Heteroscedasticity and Autocorrelation

(A few words on what these are. The failure of these assumptions affects the calculation of standard errors, but not unbiasedness of OLS.)

The same basic strategy underlies addressing both of these. First, use some method to estimate the 'irregularity', then use the estimate to transform the problem back to standard conditions.

The outline of a solution thus requires proving that the estimate of the irregularity in the first step is 'good enough.' The proof techniques for this, and a more general characterization of the remedies, is available in ML and QML estimation.

Heteroscedasticity

The basic strategy is that if you know or can estimate the form of the heteroscedasticity, you use this to transform the model back into one that *basically* obeys the OLS restrictions. (The force of *basically* is that you are now almost always in a situation where OLS does not have the finite sample properties.) Suppose

$$\begin{aligned} y_i &= X_i\beta + \varepsilon_i \\ E(\varepsilon_i^2|X) &= \sigma_i^2 = \alpha + X_{i2}^2\gamma + \eta, \end{aligned}$$

so that the variance is an unknown linear function of the square of the second regressor.

Heteroscedasticity: Remedy

$$\begin{aligned} y_i &= X_i\beta + \varepsilon_i \\ E(\varepsilon_i^2|X) &= \sigma_i^2 = \alpha + X_{i2}^2\gamma + \eta, \end{aligned}$$

The usual procedure is to estimate the first equation by OLS, and use the squared residuals from this regression to estimate the second equation. Now take the fitted values from this second equation, and transform the original equation to

$$\frac{y_i}{\sqrt{\sigma_i^2}} = \frac{X_i\beta}{\sqrt{\sigma_i^2}} + \frac{\varepsilon_i}{\sqrt{\sigma_i^2}}$$

and re-estimate by OLS on this transformed equation. (Rules of thumb re s.e. of b; why $\left[\sigma_i^2\right]^{1/2}$?; what is first column of transformed data matrix?)

Serial Correlation

More generally, dependence among residuals; the simplest example is ‘first order serial correlation’ or ’autocorrelation’.

$$
\begin{aligned}
y_t &= X_t\beta + u_t \\
u_t &= \rho u_{t-1} + \varepsilon_t,
\end{aligned}
$$

where ρ is unknown but $|\rho| < 1$. Notice that

$$
\begin{aligned}
y_t &= X_t\beta + u_t \\
E(u_t u_{t-1}) &= E[(\rho u_{t-1} + \varepsilon_t)u_{t-1}] \\
&= \rho E(u_{t-1}^2) + E(\varepsilon_t u_{t-1}) \\
&= \rho\sigma_u^2,
\end{aligned}
$$

where the last line follows from the standard assumption that ε_t is independent of past u’s, so that ε is an ‘innovation.’ This ‘ruins’ the calculation $E(\varepsilon\varepsilon') = \sigma^2 I$ that is used to compute $V(b_{ols})$.

Serial Correlation: Remedy (?)

One remedy to this form of the serial correlation problem is to run the regression $y_t = X_t\beta + u_t$ by OLS and estimate ρ by regressing u_t on u_{t-1}. Then the transformed regression

$$\begin{aligned} y_t - \hat{\rho} y_{t-1} &= (X_t - \hat{\rho} X_{t-1})\beta + (u_t - \hat{\rho} u_{t-1}) \\ &= (X_t - \hat{\rho} X_{t-1})\beta + \hat{\varepsilon}_t \end{aligned}$$

can be estimated, and asymptotically correct standard errors will be produced by the OLS formula applied to the transformed regression. Again, rule of thumb for moderate ρ.

BUT BEWARE: this will not work if lagged values of y are among the $X's$. (Why?)

Sources of 'Endogeneity'

The econometric use of 'endogeneity' is rather broad: it is used to mean any situation in which the 'disturbance' is correlated with explanatory variable(s). Three broad geneses of endogeneity in micro (cf. Wooldridge pp. 50-51):

- Omitted variables. If a variable is left out of the regression that should be in the regression, then its effect become part of the 'disturbance.' If the variable(s) thus relegated to the disturbance are not independent of all the included regressors, the result is biased or inconsistent estimation.

- Measurement error. If instead of the true variable x_k^* we observe an imperfect measure x_k and use it in our regression, then depending on exactly what we assume about the relation between x_k^* and x_k, we may have a correlation between the disturbance and an included regressor. In particular, under the classical assumption that measurement errors are independent of the true value x_k^*, endogeneity results. (Further discussion to follow.)

- Simultaneity. If x_k is determined partly as a function of y, then x_k and ε are generally correlated. E.g. 'If y is city murder rate and x_k is the size of the police force, ...[the latter] is partly determined by... [the former]. Conceptually, this is a more difficult situation to analyze, because we must be able to think of a situation where we *could* vary x_k exogenously, even though in the data we collect y and x_k are generated simutaneously.'

Further following Wooldridge, let $E(y|\mathbf{x},q)$ be the relation of interest, linear in parameters and additive in q; then while we can estimate $E(y|\mathbf{x})$ this bears no particular relationship to $E(y|\mathbf{x},q)$ when q and $\mathbf{x}$ are allowed to be correlated. One way to analyze the situation is to write q as part of the error term. Wooldridge uses this device repeatedly.

Instrumental Variables

Suppose one of the X's is correlated with ε. This is one of the most difficult problems in econometrics. In the context of the wage equation, education is often thought to be correlated with (unobservable) 'ability', where ability is one of the most important constituents of the log(wage) equation's disturbance. (Brief discussion: when agents choose some component of X this choice will depend on the disturbance or a part thereof that the agent knows but we don't.)

Suppose we have a variable that is (1) uncorrelated with the relevant disturbance and (2) correlated with the X in question. (So, stretching credulity a bit, this might be mother's education.) Let Z be a matrix of values of such instrumental variables, with Z having the same dimension as X. (Notice that if an X variable is 'exogenous', i.e. independent of ε, it can be an instrument; so take Z for now as X with each endogenous variable replaced with a single instrument, or that Z is a matrix with an instrument for each X, where some X's are their own instruments.)

Instrumental Variables (2)

If Z is independent of ε, then $E(Z'\varepsilon) = 0$. Now in OLS, $E(X'\varepsilon) = 0$ is 'exploited' by the normal estimating equations $X'u = X'(y - Xb) = 0$, so let us try the same strategy here and estimate b so as to make the resulting residuals orthogonal to Z. Thus

$$Z'u = Z'(y - Xb) = 0, \qquad whence$$

$$b_{IV} = (Z'X)^{-1}Z'y$$

Then $E(b_{IV}) = E(Z'X)^{-1}Z'(X\beta + \varepsilon) = E(Z'X)^{-1}Z'X\beta + E(Z'X)^{-1}Z'\varepsilon = \beta + E((Z'X)^{-1}Z'\varepsilon)$. At this point, we shift formal gears a bit and abandon the finite sample methods and results that characterize OLS under ideal conditions.

Some Assumptions for Asymptotic Analysis ('Large Sample Properties') of Instrumental Variables

Definition D1, p.897, Greene: **Convergence in Probability.** *The random variable* x_n *converges in probability to a constant* c *if* $\lim_{n\to\infty} \text{Prob}(|x_n - c| > \varepsilon) = 0$ *for any positive* c*.*

If x_n converges in probability to c, we write $plim\ x_n = c$.

The notation/assumptions we need are:

$$
\begin{aligned}
plim(1/n)X'X &= Q_{XX}, \text{ a finite positive definite matrix} \\
plim(1/n)Z'Z &= Q_{ZZ}, \text{ a finite positive definite matrix} \\
plim(1/n)Z'X &= Q_{ZX}, \text{ a finite, } L \times K \text{ matrix with rank } K \\
plim(1/n)X'Z &= Q_{XZ} \\
plim(1/n)Z'\varepsilon &= 0
\end{aligned}
$$

(The L in Q_{ZX}, a finite, L x K matrix with rank K is 'cheap generality' at this point; just think of this as K.)

Asymptotic Distribution of the IV Estimator

Since $b_{IV} = (Z'X)^{-1}Z'y$, $b_{IV} = \beta + (Z'X)^{-1}Z'\varepsilon$ and thus

$$\sqrt{n}(b_{IV} - \beta) = \left(\frac{Z'X}{n}\right)^{-1} \frac{1}{\sqrt{n}} Z'\varepsilon$$

and so we have

$$b_{IV} \sim^a N(\beta, \frac{\sigma^2}{n} Q_{ZX}^{-1}\ Q_{ZZ}\ Q_{XZ}^{-1})$$

Notice (and study) the forms in which n enters these statements. (Discussion; statements and notation follow Greene.)

This is $(V(b_{IV}))$ isa statement of a theoretical quantity. We estimate $V(b_{IV})$ by

$$\hat{V}(b_{IV}) = \hat{\sigma}^2 (Z'X)^{-1}(Z'Z)(X'Z)^{-1}$$

Hey!! Where did the n go???

Measurement Error

Measurement error in the dependent variable is typically benign. Let $y^* = X\beta + \varepsilon$ be the true model and $e = y - y^*$ be the measurement error, where y is the observed value. Then $y = y^* + e = X\beta + \varepsilon + e$. As long as the measurement error is not correlated with a variable in X, usual OLS theory applies.

When an explanatory variable is measured with error, the question of whether there is a bias in OLS depends on the nature of the measurement. The usual assumption, the 'classical errors in variable' (CEV) model, *does* generate a problem.

Measurement Error (2)

Following Wooldridge, pp. 73-76, write the'true' model as $y = \beta_0 + \beta_1 x_1 + ... + \beta_K x_K^* + \upsilon$, and write the measurement error as $e_K = x_K - x_K^*$. Then the equation we estimate is:

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_K x_K + (\upsilon - \beta_K e_K)$$

Now if $x_K = x_K^* + e_K$. The CEV model assumes e_K is independent of x_K^* so $cov(x_K, e_K) \neq 0$ indeed $cov(x_K, e_K) > 0$ is the natural normalization. So if $\beta_K > 0$ then OLS estimates of b_K will (often) be biased downward and if $\beta_K < 0$ the bias will (often) be upward—'bias towards zero' describes both cases. (Also called 'attenuation bias'.) (Discussion: force of 'often'; *all* coefficients biased.)

Another assumption is possible: write $x_K^* = x_K + e_K$, with $cov(x_K, e_K) = 0$ so that $cov(x_K^*, e_K) > 0$. Now there is no correlation between the included variable and the 'disturbance'. (Discussion; ambiguities when several explanatory variables are measured with error.) Solution: find an instrument for x_K.

Omitted Variables: Three Solutions (Two with IV)

Solution 1: Proxy variables. Following Wooldridge, pp. 63-67. We seek $E(y|x,q)$ and do not have q but do have z meeting two conditions. The first is that $E(y|x,q,z) = E(y|x,q)$, so z is 'redundant' in the 'structural model.' The second is that 'the correlation between the omitted variable q and each x_j be zero once we *partial out* z.': $L(q|1,x_1,x_2,...,x_k,z) = L(q|1,z)$. ($z$ has to be 'good enough for q' in a certain way: if we were to predict q on the basis of z, x wouldn't matter.) To see why this works, put $q = \theta_0 + \theta_1 z + r$ (the OLS 'linear projection', so r is orthogonal to z by construction.) Now follow the original regression:

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + ...\beta_K x_K + \gamma q + \upsilon \\ &= \beta_0 + \beta_1 x_1 + ...\beta_K x_K + \gamma(\theta_0 + \theta_1 z + r) + \upsilon \\ &= (\beta_0 + \gamma\theta_0) + \beta_1 x_1 + ...\beta_K x_K + \gamma\theta_1 z + (\gamma r + \upsilon) \end{aligned}$$

and verify that x and z are orthogonal to the resulting disturbance $(\gamma r + \upsilon)$.

Omitted Variables: ‘Leave it in the disturbance and use IV’

Solution 2: Write $y = \beta_0 + \beta_1 x_1 + ...\beta_K x_K + (\gamma q + v)$ and find an instrument for any variable in x that is correlated with q. Instruments should satisfy three requirements:

1. They are redundant in the structural model $E(y|x, q)$;
2. They are uncorrelated with the omitted variable q;
3. They are ‘sufficiently correlated’ with the endogenous elements of x. (Here ‘endogenous’ $\Longleftrightarrow$ correlated with q.)

Omitted Variables: Multiple Indicators

Solution 3: Multiple indicators, following Wooldridge, pp. 105–107. This is tricky but interesting. We have two *indicators* of q: q_1 and q_2; these depend on (or 'reflect' or 'indicate') q via e.g.

$$q_1 = \delta_0 + \delta_1 q + a_1$$

where $Cov(q, a_1) = 0$ and $Cov(x, a_1) = 0$. (This contains CEV as a special case: put $\delta_0 = 0$ and $\delta_1 = 1$). It follows that

$$q = -(\delta_0/\delta_1) + (1/\delta_1)q_1 - (1/\delta_1)a_1$$

where $\delta_1 \neq 0$ (why?) (and a_1 is correlated with q_1, so q_1 cannot be used as a 'proxy'–explain this.)

Omitted Variables: Multiple Indicators (2)

Now suppose we have a second indicator whose 'measurement equation' is written:

$$q_2 = \rho_0 + \rho_1 q + a_2$$

with $Cov(a_1, a_2) = 0$. (Interpret). The multiple indicators method will (1) put q_1 in the equation to be estimated and (2) estimate via IV with q_2 as an instrument for q_1.

The original equation in terms of q can be written:

$$\begin{aligned} y &= \beta_0 + x\beta + \gamma q + \upsilon \\ y &= \beta_0 + x\beta + \gamma[-(\delta_0/\delta_1) + (1/\delta_1)q_1 - (1/\delta_1)a_1] + \upsilon \\ &= \alpha_0 + x\beta + \gamma_1 q_1 + (\upsilon - \gamma_1 a_1) \end{aligned}$$

Now let's examine q_2 as a candidate instrument for q_1 in this equation. First, q_2 is correlated with q_1. Second q_2 is uncorrelated with υ because it is redundant in the structural equation. Finally, q_2 is uncorrelated with a_1 (since a_1 is uncorrelated with q and a_2.)

Multiple Indicators as an IV Solution to the Omitted Variables Problem: An Example

We follow example 5.5 in Wooldridge, pp. 106–107, using the NLS80 data.

The omitted variable in the $log(wage)$ equation is taken to be 'ability' and the 'measurement' equations for the indicators are:

$$
\begin{aligned}
IQ &= \delta_0 + \delta_1 ability + a_1 \\
KWW &= \rho_0 + \rho_1 ability + a_{2,}
\end{aligned}
$$

where IQ and KWW are test scores. The text show the results of estimating a structural equation with $\{exper, tenure, married, south, urban, black, educ, IQ\}$ on the right hand side and using KWW as an instrument for IQ.