References to Lecture 3

- HN: CH: 6 and 8 (matrix algebra) and multiple regression model
- DM: Chapter 3-5 (classical regression inference theory)
- Lecture note 2 on the web-page (optional about Wald and LM tests).
Classical inference theory I

- We have so far moved from simple statistical and econometric models to a matrix formulation of the multiple, k-variable, regression model.

- Although we have used the MLE principle as the “red tread”, instead of OLS, we see that we are now where many introductory courses ends: The classical regression model used for hypothesis testing and other type of inference.
Classical inference theory II

The basic result is that, if the “classical assumptions” hold for the disturbances, test statistics and confidence intervals can be based on:

\[ \hat{\beta} = (X'X)^{-1}X'y \]  

(1)

\[ \text{Var}(\hat{\beta}) = \sigma^2 (X'X)^{-1} \]  

(2)

\[ \hat{\sigma}^2 = \frac{\hat{\epsilon}'\hat{\epsilon}}{n-k} \]  

(3)
Classical inference theory III

at least asymptotically. But remember also:

<table>
<thead>
<tr>
<th>$\times$</th>
<th>Disturbances $\varepsilon$ are:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>heteroscedastic</td>
</tr>
<tr>
<td>exogenous</td>
<td>$\hat{\beta}$ unbiased consistent</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}$ unbiased consistent</td>
</tr>
<tr>
<td>predetermined</td>
<td>biased consistent</td>
</tr>
<tr>
<td></td>
<td>biased inconsistent</td>
</tr>
</tbody>
</table>
Classical inference theory IV

▶ Despite the change from scalar notation to matrices: Note the familiar role of restricted and unrestricted sum of squared residuals in many of the tests!

▶ Take care to note:
  ▶ The importance of $\varepsilon \overset{D}{=} N(0, \sigma^2 I \mid X)$ assumption for the regression model disturbances for obtaining exact tests
  ▶ and the importance of $\varepsilon \overset{D}{=} IID(0, \sigma^2 I \mid X)$ for the corresponding asymptotic tests.

▶ Note, in the table, the asymptotic tests are also valid for the case where the explanatory variables are predetermined, rather than exogenous.

▶ But here we are humping the gun: since pre-determinedness is a feature of models for time series-data.
Classical inference theory V

- Chapter 5 in DM reviews in particular the role of heteroscedasticity consistent covariance matrices (ch 5.5), in situations where $\hat{\text{Var}}(\hat{\beta})$ above is “wrong”. We use the PcGive version of these in seminar exercises.

- We have followed HN which give a clear presentation of LR-test.

- The LR tests are intuitive and are based on restricted and unrestricted estimation (maximization) of the model in question.

- There are two other classical test principles called Wald-test (example: a t-test) and Lagrange-multiplier tests.

- Lecture note 2 (optional) gives a brief exposition, for reference.
The delta method

- When we estimate a linear-in-parameter conditional expectation function, the purpose is sometimes to test hypotheses about derived parameters that are non-linear functions of the regression coefficients.

- The so called delta-method (Ch 5.6 in DM) gives the asymptotically valid estimate of the variance of the derived parameter in this case.

- It is make use of a Taylor-expansions and relatively weak assumptions.

- It is covered in modern elementary book, for example Hill, Griffiths and Lim as many of you know. Bårdsen and Nymoen (2011) has in Ch 4!)
The delta method II

- See also the slide set to Lecture 9 in ECON 4150 spring 2013
- Assume the simple regression model

\[ Y_i = \beta_1 + \beta_2 X_i + \epsilon_i \]

and that we are interested in the derived parameter

\[ \theta = \frac{\beta_2}{\beta_1} \]

The asymptotic variance is then

\[
\hat{\text{Var}}(\hat{\theta}) \approx \left( \frac{1}{\hat{\beta}_1} \right)^2 \left[ \text{Var}(\hat{\beta}_2) + \hat{\theta}^2 \text{Var}(\hat{\beta}_1) - 2\hat{\theta} \text{Cov}(\hat{\beta}_1, \hat{\beta}_2) \right]
\]

(4)

which actually covers many applications in econometrics.
The delta method III

- [cf. Ch. 5.6 in DM covers the case of $\theta = g\left(\frac{\beta_2}{\beta_1}\right)$ where $g(\cdot)$ is a known monotonic and differentiable function, as well as the vector case.]

- In BN, the formulae in (4) is used for **long-run coefficients** of dynamic models.
NLS estimation

So far the conditional expectation functions have been linear in parameters.

As we know, this allows a great deal of flexibility (through non-linear variable transformation) in the specification of the functional form.

Nevertheless: Sometimes necessary or appealing to estimate a model which is non-linear in the parameters.

The sum of squared residuals that we want to minimize is then

$$SSR(\hat{\beta}) = \sum_{i=1}^{n} (Y_i - x_i(\hat{\beta}))^2$$

(5)

where $x_i(\hat{\beta})$ is the analogue to $x_i\hat{\beta}$ in the linear case.
NLS estimation II

- The NLS estimator is consistent under mild assumptions.
- Minimization of (5) requires numerical optimization.
- PcGive has a good algorithm for computing NLS estimates.

Illustrate by estimation of Phillips curve natural rate ($U^{nat}$):

$$INF_i = \beta_1(U_i - \beta_2) + \epsilon_i, \ i = 1, 2, \ldots, n$$

- A non-linear regression function. $\beta_2$ can be interpreted as the natural rate (a parameter), since

$$E(INF \mid U_i = \beta_2) = 0$$

- Estimate $U^{nat}$ by PcGive with Norwegian annual data.
- Can we use a linear model to estimate, and test hypotheses about, $U^{nat}$?
- See extra slide about Natural rate estimation.
GLS I

We know that both heteroskedasticity and autocorrelation require a different specification than $Var(\varepsilon) = E(\varepsilon\varepsilon') = \sigma^2 I$ in the linear regression model

$$y = X\beta + \varepsilon \quad (7)$$

More generally we have:

$$Var(\varepsilon) = \sigma^2 \Omega, \sigma^2 > 0$$

where $\Omega$ is $n \times n$ is symmetric and **Positive Definite:**

$$z'\Omega z > 0 \text{ for all } z \neq 0$$

quadratic form in $n$ variables

with $z' = (Z_1, Z_2, \ldots, Z_n)$. 
Result from linear algebra: For a PD matrix $\Omega$ there exists a $n \times n$ matrix $\Psi$ which is invertible (non-singular), with properties

$$\Psi \Omega \Psi' = I$$  \hspace{1cm} (8)

$$\Psi' \Psi = \Omega^{-1} \iff \Psi \Psi' = \Omega^{-1} \text{ (symmetry of } \Omega)$$  \hspace{1cm} (9)

Multiplication from the left in (7) by $\Psi'$ gives:

$$\Psi' Y = \Psi' X \beta + \Psi' \varepsilon$$  \hspace{1cm} (10)
Because

$$Var(\varepsilon_*) = E(\varepsilon_*\varepsilon'_*) = E(\Psi'\varepsilon\varepsilon'\Psi) = \Psi'\sigma^2\Omega\Psi = \sigma^2I$$  \hspace{1cm} (11)

the OLS estimator for $\beta$ from (10) is BLUE under the assumption of exogeneity of the $X$-variables.

This estimator is the **Generalized Least Squares** estimator (GLS) and is, by reference to minimization of residuals and to method-of-moments, given by:

$$\hat{\beta}_{GLS} = (X'_* X'_*)^{-1} X'_* y_* = (X' \Psi \Psi' X)^{-1} X' \Psi \Psi' y$$

$$= (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y$$  \hspace{1cm} (12)

with covariance matrix.

$$Var(\hat{\beta}_{GLS}) = (X'_* \sigma^{-2} I X'_*)^{-1} = \sigma^2 (X' \Omega^{-1} X)^{-1}$$
Weighted Least Squares example I

Assume that the only departure from the classical assumptions is heteroskedasticity, and that it takes the form:

\[ \text{Var}(\varepsilon) = \sigma^2 \Omega = \sigma^2 
\begin{pmatrix}
X_{21} & 0 & \cdots & 0 \\
0 & X_{22} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & X_{2n}
\end{pmatrix}. \]

Then (show!):

\[ \Omega^{-1} = 
\begin{pmatrix}
\frac{1}{X_{21}} & 0 & \cdots & 0 \\
0 & \frac{1}{X_{21}} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \frac{1}{X_{2n}}
\end{pmatrix}. \]
Weighted Least Squares example II

and we can compute the GLS estimator from (12). Moreover, we see that \( \Psi \Omega \Psi' = I \) if \( \Psi = \Psi' \) is specified as:

\[
\Psi' = \begin{pmatrix}
\sqrt{\frac{1}{X_{21}}} & 0 & \cdots & 0 \\
0 & \sqrt{\frac{1}{X_{21}}} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \sqrt{\frac{1}{X_{2n}}}
\end{pmatrix}
\]

which gives the weights that we apply to obtain \( y_\star \) and \( X_\star \) in (10).
Feasible GLS I

▶ In most practical situations $\Omega$, is unknown as is replaced by a consistent estimator $\hat{\Omega}$. This is based on
  ▶ the OLS residuals $\hat{\varepsilon}_i$;
  ▶ and an **assumed form** of the heteroskedasticity or autocorrelation (see DM section 7.4 for example).

▶ In practice: an auxiliary regression between $\hat{\varepsilon}_i$ and a set of observable variables (often some of the $X$ variables).

▶ As long as this procedure gives a consistent estimator of $\Omega$, the **feasible** GLS estimator

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

is both **consistent** and **asymptotically efficient**.
Seemingly unrelated regression equations (SURE) I

- Sometimes the research purpose lead us to consider more than one regression equation.

\[
y_j = X_j \beta + \epsilon_j, \ j = 1, 2, \ldots, m
\]

(13)

where we assume that all \( \epsilon_j \) have the classical properties.

- System of Engle-equations is a classical example.

- If we want we can stack the data into large matrices, and apply OLS.

- The resulting least squares estimator has a GLS interpretation and is called SURE.
Seemingly unrelated regression equations (SURE) II

- However SURE reduces to equation-by-equation OLS if
  1. All $X_j$ matrices contain the same $k$-regressors
  2. The disturbances are uncorrelated.

We will leave this version of SURE for micro econometrics. In this course the common form of SURE systems Vector Autoregressive Models, which are dynamic.