrt{n}(\tilde{\beta}-\beta) \to_d N(0,\Sigma)$, and it is shown in a previous exercise that the estimator $\hat{\beta}$ obtained by plugging in an estimated instrument $\hat{Z}_i$ has the same limiting distribution,

$$\sqrt{n}(\hat{\beta}-\beta) \to_d N(0,\Sigma)$$

However, $\tilde{Z}_i = \mathbb{E}[X_i|Z_i]$ is not known in practice. Instead we estimate it by OLS, LASSO or Post-LASSO to obtain $\hat{\beta}$. So, a **feasible** procedure testing is:

1. Choose your prefered method (OLS, LASSO, post-LASSO) to estimate Π in the first-stage for $\mathbb{E}[X_i|Z_i] \sim Z_i'\hat{\Pi}$, and set $\hat{Z}_i = Z_i'\hat{\Pi}$.

2. Compute the second stage estimator

$$\hat{\beta} = \left(\frac{1}{n}\sum_{i=1}^{n} X_i \hat{Z}_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} Y_i \hat{Z}_i\right)$$

3. Estimate the variance matrix Σ by $\hat{\Sigma}$,

$$\hat{\Sigma} = \left(\frac{1}{n}\sum_{i=1}^{n} X_i \hat{Z}_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_i^2 \hat{Z}_i \hat{Z}_i'\right) \left(\frac{1}{n}\sum_{i=1}^{n} \hat{Z}_i X_i'\right)^{-1}$$

   where $\hat{\epsilon}_i = Y_i - X_i'\hat{\beta}$.

4. Form the $t-$ statistic

$$t = \frac{\sqrt{n}(\hat{\beta} - 5)}{\hat{\sigma}}$$

5. Under H_0, the t-statistic is asymptotically $N(0,1)$. Reject at the 5% if $|t| > z_{0.975}$.

Exercise 19

Show that $\sqrt{n}(\tilde{\beta} - \beta) \to_d Z$ along with $\sqrt{n}(\hat{\beta} - \tilde{\beta}) = o_p(1)$ implies that $\sqrt{n}(\hat{\beta} - \beta) \to_d Z$.

Starting from $\sqrt{n}(\hat{\beta} - \beta)$ we can add and subtract $\sqrt{n}\tilde{\beta}$ and obtain,

$$\sqrt{n}(\hat{\beta} - \beta) + \sqrt{n}\tilde{\beta} - \sqrt{n}\tilde{\beta} = \sqrt{n}(\hat{\beta} - \tilde{\beta}) + \sqrt{n}(\tilde{\beta} - \beta)$$

Notice that by assumption $\sqrt{n}(\tilde{\beta} - \beta) \to_d Z$ and $\sqrt{n}(\hat{\beta} - \tilde{\beta}) = o_p(1)$. By Slutsky's theorem, we have:

$$\sqrt{n}(\hat{\beta} - \beta) + \sqrt{n}\tilde{\beta} - \sqrt{n}\tilde{\beta} = \sqrt{n}(\hat{\beta} - \tilde{\beta}) + \sqrt{n}(\tilde{\beta} - \beta) \to_d Z + 0 = Z$$

Thus, $\sqrt{n}(\tilde{\beta} - \beta) \to_d Z$.