[Part1] Basics

CLRM

- $\hat{u}$ denote the residual, $y^t - \hat{y}^t$
- **RSS**: residual sum of squares
- OLS

$$\underset{\beta_0, \beta_1}{\text{minimize}} \sum_{i=1}^{N} u_i^2 = \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x_i)^2$$ (1)

F.O.C.

with respect to β_0

$$-2 \sum_{i=1}^{N} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (*)$$

with respect to β_1

$$-2 \sum_{i=1}^{N} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \cdot x_i = 0 \quad (**)$$

Rewrite $(*)$

$$\sum y_i - N\hat{\beta}_0 - \hat{\beta}_1 \sum x_i = 0 \quad \text{(divide by } N)$$

$$\frac{1}{N} \sum y_i - \hat{\beta}_0 - \hat{\beta}_1 \frac{1}{N} \sum x_i = 0$$

$$\bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x} = 0$$

$$\Rightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Rewrite $(**)$

$$\sum x_i (y_i - \bar{y} + \hat{\beta}_1 \bar{x} - \hat{\beta}_1 x_i) = 0$$

$$\sum x_i y_i - \bar{y} \sum x_i + \hat{\beta}_1 \bar{x} \sum x_i - \hat{\beta}_1 \sum x_i^2 = 0 \quad \text{divide by } N$$

$$\frac{1}{N} \sum x_i y_i - \bar{y} \bar{x} + \hat{\beta}_1 (\bar{x})^2 - \hat{\beta}_1 \frac{1}{N} \sum x_i^2 = 0$$

$$\hat{\beta}_1 = \frac{\frac{1}{N} \sum x_i y_i - \bar{x}\bar{y}}{\frac{1}{N} \sum x_i^2 - (\bar{x})^2} = \frac{\sum x_i y_i - N\bar{x}\bar{y}}{\sum x_i^2 - N(\bar{x})^2}$$

Practice exercise: show that

$$\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{N} x_i y_i - N \bar{x} \bar{y}$$

Assumptions about μ_t

- **assumptions about μ_t:**

  1. E[μ_t] = 0
  - Var[μ_t] = σ^2 < ∞
  - Cov[μ_i,μ_j] = 0
  - Cov[μ_t,x_t] = 0
  - μ_t ~ N(0,σ^2)
- BLUE: "Best Linear Unbiased Estimators" if 1-4 hold
- **Consistent**: need assumption 2 and 4
  - $\lim_{T \to \infty} Pr[\mid \hat{\beta} - \beta \mid > \delta] = 0 \quad \forall \quad \delta > 0$
- **Unbiased**: need assumption 1
  - $E[\hat{\alpha}] = \alpha$
  - $E[\hat{\beta}] = \beta$
- **Efficiency**: unbiased and no other unbiased estimator has a smaller variance

T-test

- Standard Error (for simple OLS)

$$SE\left(\hat{\beta}_0\right) = s\sqrt{\frac{\frac{1}{N}\sum x_i^2}{\sum(x_i - \overline{x})^2}}$$
$$SE\left(\hat{\beta}_1\right) = s\sqrt{\frac{1}{\sum(x_i - \overline{x})^2}}$$

where s is the estimated standard deviation of residuals u_i.

Unbiased estimator of the variance of the error term $\sigma^2 = Var(u_i)$: $s^2 = \frac{1}{N-2}\sum_{i=1}^{N}\hat{u}_i^2$

- the test of significance approach
  - 2-side test

Standard normal variates can be constructed from $\hat{\alpha}$ and $\hat{\beta}$:

$$\frac{\hat{\alpha}-\alpha}{\sqrt{Var(\alpha)}} \sim \mathcal{N}(0,1) \quad \text{and} \quad \frac{\hat{\beta}-\beta}{\sqrt{Var(\beta)}} \sim \mathcal{N}(0,1)$$

But $Var(\alpha)$ and $Var(\beta)$ are unknown, so

$$\boxed{\frac{\hat{\alpha}-\alpha}{SE(\hat{\alpha})} \sim t_{T-2} \quad \text{and} \quad \frac{\hat{\beta}-\beta}{SE(\hat{\beta})} \sim t_{T-2}}$$

TS

earch Methodology in Finance

- the confidence interval approach
  - ...
- always keeping an intercept even if it is “not statistically significantly different from zero”
- p-value
  - p < α -> reject-> significant
  - p smaller, more significant
  - p bigger, H0 holds more

multiple linear regression model

- expressing the multiple linear regression model
  - **y** is T × 1
  - **X** is T × k
  - β is k × 1
  - μ is T × 1
  - T is samples
  - k-1 is variables
  - *intercept contains
- $\hat{\beta} = (X'X)^{-1}X'y$
- variance of the errors
  - $s^2 = \frac{\hat{\mu}'\hat{\mu}}{T-k}$
- The variance-covariance matrix of $\hat{\beta}$
  - $s^2(X'X)^{-1}$
- The variances are on the leading diagonal:
  - $Var(\hat{\beta})$
- Standard Errors
  - $SE = \sqrt{Var(\hat{\beta})}$

F-test

$$\boxed{\begin{aligned} \text{test stat} &= \frac{RRSS - URSS}{URSS} \times \frac{T-k}{m} \\ \text{where URSS} &= \text{RSS from unrestricted regression} \\ \text{RRSS} &= \text{RSS from restricted regression} \\ m &= \text{number of restrictions} \\ k &= \text{total number of parameters to be estimated} \end{aligned}}$$

```
- k: intercept contains
- 1-side test
- Critical Value: F(m,T - k)
- TS > CV: reject
```

R2

- The most common goodness of fit statistic is known as R^2
  - TSS = RSS + ESS
  - $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$
- R^2 never falls if more regressors are added
- $Adjusted R^2$
  - $\bar{R}^2 = 1 - [\frac{T-1}{T-K}(1-R^2)]$
  - k increases and unless R^2 increases by a more than offsetting amount, $\bar{R}^2$ will actually fall.

test for assumptions about μ_t

- test for those assumptions about μ_t:

```
1. E[$\mu_t$] = 0
    - The mean of the residuals will always be zero provided that there is a constant term
  in the regression.
```

- 1. Var[μ_t] = σ^2 < ∞
    - White's test
    - Run the auxiliary regression $\hat{\mu}_t^2 = \alpha_1 + \alpha_2 x_{2t} + \alpha_3 x_{3t} + \alpha_4 x_{2t}^2 + \alpha_5 x_{3t}^2 + \alpha_6 x_{2t} x_{3t} + v_t$
    - $teststat = TR^2$
    - $criticalvalue = \chi^2_{1-\alpha}(m)$
    - m is the number of parameters in the auxiliary regression (without the constant term)
    - TS > CV: reject H0. assumption 2 violated
    - OLS estimation still gives unbiased coefficient estimates, but they are no longer BLUE
    - if we still use OLS in the presence of heteroscedasticity, our **standard errors could be inappropriate** and hence any inferences we make could be misleading
    - Use White's heteroscedasticity consistent standard errors

- 1. Cov[μ_i,μ_j] = 0
    - If there are patterns in the estimated residuals, they are autocorrelated
    - Breusch-Godfrey test
    - $\mu_t = \rho_1\mu_{t-1} + \rho_2\mu_{t-2} + \ldots + \rho_r\mu_{t-r} + v_t$
    - v_t~$N(0, \sigma_v^2)$
    - H0: $\rho = 0$
    - $teststat = (T - r)R^2$
    - $criticalvalue = \chi^2_{1-\alpha}(r)$
    - TS > CV: reject H0. assumption 3 violated
    - unbiased, but they are inefficient
    - R^2 is likely to be inflated relative to its “correct” value for positively correlated residuals.
    - A. Switch to a model in first differences
    - A. Use Newey-West’s heteroscedasticity and autocorrelation consistent (HAC) standard errors

- 1. Cov[μ_t ,x_t] = 0
    - The OLS estimator is consistent and unbiased in the presence of stochastic regressors, provided that the regressors are not correlated with the error term of the estimated equation
    - if one or more of the explanatory variables is contemporaneously correlated with the disturbance term, the OLS estimator will not even be consistent.

- 1. μ_t ~ N(0,σ^2)
    - Jarque and Bera test
    - Estimate the skewness b_1 and the excess kurtosis b_2 of the $\hat{u}_t$
    - $TS = T[\frac{b_1^2}{6} + \frac{(b_2-3)^2}{24}]$
    - $CV = \chi^2_{1-\alpha}(2)$
    - TS > CV: reject H0. assumption 5 violated
    - Often the case that one or two very extreme residuals causes us to reject the normality assumption.
    - An alternative is to use dummy variables.

Multicollinearity

- Measuring Multicollinearity
    - Problems if near multicollinearity is present but ignored
        - R^2 will be high but the individual coefficients will have high standard errors.
        - The regression becomes very sensitive to small changes in the specification.
        - Thus confidence intervals for the parameters will be very wide, and significance tests might therefore give inappropriate conclusions.
    - look at the matrix of correlations between the different individual variables

[Part2] Time series analysis

stationary process and white noise

- stationary process
  - **strictly stationary process**:
    - $Prob(y_{t1} <= b_1, \ldots, y_{tn} <= b_n) = Prob(y_{t1+m} <= b_1, \ldots, y_{tn+m} <= b_n)$
    - **distribution remains unchanged overtime**
  - weakly stationary process:
    1. $E(y_t) = \mu$
    2. $E[(y_t - E(y_t))(y_t - E(y_t))] = \sigma^2 < \infty$
    3. $E[(y_{t1} - E(y_{t1}))(y_{t2} - E(y_{t2}))] = \gamma_{t2-t1} \quad \forall \quad t_1, t_2$
    - **that is, mean,SD and COV stay the same**
  - autocovariance function: $E[(y_t - E(y_t))(y_{t-s} - E(y_{t-s}))] = \gamma_s \quad s = 0, 1, 2, \ldots$
  - **ACF!!!**
  - autocorrelation function: $\tau_s = \gamma_s / \gamma_0 \quad s = 0, 1, 2, \ldots$

white noise

- white noise process
  - $E(y_t) = \mu$

  1. $Var(y_t) = \sigma^2$
  2. $\gamma_{t-r} = \sigma^2$ if t=r
  3. $\gamma_{t-r} = 0$ if otherwise
  - ==> y_t~$N(\mu, \sigma^2)$
  - ==> τ~$N(0, \frac{1}{T})$, T is sample size
- Q-statistic (Box and Pierce) (one side test)
  - $Q = T\Sigma_{k=1}^{m}\hat{\tau}_k^2$ (m:lags)
  - CV~$\chi^2(m)$
- Q-statistic (Ljung-Box) (one side test)
  - $Q = T(T+2)\Sigma_{k=1}^{m}\frac{\hat{\tau}_k^2}{T-k}$ (m:lags)
  - CV~$\chi^2(m)$

MA

- MA: moving average process
  - $\mu_t = white.noise$
  - MA(q): $y_t = \mu + \mu_t + \theta_1\mu_{t-1} + \theta_1\mu_{t-1} + \ldots + + \theta_q\mu_{t-q}$
    1. $E(y_t) = \mu$
    2. $Var(y_t) = \gamma_0 = (1 + \theta_1^2 + \theta_2^2 + .. + \theta_q^2)\sigma^2$
    3. $\gamma_s = (\theta_s + \theta_{s+1}\theta_1 + \ldots + \theta_q\theta_{q-s})\sigma^2$ for s=1,2,..,q
    4. $\gamma_s = 0$ for s>q
  - MA(2): $y_t = \mu_t + \theta_1\mu_{t-1} + \theta_2\mu_{t-2}$
    - $\gamma_0 = var(y_t) = (1 + \theta_1^2 + \theta_2^2)\sigma^2$
    - $\gamma_1 = (\theta_1 + \theta_1\theta_2)\sigma^2$
    - $\gamma_2 = \theta_2\sigma^2$
    - $\gamma_3 = 0$

AR

- AR: autoregressive process
  - AR(q): $y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_q y_{t-q} + \mu_t$
  - use $L^i y_t$ as y_{t-i}
  - or use $\phi(L)y_t = \mu + \mu_t$ where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$
  - stationarity
    - the roots of $(1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p)$ all lie outside the unit circle. (solve this equation)
    - A stationary AR(p) model is required for it to have an MA(∞) representation
  - the wold's decpmposition theorem
    - ignoring the intercept
    - $y_t = \phi(L)^{-1}\mu_t$
    - $E(y_t) = \frac{\mu}{1-\phi_1-\phi_2-\ldots-\phi_p}$
  - Yule-Walker equations (solve it and got autocovariances and autocorrelation functions)

$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \vdots \\ \gamma_p \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_{-1} & \gamma_{-2} & \cdots \\ \gamma_1 & \gamma_0 & \gamma_{-1} & \cdots \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \cdots \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \\ \vdots \\ \varphi_p \end{bmatrix}$$

- - AR(1)
    - $\gamma_0 = \frac{\sigma_\mu^2}{1-\phi_1^2}$
    - $\gamma_1 = \frac{\phi_1\sigma_\mu^2}{1-\phi_1^2}$
    - $\gamma_2 = \frac{\phi_1^2\sigma_\mu^2}{1-\phi_1^2}$
    - $\gamma_s = \frac{\phi_1^s\sigma_\mu^2}{1-\phi_1^2}$
    - autocorrelation function (ACF)
    - $\tau_0 = \frac{\gamma_0}{\gamma_0} = 1$
    - $\tau_1 = \frac{\gamma_1}{\gamma_0} = \phi_1$
    - $\tau_2 = \frac{\gamma_2}{\gamma_0} = \phi_1^2$
    - $\tau_s = \phi_1^s$

PACF

- PACF (partial autocorrelation function)
  - PACF (τ_{kk}) measures the autocorrelation between the current observation (y_t) and an observation k periods ago (y_{t-k}), after controlling for observations between t and t −k
    - at lag 1 PACF = ACF
    - at lag 2 $\tau_{22} = \frac{\tau_2-\tau_1^2}{1-\tau_1^2}$
    - more lags: more complex
  - PACF helps to distinguish between AR and MA process
    - AR(p)
      - there are direct connections between yt and yt−s only for s ≤ p
      - theoretical PACF will be zero after lag p
    - MA(q)
      - can be written as AR(∞), thus there are direct connections between yt and all its previous values
      - theoretical PACF will be geometrically declining

ARMA

- ARMA
  - ARMA(P,Q):
    - $\phi(L)y_t = \mu + \theta(L)\mu_t$ where
    - $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$
    - $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots \theta_q L^q$
    - or: $y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \theta_1 \mu_{t-1} + \theta_2 \mu_{t-2} + \ldots + \theta_q \mu_{t-q} + \mu_t$
    - **Similar to the stationarity condition, we typically require the MA(q) part of the model to have roots of θ(z) = 0 greater than one in absolute value.**
    - $E(y_t) = \frac{\mu}{1-\phi_1-\phi_2-\ldots-\phi_p}$
      - the AR part have mean 0

ACF

- **summary of ACF** (stationary)
  - An autoregressive process AR(p) has
    - a geometrically declining ACF
    - zero after lag p for PACF
  - A moving average process MA(q) has
    - zero after lag q for ACF
    - a geometrically declining PACF
  - An autoregressive moving average process ARMA(p,q) has
    - a geometrically declining ACF
    - a geometrically declining PACF

B-J approach ARMA models

- Box-Jenkins approach building ARMA models
  - information criteria
    - to minimises the information criterion
    - AIC = $Ln(\hat{\sigma}^2) + \frac{2k}{T}$
    - SBIC = $Ln(\hat{\sigma}^2) + \frac{k}{T}Ln(T)$
    - HQIC = $Ln(\hat{\sigma}^2) + \frac{2k}{T}Ln(Ln(T))$
    - where $\hat{\sigma}^2 = \frac{\Sigma\hat{\mu}^2}{T-k}$
    - k=p+q+1 is the total number of parameters estimated
    - suggest
      - SBIC is strongly consistent but (inefficient).
      - AIC is not consistent, and will typically pick “bigger” models.
      - Overall, no criterion is definitely superior to others

ARIMA models

- The "I" stands for integrated.
- An ARMA(p,q) model in the variable differenced d times is equivalent to an ARIMA(p,d,q) model on the original data.

forecasting

- forecasting
  - An MA(q) only has memory of q
  - Some of the most popular criteria for assessing the accuracy of time series forecasting techniques are:
    - Mean squared error
    - Mean absolute error
    - Mean absolute percentage error
  - problems:
    - over-confidence
    - inconsistency
    - recency
    - anchoring
    - illusory
    - patterns
    - “group-think”
  - The Usually Optimal Approach
    - To use a statistical forecasting model built on solid theoretical foundations supplemented by expert judgements and interpretation.

Multivariate time series analysis

- vector autoregression (VAR) model
  - bivariate VAR(k)
  - Advantages of VAR Modelling
    - Do not need to specify which variables are endogenous or exogenous - all are endogenous
    - Allows the value of a variable to depend on more than just its own lags or combinations of white noise terms, so more general than ARMA modelling
    - Provided that there are no contemporaneous terms on the right hand side of the equations, can simply use OLS separately on each equation
    - Forecasts are often better than “traditional structural” models.
  - Problems with VAR’s
    - VAR’s are a-theoretical (as are ARMA models)
    - How do you decide the appropriate lag length?
    - So many parameters! If we have g equations for g variables and we have k lags of each of the variables in each equation, we have to estimate (g + kg^2) parameters, e.g., g = 3 and k = 3 imply 30 parameters
    - Do we need to ensure all components of the VAR are stationary?
    - How do we interpret the coefficients?

- likelihood ratio test
  - variance-covariance matrix of residuals (given by $\hat{\mu}\hat{\mu}'/T$), as Σ
  - $LR = T(log \mid \hat{\Sigma}_r \mid - log \mid \hat{\Sigma}_u \mid)$
  - restricted, unrestricted
  - $\chi^2_{1-\alpha}(df = total.number.of.restrictions)$

- Information criteria
  - MAIC = $Ln \mid \hat{\Sigma} \mid + \frac{2k'}{T}$
  - MABIC = $Ln \mid \hat{\Sigma} \mid + \frac{k'}{T} Ln(T)$
  - MHQIC = $Ln \mid \hat{\Sigma} \mid + \frac{2k'}{T} Ln(Ln(T))$
  - k is the total number of regressors in all
    - $k' = g^2 k + g$
      - k: lags
      - g: variables

- if the equations had a contemporaneous feedback term

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} \beta_{1,0} \\ \beta_{2,0} \end{pmatrix} + \begin{pmatrix} \beta_{1,1} & \alpha_{1,1} \\ \alpha_{2,1} & \beta_{2,1} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \alpha_{1,2} & 0 \\ 0 & \alpha_{2,2} \end{pmatrix} \begin{pmatrix} y_{2,t} \\ y_{1,t} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix}$$

We can take the contemporaneous terms over to the LHS and write

$$\begin{aligned} \begin{pmatrix} 1 & -\alpha_{1,2} \\ -\alpha_{2,2} & 1 \end{pmatrix} \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} &= \begin{pmatrix} \beta_{1,0} \\ \beta_{2,0} \end{pmatrix} + \begin{pmatrix} \beta_{1,1} & \alpha_{1,1} \\ \alpha_{2,1} & \beta_{2,1} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} \\ \text{or} \qquad A\, y_t &= \beta_0 + \beta_1\, y_{t-1} + u_t \end{aligned}$$

If both sides are pre-multiplied by A^{-1}

$$\begin{aligned} y_t &= A^{-1}\beta_0 + A^{-1}\beta_1\, y_{t-1} + A^{-1} u_t \\ \text{or} \qquad y_t &= A_0 + A_1\, y_{t-1} + e_t \end{aligned}$$

This is known as a standard form VAR, which we can estimate using OLS.

- However,the error terms will typically be correlated to some degree.

causality tests (Granger causality tests)

	Hypothesis	Implied restriction
1	Lags of $y_{1,t}$ do not explain current $y_{2,t}$	$\beta_{2,1} = 0$ and $\gamma_{2,1} = 0$ and $\delta_{2,1} = 0$
2	Lags of $y_{1,t}$ do not explain current $y_{1,t}$	$\beta_{1,1} = 0$ and $\gamma_{1,1} = 0$ and $\delta_{1,1} = 0$
3	Lags of $y_{2,t}$ do not explain current $y_{1,t}$	$\beta_{1,2} = 0$ and $\gamma_{1,2} = 0$ and $\delta_{1,2} = 0$
4	Lags of $y_{2,t}$ do not explain current $y_{2,t}$	$\beta_{2,2} = 0$ and $\gamma_{2,2} = 0$ and $\delta_{2,2} = 0$

- tested within the F-test framework

construct the impulse responses and variance decompositions

- Impulse responses:
  - responsiveness of the **dependent variables** shocks to the **error term**
- variance decompositions:
  - give the proportion of the movements in the dependent variables that are due to their “own” shocks, versus shocks to the other variables.
  - determining how much of the s-step ahead forecast error variance for each variable is explained innovations to each explanatory variable.
- we assumed that the VAR error terms were statistically independent of one another. (This is generally not true)
- “orthogonalise”

[Part3.1] Cointegration: Modelling long-run financial behaviour

non-stationarity

- non-stationarity
  - If the variables in the regression model are **not stationary, the usual t-ratios will not follow a t-distribution**
  - two models which have been frequently used to characterise non-stationarity
    1. random walk model: $y_t = \mu + y_{t-1} + \mu_t$
    2. deterministic time trend process: $y_t = \alpha + \beta t + \mu_t$
  - AR(1) for
    1. $\phi = 0$ white noise process
    2. $\phi = 0.8$ stationary AR(1)
    3. $\phi = 1$ (random walk)
- integrated of order d
  - If a non-stationary series yt has to be differenced d times before it becomes stationary, then it is said to be integrated of order d.
  - **if** $y_t \sim I(d)$ **then** $\Delta^d y_t \sim I(0)$
  - An I(2) series contains two unit roots and so would require differencing (get the Δ) twice to induce stationarity.
  - The majority of economic and financial series contain a single unit root

unit root (D&F)

- test a unit root (Dickey and Fuller)
  - $y_t = \phi y_{t-1} + \mu_t$
  - $\Delta y_t = \psi y_{t-1} + \mu_t$
  - we can write $\Delta y_t = \psi y_{t-1} + \mu + \lambda t + \mu_t$
  - with:
    1. $\mu = \lambda = 0$
    2. $\lambda = 0$
  - **H0: series contains a unit root (**$\phi = 1$ **or** $\psi = 0$**)**
  - **H1: series is stationary (**$\phi < 1$**)**
  - TS=$\frac{\hat{\psi}}{SE(\hat{\psi})}$
  - CV:

significance level	10%	5%	1%
CV for constant but no trend	−2.57	−2.86	−3.43
CV for constant and trend	−3.12	−3.41	−3.96

- if **TS (negative) < CV (negative), reject H0**

ADF

- augmented Dickey Fuller (**ADF**) test
  - in particular, μ_t will be autocorrelated if there was autocorrelation in the dependent variable of the regression (Δy_t) which we have not modelled.
  - $\Delta y_t = \psi y_{t-1} + \Sigma_{i=1}^{p} \alpha_i \Delta y_{t-i} + \mu_t$
  - CV as before
  - A problem now arises in determining the optimal number of lags of the dependent variable, There are 2 ways:
    1. use the frequency of the data to decide
    2. use information criteria

- testing for higher orders of integration
  - simple regression $\Delta y_t = \psi y_{t-1} + \mu_t$
  - test $\psi = 0$ or $\psi < 0$
  - if not reject, cont. test $y_t \sim I(2)$ or $y_t \sim I(1)$
  - until we reject H0, result is stationary

Phillips-Perron test

- Phillips-Perron test
  - Phillips and Perron have developed a more comprehensive theory of unit root nonstationarity. The tests are similar to ADF tests, but they incorporate an automatic correction to the DF procedure to allow for autocorrelated residuals.
  - The tests usually give the same conclusions as the ADF tests, and the calculation of the test statistics is complex.
  - the tests are poor at deciding if φ=1 or φ=0.95, especially with small sample sizes.

Stationarity test

- H0: y_t is stationary
- H1: y_t is non-stationary
- KPSS test: H0:$y_t \sim I(0)$ H1:$y_t \sim I(1)$
- ADF/PP: H0:$y_t \sim I(1)$ H1:$y_t \sim I(0)$

Cointegration

- In most cases, if we combine two variables which are I(1), then the combination will also be I(1)
- More generally, if we combine variables with differing orders of integration, the combination will have an order of integration equal to the largest.
- if $X_{i,t} \sim I(d_i)$ for i=1,2,3,...,k
- $z_t = \Sigma_{i=1}^{k} \alpha_i X_{i,t}$
- then $z_t \sim I(max d_i)$
- $X_{1,t} = \Sigma_{i=2}^{k} \beta_i X_{i,t} + z'_t$
- **z'_t is not stationary and is autocorrelated if all of the X_i are I(1)**

- definition of cointegration (Engle & Granger, 1987)
  - Let z_t be a k ×1 vector of variables, then the components of z_t are cointegrated of order (d,b) if
    1. all components of z_t are I(d)
    2. there is at least one vector of coefficients α such that $\alpha' z_t \sim I(d-b)$
  - **Many time series are non-stationary but "move together" over time.**
  - **If variables are cointegrated, it means that a linear combination of them will be stationary.**
  - There may be up to **r linearly independent cointegrating relationships** (where r ≤ k −1), also known as cointegrating vectors. r is also known as the **cointegrating rank of z_t**.
  - No cointegration implies that series could wander apart without bound in the long run.

- Examples of possible cointegrating relationships in finance:
  - spot and futures prices
  - ratio of relative prices and an exchange rate
  - equity prices and dividends

ECM

- sometimes should not use first diff for non-stationary series because they have no long run solution
  - long run: $y_t = y_{t-1} = y \quad x_t = x_{t-1} = x$
  - one way is use both first diff and the levels terms (original ones), that is:
- Equilibrium correction or error correction models (ECM)
  - $\Delta y_t = \beta_1 \Delta x_t + \beta_2 (y_{t-1} - \hat{\gamma} x_{t-1}) + u_t$
  - $y_{t-1} - \hat{\gamma} x_{t-1}$ is known as the error correction term
  - y_t and x_t are cointegrated with cointegrating coefficient ˆγ, then this term will be I(0)
  - We can thus validly use OLS on it
  - The Granger representation theorem shows that any cointegrating relationship can be expressed as an equilibrium correction model.

residuals-based approach

- **theory: u_t should be I(0) if the variables $y_t, x_{2,t}, \ldots, x_{k,t}$ are cointegrated**
- Engle and Granger (1987) have tabulated a new set of critical values and hence the test is known as the Engle Granger (E.G.) test. (book p374)
- H0 : unit root in cointegrating regression's residuals
- H1 : residuals from cointegrating regression are stationary

The Engle Granger 2-step method

- step1
  - Make sure that all the individual variables are I(1)
  - Test these residuals to ensure that they are I(0).
  - If they are I(1), estimate a model containing only first differences.
  - If they are I(0), proceed to step 2.
- step2
  - $\Delta y_t = \beta_1 \Delta x_t + \beta_2 \hat{u_{t-1}} + u_t$ where $\hat{u_{t-1}} = y_{t-1} - \hat{\gamma} x_{t-1}$ (ECM)
- This method suffers from a number of problems:
  1. Unit root and cointegration tests have low power in finite samples
  2. We are forced to treat the variables asymmetrically and to specify one as the dependent and the other as independent variables. (addressed by the Johansen approach)
  3. Cannot perform any hypothesis tests about the actual cointegrating relationship estimated at stage 1. (addressed by the Engle and Yoo approach or the Johansen approach)

The Engle and Yoo (EY) 3-step

- The Engle and Yoo (EY) 3-step procedure takes its first two steps from EG, add a third step giving updated estimates of the cointegrating vector and its standard errors.
  - most important problem: could be more than one cointegrating relationship

Johansen's method

- turn the VAR into a VECM
- $\Pi = (\sum_{i=1}^{k} \beta_i) - I_g$
- test for cointegration by looking at rank of Π via its eigenvalues λ
- Eigenvalues λ_i are sorted: $\lambda_1 >= \lambda_2 >= ... >= \lambda_g$
- if variables are not cointegrated, $rank(\Pi) \sim 0$,ie., $\lambda_i \sim 0$ for all i
- Johansen approach: test statistics incorporate $ln(1-\lambda_i)$, instead of λ_i
  - λ_i=0 inplies $ln(1-\lambda_i)$=0
- r is rank of Π

1. trace: $\lambda_{trace}(r) = -T * \sum_{i=r+1}^{g} ln(1-\hat{\lambda}_i)$
   - trace: sum of the eigenvalues
   - λ_{trace} tests the null that the number of cointegrating vectors is less than equal to r against an unspecified alternative that there are more than r.
     - H0: <= r
     - H1: > r
2. max: $\lambda_{max}(r, r+1) = -T * ln(1-\hat{\lambda}_{i+1})$
   - λ_{max} tests the null that the number of cointegrating vectors is r against an alternative of r+1
     - H0: = r
     - H1: = r+1

- rank
  - if rank (Π) = g, original yt is stationary
  - if rank (Π) = 0, Δyt depends only on Δyt−j and not on yt−1
  - if 1 < rank (Π) < g, there are multiple cointegrating vectors

- we keep increasing r until we no longer reject the null
  - Engle and Granger (1987) did not allow us to do hypothesis tests on the cointegrating relationship itself, but the Johansen approach does.
  - If there exist r cointegrating vectors, only these linear combinations will be stationary

- hypothesis tests (as a restriction on the Π matrix)
  - $TS = -T \sum_{i=1}^{r}[ln(1-\lambda_i) - ln(1-\lambda_i^*)] \sim \chi^2(m)$
    - λ_i^* are the characteristic roots of the restricted model
    - λ_i are the characteristic roots of the unrestricted model
    - r is the number of non-zero characteristic roots in the unrestricted model
    - m is the number of restrictions

[Part3.2] Panel data

- important:
  - key features
  - advantages and disadvantages
  - contrast the fixed effect and random effect
  - which is the more appropriate

Panel data

- **Panel data have both time series (t = 1, . . . , T) and cross-sectional dimensions (i = 1, . . . , N)**
  - $y_{it} = \alpha + \beta x_{it} + \mu_{it}$
  - simplest way: estimate a single, pooled regression on all the observations together
  - but: assumes that there is no heterogeneity
- advantage
  1. address broader range and more complex problems than pure time series or cross-sectional data.
  2. examine how variables and their relationships change dynamically over time.
  3. remove the omitted variables bias.
  4. extra. less collinearity and more degrees of freedom

fixed effects model

- $y_{it} = \alpha + \beta x_{it} + \mu_i + v_{it}$-
- μ_i: "firm-fixed" effect
- $\mu_{it} = \mu_i + v_{it}$
- This model could be estimated using dummy variables, which would be termed the least squares dummy variable (LSDV) approach.
  - *the formula* (dummy variables)
  - $y_{it} = \beta x_{it} + \mu_1 D1_i + \mu_2 D2_i + \ldots + \mu_N DN_i + v_{it}$
  - it can be estimated using OLS
  - the intercept is in the "martix" already

within transformation

- use demeaned variables:
- $y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + u_{it} - \bar{u}_i$

first difference operator

- When differences are taken, any variables that do not change over time will again cancel out.
- Differencing and the within transformation will produce identical(same) estimates in situations where there are only two time periods.

between estimator

- cross-sectional regression on the time-averaged values of the variables
- advantage: reduce the effect of measurement error in the variables on the estimation process.

time-fixed effects models

- $y_{it} = \alpha + \beta x_{it} + \lambda_t + v_{it}$
- where λ_t is a time-varying intercept that captures all of the variables that affect y and that vary over time but are constant cross-sectionally.

random effects model

- $y_{it} = \alpha + \beta x_{it} + \omega_{it}$
- $\omega_{it} = \epsilon_i + v_{it}$
- ϵ_i measures the random deviation of each intercept term, constant over time
- The parameters (α and the β vector) are estimated consistently but inefficiently by OLS
- generalised least squares (GLS) procedure is usually used

quasi-demeaned

- Define the 'quasi-demeaned' data as $y_{it}^* = y_{it} - \theta \bar{y}_i$ and similarly for x_{it}
- $\theta = 1 - \frac{\sigma_v}{\sqrt{T\sigma_\epsilon^2 + \sigma_v^2}}$
- required to ensure that there are no cross-correlations in the error terms

Fixed or random effects

- population
  - been randomly selected from the population: random
  - effectively constitute the entire population: fixed
- random should more efficient
- but: ω_{it} is uncorrelated, ϵ_{it} and v_{it} to be independent of all x_{it}.
- use: Hausman test
- or will: biased and inconsistent.

[Part3.3] Modelling Volatility: GARCH models

- important:
  - features of data
  - estimate conditional volatility
  - test ARCH-effects
  - estimate GARCH models

features of data

- leptokurtosis (high kurtosis)
- volatility pooling(汇集)
- leverage effects

motivations volatility

- important concepts in finance
- crude measure of risk
- VaR model
- BS formula
- **variance of the errors not be constant over time Not for financial data.**

Autoregressive conditionally heteroscedastic (ARCH)

- ARCH(1):
  - $\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2$
  - $u_t^2 \sim N(0, \sigma_t^2)$
  - The full model: $y_t = \beta_1 + \beta_2 x_{2,t} + \ldots + \beta_k x_{k,t} + u_t$

- ARCH(1) (alternative)
  - $y_t = \beta_1 + \beta_2 x_{2,t} + \ldots + \beta_k x_{k,t} + u_t$
  - $\sigma_t = \sqrt{\alpha_0 + \alpha_1 u_{t-1}^2}$
  - $u_t = v_t \sigma_t$
  - $v_t = N(0,1)$

ARCH(q):

- extend to q lags, h_t as σ_t^2
- $h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \ldots + \alpha_q u_{t-q}^2$
- $u_t^2 \sim N(0, h_t)$
- The full model: $y_t = \beta_1 + \beta_2 x_{2,t} + \ldots + \beta_k x_{k,t} + u_t$

ARCH effects test

- test if $\hat{u}_t^2$ is related to the $\hat{u}_{t-i}^2$
- $TS = TR^2$
  - R^2 of $\mu_t^2 = \gamma_0 + \gamma_1 u_{t-1}^2 + \gamma_2 u_{t-2}^2 + \ldots + \gamma_q u_{t-q}^2 + v_t$
- $CV \sim \chi_{1-\alpha}^2(q)$

Generalised ARCH (GARCH) models

- Due to Bollerslev (1986)
- GARCH(1,1) $\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta \sigma_{t-1}^2$
- can be written as an infinite order ARCH model

- GARCH(p, q)
  - $\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \ldots + \alpha_q u_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_p \sigma_{t-p}^2$
  - $= \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$

- advantage
  - more parsimonious - avoid overfitting
  - non-negativity constraints

conditional volatility

- The unconditional variance of u_t is given by
  - $Var(u_t) = \frac{\alpha 0}{1-(\alpha 1+\beta)}$ when $\alpha_1 + \beta$ <1
  - $\alpha_1 + \beta$ ≥ 1: “non-stationarity” in variance
  - $\alpha_1 + \beta$ = 1: intergrated GARCH

Estimation of ARCH/GARCH models 1

cannot use OLS, use maximum log-likelihood

1. Specify the appropriate equations for the mean and the variance
   - $y_t = \mu + \phi y_{t-1} + u_t$
   - $u_t \sim N(0, \sigma_t^2)$
   - $\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta \sigma_{t-1}^2$
2. Specify the log-likelihood function to maximise
   - $LLF = -\frac{T}{2} log(2\pi) - \frac{1}{2} \sum_{t=1}^{T} log(\sigma_t^2) - \frac{1}{2} \sum_{t=1}^{T} \frac{(y_t - \mu - \phi y_{t-1})^2}{\sigma_t^2}$
3. The computer will maximise the function and give parameter values and their standard error

Estimation of ARCH/GARCH models 2

- choosing parameter values (β1, β2 ,σ2) that maximise this function.
- $LF(\beta_1, \beta_2, \sigma^2) = \frac{1}{\sigma^T (\sqrt{2\pi})^T} exp\{-\frac{1}{2}\sum_{t=1}^{T}\frac{(y_t - \beta_1 - \beta_2 x_t)^2}{\sigma^2}\}$
- $LLF = -\frac{T}{2}log(\sigma^2) - \frac{T}{2}log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\frac{(y_t - \beta_1 - \beta_2 x_t)^2}{\sigma^2}$
- got differentials and some complex steps
  - $\hat{\beta}_1 = \bar{y} - \hat{\beta}_2\bar{x}$
  - $\hat{\beta}_2 = \frac{\sum y_t x_t - T\bar{x}\bar{y}}{\sum x_t^2 - T\bar{x}^2}$
  - $\hat{\sigma}^2 = \frac{1}{T}\sum \hat{u}_t^2$
- OLS estimator was $\hat{\sigma}^2 = \frac{1}{T-k}\sum \hat{u}_t^2$
- ML estimator of the variance of the disturbances is biased, although it is consistent.

optimisation *

- the LLF for a model with time-varying variances cannot be maximised analytical
- The way we do the optimisation is:
  1. Set up LLF.
  2. Use regression to get initial guesses for the mean parameters.
  3. Choose some initial guesses for the conditional variance parameters.
  4. Specify a convergence criterion - either by criterion or by value.

Non-Normality and Maximum Likelihood

- are $\hat{v}_t = \frac{\hat{u}_t}{\hat{\sigma}_t}$ normal
- Typically are still leptokurtic
- We can use the ML with a robust variance/covariance estimator, called Quasi-Maximum Likelihood or QML

[Extra1] Fama-French

- $R^e_{i,t} = \beta_0 + \beta_1 * MKT_t + \beta_2 * SUB_t + \beta_3 * HUL_t + u_{it}$
- $MKT_t = R_{\mu,t} - R_{f,t}$
- BM ration: Book-to-market ratio

.	.	**lowBM(<30th percentile) (growth)**	**midBM(>=30th & <70th) (mid)**	**highBM(>=70th percentile) (value)**
size	low(<mediam)	(1)	(2)	(3)
size	high(>=mediam)	(4)	(5)	(6)

- $SUB = \frac{1}{3}[(1) + (2) + (3)] - \frac{1}{3}[(4) + (5) + (6)]$
- $HUL = \frac{1}{2}[(3) + (6)] - \frac{1}{2}[(1) + (4)]$

[Extra2] Guest lecture: Global Factor Premiums

- P-hacking
- Portfolio construction
  - Zero-investment portfolio construction method for each asset-class factor:
    - $w^i_t = z_t * (Rank(S^i_t) - \frac{N_t+1}{2})$
  - To create equal-risk positions, we scale each asset by its 3-year return volatility and each asset-class factor by its 10-year return volatility
  - For multi-asset factor portfolios, we equally weight the asset-class factors
  - We rebalance the factor portfolios each month
  - Returns of all long-short portfolios are expressed in US dollars
  - We exclude hyperinflation periods (ex-ante > 50% p.m. inflation, Cagan 1956)
  - We exclude assets with more than 2 missing or zero observations in past 12m
- Summary and conclusion
  - When we critically examine the evidence for factors published in leading finance journals, replicate it removing some degrees of freedom, and take the possibility of p-hacking into account, it is positive but rather weak.
  - As a remedy, we construct a deep historical sample starting in 1800, and test the performance of original factors on this previously unused sample. It turns out that most are statistically significant, even accounting for phacking
  - We try to explain why these global factor premiums exist, but cannot find evidence that they are due to market risks, downside risks, or macro risks

[Summary1] all the TS and CV

- T-test
  - $TS = \frac{\hat{\beta}-\beta}{SE(\hat{\beta})}$
  - $CV \sim t_{1-\alpha/2}(T-2)$
    - for MLRM, 2 is number of **varibles + 1(intercept)**

- F-test
  - $TS = \frac{RRSS-URSS}{URSS} * \frac{T-k}{m}$
  - $CV \sim F_{1-\alpha}(m, T-k)$
    - m: no. of restrictions

- White's test
  - $TS = TR^2$
    - R^2 of the auxiliary regression $\hat{\mu}_t^2 = \alpha_1 + \alpha_2 x_{2t} + \alpha_3 x_{3t} + \alpha_4 x_{2t}^2 + \alpha_5 x_{3t}^2 + \alpha_6 x_{2t} x_{3t} + v_t$
  - $CV \sim \chi^2_{1-\alpha}(m)$$
    - m is the no. of para. in the auxiliary regression (without the constant term)

- Breusch-Godfrey test
  - $TS = (T-r)R^2$
    - R^2 of $\mu_t = \rho_1\mu_{t-1} + \rho_2\mu_{t-2} + \ldots + \rho_r\mu_{t-r} + v_t$, v_t~$N(0, \sigma_v^2)$
  - $CV \sim \chi^2_{1-\alpha}(r)$

- Jarque and Bera test
  - $TS = T[\frac{b_1^2}{6} + \frac{(b_2-3)^2}{24}]$
    - skewness b_1 and the excess kurtosis b_2 of the $\hat{u}_t$
  - $CV \sim \chi^2_{1-\alpha}(2)$

- Q-statistic (Box-Pierce)
  - $Q = T\Sigma_{k=1}^m \hat{\tau}_k^2$
    - m: lags
    - correlation: $\tau_s = \gamma_s/\gamma_0$
      - autocovariance function: $E[(y_t - E(y_t))(y_{t-s} - E(y_{t-s}))] = \gamma_s$
  - $CV \sim \chi^2_{1-\alpha}(m)$

- Q-statistic (Ljung-Box)
  - $Q = T(T+2)\Sigma_{k=1}^{m} \frac{\hat{\tau}_k^2}{T-k}$
  - $CV \sim \chi^2_{1-\alpha}(m)$

- likelihood ratio test
  - $LR = T(log \mid \hat{\Sigma}_r \mid - log \mid \hat{\Sigma}_u \mid)$
    - Denote the variance-covariance matrix of residuals (given by $\hat{\mu}\hat{\mu}'/T$), as Σ
    - r, u: restricted, unrestricted
  - $CV \sim \chi^2_{1-\alpha}$ (total number of restrictions)

- Dickey-Fuller (**DF**) test (unit root test)
  - TS=$\frac{\hat{\psi}}{SE(\hat{\psi})}$
    - use $\Delta y_t = \psi y_{t-1} + \mu_t$
  - CV is a table

- Augmented Dickey Fuller (**ADF**) test (unit root test)
  - TS=$\frac{\hat{\psi}}{SE(\hat{\psi})}$
    - use $\Delta y_t = \psi y_{t-1} + \Sigma_{i=1}^{p} \alpha_i \Delta y_{t-i} + \mu_t$
  - CV is a table
    - if TS (negative) < CV (negative), reject H0
    - H0: a unit root (non-stationary)
    - H1: stationary

- KPSS test (stationarity test)
  - $TS = \quad ??$
  - $CV \sim \quad ??$
    - H0: stationary
    - H1: non-stationary

- Stationarity tests
  - use ADF and KPSS together

- ARCH effects test
  - $TS = TR^2$
    - R^2 of $\mu_t^2 = \gamma_0 + \gamma_1 u_{t-1}^2 + \gamma_2 u_{t-2}^2 + \ldots + \gamma_q u_{t-q}^2 + v_t$
  - $CV \sim \chi^2_{1-\alpha}(q)$