ECMT2150 FINAL EXAM NOTES

Week 1

Cross section
- Same point in time
- Random sampling (independent observations)
- Order doesn't matter

Time series
- Collected observations over time
- Chronological ordering important
- Observation frequency important
- Seasonality needs to be accounted for

Pooled cross sections
- 2 or more sets of cross sectional data at diff points in time
- Same variables but diff units
- Useful to look at relationships before and after introduction of something (e.g. govt policy)

Panel/longitudinal data
- Time series for each cross sectional variable
- Same units over time
- Difficult/expensive to obtain than pooled cross section

Week 2

Zero Conditional Mean (ZCM) assumption

$$E(u|x_1, \dots, x_k) = 0$$

For the multiple regression model
- It requires the average of u to be the same for any given values of x's
- It implies that the factors in u are uncorrelated with all x's
- It is a key condition for the estimators to be unbiased and consistent
- It defines the pop regression function
  - $E(y|x_1, \dots, x_k) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$
    U is estimated to equal zero

$E(u) = 0$ normalises the effect of the unobserved factors on the dependent variable

- The sum of residuals is zero
- Sample covar(u,x)=0
- The mean point $(\bar{x}, \bar{y})$ is always on the SRF (or OLS regression line)

PRF vs SRF
SRF = PRF (population regression function)
"on average" or "when $n \to \infty$"

Sums of squares
Each y_i may be decomposed into $y_i = \hat{y}_i + \hat{u}_i$
Measuring variations from $\bar{y}$

- Total sum of squares (total variation in y_i):
  $SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$,
- Explained sum of squares (variation in $\hat{y}_i$):
  $SSE = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$,
- sum of squared **R**esiduals (variation in $\hat{u}_i$):
  $SSR = \sum_{i=1}^{n} \hat{u}_i^2$.
- It can be shown that $SST = SSE + SSR$.

Coefficient of determination/R-squared

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}.$$

- larger R^2, better fit;
- $0 \le R^2 \le 1$.

Can only be negative if the model doesn't contain an intercept.

- It is not advisable to put too much weight on this measure when comparing models
- If R-squared = 0.165 then 16.5% of y ix explained by x

Nonlinear relationships

- The parameters need to be linear for OLS
- Beta is then interpreted differently

▸ Linear: $\hat{y} = b_1 + b_2 x$, where b_2 is the partial effect $\frac{\Delta\hat{y}}{\Delta x}$. If x goes up by 1 unit, $\hat{y}$ goes up by b_2 units

▸ Linear–log: $\hat{y} = b_1 + b_2 \ln x$, where $b_2 \approx \frac{\Delta\hat{y}}{\Delta x / x}$, and the partial effect is b_2/x. If x goes up by 1%, $\hat{y}$ goes up by $b_2/100$ units

▸ Log–linear: $\widehat{\ln y} = b_1 + b_2 x$, where $b_2 \approx \frac{\Delta\hat{y}/\hat{y}}{\Delta x}$, and the partial effect is $b_2 \cdot \hat{y}$. If x goes up by 1 unit, $\hat{y}$ goes up by $b_2 \cdot 100\%$

▸ Log–log: $\widehat{\ln y} = b_1 + b_2 \ln x$, where $b_2 \approx \frac{\Delta\hat{y}/\hat{y}}{\Delta x / x}$, and the partial effect is $b_2 \cdot \hat{y}/x$. If x goes up by 1%, $\hat{y}$ goes up by $b_2\%$

Model	Dependent var	Independent var	Interpretation of β_1
Linear-linear	Y	X	*When* $\Delta x = 1, \Delta y = \beta_1$
Linear-log	Y	Log(x)	$\text{when } \Delta x = 1\%, \Delta y = \left(\frac{\beta_1}{100}\right)$
Log-linear	Log(y)	X	$\text{when } \Delta x = 1, \Delta y = (100\beta_1)\%$
Log-log	Log(y)	Log(x)	$\text{when } \Delta x = 1\%, \Delta y = \beta_1\%$

Underlying assumptions of simple regression model

1. (linear in parameters) In the population model, y is related to x by $y = \beta_0 + \beta_1 x + u$, where ($\beta_0$, β_1) are population parameters and u is disturbance.

2. (random sampling) $\{(x_i, y_i), i = 1,2,\ldots,n\}$ is a random sample drawn from the population model.

3. (sample variation in the explanatory variable) The sample outcomes on x are not of the same value.

4. (zero conditional mean) The disturbance u satisfies $E(u \mid x) = 0$ for any given value of x. For the random sample, $E(u_i \mid x_i) = 0$ for $i = 1,2,\ldots,n$.

5. (Homoscedasticity) $Var(u_i \mid x_i) = \sigma^2 \quad i = 1,2,\ldots,n$.

Under 1-4, the estimators are unbiased

unbiased: $E(\hat{\beta}_1) = \beta_1, \quad E(\hat{\beta}_0) = \beta_0.$

Under 5,
The estimators are homoscedastic, and the variances are constant

– the larger is σ^2, the greater are the variances.

– the larger the variation in x, the smaller the variances.

– As the residual approximates u, the estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{SSR}{n-2} = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-2}.$$

"2" is the number of estimated coefficients

– $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ is known as the **standard error of the regression**, useful in forming the standard errors of $(\hat{\beta}_0, \hat{\beta}_1)$.

Theorem 2.3 (unbiased estimator of σ^2)
Under SLR1 to SLR5, $E(\hat{\sigma}^2) = \sigma^2$.

---

week 3

In general, regression models with multiple x's :

- Allow us to explicitly control for (hold fixed) many factors that affect the dependent variable in order to draw ceteris paribus conclusions
- Provide better explanation of the dependent variable by accommodating flexible functional forms

– The OLS regression line or SRF can be written in the form of changes, holding u fixed:

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1 + \cdots + \hat{\beta}_k \Delta x_k.$$

- The coefficient on x_1 is the partial effect x_1 on y, holding u and the rest of x's fixed $\Delta\hat{y} = \beta_1 \Delta x_1$
- We are able to control (hold fixed) x variables when considering effect of $x_1 on\ y$
- β_1 has a ceteris paribus interpretation when ZCM holds for u

Use *educ, exper, tenure* (years with current employer) to explain hourly *wage*:

$$\widehat{\log(wage)} = .284 + .092\ educ + .004\ exper + .022\ tenure$$

- the coefficient on *educ* : holding *exper* and *tenure* fixed, an extra year of education is **predicted** to increase log(*wage*) by 0.092 (or 9.2% increase in *wage*), which is the ceteris paribus effect under ZCM.
- holding *educ* fixed, the effect of an individual staying at the same firm for an extra year on log(*wage*) :

$$\Delta\widehat{\log(wage)} = .004 + .022 = .026$$

Predicted value and residual

– The *fitted* value

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik}$$

is also known as *predicted* value.

– The residual $\hat{u}_i = y_i - \hat{y}_i$ can be regarded as prediction error.

assumptions of Multiple Regression Model

1. **Linear in parameters**
   In the pop model, y is related to x's by $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$
2. **Random sample**
   With n > k+1
3. **No perfect collinearity**
   None of the x's are constant and there is no perfect linear relationship among x's
4. **ZCM**
   The disturbance (u) satisfies $E(u|x_1, \ldots, x_k) for\ any\ x$

Unbiasedness of OLS estimators

Under assumptions 1-4, estimators are unbiased.

- These are centred around the parameters
- Correctly estimate the parameters, on average
- Will be near the population parameters for a typical sample

Irrelevant explanatory variables

If an irrelevant x is included:

- Means the pop coefficient of that variable is 0
- The OLS estimators are unbiased, so the estimate of that coefficient will typically be near 0
- The inclusion of irrelevant variables has undesirable effects on the variances of OLS estimators

Explanatory variables

If a relevant x is omitted:

- The OLS estimators will be biased
- The direction and size of bias depend on how the omitted is related to the included

The estimated model is $\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$.
It can be shown that OLS is biased: $E(\tilde{\beta}_1) = \beta_1 + \beta_2 \tilde{\delta}$,
where $\beta_2 \tilde{\delta}$ is known as ***omitted variable bias*** and $\tilde{\delta}$ is the coefficient of regressing x_2 on x_1.

Omitted variable bias is zero when:

- $\beta_2 = 0$ $(irrelevant\ variable)$
- $\tilde{\delta} = 0$ $(uncorrelated\ \ x\ variables)$

	$cov(x_1, x_2) > 0$	$cov(x_1, x_2) < 0$
$\beta_2 > 0$	+ ve bias	− ve bias
$\beta_2 < 0$	− ve bias	+ ve bias

Variance of OLS estimators

5. **Homoskedasticity**

$$Var(u_i|x_{i1}, \dots, x_{ik}) = \sigma^2\ for\ i = 1,2,\dots,n$$

Implies $Var(u_i) = \sigma^2$
Requires that the conditional variance of u be unrelated to x's

Gauss-Markov theorem

Assumptions 1-5 are collectively known as **Gauss-Markov assumptions**

5. Is needed to derive a 'simple' formula for the variances of the OLS estimators

Theorem 3.2

Strictly, Theorem 3.2 is about the variances of OLS estimators, **conditional on given *x*.**

Under MLR1 to MLR5, the variances of the OLS estimators are given by:

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1-R_j^2)}, \qquad j = 1,\dots,k,$$

where $SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$, $\bar{x}_j = n^{-1}\sum_{i=1}^{n} x_{ij}$
and R_j^2 is the R-squared from regressing x_j on all other independent variables.

– the larger is σ^2, the greater is $Var(\hat{\beta}_j)$.
– the larger is R_j^2, the greater is $Var(\hat{\beta}_j)$.
– the larger the variation in x_j, the smaller $Var(\hat{\beta}_j)$.

Multicollinearity

- The larger is R_j^2, the greater is $Var(\beta_j)$

− R_j^2 is the R-squared from regressing x_j on all other x's

- The larger R_j^2, the strong x_j is associated with other x's, the less informative x_j
- R_j^2, = 1 (ruled out by assumption 3) implies there is a perfect linear relationship between x_j and other x's (so x_j is redundant)
- High but not perfect correlation between 2 or more independent variables is known as multicollinearity which does not violate ass 3