Applied Econometrics II
Final Review

Haoling Wang

Contents

Preface

Applied Econometrics II is also a practical course. It is similar to Applied Econometrics I, in that both require a mindset that translates knowledge into analytical tools. Similarly, it is closely connected with other courses. Personally, I believe its foundation lies in the combination of linear algebra and multivariable regression from ECON1049 Mathematical Economics and Econometrics (now renamed ECON1057 Introductory Econometrics and ECON1058 Mathematical Economics), as well as the linear algebra and multivariable regression from ECON2051 Econometric Theory I. (Perhaps this may not be immediately obvious during your studies, as Applied Econometrics 2 involves writing extensive text during the exams, and you might think that some theoretical courses are not useful. However, without theoretical support, you are unlikely to truly understand analysis and how to interpret data in Stata. Economic analysis must be grounded in the understanding of mathematical courses). Moreover, if you are interested in Econometrics, I suggest that you also learn ECON2052 Econometrics Theory II, it demonstrate the theoretical aspect of the model you will learn in Applied Econometrics II such as time series. In my perspective, they are complimentary. The theoretical aspect is abstract so you need real world data analysis to understand. And only analysis is not enough you should know how to use theory to explain it.

Once again, aside from the exams, the most important aspect of this course is hands-on skills. Be sure to practice Stata frequently, as you will use it in your future papers. Meanwhile, you should read more papers to see if you can identify models that you have already learned. This will be immensely helpful when reading literature in the future, as you'll begin to recognize the structure of articles, rather than just skimming the introduction and vaguely understanding the conclusion. This deeper approach to reading will foster critical thinking. Additionally, make sure to look for extra materials to supplement the knowledge that was briefly covered in the slides. You cannot expect your professors to teach every single point. It is common to self-study additional content from books and resources on your own, as this will help you when writing papers and searching for knowledge that you have not yet learned.

This note is just for your glance during your final exam preparation, it will not be a very good complementary of your ppt because all the content are in the ppt. And there maybe some mistake in the note, go through ppt and find where I wrote wrong. Do not doubt yourself, you can correct my mistake. And if you are interested to make a friend with me, welcome to email me first: hmyhw8@nottingham.edu.cn (or the personal email, wang95483@gmail.com). I will reply you as soon as possible.

I will upload another note for this course in my notebook webpage, it is the answer that I used to post on the forum in the moodle page of this course. Again, the most important thing is not answer but the logic or the ability to search knowledge and try to organize and answer the question by yourself.

Good luck with your Applied Econometric II study!

1 Model

1.1 Time series

1.1.1 Finite distributed lag (DL) model

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + \varepsilon_t$$

Interpretation

An unit increase in x_t leads to a change in y_t by β_0 unit.

Long-run multiplier

An unit increase in x_t for every period, y_t increases by $\beta_0 + \beta_1 + \beta_2 + \beta_3$

1.1.2 Autoregressive (AR) model

AR(1)

$$y_t = \alpha + \alpha_1 y_{t-1} + \varepsilon_t$$

Long-run multiplier

$$\alpha_1 + \alpha_1^2 + \alpha_1^3 + ... = \frac{\alpha_1}{1 - \alpha_1}$$

AR(p)

$$y_t = \alpha + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + ... + \alpha_p y_{t-p} + \varepsilon_t$$

Long-run multiplier

$$\frac{\alpha_1 + \alpha_2 + ... + \alpha_p}{1 - \alpha_1 - \alpha_2 - ... - \alpha_p}$$

1.1.3 Random Walk model

$$y_t = \rho y_{t-1} + \varepsilon_t$$
$$y_t = \alpha + \rho y_{t-1} + \varepsilon_t$$
$$y_t = \alpha + \beta t + \rho y_{t-1} + \varepsilon_t$$

1.1.4 Autoregressive distributed lags (ARDL)

$$y_t = \alpha + \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + \beta_0 x_t + \beta_1 x_{t-1} + \cdots + \beta_q x_{t-q} + \varepsilon_t$$

Long-run multiplier

$$\frac{\beta_0 + \beta_1 + ... + \beta_q}{1 - \alpha_1 - \alpha_2 - ... - \alpha_p}$$

1.1.5 Error Correction Model (ECM)

Captures both long-run (cointegration) and short-run effect.

ARDL model with unit root

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$$

Building process

$$y_t - y_{t-1} = \alpha_0 + \alpha_1 y_{t-1} - y_{t-1} + \beta_0 x_t - \beta_0 x_{t-1} + \beta_0 x_{t-1} + \beta_1 x_{t-1} + \varepsilon_t$$

The formal Error Correction Model (ECM) model

$$\Delta y_t = \beta_0 \Delta x_t + (\alpha_1 - 1)\left(y_{t-1} - \frac{\alpha_0}{1-\alpha_1} - \frac{\beta_0 + \beta_1}{1-\alpha_1} x_{t-1}\right) + \varepsilon_t$$

- $1 - \alpha_1$ Rate of adjustment back to equilibrium
- Long-run movements

$$y_{t-1} - \frac{\alpha_0}{1-\alpha_1} - \frac{\beta_0 + \beta_1}{1-\alpha_1} x_{t-1} = u_{t-1}$$

- Short-run movements

$$\beta_0 \Delta x_t$$

1.2 Binary (Limited) dependent variable

1.2.1 Linear probability model

$$y_i = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i} + u_i = \mathbf{x_i}\beta + u_i$$
$$y_i = 1 \quad or \quad 0$$
$$\mathbf{x_i} = \begin{bmatrix} 1 & x_{1i} & \cdots & x_{ki} \end{bmatrix}$$
$$\beta = \begin{bmatrix} \beta_0 & \beta_1 & \cdots & \beta_k \end{bmatrix}'$$

Assumption

$$\mathbb{E}(u|\mathbf{x_i}) = 0$$

The regression model corresponds to the probability that $y_i = 1$ given $\mathbf{x_i}$

$$P(y_i = 1|\mathbf{x_i}) = \mathbb{E}(y_i|\mathbf{x_i}) = \mathbf{x_i}\beta$$

Proof 1.1 *Expectation of linear probability model*

Take the conditional expectation of both sides of the model with respect to x_i:

$$\mathbb{E}(y_i \mid \mathbf{x_i}) = \mathbb{E}(\mathbf{x_i}\beta + u_i \mid x_i)$$
$$\mathbb{E}(y_i \mid \mathbf{x_i}) = \mathbf{x_i}\beta + \mathbb{E}(u_i \mid \mathbf{x_i})$$

Now apply the assumption:

$$\mathbb{E}(u_i \mid x_i) = 0$$

So:

$$\mathbb{E}(y_i \mid x_i) = x_i\beta$$

Interpretation of coefficient β_j

The change in probability that $y_i = 1$ from a unit change in x_{ij}.

Marginal effect for continuous variable

$$\frac{\partial P(y_i = 1 \mid x_i)}{\partial x_{ij}} = \beta_j$$

Goodness-of-fit

Measured by the **percent** of within-sample predictions that are **correct**

- Define $\tilde{y}_i = 1$ if $\hat{y}_i \geq 0.5$, or else $\tilde{y}_i = 0$.
- If $\tilde{y}_i = y_i$, the prediction for i is correct.

Disadvantage

1. Negative probability and probability greater than 1.
2. Heteroskedasticity

   The estimated β are still unbiased, but the estimated standard error of β from OLS is bias and this lead to wrong inferences, need to correct it in STATA.
   $var(u_i)$ depends on xi, which changes from person to person.

   Proof 1.2 *Heteroskedasticity*

$$y_i = \mathbf{x_i}\beta + u_i$$
$$u_i = y_i - \mathbf{x_i}\beta = y_i - P(y_i = 1|\mathbf{x_i})$$

- If $y_i = 1$, $u_i = 1 - P(y_i = 1|\mathbf{x_i})$
- If $y_i = o$, $u_i = -P(y_i = 1|\mathbf{x_i})$

$$var(u_i) = -P(y_i = 1|\mathbf{x_i})[1 - P(y_i = 1|\mathbf{x_i})] = -\mathbf{x_i}\beta(1 - \mathbf{x_i}\beta)$$

1.2.2 Underlying latent variable models

$$y_i^\star = \beta_0 + \beta_1 x_{1,i} + ... + \beta_6 x_{6,i} + \varepsilon_i = \mathbf{x_i}'\beta + \varepsilon_i$$

- $y_i^\star$ an unobserved (latent) variable, a linear function of x_i .
- y_i is a non-linear function of $y_i^\star$ and, hence, x_i.

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \leq 0 \end{cases}$$

- ε_i
  1. symmetrically distributed around zero, $\mathbb{E}(\varepsilon_i) = 0$
  2. has either the standard logistic distribution or the standard normal distribution

1.2.3 Logistic model

$$P(y_i = 1 \mid x_i) = \Phi(x_i\beta) = \frac{1}{1 + \exp(-(x_i\beta))} = \frac{\exp(x_i\beta)}{1 + \exp(x_i\beta)}$$

- non-linear function
- $\Phi(.)$ cumulative distribution function, a standard logistic random variable. Is a monotonically increasing function.

$$\frac{\partial y_i^*}{\partial x_{ij}} > 0 \Rightarrow \frac{\partial P(y_i = 1 \mid x_i)}{\partial x_{ij}} > 0$$

$$\frac{\partial y_i^*}{\partial x_{ij}} < 0 \Rightarrow \frac{\partial P(y_i = 1 \mid x_i)}{\partial x_{ij}} < 0$$

- $P(y_i = 1 \mid x_i) \in [0, 1]$

Marginal effect for continuous variable

$$\frac{\partial P(y_i = 1 \mid x_i)}{\partial x_{ij}} = \beta_j \cdot \frac{\exp(x_i\beta)}{(1 + \exp(x_i\beta))^2}$$

Partial effect for discrete variable, i.e. binary independent variable

If $x_{1,i}$ is binary, $x_{2,i}$ to $x_{k,i}$ are continuous, for the model

$$P(y_i = 1 \mid x_i) = \Phi(x_i\beta)$$

1. When $x_{1i} = 1$,
$$P(y_i = 1 \mid x_i) = \Phi(\beta_0 + \beta_1 + \beta_2 x_{2i} + \cdots + \beta_k x_{ki}).$$
2. When $x_{1i} = 0$,
$$P(y_i = 1 \mid x_i) = \Phi(\beta_0 + \beta_2 x_{2i} + \cdots + \beta_k x_{ki}).$$
3. Partial effect of x_{1i}
$$\Phi(\beta_0 + \beta_1 + \beta_2 x_{2i} + \cdots + \beta_k x_{ki}) - \Phi(\beta_0 + \beta_2 x_{2i} + \cdots + \beta_k x_{ki})$$

Estimation of logit models

Using maximum likelihood estimation (MLE).

$$L(\beta \mid y, x) = \prod_{i=1}^{N} (P(y_i = 1 \mid x_i))^{y_i} (1 - P(y_i = 1 \mid x_i))^{1-y_i} \in [0, 1]$$

Using the form of sum of log-likelihoods

$$l(\beta \mid y, x) = \log L(\beta \mid y, x) = log(\prod_{i=1}^{N} (P(y_i = 1 \mid x_i))^{y_i} (1 - P(y_i = 1 \mid x_i))^{1-y_i})$$

$$= \sum_{i=1}^{N} \{y_i \log (P(y_i = 1 \mid x_i)) + (1 - y_i) \log (1 - P(y_i = 1 \mid x_i))\} < 0$$

F.O.C

$$\frac{\partial L(\beta \mid y, x)}{\partial \beta} = 0$$

Solution (estimation of $\hat{\beta}$)

$$\hat{\beta} = [\hat{\beta_1} \quad \hat{\beta_2} \quad \hat{\beta_3} \quad ... \quad \hat{\beta_N}]$$

Interpretation of the estimated coefficients

The sign of an estimated coefficient indicates the sign of the marginal or partial effect of a change in the corresponding variable.

Null hypothesis: $x_{(.,j)}$ has no effect on y

$$z - statistic = \frac{\hat{\beta}}{S.E.(\hat{\beta})}$$

Goodness-of-fit measure

Per cent of observations correctly predicted.

1. **Sample proportion of successes**

$$\hat{P}(y_i = 1 \mid x_i) \quad for \quad i = 1, ..., N$$

2. **pseudo** R^2 Compares between goodness-of-fit of the estimated model and a model that contains only an intercept.

$$pseudoR^2 = 1 - \frac{l_u}{l_0}$$

- l_u the log-likelihood of the estimated model
- l_0 the log-likelihood of a model that contains only an intercept

Table 1: Goodness-of-fit using pseudo R^2

$l_u = l_0$	$pseudoR^2 = 0$	all $x_1, ..., x_k$ explain none of the variation in y.
$l_u > l_0, \frac{l_u}{l_0} \to 0$	$pseudoR^2 = 1$	all $x_1, ..., x_k$ can fully explain the variation in y.

1.2.4 Probit model

$$P(y_i = 1 \mid x_i) = \Phi(x_i\beta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_i} \exp\left\{-\frac{(x_i\beta)^2}{2}\right\} d(x_i\beta)$$

- $\Phi(.)$ a standard normal cumulative distribution function.

Marginal effect for continuous variable

$$\frac{\partial P(y_i = 1 \mid x_i)}{\partial x_{ij}} = \beta_j \cdot \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x_i\beta)^2}{2}\right\} d(x_i\beta)$$

Partial effect is similar to the Logit models

Estimation is similar to logit models

Interpretation of estimated coefficient is similar to logit models

Goodness-of-fit measurement is similar to logit models

1.3 Panel data

1.3.1 Use OLS separately for each year

$$y_i = \beta_0 + \beta_1 x_i + u_i \quad \text{for } \textit{year} = \mathbf{1982}$$

$$y_i = \beta_0 + \beta_1 x_i + u_i \quad \text{for } \textit{year} = \mathbf{1987}$$

1.3.2 A model for two-period panel data

$$y_{it} = \beta_0 + \delta d_{2t} + \beta_1 x_{it} + a_i + u_{it}, \quad t = 1, 2$$

- a_i individual fixed effect, captures all the unobserved, time-invariant (i.e. constant) factors that affect y_{it}.
- d_{2t} time dummy
- x_{it} observable characteristics

1.3.3 Pooled ordinary least squares

$$y_{it} = \beta_0 + \delta d_{2t} + \beta_1 x_{it} + v_{it}, \quad t = 1, 2$$

- ignore the fixed effect a_i
- new idiosyncratic error $v_{it} = a_i + u_{it}$

Advantage

A bigger sample size, and hence more accurate estimates.

Disadvantage

If a_i and x_{it} are correlated, v_{it} and x_{it} will also be correlated, $E(v_{it}|x_{it}) \neq 0$.
In this case, it lead to biased estimates of β.

1.3.4 First differencing estimator

$$\Delta y_{i2} = \delta + \beta_1 \Delta x_{i2} + \Delta u_{i2}$$

- $\mathbb{E}(\Delta u_{i2} \mid \Delta x_{i2}) = 0$
- **Advantage**
  1. reduced omitted variable bias
  2. increased the accuracy of our estimates
  3. increased the power of our hypothesis tests
- **Problem**
  1. heteroskedasticity
     - Breusch-Pagan test: hettest
     - White test: imtest, white
     - White-corrected standard errors: reg. . . , robust
  2. cannot investigate the effects of explanatory variables that are **time-invariant**

Proof 1.3 *First differencing estimator*

$$y_{i2} = \beta_0 + \delta + \beta_1 x_{i2} + a_i + u_{i2} \quad \text{individual i in period 2}, d_{2t} = 1$$

$$y_{i1} = \beta_0 + \beta_1 x_{i1} + a_i + u_{i1} \quad \text{individual i in period 1}, d_{2t} = 0$$

$$y_{i2} - y_{i1} = \delta + \beta_1(x_{i2} - x_{i1}) + (u_{i2} - u_{i1})$$
$$\Delta y_{i2} = \delta + \beta_1 \Delta x_{i2} + \Delta u_{i2}$$

1.3.5 Least squares dummy variable estimator

1. Use dummy for each individual i
2. Works fine if number of individual is small. Difficult to estimate when the number of individual is large.

1.3.6 Fixed Effects Model (Within-groups estimator)

$$\tilde{y}_{it} = \beta_1 \tilde{x}_{it} + \tilde{u}_{it}$$

- $\tilde{y}_{it}$ time demeaned dependent variable
- $\tilde{x}_{it}$ time-demeaned explanatory variable
- $\tilde{x}_{it}$ and $\tilde{u}_{it}$ is uncorrelated across all time periods
- apply OLS to this model, get **fixed effects estimator** of β_1
- **Assumption**

  The unobserved effect, a_i , is assumed to be **correlated** with one or more of the explanatory variables, x_{jit}.

  $$cov(a_i, x_i) \neq 0$$

- **Problem**
  1. Heteroskedasticity
  2. Serially correlated idiosyncratic u_{it}
  3. Cannot use them to investigate the effects of explanatory variables that are time-invariant

Proof 1.4 *Fixed effects transformation (Within transformation)*

For the model

$$y_{it} = \beta x_{it} + a_i + u_{it}, \quad t = 1, 2, ...T$$

Compute the **within mean** of y and x

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$$

$$\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$$

So the model can be written as

$$\bar{y}_i = \beta_1 \bar{x}_i + a_i + \bar{u}_i$$

Now remove a_i by:

$$y_{it} - \bar{y}_i = \beta_1(x_{it} - \bar{x}_i) + (u_{it} - \bar{u}_i) \quad t = 1, 2, \ldots, T$$

1.3.7 Random effects estimates

$$y_{it} = \beta_0 + \beta_1 x_{1it} + \ldots + \beta_3 x_{3it} + \beta_4 x_{4it} + v_{it}$$

The expression here is the same as the pooled OLS, but it will be changed later on.

- composite error term $v_{it} = a_i + u_{it}$
- consistent but inefficient estimated β, leading to incorrect test statistics and incorrect confidence inference.
- a_i is part of v_{it} in each time period, v_{it} will be **serially correlated** across time.

$$Corr(v_{it}, v_{is}) = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_u^2}, \quad t \neq s$$

Assumption

1. a_i is uncorrelated with all explanatory variables.

$$cov(a_i, x_i) = 0$$

2. $\mathbb{E}(a_i) = 0$

NOTE The original expression is the same as fixed effects model, but the assumption changed.

$$y_{it} = \beta_0 + \beta_1 x_{1it} + \ldots + \beta_3 x_{3it} + \beta_4 x_{4it} + a_i + u_{it}$$

Correct approach: generalized least squares (GLS)

1. Transforming the data to eliminate the serial correlation issue.
First stage produces partially demeaned dependent and explanatory variables.

$$y_{it} - \lambda \bar{y}_i = \beta_0(1-\lambda) + \beta_1(x_{1it} - \lambda \bar{x}_{1i}) + \cdots + \beta_k(x_{kit} - \lambda \bar{x}_{ki}) + (\nu_{it} - \lambda \bar{\nu}_i)$$

$$\lambda = 1 - \left(\frac{\sigma_u^2}{T\sigma_a^2 + \sigma_u^2}\right)^{\frac{1}{2}}$$

   - $\lambda = 0$ this transformation brings us back to OLS on the pooled sample
   - $\lambda = 1$ this transformation is the fixed effects (within) transformation

2. using OLS to estimate transformed data.

1.4 Instrumental variables

Treats the **omitted variable** as part of the **error term**, and acknowledges the presence of the omitted variable.

2 requirements

1. **Instrumental exogeneity**

$$Cov(z_{1i}, u_i) = 0$$

The instrument must not be correlated with the unobserved variable. i.e. z_{1i}, is not correlated with the ability as u_i is unobservable.

2. **Instrumental relevance**

$$Cov(z_{1i}, y_{2i}) \neq 0$$

The instrument must be correlated with the endogenous explanatory variable, y_{2i} .

1.4.1 Two stage least square

1. First stage

Regression y_{2i} on z_{1i} (instrumental variable) and x_{1i} (explanatory variable) , and obtain a prediction for y_{2i} (endogenous explanatory variable).

$$y_{2i} = \gamma_0 + \gamma_1 \mathbf{x}_{1i} + \gamma_2 \mathbf{x}_{1i}^2 + \gamma_3 \mathbf{z}_{1i} + \boldsymbol{\nu}_i$$

2. Second stage

Replace y_{2i} with its prediction, $\hat{y}_{2i}$ from the first stage.
Perform a least square regression.

$$\mathbf{y}_{1i} = \beta_0 + \beta_1 \hat{y}_{2i} + \beta_2 \mathbf{x}_{1i} + \beta_3 \mathbf{x}_{1i}^2 + \mathbf{u}_i$$

The standard error in the IV estimator is always larger than in the OLS case.

It can be much larger if the relationship between y_{2i} and z_{1i} is weak.

2 Formula

2.1 Time series

2.1.1 Introduction

k^{th} sample autocovariance of a series Y_t

$$\operatorname{cov}(Y_t, Y_{t-k}) = \mathbb{E}\left[(Y_t - \mathbb{E}(Y_t))(Y_{t-k} - \mathbb{E}(Y_{t-k}))\right]$$

k^{th} sample autocorrelation of a series Y_t

$$\rho_k = \frac{\operatorname{cov}(Y_t, Y_{t-k})}{\sqrt{\operatorname{var}(Y_t)\operatorname{var}(Y_{t-k})}}, \quad k = 0, 1, 2, \ldots$$

2.1.2 Forecast

1-step ahead point forecast of y_{T+1}

$$\hat{y}_{T+1|T} = E(y_{T+1}|I_T)$$

Forecast error

$$y_{T+1} - \hat{y}_{T+1|T}$$

Forecasting AR(1) model

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t, \varepsilon_t \sim \mathcal{N}(0, \sigma^2)$$

Iterating forward one period

$$y_{T+1} = a_0 + a_1 y_T + \varepsilon_{T+1}$$

One-step-ahead point forecast

$$E(y_{T+1}|I_T) = a_0 + a_1 y_T$$

Two-step-ahead point forecast

$$E(y_{T+2}|I_T) = a_0 + a_0 a_1 + a_1^2 y_T$$

Forecast for forecast horizon h

$$E(y_{T+h} \mid I_T) = a_0\left(1 + a_1 + a_1^2 + \cdots + a_1^{h-1}\right) + a_1^h y_T \to \frac{a_0}{1 - a_1} \quad \text{as} \quad h \to \infty$$

Forecast error for the h-step-ahead forecast

$$y_{T+h} - E(y_{T+h}|I_T) = \varepsilon_{T+h} + a_1\varepsilon_{T+h-1} + \ldots + a_1^{h-1}\varepsilon_{T+1}$$

$$E[y_{T+h} - E(y_{T+h}|I_T)] = 0$$

$$V(E[y_{T+h} - E(y_{T+h}|I_T) = E[(E[y_{T+h} - E(y_{T+h}|I_T)^2] = \sigma^2\left(1 + a_1^2 + a_1^4 + \cdots + a_1^{2(h-1)}\right)$$

$$\to \sigma^2 \frac{1}{1 - a_1^2} \quad \text{as} \quad h \to \infty$$

95% confidence interval for the one-step ahead forecasts

$$a_0 + a_1 y_T \pm 1.96\sigma$$

95% confidence interval for the h-step-ahead forecast

$$a_0\left(1+a_1+a_1^2+\cdots+a_1^{h-1}\right)+a_1^h y_T \pm 1.96\sigma\left(1+a_1^2+a_1^4+\cdots+a_1^{2(h-1)}\right)^{\frac{1}{2}}$$

$$\rightarrow \frac{a_0}{1-a_1} \pm 1.96\sigma\left(\frac{1}{1-a_1^2}\right)^{\frac{1}{2}} \quad \text{as} \quad h \rightarrow \infty$$

Forecast error

Root Mean Squared Error (RMSE) for point forecast

$$\mathrm{RMSE}_i=\sqrt{\frac{1}{n} \sum_{i=1}^n\left(y_{T+i}-E\left(y_{T+i} \mid I_{T+i-1}\right)\right)^2}$$

Decision rule Select the model with the lowest RMSE value.

Appropriate model selection criteria

$$AIC(p, q)=\frac{2k}{T}+\ln\left(\frac{SSR(p, q)}{T}\right)$$

$$BIC(p, q)=\frac{k \ln(T)}{T}+\ln\left(\frac{SSR(p, q)}{T}\right)$$

- $k=p+q+1$ number of regressors
- T number of observations
- $SSR(p, q)$ the sum of square residuals

$$\hat{\varepsilon}_t=y_t-\hat{a}_0-\sum_{i=1}^p \hat{a}_i y_{t-i}-\sum_{j=1}^q \hat{\beta}_j x_{t-j}$$

$$SSR(p, q)=\sum_{t=1}^T \hat{\varepsilon}_t^2$$

Decision rule Models with the lowest value of AIC/BIC should be preferred.

3 Tests

3.1 Time series

3.1.1 Box-Pierce Q

Null hypothesis: the correlations up to lag k are jointly equal to 0 (all ACs up to lag k are not significant).

If reject the null hypothesis, then find evidence in favour of significant autocorrelations.

3.1.2 Classical Durbin-Watson test

First-order autocorrelation

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t \quad -1 < \rho < 1$$

Table 2: Different serial correlation

$\rho = 0$	no serial correlation	$cov(\varepsilon_t, \varepsilon_{t-k}) = 0$
$\rho > 0$	positive serial correlation	$cov(\varepsilon_t, \varepsilon_{t-k}) > 0$
$\rho < 0$	negative serial correlation	$cov(\varepsilon_t, \varepsilon_{t-k}) < 0$

Test for testing first-order autocorrelation in ε_t.

The builded model: AR(1) model

$$y_t = \alpha + \alpha_1 y_{t-1} + \varepsilon_t$$

Test for:

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t \quad -1 < \rho < 1$$

Null hypothesis

$$\rho = o$$

Alternative hypothesis

$$0 < \rho < 1 \quad \text{positive correlation}$$

$$-1 < \rho < 0 \quad \text{negative correlation}$$

Steps

1. Estimate parameters in using OLS and obtain the estimated residual

$$\hat{\varepsilon}_t = y_t - \hat{\alpha} - \hat{\alpha}_1 y_{t-1} \quad \text{for} \quad t = 1, \ldots, T$$

2. Compute the Durbin–Watson test statistic

$$DW = \frac{\sum_{t=2}^{T} (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=2}^{T} \hat{\varepsilon}_{t-1}^2}$$

3. The calculated DW statistic is benchmarked with the tabulated critical values to decide whether to reject the null hypothesis or accept the alternative hypothesis.
To formally test for serial correlation, we rely on a critical lower bound, D_L, and a critical upper bound, D_U.

Table 3: DW test statistics and interpretation

$\rho \to 2$	no serial correlation	$cov(\varepsilon_t, \varepsilon_{t-k}) = 0$
$\rho \to 0$	positive serial correlation	$cov(\varepsilon_t, \varepsilon_{t-k}) > 0$
$\rho \to 4$	negative serial correlation	$cov(\varepsilon_t, \varepsilon_{t-k}) < 0$

Decision rule

Table 4: DW test statistics and interpretation

$DW < D_L$	Evidence of positive serial correlation
$DW > D_U$	No enough evidence to conclude that there is positive serial correlation. Do not reject the null hypothesis yet
$DW \in [D_L, D_U]$	test is inconclusive
$DW > 4 - D_L$	Evidence of negative serial correlation
$DW < 4 - D_U$	No enough evidence to conclude that there is negative serial correlation. do not reject the null hypothesis
$DW \in [4 - D_U, 4 - D_L]$	test is inconclusive

3.1.3 Augmented Durbin-Watson test

Still can only be used to test the first-order sutocorrelation. But can be carried out even when the regressors contain **lagged dependent variables**.

First-order serial correlations

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t \qquad -1 < \rho < 1$$

Null hypothesis

$$\rho = 0$$

Alternative hypothesis

$$\rho \neq 0$$

Steps

1. Estimate parameters in using OLS and obtain the estimated residual
$$\hat{\varepsilon}_t = y_t - \hat{\alpha} - \hat{\beta}x_t \quad \text{for} \quad t = 1, \ldots, T$$
2. test statistic
$$h = \hat{\rho} \cdot \sqrt{\frac{T}{1 - T \cdot \mathrm{Var}(\hat{\beta}_1)}}$$
follows an **asymptotic standard normal distribution** under the null hypothesis of no first-order autocorrelation.
$$h \sim \mathcal{N}(0, 1)$$
(approximately, as $T \to \infty$
3. If $\mid h \mid > 1.645$, reject H_0 at the 5% level (two-sided). This is similar to a z-test for large samples.

3.1.4 Breusch-Godfrey test

It can be used to test the null hypothesis of no serial correlation against **higher-order serial correlations** and not just a first-order one. Can be carried out even when the regressors contain **lagged dependent variables**.

Null hypothesis (no serial correlation)

$$\rho_1 = \rho_2 = ... = \rho_p = 0$$

Alternative hypothesis (serial correlation)

$$\rho_1 \neq 0 \quad \text{and/or} \quad \rho_2 \neq 0 \quad \text{and/or...} \quad \rho_p \neq 0$$

Steps

1. Estimate parameters using OLS and obtain the estimated residual

$$\hat{\varepsilon}_t = y_t - \hat{\alpha} - \hat{\beta} x_t \quad \text{for} \quad t = 1, \ldots, T$$

2. Regress the following auxiliary regression model and collect the $\mathbf{R^2}$.

$$\hat{\varepsilon}_t = \gamma + \rho_1 \hat{\varepsilon}_{t-1} + \rho_2 \hat{\varepsilon}_{t-2} + \cdots + \rho_p \hat{\varepsilon}_{t-p} + \delta x_t + u_t,$$

3. Calculate the Breusch-Godfrey test statistic:

$$T \times R^2 \sim \chi^2_p$$

- T the number of observations
- p the number of error lags in the auxiliary regression model.

Decision rule
Reject the null if the calculated value exceeds the tabulated chi-squared critical value at the desired level of significance.

3.1.5 Dickey Fuller unit root test

Models that can be used for testing

$$y_t = \rho y_{t-1} + \varepsilon_t \quad \text{random walk}$$

$$y_t = \alpha + \rho y_{t-1} + \varepsilon_t \quad \text{random walk with drift}$$

$$y_t = \alpha + \beta t + \rho y_{t-1} + \varepsilon_t \quad \text{unit root series with drift and linear time trend}$$

Transformed model

$$y_t - y_{t-1} = \Delta y_t = \gamma y_{t-1} + \varepsilon_t$$

$$\Delta y_t = \alpha + \gamma y_{t-1} + \varepsilon_t$$

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \varepsilon_t$$

Null hypothesis

$$\gamma = 0 \quad (\rho = 1)$$

Alternative hypothesis

$$\gamma < 0 \quad (\rho < 1)$$

3.1.6 Augmented Dickey Fuller test

Use information criteria to select lag, p.

Model that can be used to test

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \theta_1 \Delta y_{t-1} + \cdots + \theta_p \Delta y_{t-p} + \varepsilon_t$$

Null hypothesis

$$\gamma = 0 \quad (\rho = 1)$$

Alternative hypothesis

$$\gamma < 0 \quad (\rho < 1)$$

Test statistics

Use t-ratio and F-test

$$\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$$

Use the critical values provided by Dickey and Fuller.

3.1.7 Engle-Granger test

Select the appropriate lag length using information criteria.

Model (example)

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$$

Steps

1. First run a linear regression of y_t on x_t.
2. Test unit root in the residuals using the Augmented Dickey-Fuller test statistic.
3. **The critical values in the Engle-Granger test is unique**, not the Dickey fuller critical value.
   Reasons:
   - $\hat{u}_t$ are **not observed** variables but estimated ones, they **introduce extra uncertainty** into the test.
   - This estimation error changes the distribution of the test statistic under the null hypothesis.
4. If the residuals are **stationary**, then there is long run equilibrium relationship between x_t and y_t, they are cointegrated.
5. Build the Error Correction Model (ECM)

$$\Delta y_t = \beta_0 \Delta x_t + (\alpha_1 - 1)\left(y_{t-1} - \frac{\alpha_0}{1-\alpha_1} - \frac{\beta_0 + \beta_1}{1-\alpha_1} x_{t-1}\right) + \varepsilon_t$$

3.2 Limited dependent variable

3.2.1 Likelihood ratio (LR) test

Example

Purpose Testing the joint significance of x_1 and x_2.

Null hypothesis

$$\beta_1 = \beta_2 = 0$$

Use MLE to estimate:

1. **unrestricted model**

$$P(y_i = 1 \mid x_i) = \Phi(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}) \quad \text{for } i = 1, 2, \ldots, N$$

   obtain the unrestricted log-likelihood value, $\mathbf{l_u}$

2. **restricted model**

$$P(y_i = 1 \mid x_i) = \Phi(\beta_0 + \beta_3 x_{i3} + \beta_4 x_{i4} + \cdots + \beta_k x_{ik}) \quad \text{for } i = 1, 2, \ldots, N$$

   obtain the restricted likelihood value, $\mathbf{l_r}$

LR test statistic

$$LR = 2(\mathbf{l_u} - \mathbf{l_r})$$

LR is an asymptotic χ^2 (chi-squared) statistic with degrees of freedom equal to the number of restrictions, i.e. 2 restrictions in our case.

3.3 Panel data

3.3.1 F-test for Fixed Effects

Purpose

This test checks whether the **fixed effects** (individual-specific intercepts) are jointly significant — i.e., whether including them improves the model fit significantly compared to a **pooled OLS regression** (which assumes no unobserved heterogeneity across individuals).

For the model

$$y_{it} = \beta x_{it} + a_i + u_{it}, \quad t = 1, 2, ...T$$

Null hypothesis

All individual-specific effects are zero
(pooled OLS is adequate)

Alternative hypothesis

At least one $a_i \neq 0$
(fixed effects model is necessary due to unobserved heterogeneity)

3.3.2 Hausman test

Choose between random effects model and fixed effects model.

Null hypothesis

Random effects model

Alternative hypothesis

Fixed effects model

3.4 Instrumental variables

3.4.1 Test for endogeneity - Hausman test

Purpose Compare IV with OLS estimates since when the explanatory variable, y_{2i} , is exogenous, the IV estimator is less efficient than OLS (has larger standard errors).

Table 5: Principle

Exogenous y_{2i}	OLS and IV estimates are consistent
Endogenous y_{2i}	IV and OLS differ significantly

Null hypothesis

OLS is unbiased/ No endogeneity issue/ IV and OLS estimates are consistent

Alternative hypothesis

OLS is biased/ endogeneity issue/ IV are consistent

3.4.2 Test for over-identification

Purpose

Test if all IV satisfy the Instrumental Exogeneity condition.

$$\begin{aligned}\text{Cov}(z_{1i}, u_i) &= 0\\ \text{Cov}(z_{2i}, u_i) &= 0\\ &\vdots\\ \text{Cov}(z_{qi}, u_i) &= 0\end{aligned}$$

Null Hypothesis

All IVs are uncorrelated with u_i (Instrumental exogeneity)

Alternative Hypothesis

Some of IVs are correlated with u_i

Steps

1. Obtain the predicted residuals from the second stage regression.
2. Regress the **predicted residuals** on all **exogenous variables**, and obtain the R-square.
3. Test statistic

$$\text{number of observations} \times \text{R-square}$$

4. compare with the critical value of a χ^2 distribution with $q - k$ degrees of freedom where k is the number of endogenous explanatory variables.

3.4.3 Test for weak instrument

Problem

Poor predictors, which means the instrument is only weakly correlated with the endogenous explanatory variable, of the endogenous explanatory variable, y_{2i}.

Approach

Examine the significant of the coefficients of the instrument variables in the first stage regression — if they are **not significant** then there is **weak instrument** problem.

Null hypothesis

All instrument variable coefficients are zeros

Test Statistic

joint F-statistic

Decision rule

Reject the null if F-statistic > 10.