# Econometrics 1, Class Week 5

## Back-up Video: Class week 5 video (click here)

Learning Outcomes

(1) Diagnostic tests:
   (a) Heteroskedasticity
   (b) Serial Correlation
   (c) Non-normality
   (d) Non-linearity
   (e) Exact Multicollinearity.

(2) Robust SEs vs. GLS.

Prerequisites

None.

## (1) Overview

Under Gauss-Markov (GM) assumptions, OLS estimator is BLUE.

Diagnostic tests attempt to check whether GM assumptions are valid.

Negative test results point to problems, not to solutions.

Tests relate to GM assumptions on (specification of) conditional mean and on conditional homoskedasticity.

General principle: If a GM assumption that relates to first (second) conditional moment – i.e. linearity of conditional mean and regressors exogenous (conditional homoskedasticity of errors) – is not satisfied, then the Gauss-Markov Theorem's conclusion about the first (second) conditional moment – i.e. conditional unbiasedness (minimum variance) – is not valid.

## (a) Heteroskedasticity

If present, failure of conditional second moment GM assumption.

Test based on $T$-vector of regression residuals $\hat{\epsilon} = \mathbf{y} - \mathbf{X}\hat{\beta}$, or rather their squares $\hat{\epsilon}_t^2$, $t = 1, \cdots, T$.

Need to have some sort of "model" of what drives potential heteroskedasticity:

$$\hat{\epsilon}_t^2 = \mathbf{z}_t'\gamma + \nu_t, \tag{1}$$

where $\mathbf{z}$ is a vector of possible drivers of heteroskedasticity.

Test null hypothesis of absence of heteroskedasticity, i.e. whether $\gamma \neq \mathbf{0}$ (joint hypothesis).

Here, $\mathbf{z}_t$ could include elements (and functions) of the regressors $\mathbf{x}_t$, in time-series models lagged squared residuals ($\hat{\epsilon}_{t-s}^2$, $s = 1, \cdots$).

Often, presence of heteroskedasticity is interpreted as model misspecification; but there exist also econometric models that are intrinsically heteroskedastic.

In general, economic theory should guide model of data generating process, model specification and model testing.

## (b) Serial Correlation

Only applicable in time-series context.

Consider regression

$$\hat{\epsilon}_t = \alpha_0 + \alpha_1 \hat{\epsilon}_{t-1} + \alpha_2 \hat{\epsilon}_{t-2} + \cdots + \nu_t. \tag{2}$$

Test null hypothesis of absence of serial correlation, i.e. that $\alpha_1 = \alpha_2 = \cdots = 0$ (joint hypothesis).

If test rejects, then should reconsider specification of conditional mean function in main regression, i.e. include lags of dependent variable among regressors $\mathbf{x}_t$.

## Non-Normality

Rarely relevant.

Test based on moments of the normal distribution $Z \sim N(\mu, \sigma^2)$. Know: Normal distribution is *completely* characterised by its mean and its variance. They determine every other moment of the normal. E.g.

$$
\begin{aligned}
\mathbb{E}[(Z-\mu)^3] &= 0 \quad \text{b/c of symmetry} \\
\mathbb{E}\left[\left(\frac{Z-\mu}{\sigma}\right)^4\right] &= 3. \qquad (3)
\end{aligned}
$$

Jarque-Bera test uses these: Test statistic

$$
J = T\left(\frac{\hat{\mu}_3^2}{6\hat{\mu}_2^3} + \frac{1}{24}\left(\frac{\hat{\mu}_4}{\hat{\mu}_2^2} - 3\right)\right), \qquad (4)
$$

where $\hat{\mu}_k$ is the $k$th sample moment, $k = 1, \cdots, 4$,

$$
\hat{\mu}_k = \frac{1}{T}\sum_{t=1}^{T}(\hat{\epsilon}_t - \bar{\hat{\epsilon}})^k. \qquad (5)
$$

Under null hypothesis of normality,

$$
J \sim \chi_2^2. \qquad (6)
$$

## Non-Linearity

This is Reset version of (1).

I.e.

$$\hat{\epsilon}_t^2 = \gamma_0 + \mathbf{x}_t'\gamma_1 + \gamma_2\hat{y}_t^2 + \gamma_3\hat{y}_t^3 + \cdots + \nu_t. \tag{7}$$

Test null hypothesis of absence of non-linearity, i.e. $\gamma_2 = \gamma_3 = \cdots = 0$ (joint hypothesis).

## Exact Multicollinearity

Exact multicollinearity implies that $\text{rk}(\mathbf{X}) < k$, where $k$ is the number of columns of $\underset{T \times k}{\mathbf{X}}$.

Statistical software (e.g. Stata) will simply drop one (or more) regressors.

## (2) Robust SEs vs. GLS

For GLS, need to have a model for $\Omega$ in $\text{var}(\mathbf{y}|\mathbf{X}) = \sigma_0^2\Omega$ (i.e. $\frac{1}{2}T(T-1)$ additional parameters), to estimate

$$\text{var}(\hat{\beta}_{GLS}) = \sigma_0^2(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}. \tag{8}$$

Instead, Eicker-Huber-White robust SEs can be obtained from

$$\text{var}_{EHW}(\hat{\beta}_{OLS}) = \sigma_0^2(\mathbf{X'X})^{-1}\left(\sum_{t=1}^{T}\hat{\epsilon}_t^2\mathbf{x}_t\mathbf{x}_t'\right)(\mathbf{X'X})^{-1}. \tag{9}$$

There is no assumption on specific form of heteroskedasticity. Also, no additional parameters to estimate.