The outline for Unit 4

UNIT 1. Introduction: The regression model. ✓

UNIT 2. Estimation principles. ✓

UNIT 3: Hypothesis testing principles. ✓

UNIT 4: Heteroscedasticity in the regression model.

4.1 Ordinary Least Squares.
4.2 Generalized Least Squares.
4.3 Inefficiency of OLS.
4.4 Testing for heteroscedasticity.

UNIT 5: Endogeneity of regressors.

Introduction

One of the assumptions we've made up to now is that

$$\varepsilon_t \sim IID(0, \sigma^2).$$

Now we'll investigate the consequences of nonidentically and/or dependently distributed errors. We'll assume strong exogeneity. The model is

$$\begin{aligned} y &= X\beta + \varepsilon \\ \mathrm{E}(\varepsilon|X) &= 0 \\ V(\varepsilon|X) &= \Sigma \end{aligned}$$

where Σ is a general symmetric positive definite matrix.

Introduction - 2

- The case where Σ is a diagonal matrix gives uncorrelated, nonidentically distributed errors. This is known as *heteroscedasticity.*

- The case where Σ has the same number on the main diagonal but nonzero elements off the main diagonal gives identically dependently distributed errors. This is known as *autocorrelation*.

- The general case combines heteroscedasticity and autocorrelation. This is known as "*nonspherical*" disturbances.

Effects of nonspherical disturbances on the OLS estimator

The least square estimator $\hat{\beta} = (X'X)^{-1}X'y$:

- Is unbiased, as before.
- The variance of $\hat{\beta}$, supposing X is nonstochastic, is

$$\begin{aligned} \mathrm{E}\left[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\right] &= \mathrm{E}\left[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}\right] \\ &= (X'X)^{-1}X'\Sigma X(X'X)^{-1} \end{aligned}$$

  Due to this, any test statistic that is based upon $\widehat{\sigma^2}$ or the probability limit $\widehat{\sigma^2}$ is invalid.

  In particular, the formulas for the t, F, χ^2 based tests given above do not lead to statistics with these distributions.

- $\hat{\beta}$ is still consistent.

- If ε is normally distributed, then, conditional on X

$$\hat{\beta} \sim N\left(\beta, (X'X)^{-1}X'\Sigma X(X'X)^{-1}\right)$$

- Without normality, and unconditional on X we still have

$$\sqrt{n}\left(\hat{\beta} - \beta\right) = \sqrt{n}(X'X)^{-1}X'\varepsilon = \left(\frac{X'X}{n}\right)^{-1} n^{-1/2}X'\varepsilon$$

Define the limiting variance of $n^{-1/2}X'\varepsilon$ as

$$\lim_{n\to\infty} \mathrm{E}\left(\frac{X'\varepsilon\varepsilon'X}{n}\right) = \Omega$$

so we obtain $\sqrt{n}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N\left(0, Q_X^{-1}\Omega Q_X^{-1}\right)$.

The GLS estimator

Suppose Σ were known.

Then one could form the Cholesky decomposition

$$PP' = \Sigma^{-1}$$

We have $PP'\Sigma = I_n$ so $P'\left(P\Sigma P'\right) = P'$, which implies that $P'\Sigma P = I_n$

Consider the model

$$P'y = P'X\beta + P'\varepsilon,$$

or,

$$y^* = X^*\beta + \varepsilon^*.$$

The variance of $\varepsilon^* = P'\varepsilon$ is $\mathrm{E}(P'\varepsilon\varepsilon' P) = P'\Sigma P = I_n$

The GLS estimator - 2

Therefore, the model

$$\begin{aligned} y^* &= X^*\beta + \varepsilon^* \\ \mathrm{E}(\varepsilon^*) &= 0 \\ V(\varepsilon^*) &= I_n \\ \mathrm{E}(X^{*\prime}\varepsilon^*) &= 0 \end{aligned}$$

satisfies the classical assumptions (with modifications to allow stochastic regressors and nonnormality of ε). The GLS estimator is simply OLS applied to the transformed model:

$$\begin{aligned} \hat{\beta}_{GLS} &= (X^{*\prime}X^*)^{-1}X^{*\prime}y^* \\ &= (X'PP'X)^{-1}X'PP'y \\ &= (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y \end{aligned}$$

The GLS estimator - 3

The GLS estimator is unbiased in the same circumstances under which the OLS estimator is unbiased. For example, assuming X is nonstochastic

$$\begin{aligned} \mathrm{E}(\hat{\beta}_{GLS}) &= \mathrm{E}\left\{(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y\right\} \\ &= \mathrm{E}\left\{(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}(X\beta+\varepsilon\right\} \\ &= \beta. \end{aligned}$$

The variance of the estimator, conditional on X can be calculated using

$$\begin{aligned} \hat{\beta}_{GLS} &= (X^{*\prime}X^{*})^{-1}X^{*\prime}y^{*} \\ &= (X^{*\prime}X^{*})^{-1}X^{*\prime}\left(X^{*}\beta+\varepsilon^{*}\right) \\ &= \beta+(X^{*\prime}X^{*})^{-1}X^{*\prime}\varepsilon^{*} \end{aligned}$$

The GLS estimator - 4

So

$$
\begin{aligned}
\mathrm{E}\left\{\left(\hat{\beta}_{GLS}-\beta\right)\left(\hat{\beta}_{GLS}-\beta\right)'\right\} &= \mathrm{E}\left\{(X^{*\prime}X^{*})^{-1}X^{*\prime}\varepsilon^{*}\varepsilon^{*\prime}X^{*}(X^{*\prime}X^{*})^{-1}\right\} \\
&= (X^{*\prime}X^{*})^{-1}X^{*\prime}X^{*}(X^{*\prime}X^{*})^{-1} \\
&= (X^{*\prime}X^{*})^{-1} \\
&= (X'\Sigma^{-1}X)^{-1}
\end{aligned}
$$

Either of these last formulas can be used.

- All the previous results regarding the desirable properties of the OLS estimators hold, when dealing with the transformed model.

- Tests are valid, using the previous formulas, as long as we substitute X^{*} in place of X. Furthermore, any test that involves σ^2 can set it to 1. This is preferable to re-deriving the appropriate formulas.

The GLS estimator - 5

- The GLS estimator is more efficient than the OLS estimator. This is a consequence of the Gauss-Markov theorem, since the GLS estimator is based on a model that satisfies the classical assumptions but the OLS estimator is not.

- As one can verify by calculating fonc, the GLS estimator is the solution to the minimization problem

$$\hat{\beta}_{GLS} = \arg\min (y - X\beta)' \Sigma^{-1} (y - X\beta)$$

so the *metric* Σ^{-1} is used to weight the residuals.

Feasible GLS

The problem is that Σ isn't known usually, so this estimator isn't available.

- Consider the dimension of Σ : it's an $n \times n$ matrix with $\left(n^2 - n\right)/2 + n = \left(n^2 + n\right)/2$ unique elements.
- The number of parameters to estimate is larger than n and increases faster than n. There's no way to devise an estimator that satisfies a LLN without adding restrictions.
- The *feasible GLS estimator* is based upon making sufficient assumptions regarding the form of Σ so that a consistent estimator can be devised.

Feasible GLS - 2

Suppose that we *parameterize* Σ as a function of X and θ, where θ may include β as well as other parameters, so that

$$\Sigma = \Sigma(X, \theta).$$

If we can consistently estimate θ, we can consistently estimate Σ, as long as $\Sigma(X, \theta)$ is a continuous function of θ (by the Slutsky theorem).

In this case,

$$\widehat{\Sigma} = \Sigma(X, \hat{\theta}) \xrightarrow{p} \Sigma(X, \theta)$$

If we replace Σ in the formulas for the GLS estimator with $\widehat{\Sigma}$, we obtain the *FGLS estimator*.

The FGLS estimator shares the same asymptotic properties as GLS.

Feasible GLS - 3

In practice, the usual way to proceed is:

1. Define a consistent estimator of θ. This is a case-by-case proposition, depending on the parameterization $\Sigma(\theta)$.

2. Form $\widehat{\Sigma} = \Sigma(X, \hat{\theta})$

3. Calculate the Cholesky factorization $\widehat{P} = Chol(\hat{\Sigma}^{-1})$.

4. Transform the model using

$$\hat{P}'y = \hat{P}'X\beta + \hat{P}'\varepsilon$$

5. Estimate using OLS on the transformed model.

OLS with heteroscedastic consistent varcov estimation

Eicker (1967) and White (1980) showed how to modify test statistics to account for heteroscedasticity of unknown form.

The OLS estimator has asymptotic distribution

$$\sqrt{n}\left(\hat{\beta}-\beta\right)\overset{d}{\rightarrow}N\left(0,Q_X^{-1}\Omega Q_X^{-1}\right)$$

Recall that we defined $\lim_{n\rightarrow\infty}\mathcal{E}\left(\frac{X'\varepsilon\varepsilon'X}{n}\right)=\Omega.$

This matrix has dimension $K\times K$ and can be consistently estimated, even if we can't estimate Σ consistently. The consistent estimator, under heteroscedasticity but no autocorrelation is

$$\widehat{\Omega}=\frac{1}{n}\sum_{t=1}^{n}x_t'x_t\hat{\varepsilon}_t^2$$

OLS with heteroscedastic consistent varcov estimation - 2

One can then modify the previous test statistics to obtain tests that are valid when there is heteroscedasticity of unknown form.

For example, the Wald test for $H_0 : R\beta - r = 0$ would be

$$n\left(R\hat{\beta} - r\right)'\left(R\left(\frac{X'X}{n}\right)^{-1}\hat{\Omega}\left(\frac{X'X}{n}\right)^{-1}R'\right)^{-1}\left(R\hat{\beta} - r\right) \overset{a}{\sim} \chi^2(q)$$

Detecting heteroscedasticity

Goldfeld-Quandt's test:

The sample is divided in to three parts, with n_1, n_2 and n_3 observations, where $n_1 + n_2 + n_3 = n$. The model is estimated using the first and third parts of the sample, separately, so that $\hat{\beta}^1$ and $\hat{\beta}^3$ will be independent.

Then we have

$$\frac{\hat{\varepsilon}^{1\prime}\hat{\varepsilon}^1}{\sigma^2} = \frac{\varepsilon^{1^\prime} M^1 \varepsilon^1}{\sigma^2} \xrightarrow{d} \chi^2(n_1 - K)$$

and

$$\frac{\hat{\varepsilon}^{3\prime}\hat{\varepsilon}^3}{\sigma^2} = \frac{\varepsilon^{3^\prime} M^3 \varepsilon^3}{\sigma^2} \xrightarrow{d} \chi^2(n_3 - K)$$

so

$$\frac{\hat{\varepsilon}^{1\prime}\hat{\varepsilon}^1/(n_1 - K)}{\hat{\varepsilon}^{3\prime}\hat{\varepsilon}^3/(n_3 - K)} \xrightarrow{d} F(n_1 - K, n_3 - K).$$

Detecting heteroscedasticity - 2

Goldfeld-Quandt's test: (cont.)

- The motive for dropping the middle observations is to increase the difference between the average variance in the subsamples, supposing that there exists heteroscedasticity. This can increase the power of the test.

- On the other hand, dropping too many observations will substantially increase the variance of the statistics $\hat{\varepsilon}^{1\prime}\hat{\varepsilon}^{1}$ and $\hat{\varepsilon}^{3\prime}\hat{\varepsilon}^{3}$.

- A rule of thumb, based on Monte Carlo experiments is to drop around 25% of the observations.

- If one doesn't have any ideas about the form of the heteroscedasticity the test will probably have low power since a sensible data ordering isn't available.

Detecting heteroscedasticity - 3

White's test:

When one has little idea if there exists heteroscedasticity, and no idea of its potential form, the White test is a possibility.

The idea is that if there is homoscedasticity, then $\mathrm{E}(\varepsilon_t^2|x_t) = \sigma^2, \forall t$, so that x_t or functions of x_t shouldn't help to explain $\mathrm{E}(\varepsilon_t^2)$.

The test works as follows:

1. Since ε_t isn't available, use the consistent estimator $\hat{\varepsilon}_t$ instead.

2. Regress
$$\hat{\varepsilon}_t^2 = \sigma^2 + z_t'\gamma + v_t$$
where z_t is a P -vector. z_t may include some or all of the variables in x_t, as well as other variables. White's original suggestion was to use x_t, plus the set of all unique squares and cross products of variables in x_t.

Detecting heteroscedasticity - 4

White's test: (cont.)

3. Test the hypothesis that $\gamma = 0$. The qF statistic in this case is

$$qF = \frac{P\left(ESS_R - ESS_U\right)/P}{ESS_U/\left(n-P-1\right)}$$

and dividing both numerator and denominator by $ESS_R = TSS_U$, we get

$$qF = \left(n-P-1\right)\frac{R^2}{1-R^2}.$$

Note that this is the R^2 of the artificial regression used to test for heteroscedasticity, not the R^2 of the original model.

An asymptotically equivalent statistic, under the null of no heteroscedasticity is

$$nR^2 \overset{a}{\sim} \chi^2(P).$$