LECTURE 7: STILL MORE, BUT LAST, ON THE K-VARIABLE LINEAR MODEL

SPECIFICATION ERROR:

Suppose the model generating the data is $y = X\beta + \epsilon$

However, the model fitted is $y = X^*\beta^* + \epsilon$, with the LS estimator

$$b^* = (X^*X^*)^{-1}X^*y = (X^*X^*)^{-1}X^*X\beta + (X^*X^*)^{-1}X^*\epsilon.$$

Then $Eb^* = (X^*X^*)^{-1}X^*X\beta$ and $V(b^*) = \sigma^2(X^*X^*)^{-1}$

Application 1: Excluded variables

Let $X = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$ and $X^* = X_1$.

That is, the model that generates the data is $y = X_1\beta_1 + X_2\beta_2 + \epsilon$.

Consider $b^*$ as an estimator of $\beta_1$.

Proposition: $b^*$ is biased.

Proof:

$$b^* = (X^*X^*)^{-1}X^*y = (X^*X^*)^{-1}X^*_1(X_1\beta_1 + X_2\beta_2 + \epsilon)$$

$$b^* = \beta_1 + (X^*_1X_1)^{-1}X^*_1X_2\beta_2 + (X^*_1X_1)^{-1}X^*_1\epsilon$$

$$Eb^* = \beta_1 + (X^*_1X_1)^{-1}X^*_1X_2\beta_2$$

The second expression on the right hand side is the bias. ■

Interpretation?
A classic example:

Suppose that the model generating the data is \( y_i = \beta_0 + \beta_1 S_i + a_i + \varepsilon_i \) where \( y \) is the natural logarithm of earnings, \( S \) is schooling and \( a \) is ability. \( a \) is unobserved and omitted, but it is positively correlated with \( S \). Then

\[
Eb^* = \left[ \begin{array}{c} \beta_0 \\ \beta_1 \end{array} \right] + \left[ \begin{array}{c} \sum S \\ \sum S^2 \end{array} \right]^{-1} \left[ \sum aS \right]
\]

supposing \( a \) is measured so that its coefficient is 1.

If we suppose that \( \sum a = 0 \), then the bias in the coefficient of schooling is positive. Generally, we cannot sign the bias, it depends not only on \( \beta_2 \) but also on \( (X'_1X_1)^{-1}X'_1X_2 \), which of course can be positive or negative.

Note that \( Vb^* = \sigma^2(X'_1X_1)^{-1} \). So if \( \beta_2 = 0 \), there is an efficiency gain from imposing the restriction and leaving out \( X_2 \). This confirms our earlier results.

**Estimation of \( \sigma^2 \):**

\[
e^* = M_1 y = M_1(X_1\beta_1 + X_2\beta_2 + \varepsilon) = M_1X_2\beta_2 + M_1\varepsilon \neq M_1\varepsilon
\]

\[
\Rightarrow e^*e^* = \beta_2'X_2'M_1X_2\beta_2 + \varepsilon'M_1\varepsilon + 2\beta_2'X_2'M_1\varepsilon
\]

Note the expected value of the last term is 0.

Clearly, we cannot estimate \( \sigma^2 \) by usual methods even if \( X'_2X_2 = 0 \) (no bias) since still \( M_1X_2 \neq 0 \).

There is hope of detecting misspecification from the residuals since \( Ee^*e^{*'} = \sigma^2M_1 \) under correct specification and \( Ee^*e^{*'} = \sigma^2M_1 + M_1X'_2\beta_2X_2M_1 \) under misspecification. Look for departures...

**Application 2: Inclusion of unnecessary variables**

Let \( X = X_1 \) and \( X^* = [X_1 \ X_2] \) where \( X_1 \) is \( NxK_1 \) and \( X_2 \) is \( NxK_2 \).
That is, the "true" model is \( y = X_1\beta + \epsilon \).

From the discussion on specification error in the beginning of this lecture, we know that

\[
Eb^* = (X^* X^*)^{-1} X^* \beta. \quad (What \ is \ the \ size \ of \ Eb^*?)
\]

Each column of \((X^* X^*)^{-1} X^* X\) is the coefficient for the regression of a column of \(X_1\) on \(X_1\) and \(X_2\). For example:

\[
x_{ij} = \beta_0 + \beta_{1} x_{i1} + \beta_{2} x_{i2} + \ldots + \beta_{j} x_{ij} + \ldots + \tilde{\beta} X_2 + \epsilon
\]

where \(x_{ij}\) is the \(j\)th column of \(X_1\).

With this regression, the coefficient of \(x_{ij}\) is one, and all others are zero. The fit is perfect. Thus

\[
Eb^* = \begin{bmatrix} \beta \\ 0 \end{bmatrix}
\]

Let's check this argument.

**Proposition:** \(b^*\) is unbiased.

**Proof:**

\[
Eb^* = (X^* X^*)^{-1} X^* \beta
\]

\[
= \begin{bmatrix} X_1' X_1 & X_1' X_2 \\ X_2' X_1 & X_2' X_2 \end{bmatrix}^{-1} \begin{bmatrix} X_1' X_1 \\ X_2' X_1 \end{bmatrix} \beta
\]

The partitioned inversion formula gives

\[
\begin{bmatrix} (X_1' X_1)^{-1} + (X_1' X_1)^{-1} X_1' X_2 DX_2' X_1 (X_1' X_1)^{-1} & -DX_2' X_1 (X_1' X_1)^{-1} \\ -DX_2' X_1 (X_1' X_1)^{-1} & D \end{bmatrix}
\]

for \((X^* X^*)^{-1}\) where \(D = (X_2' M X_2)^{-1}\). This is a symmetric matrix.

Multiplying this out verifies that
Note that the variance of $b^*$ is $V(b^*) = \sigma^2(X^*X^*)^{-1}$.

**Proposition:** The variance of the coefficients of $X_i$ in the unrestricted model (where the matrix of explanatory variables is $X^*$) is greater than the variance of the coefficients in the restricted model (where the matrix of explanatory variables is $X_i$).

**Proof:** Using partitioned inversion (it might help to consider the alternative expression rather than the one given above), the variance of the first $K_1$ elements (coefficients on $X_i$) is $\sigma^2(X'_iM_iX_i)^{-1} \geq \sigma^2(X'_iX_i)^{-1} = \text{variance of the restricted estimator. (why?) } $

**Estimation of $\sigma^2$:**

$e^* = M^*y = M^*\varepsilon$

Under normality, $(e^*e^*/\sigma^2) = (\varepsilon'M^*\varepsilon/\sigma^2) \sim \chi^2(N - K_i - K_2)$.

$\Rightarrow s^2$ has higher variance than in the restricted model. (why?)

**Note on the interpretation of bias:**

$Ey = X\beta$ defines $\beta$ and LS gives unbiased estimates of that $\beta$. Questions of bias really require a model.

Further statistical assumptions like $V\gamma = \sigma^2I$ allow some sorting out of specifications, but is this assumption really attractive?

**Cross products matrix:**

In LS, "everything" comes from the cross products matrix.

**Definition:** The cross products matrix is
\[
\begin{bmatrix}
y'y & y'X \\
x'y & X'X \\
\end{bmatrix} = 
\begin{bmatrix}
\sum y_i^2 & \sum y_i & \sum y_i x_{2i} & \cdots & \sum y_i x_{ki} \\
\sum 1 & \sum x_{2i} & \cdots & \sum x_{ki} \\
\sum x_{2i}^2 & \sum x_{ki} & \cdots & \sum x_{2i} x_{ki} \\
\vdots & \vdots & \ddots & \vdots \\
\sum x_{ki}^2 & \cdots & \cdots & \sum x_{ki}^2 \\
\end{bmatrix}
\]

with a column of ones in X. It is a symmetric matrix. (Note that \(x_j\) refers to the \(j^{th}\) column of X.)
It is a good exercise to try to figure out \(\beta, s^2, N\) and \(R^2\) from the cross products matrix.

**HETEROSKEDASTICITY:**

\(V(y) = V \neq \sigma^2 I\)

Is the LS estimator unbiased? Is it BLUE?

**Proposition:** Under the assumption of heteroskedasticity, \(V(\hat{\beta}) = (X'X)^{-1} X' VX (X'X)^{-1}\).

**Proof:**

\[
V(\hat{\beta}) = E(X'X)^{-1} X' \varepsilon X (X'X)^{-1} = (X'X)^{-1} X' VX (X'X)^{-1}
\]

Suppose \(\Sigma = V(\varepsilon)\) is a diagonal matrix. Then \(X' \Sigma X = E \sum X_i' \varepsilon_i^2 X_i\) where \(X_i\) is the \(i^{th}\) row of X.

Note that the cross products have expectation 0. This suggests using \(\sum X_i' \varepsilon_i^2 X_i\). So we can estimate standard errors under the assumption that \(V(y)\) is diagonal.

**Testing for heteroskedasticity:**

1. **Goldfeld-Quandt test:**

Suppose we suspect that \(\sigma_i^2\) varies with \(x_i\) (one of the X variables or with time in a time series).

Then reorder the observations in the order of \(x_i\).

Suppose \(N\) is even. If \(\varepsilon\) was observed, then

\[
\frac{\varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_{N/2}^2}{\varepsilon_{[N/2]+1}^2 + \varepsilon_{([N/2]+2]}^2 + \ldots + \varepsilon_N^2} \sim F(N/2, N/2)
\]
could be used. We are tempted to use e, but we can't because the first N/2 e’s are not
independent of the last and thus the ratio of sums of square will not be distributed as F.

Here comes the Goldfeld-Quandt trick:

Estimate e separately for each half of the sample with K parameters. The statistic is
\[ F((N/2)-K, (N/2)-K). \]

It turns out that this "works" better if you delete the middle N/3 observations. Sampling
experiments show that the power is increased if the middle third of the observations is eliminated.

Of course, by dropping part of the sample and by doing separate estimation some power is lost
relative to an optimal test (see lecture 10).

2. Breusch-Pagan test:

The disturbances \( \varepsilon_i \) are assumed to be normally and independently distributed with variance
\[ \sigma_i^2 = h(z'_i\alpha) \] where \( h \) denotes a function, and \( z'_i \) is a 1xP vector of variables influencing
heteroskedasticity.

Let \( Z \) be an NxP matrix with row vectors \( z'_i \). Some of the variables in \( Z \) could be same as the
variables in \( X \).

Regress \( e^2 / \sigma^2_{\text{ML}} \) (which is an Nx1 vector) on \( Z \), including an intercept term.

Note that (Sum of squares due to \( Z \))/2 \( \sim \chi^2(P-1) \) approximately. The factor 1/2 appears here
since under normality the variance of \( e^2/\sigma^2 \) is 2 \( (Ee^4 = 2\sigma^4) \). An alternative approach (Koenker)
drops normality and estimates the variance of \( e_i^2 \) directly by \( N^{-1} \sum (e_i^2 - \hat{\sigma}^2)^2 \). The resulting
statistic can be obtained by regressing \( e^2 \) on \( z \) and looking at \( NR^2 \) from this regression.

Other tests are available for time series.