Instrumental Variables

Previously, the OLS estimator was shown to be consistent under the predetermination assumption, $\mathbb{E}(u_t|\mathbf{X}_t)$. As we also saw with LLN and CLT theorems, this condition can also be expressed by saying that the error terms u_t are innovations. When the predetermined condition does not hold, the consistency proof that we saw when finding the probability limit of $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{u}$ is not applicable, and in general, the OLS estimator is both biased and inconsistent.

It is not always reasonable to assume that the error terms are innovations. In fact, as we will see in the next section, there are commonly encountered situations in which the error terms are necessarily correlated with some of the regressors for the same observation. Such examples include:

1. Error in variables;
2. Simultaneous equations;
3. Omitted variables;
4. Lag dependent variable with autoregressive errors.

Even in these circumstances, however, it is usually possible, although not always easy, to define an information set Ω_t for each observation such that:

$$\mathbb{E}(u_t|\Omega_t) = \mathbf{0}. \tag{1}$$

i.e. Any regressor of which the value in period t is correlated with u_t, cannot belong to Ω_t.

Although this course didn't take particular focus into it, we briefly discussed method-of-moments (MM) estimators for linear regression models. Such estimators are defined by the moment condition in terms of a matrix $\mathbf{W}$ of variables, with one row for each observation:

$$\mathbf{W}^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}$$

They were shown to be consistent provided that the t^{th} row $\mathbf{W}_t$ of $\mathbf{W}$ belongs to Ω_t, and provided that an asymptotic identification condition is satisfied. In econometrics, these MM estimators are usually called Instrumental Variables estimators, or IV estimators,

1 Errors in Variables

For a variety of reasons, many economics variables are measured with error. For example, macroeconomics time series are often based on surveys, and they must therefore suffer from sampling variability. Whenever there are measurement errors, the values economists observe inevitably differ from the true values that economic agents presumably act upon. As we will see, errors in the dependent variable, so long as they are not too large, are not a big issue. However, errors in the independent variables cause the error terms to be correlated with the regressors that are measured with error, and this causes OLS to be inconsistent.

The problems caused by errors in variables can be seen quite clearly. Consider:

$$\tilde{y}_t = \beta_1 + \beta_2 \tilde{x}_t + \tilde{u}_t,\ \tilde{u}_t \sim \text{IID}(0, \sigma^2), \tag{2}$$

where the variables $\tilde{x}_t$ and $\tilde{y}_t$ are not actually observed. Instead, we observe:

$$\begin{aligned} x_t &= \tilde{x}_t + v_{1t} \\ y_t &= \tilde{y}_t + v_{2t}. \end{aligned}$$

Here v_{1t} and v_{2t} are measurement errors which are assumed, perhaps not realistically in some cases, to be IID with variances ω_1^2 and ω_2^2, respectively, and independent of the "real" dependent, independent and error terms.

If we suppose that the true DGP is a special case of (2) along with the above variables, we see that:

$$\begin{aligned} \tilde{x}_t &= x_t - v_{1t} \\ \tilde{y}_t &= y_t + v_{2t}. \end{aligned}$$

If we substitute these into (2), we get:

$$\begin{aligned} y_t &= \beta_1 + \beta_2(x_t - v_{1t}) + \tilde{u}_t + v_{2t} \\ &= \beta_1 + \beta_2 x_t + \tilde{u}_t + v_{2t} - \beta_2 v_{1t} \\ &= \beta_1 + \beta_2 x_t + u_t, \end{aligned}$$

where

$$u_t = \tilde{u}_t + v_{2t} - \beta_2 v_{1t}.$$

Thus the variance of this new error term is:

$$\text{Var}(u_t) = \sigma^2 + \omega_2^2 + \beta_2^2 \omega_1^2.$$

The effect of the measurement error in the dependent variable is simply to increase the variance of the error terms. Unless the increase is substantial, this generally not a serious problem.

The measurement error in the independent variable also increases the variance of the error terms, but it has another, much more serious consequence as well. Because $x_t = \tilde{x}_t + v_{1t}$, and u_t depends on v_{1t}, u_t must be correlated with x_t whenever $\beta_2 \neq 0$. In fact, since the random part of x_t is v_{1t}, we see that:

$$\mathbb{E}(u_t | x_t) = \mathbb{E}(u_t | v_{1t}) = -\beta_2 v_{1t}, \tag{3}$$

because we assume v_{1t} is independent of $\tilde{u}_t$ and v_{2t}. From (3) we can see, using the fact that $\mathbb{E}(u_t) = 0$ unconditionally, that:

$$\begin{aligned}\text{Cov}(x_t, u_t) &= \mathbb{E}(x_t u_t) = \mathbb{E}(x_t \mathbb{E}(u_t|x_t)) \\ &= -\mathbb{E}((\tilde{x}_t + v_{1t})\beta_2 v_{1t}) = -\beta_2 \omega_1^2.\end{aligned}$$

This covariance is negative if $\beta_2 > 0$ and it is positive for when $\beta_2 < 0$, and, since it does not depend on the sample size n, it does not go away as n becomes large. An exactly similar argument shows that the assumption that $E(u_t|\mathbf{X}_t) = 0$ is false whenever any element of $\mathbf{X}_t$ is measured with error. In consequence, the OLS estimator is biased and inconsistent.

2 Simultaneous Equations

Economic theory often suggest that two or more endogenous variables are determined simultaneously. In this situation, as we will see shortly, all of the endogenous variables must necessarily be correlated with the error terms in all of the equations. This means that none of them may validly appear in the regression functions of models that are to be estimated by OLS.

A classic example is the demand-supply problem of a certain type of good, both of which would be in log form. A linear (or log-linear) model of demand and supply is:

$$Q_t = \gamma_d P_t + \mathbf{X}_t^d \boldsymbol{\beta}_d + u_t^d \tag{4}$$

$$Qt = \gamma_s P_t + \mathbf{X}_t^s \boldsymbol{\beta}_s + u_t^s. \tag{5}$$

Where (4) gives us the demand side relation and (5) gives us the supply side relation. Here, $\mathbf{X}_t$ are row vectors of observations on exogenous or predetermined variables that appear in the demand and supply function. Economic theory also tells is that usually, $\gamma_d < 0$ and $\gamma_s > 0$, which is equivalent in saying that the demand curve slopes downward and the supply slopes upward.

To solve, it is straightforward to solve (4) and (5) by rewriting them in matrix notation as:

$$\begin{bmatrix} 1 & -\gamma_d \\ 1 & -\gamma_s \end{bmatrix} \begin{bmatrix} Q_t \\ P_t \end{bmatrix} = \begin{bmatrix} \mathbf{X}_t^d \boldsymbol{\beta}_d \\ \mathbf{X}_t^s \boldsymbol{\beta}_s \end{bmatrix} + \begin{bmatrix} u_t^d \\ u_t^s \end{bmatrix}.$$

The solution exists whenever $\gamma_d \neq \gamma_s$, so that the matrix on the LHS, above, is nonsingular:

$$\begin{bmatrix} Q_t \\ P_t \end{bmatrix} = \begin{bmatrix} 1 & -\gamma_d \\ 1 & -\gamma_s \end{bmatrix}^{-1} \left(\begin{bmatrix} \mathbf{X}_t^d \boldsymbol{\beta}_d \\ \mathbf{X}_t^s \boldsymbol{\beta}_s \end{bmatrix} + \begin{bmatrix} u_t^d \\ u_t^s \end{bmatrix} \right)$$

or:

$$\begin{cases} Q_t = & \frac{1}{\gamma_s - \gamma_d}(\gamma_s \mathbf{X}_t^d \boldsymbol{\beta}_d - \gamma_d \mathbf{X}_t^s \boldsymbol{\beta}_s + \gamma_s u_t^d - \gamma_d u_t^s) \\ P_t = & \frac{1}{\gamma_s - \gamma_d}(\mathbf{X}_t^d \boldsymbol{\beta}_d - \mathbf{X}_t^s \boldsymbol{\beta}_s + u_t^d - u_t^s). \end{cases}$$

We can also see, taking the covariance between price and the error terms of demand and supply, that:

$$\begin{cases} \mathbb{E}(P_t u_t^d) = & \frac{1}{\gamma_s - \gamma_d}\mathbb{E}((u_t^d)^2) = \frac{\sigma^2_{u^d}}{\gamma_s - \gamma_d} \\ \mathbb{E}(P_t u_t^s) = & -\frac{1}{\gamma_s - \gamma_d}\mathbb{E}((u_t^s)^2) = -\frac{\sigma^2_{u^s}}{\gamma_s - \gamma_d}. \end{cases} \tag{6}$$

Applying OLS separately in each equation will produce inconsistent estimators for the parameters of the demand and supply functions. We can see that price and quantity depend on the both of the error terms in the demand and supply relations, and on every exogenous and predetermined variable that appears in either the demand function, supply function, or both. Therefore, price, which appears on the RHS of (4) and (5), must be correlated with the error terms in both of those equations.

We have covered two cases where our exogeneity and predetermination assumptions have been violated. See lecture notes for omitted variables and time series models with a lagged dependent variable.

3 Instrumental Variable Estimation

Consider the following linear regression model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u},\ \mathbb{E}(\mathbf{X}^\top \mathbf{u}) \neq \mathbf{0},\ \mathbb{E}(\mathbf{u}\mathbf{u}^\top) = \sigma^2 \mathbf{I} \tag{7}$$

Where $\mathbf{y}$ and $\mathbf{u}$ are $n \times 1$ vectors and $\mathbf{X}$ is an $n \times k$ matrix of explanatory variables. For a slightly lighter assumption, we could expect that at least one of the explanatory variables in X is assumed not to be predetermined with respect to the error terms. Suppose that, for each $t = 1, 2, \dots, n$, condition (1) is satisfied for some suitable information set Ω_t, and that we can form an $n \times k$ matrix $\mathbf{W}$ with typical row $\mathbf{W}_t$ such that all its elements belong to Ω_t. The k variables given by the k columns of $\mathbf{W}$ are called instrumental variables. Let $\mathbf{W}_t$ be a $1 \times k_w$ random variable with $k_w \geq k_x$. Note that the non-observable component u_t satisfies:

$$\begin{aligned} \mathbb{E}(u_t | W_t) &= 0 \qquad (8) \\ \mathbb{E}(u_t^2 | W_t) &= \sigma_u^2. \qquad (9) \end{aligned}$$

These assumptions allow us to derive:

$$\mathbb{E}(\mathbf{W}^\top \mathbf{u}) = \mathbb{E}[\mathbf{W}_t^\top \mathbb{E}(u_t | \mathbf{W}_t)] = 0$$

,and:

$$\mathrm{Var}(\mathbf{W}_t^\top u_t) = \mathbb{E}[\mathbb{E}(u_t^2 | \mathbf{W}_t)\mathbf{W}_t^\top \mathbf{W}_t] = \sigma_u^2 \mathbb{E}(\mathbf{W}_t^\top \mathbf{W}_t).$$

From here, we can derive IV estimator in two ways: Through a MM estimation technique, or by using projection matrices:

The MM method:
The moment conditions simplify to:

$$\mathbf{W}^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}.$$

Since there are k equations and k unknowns, we can solve the above problem directly to obtain the simple IV estimator:

$$\hat{\boldsymbol{\beta}}_{\text{IV}} = (\mathbf{W}^\top\mathbf{X})^{-1}\mathbf{W}^\top\mathbf{y}. \tag{10}$$

3.1 Generalised Instrumental Variable Estimation

Alternatively, we could derive the IV estimator using a geometric approach (or two stage least squares approach). Consider the following $n \times n$ projection matrix:

$$\mathbf{P}_{\mathbf{W}} = \mathbf{W}(\mathbf{W}^\top\mathbf{W})^{-1}\mathbf{W}^\top.$$

We know that the IV estimator solves the following problem:

$$\hat{\boldsymbol{\beta}}_{\text{IV}} = \arg\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_{\mathbf{P}_{\mathbf{W}}} = \arg\min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top\mathbf{P}_{\mathbf{W}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

The first order condition (FOC):

$$-2\mathbf{X}^\top\mathbf{P}_{\mathbf{W}}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{0}, \tag{11}$$

and so the Generalised IV Estimator (sometimes referred to as GIVE) is:

$$\hat{\boldsymbol{\beta}}_{\text{IV}} = (\mathbf{X}^\top\mathbf{P}_{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{P}_{\mathbf{W}}\mathbf{y}. \tag{12}$$

Note from equation (11) $\hat{\boldsymbol{\beta}}_{\text{IV}}$ is the OLS estimator of the following auxiliary regression:

$$\mathbf{y} = \hat{\mathbf{X}}\boldsymbol{\beta} + \text{residuals},$$

where:

$$\hat{\mathbf{X}} = \mathbf{P}_{\mathbf{W}}\mathbf{X} = \mathbf{W}\underbrace{(\mathbf{W}^\top\mathbf{W})^{-1}\mathbf{W}^\top\mathbf{X}}_{\hat{\boldsymbol{\Pi}}},$$

and $\hat{\boldsymbol{\Pi}}$ is the OLS estimator of the following model:

$$\mathbf{X} = \mathbf{W}\boldsymbol{\Pi} + \mathbf{V}.$$

Thus we often say that $\hat{\boldsymbol{\beta}}_{\text{IV}}$ is also known as a two-stage least squares estimator (2SLS).

When $k_w = k_x$, then

$$\mathbf{W}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{IV}}) = \mathbf{0},$$

and we say that the system of equations in (11) is exactly identified. And we can easily think of the GIVE estimator as being the simple IV estimator based on MM estimation.

If $k_w > k_x$, then we say the model is overidentified:

$$\mathbf{W}^\top(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{IV}}) \neq \mathbf{0},$$

since we have more equations than parameters to be estimated. However, if $\mathbf{W}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{IV}})$ is very different from 0 in a statistical sense, it must be that $\mathbb{E}(\mathbf{W}^\top\mathbf{u}) = \mathbf{0}$ is not valid. So we must test if $\mathbb{E}(\mathbf{W}^\top\mathbf{u}) = \mathbf{0}$.

Therefore we drop the assumption that the number of instruments is equal to the number of parameters and let $\mathbf{W}$ denote an $n \times l$ matrix of instruments. Often, l is greater than k, the number of regressors.

4 Large Sample Properties of the IV Estimator

Consider:

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{\text{IV}} &= \boldsymbol{\beta} + (\mathbf{X}^\top\mathbf{P}_{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{P}_{\mathbf{W}}\mathbf{u} &(13)\\
\hat{\boldsymbol{\beta}}_{\text{IV}} - \boldsymbol{\beta} &= (\mathbf{X}^\top\mathbf{P}_{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{P}_{\mathbf{W}}\mathbf{u}, &(14)
\end{aligned}$$

and with our assumptions:

$$\begin{aligned}
\mathbb{E}(u_t|\mathbf{W}_t) &= 0\\
\mathbb{E}(u_t^2|\mathbf{W}_t) &= \sigma_u^2
\end{aligned}$$

Combine this with the LLN we find:

$$\operatorname*{plim}_{n\to\infty}\frac{1}{n}\mathbf{W}^\top\mathbf{u} = \operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathbf{W}_t u_t = 0$$

which we refer to as asymptotic uncorrelation. And, by the CLT we have:

$$\begin{aligned}
\frac{1}{\sqrt{n}}\mathbf{W}^\top\mathbf{u} &\overset{a}{\sim} \mathrm{N}\left(0, \lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left(\sum_{t=1}^{n}u_t^2\mathbf{W}_t^\top\mathbf{W}_t\right)\right)\\
&\overset{a}{\sim} \mathrm{N}(0, \sigma_u^2\mathbf{S}_{\mathbf{W}^\top\mathbf{W}}).
\end{aligned}$$

This is because:

$$\begin{aligned}
\operatorname*{plim}_{n\to\infty}\frac{1}{n}\mathbf{W}^\top\mathbf{W} &= \operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathbf{W}_t^\top\mathbf{W}_t\\
&= \mathbf{S}_{\mathbf{W}^\top\mathbf{W}},
\end{aligned}$$

and we know from our LLN and CLT analysis that this matrix is deterministic and non-singular.

For future reference, let's also assume that

$$\begin{aligned}
\operatorname*{plim}_{n\to\infty}\frac{1}{n}\mathbf{X}^\top\mathbf{W} &= \operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathbf{X}_t^\top\mathbf{W}_t \\
&= \mathbf{S}_{\mathbf{X}^\top\mathbf{W}},
\end{aligned}$$

which is also deterministic and non-singular.

Let us focus on $(\mathbf{X}^\top\mathbf{P}_\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{P}_\mathbf{W}\mathbf{u}$ and its two components:

$$\frac{1}{n}\mathbf{X}^\top\mathbf{P}_\mathbf{W}\mathbf{X} = \frac{\mathbf{X}^\top\mathbf{W}}{n}\left(\frac{\mathbf{W}^\top\mathbf{W}}{n}\right)^{-1}\frac{\mathbf{W}^\top\mathbf{X}}{n} \xrightarrow{p} \mathbf{S}_{\mathbf{X}^\top\mathbf{W}}\mathbf{S}_{\mathbf{W}^\top\mathbf{W}}^{-1}\mathbf{S}_{\mathbf{W}^\top\mathbf{X}},$$

and

$$\frac{1}{n}\mathbf{X}^\top\mathbf{P}_\mathbf{W}\mathbf{u} = \frac{\mathbf{X}^\top\mathbf{W}}{n}\left(\frac{\mathbf{W}^\top\mathbf{W}}{n}\right)^{-1}\frac{\mathbf{W}^\top\mathbf{u}}{n} \xrightarrow{p} \mathbf{S}_{\mathbf{X}^\top\mathbf{W}}\mathbf{S}_{\mathbf{W}^\top\mathbf{W}}^{-1}\underbrace{\mathbf{S}_{\mathbf{W}^\top\mathbf{u}}}_{=0},$$

as $n\to\infty$. Therefore, we can see that the IV estimator is consistent.

4.0.1 Deriving the Variance of the IV Estimator

We first find the asymptotic distribution of $\sqrt{n}(\widehat{\beta}_{IV}-\beta)$, which can be rewritten as:

$$\begin{aligned}
\sqrt{n}(\widehat{\boldsymbol{\beta}}_{\text{IV}}-\boldsymbol{\beta}) &= (\frac{1}{n}\mathbf{X}^\top\mathbf{P}_\mathbf{W}\mathbf{X})^{-1}\frac{1}{\sqrt{n}}\mathbf{X}^\top\mathbf{P}_\mathbf{W}\mathbf{u} \\
&= (\frac{1}{n}\mathbf{X}^\top\mathbf{P}_\mathbf{W}\mathbf{X})^{-1}\frac{1}{\sqrt{n}}\mathbf{X}^\top\mathbf{W}(\mathbf{W}^\top\mathbf{W})^{-1}\mathbf{W}^\top\mathbf{u} \\
&= (\frac{1}{n}\mathbf{X}^\top\mathbf{P}_\mathbf{W}\mathbf{X})^{-1}(\frac{1}{n}\mathbf{X}^\top\mathbf{W})(\frac{1}{n}\mathbf{W}^\top\mathbf{W})^{-1}(\frac{1}{\sqrt{n}}\mathbf{W}^\top\mathbf{u}),
\end{aligned}$$

and so

$$\begin{aligned}
\text{Var}(\sqrt{n}(\hat{\boldsymbol{\beta}}_{\text{IV}}-\hat{\boldsymbol{\beta}}) &= \mathbb{E}\left[n(\hat{\boldsymbol{\beta}}_{\text{IV}}-\hat{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}_{\text{IV}}-\hat{\boldsymbol{\beta}})^\top\right] \\
&= \mathbb{E}[(\mathbf{S}_{\mathbf{X}^\top\mathbf{W}}\mathbf{S}_{\mathbf{W}^\top\mathbf{W}}^{-1}\mathbf{S}_{\mathbf{W}^\top\mathbf{X}})^{-1}\mathbf{S}_{\mathbf{X}^\top\mathbf{W}}\mathbf{S}_{\mathbf{W}^\top\mathbf{W}}^{-1}\frac{1}{\sqrt{n}}\mathbf{W}^\top\mathbf{u} \\
&\times \mathbf{u}^\top\mathbf{W}\frac{1}{\sqrt{n}}\mathbf{S}_{\mathbf{W}^\top\mathbf{W}}^{-1}\mathbf{S}_{\mathbf{W}^\top\mathbf{X}}(\mathbf{S}_{\mathbf{X}^\top\mathbf{W}}\mathbf{S}_{\mathbf{W}^\top\mathbf{W}}{}^{-1}\mathbf{S}_{\mathbf{W}^\top\mathbf{X}})^{-1}] \\
&= \sigma_\mathbf{u}^2(\mathbf{S}_{\mathbf{X}^\top\mathbf{W}}\mathbf{S}_{\mathbf{W}^\top\mathbf{W}}{}^{-1}\mathbf{S}_{\mathbf{W}^\top\mathbf{X}})^{-1},
\end{aligned}$$

where we use the assumption that $\mathbb{E}[\mathbf{u}^\top\mathbf{u}]=\sigma_\mathbf{u}^2\mathbf{I}$. Thus, by combining the LLN and CLT we obtain that $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{\text{IV}}-\boldsymbol{\beta})$ asymptotically behaves as:

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_{\text{IV}}-\boldsymbol{\beta}) \overset{a}{\sim} N\left(0,\sigma_\mathbf{u}^2(\mathbf{S}_{\mathbf{X}^\top\mathbf{W}}\mathbf{S}_{\mathbf{W}^\top\mathbf{W}}{}^{-1}\mathbf{S}_{\mathbf{W}^\top\mathbf{X}})^{-1}\right).$$

In practice, since $\sigma^2_{\mathbf{u}}$ is unknown, we use:

$$\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_{\text{IV}}) = \hat{\sigma}^2_{\mathbf{u}}(\mathbf{X}^\top \mathbf{P}_{\mathbf{W}} \mathbf{X})^{-1}, \tag{15}$$

where

$$\hat{\sigma}^2_{\mathbf{u}} = \frac{\hat{\mathbf{u}}^\top \hat{\mathbf{u}}}{n}, \ \hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{IV}}.$$

Note, if we assume that:

$$\text{Var}(\mathbf{W}_t u_t) = \mathbb{E}\left[\mathbb{E}\left[u_t^2 | \mathbf{W}_t\right] \mathbf{W}_t^\top \mathbf{W}_t\right] = \sigma_t^2 \mathbb{E}[\mathbf{W}_t^\top \mathbf{W}],$$

then we can show that the asymptotic variance of $\frac{\mathbf{W}^\top \mathbf{u}}{\sqrt{n}}$ is:

$$\operatorname*{plim}_{n \to \infty} \frac{1}{n} \left(\sum_{t=1}^n u_t^2 \mathbf{W}_t^\top \mathbf{W}_t \right) = \operatorname*{plim}_{n \to \infty} \frac{1}{n} \mathbf{W}^\top \boldsymbol{\Omega} \mathbf{W},$$

where

$$\boldsymbol{\Omega} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \sigma_n^2 \end{bmatrix}.$$

Therefore, the asymptotic covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\text{IV}} - \hat{\boldsymbol{\beta}})$ is

$$\left(\operatorname*{plim}_{n \to \infty} \frac{\mathbf{X}^\top \mathbf{P}_{\mathbf{W}} \mathbf{X}}{n} \right)^{-1} \left(\operatorname*{plim}_{n \to \infty} \frac{\mathbf{X}^\top \mathbf{P}_{\mathbf{W}} \boldsymbol{\Omega} \mathbf{P}_{\mathbf{W}} \mathbf{X}}{n} \right) \left(\operatorname*{plim}_{n \to \infty} \frac{\mathbf{X}^\top \mathbf{P}_{\mathbf{W}} \mathbf{X}}{n} \right)^{-1}$$

which is the sandwich covariance matrix in the IV case. Similar to the OLS estimator with heteroskedastic errors. A consistent estimator of $\text{Var}(\hat{\boldsymbol{\beta}}_{\text{IV}})$ is:

$$(\mathbf{X}^\top \mathbf{P}_{\mathbf{W}} \mathbf{X})^{-1} (\mathbf{X}^\top \mathbf{P}_{\mathbf{W}} \hat{\boldsymbol{\Omega}} \mathbf{P}_{\mathbf{W}} \mathbf{X}) (\mathbf{X}^\top \mathbf{P}_{\mathbf{W}} \mathbf{X})^{-1}, \tag{16}$$

where

$$\hat{\boldsymbol{\Omega}} = \text{diag}(\{\hat{u}_t^2\}_{t=1}^n).$$