# EC226 ECONOMETRICS 1

# FORMULA SHEET

# FOR EXAMS

# 1 2-Variable Regression Model

$Y_i = \alpha + \beta X_i + \varepsilon_i$ where $i = 1, \ldots, n$

## 1.1 Normal Equations

$\sum_{i=1}^{n} e_i = 0$; $\sum_{i=1}^{n} e_i X_i = 0$

## 1.2 Least Squares estimates

$b = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$, $a = \bar{Y} - b\bar{X}$

## 1.3 Estimation of the error variance

$s_e^2 = \frac{RSS}{n-2} = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2 - b^2 \sum_{i=1}^{n}(X_i - \bar{X})^2}{n-2}$

## 1.4 Test on the regression slope coefficient

$H_0 : \beta = \beta_0$, $H_1 : \beta \neq \beta_0$

$t = \frac{(b - \beta_0)}{se(b)}$, where $se(b)^2 = \frac{s_e^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$

## 1.5 Standard error of prediction of $Y_{n+1}$ given $X_{n+1}$

$se(Y_{n+1}) = s_e \sqrt{1 + \frac{1}{n} + \frac{(X_{n+1} - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}$

# 2 Multiple Regression Model

$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki} + \varepsilon_i$ where $i = 1, \ldots, n$

## 2.1 Normal Equations

$\sum_{i=1}^{n} e_i = 0$; $\sum_{i=1}^{n} e_i X_{1i} = 0$, $\sum_{i=1}^{n} e_i X_{2i} = 0$; $\ldots$ $\sum_{i=1}^{n} e_i X_{ki} = 0$

## 2.2 Estimation of the error variance

$s_e^2 = \frac{RSS}{n-k-1} = \frac{\sum_{i=1}^{n} e_i^2}{n-k-1}$

## 2.3 R-squared

$R^2 = 1 - \frac{RSS}{TSS}$;
R-bar-squared: $\bar{R}^2 = 1 - \frac{RSS/(n-k-1)}{TSS/(n-1)}$

## 2.4 Test on regression coefficients

(i) Single coefficient is equal to some hypothesised value:
$H_0 : \beta_i = \beta_{i0}$, $H_1 : \beta_i \neq \beta_{i0}$,
$t = \frac{(b_i - \beta_{i0})}{se(b_i)} \sim t_{n-k-1}$

(ii) All slope coefficients are equal to zero:
$H_0 : \beta_1 = \beta_2 ... = \beta_k = 0$, $H_1 :$ Any $\beta_j \neq 0$ $j = 1, \ldots, k$
$F = \frac{R^2}{1-R^2} . \frac{n-k-1}{k} \sim F_{k,n-k-1}$

(iii) A subset of coefficients are equal to zero:
$H_0 : \beta_1 = \beta_2 ... = \beta_j = 0$, $H_1 :$ Any $\beta_i \neq 0$ $i = 1, \ldots, j$
$F = \frac{(RSS^R - RSS^U)}{RSS^U} . \frac{n-k-1}{j} \sim F_{j,n-k-1}$

(iv) The coefficients from some sub-set of observations are equal to those of some other sub-set of observations:
$H_0 : \beta_0^1 = \beta_0^2, \beta_1^1 = \beta_2^2 ... = \beta_k^1 = \beta_k^2$, $H_1 :$ Any $\beta_j^1 \neq \beta_j^2$, $j = 0, \ldots, k$
$F = \frac{[RSS^R - (RSS^1 + RSS^2)]/(k+1)}{(RSS^1 + RSS^2)/[n-2(k+1)]} \sim F_{k+1,n-2(k+1)}$

## 2.5 Omitted Relevant Variables

$Y_i = \delta_0 + \delta_1 X_i + u_i$
$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \varepsilon_i$
$E(\hat{\delta}_1) = \beta_1 + \beta_2 \frac{cov(X_i, Z_i)}{Var(X_i)}$

## 2.6 Measurement Error

$Y_i = \beta_0 + \beta_1 X_i^T + \varepsilon_i$
$X_i^R = X^T + u_i$
$E(\hat{\beta}_1) = \beta_1 - \beta_1 \frac{cov(X_i^R, u_i)}{Var(X_i^R)}$

## 2.7 Selection Criteria

(i) AIC: $\ln(RSS)/n + 2k/n$

(ii) HQ: $\ln(RSS)/n + 2k \ln\ln(n)/n$

(iii) BIC: $\ln(RSS)/n + k\ln(n)/n$

# 3 Limited Dependent Variable Model

## 3.1 Linear Probability Model

$P[Y_i = 1] = X_i'\beta$

## 3.2 Logit Model

$P[Y_i = 1] = F(X_i'\beta) = \frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)} = \Lambda(X_i'\beta)$
pdf for logit model:
$f(X_i'\beta) = \lambda(X_i'\beta) = \Lambda(X_i'\beta)[1 - \Lambda(X_i'\beta)]$

## 3.3 Probit Model

$P[Y_i = 1] = F(X_i'\beta) = \int\limits_{-\infty}^{X_i'\beta} (2\pi)^{-\frac{1}{2}} \exp(-z/2)dz = \Phi(X_i'\beta)$
pdf for probit model:
$f(X_i'\beta) = \phi(X_i'\beta) = (2\pi)^{-\frac{1}{2}} \exp(-(X_i'\beta)^2/2)$

## 3.4 Interpreting Coefficients (for continuous variable)

$\frac{\partial E(Y_i)}{\partial X_{ji}} = \frac{\partial F(X_i'\beta)}{\partial (X_i'\beta)}.\beta_j$
where $\frac{\partial F(X_i'\beta)}{\partial (X_i'\beta)} = f(X_i'\beta)$ and $f(X_i'\beta)$ is the pdf.

# 4 Time Series Models

$Y_t = \alpha + \beta_1 X_{1t} + \beta_2 X_{2t} + ... + \beta_k X_{kt} + \varepsilon_t \;\; t = 1, \ldots, T$

Normal Equations: $\sum_{t=1}^{T} e_t = 0$; $\sum_{t=1}^{T} e_t X_{1t} = 0$, $\sum_{t=1}^{T} e_t X_{2t} = 0$; $\ldots$ $\sum_{t=1}^{T} e_t X_{kt} = 0$

Durbin-Watson Test statistic: $d = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$

Durbin's h-statistic: $h = \phi \sqrt{\frac{T}{1 - T s_c^2}}$, where $s_c^2$ = Estimated variance of coefficient on lagged dependent variable and $\phi$ = Estimated 1st autocorrelation term.

## 4.1 Serial Correlation

### 4.1.1 AutoRegressive Model (AR(p))

$Y_t = \mu + \sum_{j=1}^{p} \phi_j Y_{t-j} + \varepsilon_t$
$E(Y_t) = \mu / (1 - \phi_1 - \phi_2 - \ldots - \phi_p)$
Yule Walker Equations:
$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} \ldots + \phi_p \gamma_{j-p}$
where $\gamma_k = cov(Y_t, Y_{t-k})$ and $\gamma_0 = var(Y_t)$
$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} \ldots + \phi_p \rho_{j-p}$
where $\rho_k = corr(Y_t, Y_{t-k})$ and $\rho_0 = 1$

### 4.1.2 Moving Average Model (MA(q))

$Y_t = \mu + \sum_{j=0}^{q} \theta_j \varepsilon_{t-j}$, where $\theta_0 = 1$
$E(Y_t) = \mu$
$\gamma_0 = \sigma^2 (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2)$

# 5 Non-Stationarity

## 5.1 Unit Root Tests

### 5.1.1 ADF

Model C: $\Delta Y_t = \mu + \alpha t + \gamma Y_{t-1} + \sum_{j=1}^{q} \Delta Y_{t-j} + \epsilon_t$
where Model A: $\mu = 0, \alpha = 0$; Model B: $\alpha = 0$
$H_0 : \gamma = 0$, $H_1 : \gamma < 0$; Test statistic: $\frac{\hat{\gamma}-0}{se(\hat{\gamma})}$ ∼MacKinnon critical values

### 5.1.2 DF-GLS

For trend stationarity, estimate the model $\tilde{Y}_t = \delta_0 X_t + \delta_1 Z_t + \varepsilon_t$
where $\tilde{Y}_1 = Y_1$, $X_1 = 1$, $Z_1 = 1$, $\tilde{Y}_t = Y_t - \alpha^* Y_{t-1}$, $X_t = 1 - \alpha^*$, $Z_t = t - \alpha^*(t-1)$ for $t = 2, \ldots, T$ and $\alpha^* = 1 - (13.5/T)$.
Obtain the residuals: $Y_t^* = Y_t - (\hat{\delta}_0 + \hat{\delta}_1 t)$ and undertake the ADF test on $Y^*$:
$\Delta Y_t^* = \alpha + \gamma Y_{t-1}^* + \sum_{j=1}^{p^*} \delta_j \Delta Y_{t-j}^* + \varepsilon_t$
$H_0 : \gamma = 0$, $H_1 : \gamma < 0$

### 5.1.3 KPSS

Estimate the following model (for trend stationary series):
$Y_t = \mu + \alpha t + \eta_t$ and save the residuals, $\hat{\eta}_t$
$KPSS = \frac{1}{T^2} \frac{1}{\lambda^2} \sum_{t=1}^{T} S_t^2$
where $S_t = \sum_{j=1}^{t} \hat{\eta}_j$ and $\lambda^2$ is long-run error variance.

## 5.2 Engle-Granger Two Step Procedure

Assuming $Y_t \sim I(1)$ and $X_t \sim I(1)$ then the long-run equation is:
$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$
is cointegrated if $\hat{\varepsilon}_t \sim I(0)$.
In which case there are potential two error correction models (ECMs)
$\Delta Y_t = \alpha_1 + \sum_{j=1}^{p_1} \gamma_{1j} \Delta X_{t-j} + \sum_{j=1}^{q_1} \beta_{1j} \Delta Y_{t-j} + \delta_1 \hat{\varepsilon}_{t-1} + \nu_{1t}$
$\Delta X_t = \alpha_1 + \sum_{j=1}^{p_2} \gamma_{2j} \Delta Y_{t-j} + \sum_{j=1}^{q_2} \beta_{2j} \Delta X_{t-j} + \delta_2 \hat{\varepsilon}_{t-1} + \nu_{2t}$

# 6 Panel Data

The basic regression model is of the form:
$Y_{it} = a_i + d_t + \beta X_{it} + \varepsilon_{it}$
where $i = 1, \ldots, n$, $t = 1, \ldots, T$

## 6.1 POLS: $a_i = 0$, $d_t = 0$

$Y_{it} = \alpha + \beta X_{it} + \varepsilon_{it}$ where $i = 1, \ldots, n,\ t = 1, \ldots, T$

$$b = \frac{\sum\limits_{i=1}^{N}\sum\limits_{t=1}^{T}(X_{it} - X_{..})(Y_{it} - Y_{..})}{\sum\limits_{i=1}^{N}\sum\limits_{t=1}^{T}(X_{it} - X_{..})^2}$$

where: $X_{..} = \frac{\sum\limits_{i=1}^{N}\sum\limits_{t=1}^{T} X_{it}}{NT}$; $Y_{..} = \frac{\sum\limits_{i=1}^{N}\sum\limits_{t=1}^{T} Y_{it}}{NT}$

## 6.2 1-way Fixed Effects: $a_i \neq 0$, $d_t = 0$

$Y_{it} = a_i + \beta X_{it} + \varepsilon_{it}$ where $i = 1, \ldots, n,\ t = 1, \ldots, T$

$$b = \frac{\sum\limits_{i=1}^{N}\sum\limits_{t=1}^{T}(X_{it} - X_{i.})(Y_{it} - Y_{i.})}{\sum\limits_{i=1}^{N}\sum\limits_{t=1}^{T}(X_{it} - X_{i.})^2}$$

where: $X_{i.} = \frac{\sum\limits_{t=1}^{T} X_{it}}{T}$, $Y_{i.} = \frac{\sum\limits_{t=1}^{T} Y_{it}}{T}$

## 6.3 2-way Fixed Effects: $a_i \neq 0$, $d_t \neq 0$

$Y_{it} = a_i + d_t + \beta X_{it} + \varepsilon_{it}$ where $i = 1, \ldots, n,\ t = 1, \ldots, T$

$$b = \frac{\sum\limits_{i=1}^{N}\sum\limits_{t=1}^{T}\tilde{X}_{it}\tilde{Y}_{it}}{\sum\limits_{i=1}^{N}\sum\limits_{t=1}^{T}\tilde{X}_{it}^2}$$

where $\tilde{X}_{it} = X_{it} - X_{i.} - X_{.t} + X_{..}$ and $\tilde{Y}_{it} = Y_{it} - Y_{i.} - Y_{.t} + Y_{..}$ and

$$X_{i.} = \frac{\sum_{t=1}^{T} X_{it}}{T},\ Y_{i.} = \frac{\sum_{t=1}^{T} Y_{it}}{T}, X_{.t} = \frac{\sum_{i=1}^{N} X_{it}}{N},\ Y_{.t} = \frac{\sum_{i=1}^{N} Y_{it}}{N}, X_{..} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}}{NT},\ Y_{..} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} Y_{it}}{NT}$$

## 6.4 Dynamic Panel Data Model

$Y_{it} = a_i + \beta_1 X_{it} + \beta_2 Y_{it-1} + \varepsilon_{it}$ where $i = 1, \ldots, n$, $t = 2, \ldots, T$

Estimated as:

$\Delta Y_{it} = \beta_1 \Delta X_{it} + \beta_2 \Delta Y_{it-1} + \Delta\varepsilon_{it}$