Econometrics
Part II: Further topics. Newey-West and fixed-b approach to serial correlation.
Session 6.

References on Bruce Hansen's textbook will be abbreviated with H. Say, H 1.1 means chapter 1.1 from Bruce Hansen's textbook.

Serial correlation. HAC standard errors. (H 14.1-14.4 and these notes)
Fixed-b asymptotics. (These notes)
Errors in variables. Endogeneity. (H 12.1-12.3)

Serial correlation

Previously, we have assumed that the data are independent: in particular, Ω is diagonal. If the data are dependent, then Ω is typically no longer diagonal.

Strict Stationarity. A sequence of random variables $\{Z_t\}_{t=-\infty}^{\infty}$ is strictly stationary if, for any finite nonnegative integer m,

$$f_{Z_t,Z_{t+1},...,Z_{t+m}}(x_0,x_1,...,x_m) = f_{Z_s,Z_{s+1},...,Z_{s+m}}(x_0,x_1,...,x_m)$$

which is to say that the joint distribution, f, does not depend on the index, t.

Covariance Stationarity. A sequence of random variables $\{Z_t\}_{t=-\infty}^{\infty}$ is covariance (weakly) stationary if just the first two moments do not depend on t, e.g.

$$\begin{aligned} EZ_1 &= EZ_2 = ... \\ Var(Z_1) &= Var(Z_2) = ... \\ Cov(Z_1, Z_{1+m}) &= Cov(Z_2, Z_{2+m}) = ... \text{ for any } m \end{aligned}$$

Serial correlation, OLS

Consider a new set of OLS assumptions:

(SC0) $\{(y_t, x_t)\}_{t=1}^{T}$ is strictly stationary

(SC1) $\{(x_t x_t')\}$ satisfies LLN: $\frac{1}{T}\sum x_t x_t' \xrightarrow{p} E(x_t x_t') < \infty$, positive definite

(SC2) $\{(x_t \varepsilon_t)\}$ satisfies LLN: $\frac{1}{T}\sum x_t \varepsilon_t \xrightarrow{p} E(x_t \varepsilon_t) = 0$

(SC3) $\{(x_t \varepsilon_t)\}$ satisfies CLT: $\frac{1}{\sqrt{T}}\sum x_t \varepsilon_t \xrightarrow{d} N(0, V)$, where

$$V = E\left(\varepsilon_t^2 x_t x_t'\right) + \sum_{l=1}^{\infty}\left(E\left(\varepsilon_t \varepsilon_{t-l} x_t x_{t-l}'\right) + E\left(\varepsilon_t \varepsilon_{t-l} x_{t-l} x_t'\right)\right)$$

Note that the above assumptions generalize further our GM/OLS conditions, such that if the data were independent, we would have $V = E\left(\varepsilon_t^2 x_t x_t'\right)$ as before, in (OLS3'). Various versions of LLN and CLT applicable to time series can be found e.g. in Hayashi (2000, Ch. 2.1-2.2). Key conditions require the dependence between variables separated in time to disappear sufficiently fast as the time gap increases.

Let us examine how OLS performs under the SC assumptions.

Theorem 1 *Under SC0-1-2-3*
(i) $\hat{\beta}_{OLS} \xrightarrow{p} \beta$ *(OLS is consistent)*
(ii) $\sqrt{T}\left(\hat{\beta}_{OLS} - \beta\right) \xrightarrow{d} N\left(0, E\left(x_t x_t'\right)^{-1} V E\left(x_t x_t'\right)^{-1}\right)$

The proof is straightforward. How to estimate the asymptotic variance of OLS?

Newey-West Method

Define $V_T = Var\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t \varepsilon_t\right)$.

$$\begin{aligned}
V_T &= E\left[\frac{1}{T}\left(\sum_{t=1}^{T} x_t \varepsilon_t\right)\left(\sum_{t=1}^{T} x_t \varepsilon_t\right)'\right] \\
&= E\left[\frac{1}{T}\sum_{t=1}^{T} \varepsilon_t^2 x_t x_t' + \frac{1}{T}\sum_{l=1}^{T-1}\sum_{t=l+1}^{T}\left(\varepsilon_t \varepsilon_{t-l} x_t x_{t-l}' + \varepsilon_t \varepsilon_{t-l} x_{t-l} x_t'\right)\right] \\
&= \frac{1}{T}\sum_{t=1}^{T} E\left(\varepsilon_t^2 x_t x_t'\right) + \frac{1}{T}\sum_{l=1}^{T-1}\sum_{t=l+1}^{T}\left(E\left(\varepsilon_t \varepsilon_{t-l} x_t x_{t-l}'\right) + E\left(\varepsilon_t \varepsilon_{t-l} x_{t-l} x_t'\right)\right) \\
&= \frac{1}{T}\sum_{t=1}^{T} E\left(\varepsilon_t^2 x_t x_t'\right) + \sum_{l=1}^{T-1}\frac{T-l}{T}\left(E\left(\varepsilon_t \varepsilon_{t-l} x_t x_{t-l}'\right) + E\left(\varepsilon_t \varepsilon_{t-l} x_{t-l} x_t'\right)\right) \\
&= E\left(\varepsilon_t^2 x_t x_t'\right) + \sum_{l=1}^{T-1}\frac{T-l}{T}\left(E\left(\varepsilon_t \varepsilon_{t-l} x_t x_{t-l}'\right) + E\left(\varepsilon_t \varepsilon_{t-l} x_{t-l} x_t'\right)\right)
\end{aligned}$$

So as T gets large, $V_T \approx V$. Since we have T data points, we only estimate $G < T$ autocovariances of $x_t \varepsilon_t$, where G is called the truncation lag. Newey and West (1987) propose to estimate V as follows:

(1) Choose G such that:

$$G = O\left(T^{\alpha}\right) \text{ for } 0 < \alpha < 1/4$$

(2) Estimate autocovariances of $x_t \varepsilon_t$ of order l by

$$\hat{\Gamma}_l = \frac{1}{T}\sum_{t=l+1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-l} x_t x_{t-l}'$$

(3) Estimate V by

$$\hat{V} = \hat{\Gamma}_0 + \sum\nolimits_{l=1}^{G} \frac{G+1-l}{G+1} \left(\hat{\Gamma}_l + \hat{\Gamma}_l' \right).$$

Newey-West proved that this estimator is consistent, and the weight $\frac{G+1-l}{G+1}$ ensures that $\hat{V}$ is positive semi-definite.

Now we can estimate the variance-covariance matrix of $\hat{\beta}_{OLS}$ as

$$\frac{1}{T} \left[\frac{1}{T} \sum_{t=1}^{T} x_t x_t' \right]^{-1} \hat{V} \left[\frac{1}{T} \sum_{t=1}^{T} x_t x_t' \right]^{-1}.$$

Such an estimator is often referred to as the HAC estimator of the variance of OLS (Heteroskedasticity and Autocorrelation Consistent).

Fixed-b asymptotics.

The choice of the truncation lag G in the Newey-West method is somewhat ambiguous. Many ways of choosing it optimally have been proposed, see e.g. Andrews (1991). Usually, such optimal choices are quite complicated. Kiefer-Vogelsang-Bunzel (2000) pointed out that the accuracy of the tests based on the Newey-West variance estimate in finite samples may be quite poor, with tests often overrejecting the null hypothesis (variance is 'too small'). They proposed an alternative to the Newey-West technique where instead of choosing G so that $b \equiv (G+1)/T \to 0$ as $T \to \infty$, b (called the bandwidth) is kept fixed. For example, when $G+1 = T$, b is fixed at 1.

Kiefer-Vogelsang-Bunzel (KVB) showed that under the fixed b assumption, $\hat{V}$ converges to a limiting random matrix that is proportional to V. The distribution of the HAC robust tests based on $\hat{V}$ are shown to not depend on the model's parameters (we say that the distribution is pivotal). The distribution can be tabulated and the table can be used for testing.

To give you an idea of the techniques involved, we consider here the simplest possible situation, where the regression does not contain any explanatory variables except constant

$$Y_t = \beta + \varepsilon_t.$$

In such a simple case, $\hat{\beta}_{OLS} = \bar{Y}$, but of course, under serial correlation,

$$Var\left(\hat{\beta}_{OLS}\right) = \frac{1}{T} V_T = \frac{1}{T} \left(E\varepsilon_t^2 + \sum\nolimits_{l=1}^{T-1} \frac{T-l}{T} 2E\left(\varepsilon_t \varepsilon_{t-l}\right) \right) \neq \frac{1}{T} Var\left(\varepsilon_t\right).$$

The Newey-West method would recommend to compute

$$\hat{V} = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2 + \sum_{l=1}^{G}\frac{G+1-l}{G+1}\frac{2}{T}\sum_{t=1+l}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_{t-l}.$$

Instead, KVB obtain an inconsistent estimate of V_T

$$\hat{V}_{KVB} = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2 + \sum_{l=1}^{T-1}\frac{T-l}{T}\frac{2}{T}\sum_{t=1+l}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_{t-l}.$$

Here $\hat{\varepsilon}_t = Y_t - \bar{Y}$.

Note that by changing index l to $|t-s|$, $\hat{V}_{KVB}$ can be written as

$$\hat{V}_{KVB} = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(1-\frac{|t-s|}{T}\right)\hat{\varepsilon}_t\hat{\varepsilon}_s$$

Consider the identity (similar to the integration by parts)

$$\sum_{t=1}^{T}a_t b_t = \sum_{t=1}^{T-1}\left[(a_t - a_{t+1})\sum_{i=1}^{t}b_i\right] + a_T\sum_{i=1}^{T}b_i$$

This identity yields (because $\sum_{s=1}^{T}\hat{\varepsilon}_s = 0$)

$$\sum_{s=1}^{T}\left(1-\frac{|t-s|}{T}\right)\hat{\varepsilon}_s = \sum_{s=1}^{T-1}\left[\left(1-\frac{|t-s|}{T}-1+\frac{|t-s-1|}{T}\right)\sum_{i=1}^{s}\hat{\varepsilon}_i\right]$$

and

$$\begin{aligned}
&\sum_{t=1}^{T}\sum_{s=1}^{T}\left(1-\frac{|t-s|}{T}\right)\hat{\varepsilon}_t\hat{\varepsilon}_s = \sum_{t=1}^{T}\sum_{s=1}^{T-1}\left[\left(-\frac{|t-s|}{T}+\frac{|t-s-1|}{T}\right)\hat{\varepsilon}_t\sum_{i=1}^{s}\hat{\varepsilon}_i\right]\\
=\ &\sum_{s=1}^{T-1}\left[\sum_{i=1}^{s}\hat{\varepsilon}_i\sum_{t=1}^{T}\left(-\frac{|t-s|}{T}+\frac{|t-s-1|}{T}\right)\hat{\varepsilon}_t\right]\\
=\ &\sum_{s=1}^{T-1}\left[\sum_{i=1}^{s}\hat{\varepsilon}_i\sum_{t=1}^{T-1}\left\{\left(-\frac{|t-s|}{T}+\frac{|t-s-1|}{T}+\frac{|t+1-s|}{T}-\frac{|t-s|}{T}\right)\sum_{j=1}^{t}\hat{\varepsilon}_j\right\}\right]\\
=\ &\frac{2}{T}\sum_{s=1}^{T-1}\left[\sum_{i=1}^{s}\hat{\varepsilon}_i\sum_{j=1}^{s}\hat{\varepsilon}_j\right]
\end{aligned}$$

so

$$\hat{V}_{KVB} = 2\times\frac{1}{T}\sum_{s=1}^{T-1}\left[\frac{1}{\sqrt{T}}\sum_{i=1}^{s}\hat{\varepsilon}_i\frac{1}{\sqrt{T}}\sum_{j=1}^{s}\hat{\varepsilon}_j\right]$$

To proceed, we need to introduce the concept of Brownian motion and of the Functional Central Limit Theorem. We will be brief. Details can be found in Stock (1994). Recall that the standard Brownian motion $W(\lambda)$ $\lambda\in[0,1]$ is a continuous time stochastic process (a set of random variables indexed by λ, or, alternatively, a random function in $C[0,1]$ the space of the continuous functions on the segment $[0,1]$), such that $W(\lambda_1), ..., W(\lambda_k)$ jointly normally distributed for fixed $\lambda_1, ..., \lambda_k$ with

$$\begin{aligned} EW(\lambda_i) &= 0 \text{ and} \\ Cov(W(\lambda_1), W(\lambda_2)) &= \min(\lambda_1, \lambda_2). \end{aligned}$$

The FCLT is a generalization of the conventional CLT to function-valued random variables. Let us extend the notions of consistency and convergence in distribution to the space $C[0,1]$ equipped with sup-norm metric:

$$d(f,g) = \sup_{x \in [0,1]} |f(x) - g(x)|.$$

Consistency. A random element $\xi_T \in C[0,1]$ converges in probability to f if $\Pr(d(\xi_T, f) > \delta) \to 0$ for any $\delta > 0$.

Convergence in distribution. Let $\{\xi_T\}$ be a sequence of random elements in $C[0,1]$ with induced probability measures $\{\pi_T\}$. Then π_T converges weakly to π, or equivalently $\xi_T \Rightarrow \xi$ where ξ has probability measure π, if and only if $\int f d\pi_T \to \int f d\pi$ for all bounded continuous functionals $f : C[0,1] \to \mathbb{R}$.

Continuous Mapping Theorem. If h is a continuous functional mapping $C[0,1]$ to some metric space and $\xi_T \Rightarrow \xi$, then $h(\xi_T) \Rightarrow h(\xi)$.

Let ζ_t $t = 1, 2, ..$ be zero mean i.i.d. random variables with variance 1. Consider the following random element of $C[0,1]$:

$$\xi_T(\lambda) = T^{-1/2} \left(\sum_{t=1}^{[T\lambda]} \zeta_t + (T\lambda - [T\lambda]) \zeta_{[T\lambda]+1} \right).$$

A CLT for vector-valued processes ensures that, if $\lambda_1, ..., \lambda_k$ are fixed constants, then $[\xi_T(\lambda_1), ..., \xi_T(\lambda_k)]$ converges jointly to a k-dimensional random variable. The FCLT extends this result.

Theorem (FCLT) We have $\xi_T \Rightarrow W$, where W is a standard Brownian motion on the unit interval.

Let us now consider an arbitrary linear process

$$\varepsilon_t = c(L)\zeta_t \equiv c_0\zeta_t + c_1\zeta_{t-1} + c_2\zeta_{t-2} + ...$$

such that $c(1) = \sum_{j=0}^{\infty} c_j \neq 0$ (no unit roots) and $\sum_{j=0}^{\infty} j\,|c_j| < \infty$. Define

$$\xi_T(\lambda) = T^{-1/2}\left(\sum_{t=1}^{[T\lambda]} \varepsilon_t + (T\lambda - [T\lambda])\,\varepsilon_{[T\lambda]+1}\right)$$

Theorem. $\xi_T \Longrightarrow c(1)\,W$

Returning to KVB, assume that $\varepsilon_t = c(L)\,\zeta_t$. Consider

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum\nolimits_{i=1}^{s}\hat{\varepsilon}_i &= \frac{1}{\sqrt{T}}\sum\nolimits_{i=1}^{s}\left(\varepsilon_i - \left(\hat{\beta}_{OLS} - \beta\right)\right) \\
&= \frac{1}{\sqrt{T}}\sum\nolimits_{i=1}^{s}\left(\varepsilon_i - \frac{1}{T}\sum\nolimits_{j=1}^{T}\varepsilon_j\right) \\
&= \frac{1}{\sqrt{T}}\sum\nolimits_{i=1}^{s}\varepsilon_i - \frac{s}{T}\frac{1}{\sqrt{T}}\sum\nolimits_{j=1}^{T}\varepsilon_j \\
&= \xi_T\left(\frac{s}{T}\right) - \frac{s}{T}\xi_T(1).
\end{aligned}$$

We can approximate

$$\begin{aligned}
V_{KVB} &= 2 \times \frac{1}{T}\sum\nolimits_{s=1}^{T-1}\left[\frac{1}{\sqrt{T}}\sum\nolimits_{i=1}^{s}\hat{\varepsilon}_i\frac{1}{\sqrt{T}}\sum\nolimits_{j=1}^{s}\hat{\varepsilon}_j\right] \\
&= 2 \times \frac{1}{T}\sum\nolimits_{s=1}^{T-1}\left(\xi_T\left(\frac{s}{T}\right) - \frac{s}{T}\xi_T(1)\right)^2 \\
&\approx 2\int_0^1\left(\xi_T(\lambda) - \frac{s}{T}\xi_T(1)\right)^2 d\lambda
\end{aligned}$$

and show that the error of the approximation converges in probability to zero. The integral is a continuous functional on $C[0,1]$ so CMT can be used together with FCLT to show that

$$V_{KVB} \xrightarrow{d} 2\,[c(1)]^2\int_0^1 (W(\lambda) - \lambda W(1))^2\, d\lambda$$

The right hand side is a random quantity proportional to $V = [c(1)]^2$.

If we now consider the t-statistic (based on V_{KVB}) for testing hypothesis that

$\beta = 0$, we get

$$
\begin{aligned}
t &= \frac{\hat{\beta}}{\sqrt{\frac{1}{T}V_{KVB}}} \overset{H_0}{=} \frac{\frac{1}{\sqrt{T}}\sum_{j=1}^{T}\varepsilon_j}{\sqrt{V_{KVB}}} \xrightarrow{d} \frac{c(1)\,W(1)}{\sqrt{2\left[c(1)\right]^2\int_0^1 \left(W(\lambda)-\lambda W(1)\right)^2 d\lambda}} \\
&= \frac{c(1)\,W(1)}{c(1)\sqrt{2\int_0^1 \left(W(\lambda)-\lambda W(1)\right)^2 d\lambda}} \\
&= \frac{W(1)}{\sqrt{2\int_0^1 \left(W(\lambda)-\lambda W(1)\right)^2 d\lambda}},
\end{aligned}
$$

which does not depend on $c(1)$ (parameter of the model). The latter distribution can be simulated and its critical values recorded. KVB show that in finite sample, their tests may outperform those based on the Newey-West robust standard errors.

Classical measurement error.

Measurement errors plague most data sets to one degree or another. Suppose that we observe noisy versions of the variables we would like to observe. We obtain data y_i and x_i for $i = 1, ..., n$, while the true values are y_i^* and x_i^*. Also assume the following

$$
\begin{aligned}
x_i &= x_i^* + \nu_i \\
y_i &= y_i^* + \eta_i
\end{aligned}
$$

where the errors in measurement are such that the following hold:

$$
\begin{aligned}
E(\nu_i) = 0, &\quad E(\eta_i) = 0 \\
E(x_i^*\nu_i) = 0, &\quad E(y_i^*\eta_i) = 0 \\
E(x_i^*\eta_i) = 0, &\quad E(y_i^*\nu_i) = 0 \\
E(\nu_i\eta_i) = 0. &
\end{aligned}
$$

Given that $E(y_i^*|x_i^*) = x_i^{*\prime}\beta$, if we proceeded as if there were no measurement error, we would estimate the following by OLS

$$
\begin{aligned}
\hat{\beta}_{OLS} &= \quad (X'X)^{-1} \quad X'Y \\
&= \left(\frac{1}{n}\sum x_i x_i'\right)^{-1} \quad \frac{1}{n}\sum x_i y_i' \\
&= \quad \downarrow p \quad \downarrow p \\
&= \quad (E x_i x_i')^{-1} \quad E(x_i y_i)
\end{aligned}
$$

Note that

$$\begin{aligned} E\left(x_i x_i'\right) &= E\left[\left(x_i^* + \nu_i\right)\left(x_i^* + \nu_i\right)'\right] \\ &= E\left[x_i^* x_i^{*\prime}\right] + E\left[\nu_i \nu_i'\right] + E\left[x_i^* \nu_i'\right] + E\left[\nu_i x_i^{*\prime}\right] \\ &= E\left[x_i^* x_i^{*\prime}\right] + E\left[\nu_i \nu_i'\right] \end{aligned}$$

$$\begin{aligned} E\left(x_i y_i\right) &= E\left[\left(x_i^* + \nu_i\right)\left(y_i^* + \eta_i\right)\right] \\ &= E\left[x_i^* y_i^*\right] + E\left[\nu_i \eta_i\right] + E\left[x_i^* \eta_i\right] + E\left[\nu_i y_i^*\right] \\ &= E\left[x_i^* y_i^*\right] \\ &= E\left[x_i^* x_i^{*\prime}\right]\beta \end{aligned}$$

Thus, when measurement error in the independent variables exists, so that $Var\left(\nu_i\right) \neq 0$, OLS yields an inconsistent estimator:

$$\hat{\beta}_{OLS} \xrightarrow{p} \left(E\left[x_i^* x_i^{*\prime}\right] + E\left[\nu_i \nu_i'\right]\right)^{-1} E\left[x_i^* x_i^{*\prime}\right]\beta$$

Note that when we observe x^* directly, so that $E\left[\nu_i \nu_i'\right] = 0$, then $\hat{\beta}_{OLS} \xrightarrow{p} \beta$ as expected.

In the univariate case, the above reduces to:

$$\hat{\beta}_{OLS} \xrightarrow{p} \underbrace{\frac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_\nu^2}}_{\text{signal/noise ratio}} \beta$$

so the estimate $\hat{\beta}_{OLS}$ is biased towards 0 and inconsistent.

If there were no measurement errors, we would be using

$$y_i^* = x_i^{*\prime}\beta + \varepsilon_i$$

Any measurement error in the dependent variables is subsumed in the error term as follows:

$$y_i = y_i^* + \eta_i = x_i^{*\prime}\beta + \left(\varepsilon_i + \eta_i\right)$$

This new error term, $\varepsilon_i + \eta_i$ creates no problem for estimation, as η_i is uncorrelated with x_i^*. However, if there were measurement errors in x_i^* as well, then we have

$$y_i = x_i'\beta + \underbrace{(\varepsilon_i + \eta_i - \nu_i'\beta)}_{u_i}$$

In this case, the new error term, denoted u_i, is correlated with x_i.

$$E[x_i\nu_i'] = E[(x_i^* + \nu_i)\nu_i'] = E[x_i^*\nu_i'] + E[\nu_i\nu_i']$$

Because the error term is correlated with x_i, (OLS2') $Ex_iu_i = 0$ does not hold, and OLS is inconsistent.

There are two potential solutions to the problem of measurement error.

Solution 1. If we can estimate $E[\nu_i\nu_i']$, we can undo the error in estimation by using the following

$$E[x_ix_i'] = E[x_i^*x_i^{*\prime}] + E[\nu_i\nu_i']$$

Unfortunately, this is not usually possible.

Solution 2. Suppose we get another independent measure of x^* such that

$$w_i = x_i^* + \tau_i$$

and τ_i is uncorrelated with any of $y_i^*, x_i^*, \eta_i, \nu_i$.

$$\begin{aligned} E[w_ix_i'] &= E\left[(x_i^* + \tau_i)(x_i^* + \nu_i)'\right] = E[x_i^*x_i^{*\prime}] \\ E[w_iy_i] &= E[(x_i^* + \tau_i)(y_i^* + \eta_i)] = E[x_i^*y_i^*] \end{aligned}$$

Then if $E[w_ix_i']$ is invertible, we have

$$E[w_ix_i']^{-1}E[w_iy_i] = E[x_i^*x_i^{*\prime}]^{-1}E[x_i^*y_i^*] = E[x_i^*x_i^{*\prime}]^{-1}E[x_i^*x_i^{*\prime}]\beta = \beta$$

So $\hat{\beta}_{IV} = (W'X)^{-1}W'Y$ is consistent.

The idea behind this repeated observation estimator extends well beyond the repeated observation case and even beyond the measurement error case. Next we will discuss instrumental variables estimation.

Endogeneity

Consider the following model

$$y_i = x_i'\beta + \varepsilon_i \text{ where } E(x_i\varepsilon_i) \neq 0 \text{ in violation of OLS2 and OLS2'}.$$

This is a core problem in econometrics that distinguishes it from statistics. To distinguish this model from the regression model, we will call the above equation a structural equation, and β a structural parameter. By structural equation we understand one that represents a causal link rather than just an empirical association. When $E(x_i\varepsilon_i) \neq 0$ holds we say that x_i is endogenous for β.

To be clear, usually $E(x_i\varepsilon_i) \neq 0$ is caused by only a few components of x_i being correlated with ε_i. The components causing the problems are then called endogenous and the rest are called exogenous. By rearranging explanatory variables, we may partition x_i into the exogenous part x_{1i} and endogenous part x_{2i}.

The endogeneity may happen in many ways. We have already seen how it arises in the context of measurement error. Another classical example is the Supply and Demand system, where q_i and p_i (quantity and price) are determined jointly by the demand equation

$$q_i = -\beta_d p_i + \varepsilon_{di}$$

and the supply equation

$$q_i = \beta_s p_i + \varepsilon_{si}.$$

In matrix notation,

$$\begin{pmatrix} 1 & \beta_d \\ 1 & -\beta_s \end{pmatrix} \begin{pmatrix} q_i \\ p_i \end{pmatrix} = \begin{pmatrix} \varepsilon_{di} \\ \varepsilon_{si} \end{pmatrix}$$

so that

$$\begin{aligned} \begin{pmatrix} q_i \\ p_i \end{pmatrix} &= \frac{1}{-\beta_s - \beta_d} \begin{pmatrix} -\beta_s & -\beta_d \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{di} \\ \varepsilon_{si} \end{pmatrix} \\ &= \begin{pmatrix} (\beta_s \varepsilon_{di} + \beta_d \varepsilon_{si}) / (\beta_s + \beta_d) \\ (\varepsilon_{di} - \varepsilon_{si}) / (\beta_s + \beta_d) \end{pmatrix} \end{aligned}$$

so neither $E[p_i\varepsilon_{di}] = 0$ nor $E[p_i\varepsilon_{si}] = 0$. If we run an OLS of q_i on p_i we estimating $Cov(q_i, p_i)/Var(p_i)$. Assuming the demand and supply shocks are uncorrelated, this quantity equals

$$\begin{aligned} \frac{Cov(q_i, p_i)}{Var(p_i)} &= \frac{\frac{\beta_s}{(\beta_s+\beta_d)^2} Var(\varepsilon_{di}) - \frac{\beta_d}{(\beta_s+\beta_d)^2} Var(\varepsilon_{si})}{\frac{1}{(\beta_s+\beta_d)^2} Var(\varepsilon_{di}) + \frac{1}{(\beta_s+\beta_d)^2} Var(\varepsilon_{si})} \\ &= \beta_s \frac{Var(\varepsilon_{di})}{Var(\varepsilon_{di}) + Var(\varepsilon_{si})} - \beta_d \frac{Var(\varepsilon_{si})}{Var(\varepsilon_{di}) + Var(\varepsilon_{si})}, \end{aligned}$$

that is, some linear combination of the slopes of the demand and the supply curves.

Our last example of endogeneity would be a structural equation connecting two variables that are both chosen by economic agents, say, wage and education

$$wage_i = \beta_1 + \beta_2 educ_i + \varepsilon_i.$$

Both $wage_i$ and $educ_i$ may be affected by person i's ability or some other factor belonging to ε_i. Here the structural equation can be thought of as reflecting a causal relationship between education and wage, which would be observed if we were able to completely randomly assign education levels to people, independent on their abilities or anything else. In real situation, education is not assigned. It is chosen by people, and this choice may be affected by other factors influencing wage.

References

[1] Andrews, D.W.K. (1991) "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometric*a 59, 817–858.

[2] Hayashi, F. (2000) *Econometrics*, Princeton University Press.

[3] Kiefer, N.M.,Vogelsang, T.J., Bunzel, H. (2000) "Simple Robust Testing of Regression Hypothesis," *Econometrica* 68, 695–714.

[4] Newey, W. K. and West, K.D. (1987) "A Simple, Positive-definite, Heteroskedasticity and Autocorrelation Consistent Covariance matrix," *Econometrica* 55, 703-708.

[5] Stock, J.H. (1994) "Unit Roots, Structural Breaks and Trends," in Engle, R.F. and McFadden, D.L. (eds.) Handbook of Econometrics Vol4, Ch. 46, 2739–2841.