Lecture Topics

Maximum Likelihood

Linear models can be estimated using OLS

Maximum Likelihood

Linear models can be estimated using OLS

Maximum likelihood is a more general estimator allowing non linear models

Maximum Likelihood

Linear models can be estimated using OLS

Maximum likelihood is a more general estimator allowing non linear models

Requires knowledge or assumption of the distribution the data is drawn from

Maximum Likelihood

Linear models can be estimated using OLS

Maximum likelihood is a more general estimator allowing non linear models

Requires knowledge or assumption of the distribution the data is drawn from

For example that returns have a normal or t distribution

Maximum Likelihood

Linear models can be estimated using OLS

Maximum likelihood is a more general estimator allowing non linear models

Requires knowledge or assumption of the distribution the data is drawn from

For example that returns have a normal or t distribution

For this course you just need to be able to understand broadly the idea and be able to interpret the estimated coefficients and tests

Volatility Modelling

The last lecture looked at modelling the conditional mean $E[R_t|I_{t-1}]$

Volatility Modelling

The last lecture looked at modelling the conditional mean $E[R_t|I_{t-1}]$

E.g. this may be an AR(1), i.e. $E[R_t|I_{t-1}] = \mu + \rho R_{t-1}$

Volatility Modelling

The last lecture looked at modelling the conditional mean $E[R_t|I_{t-1}]$

E.g. this may be an AR(1), i.e. $E[R_t|I_{t-1}] = \mu + \rho R_{t-1}$

MA models are non-linear- can be estimated via ML

Volatility Modelling

The last lecture looked at modelling the conditional mean $E[R_t|I_{t-1}]$

E.g. this may be an AR(1), i.e. $E[R_t|I_{t-1}] = \mu + \rho R_{t-1}$

MA models are non-linear- can be estimated via ML

We move on to look at modelling the conditional variance of returns $Var(R_t|I_{t-1})$

Apple Share Price Return

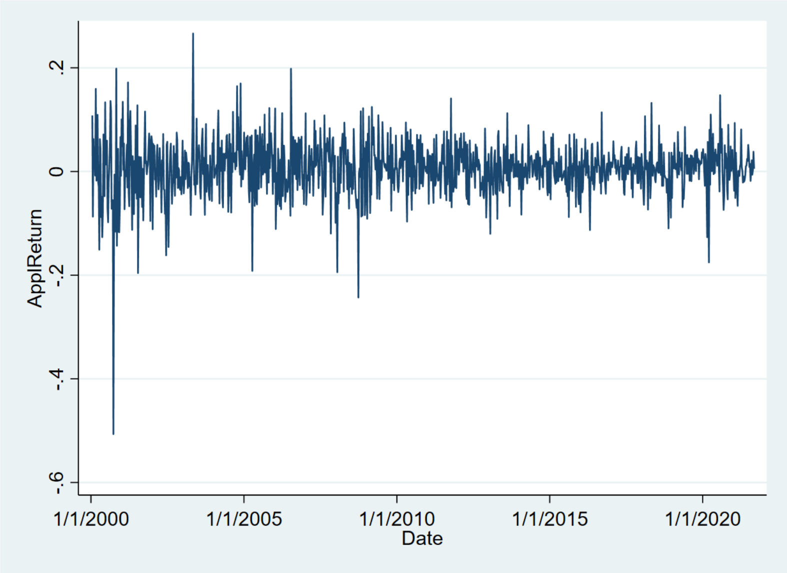

Figure 1: Weekly Apple Stock Return 2000-2021

Apple Share Price Return Correlogram

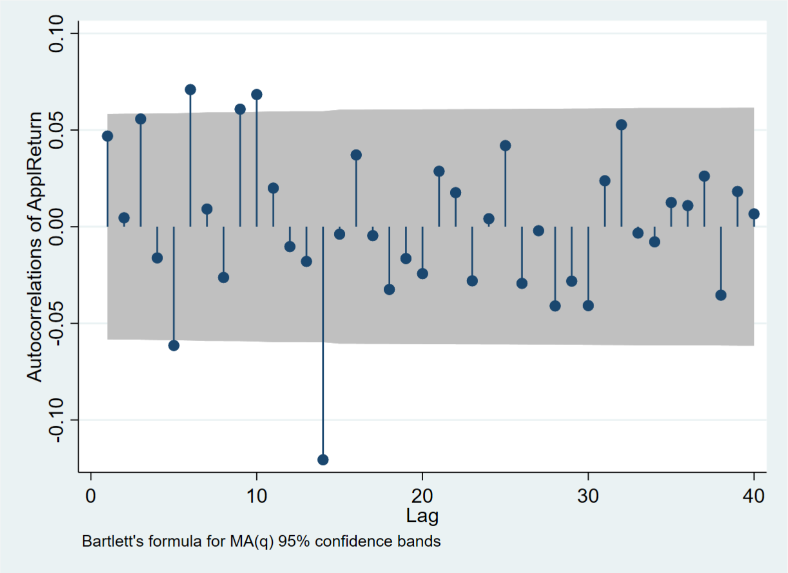

Figure 2: Weekly Apple Price Return Correlogram

Apple Share Price Return Correlogram

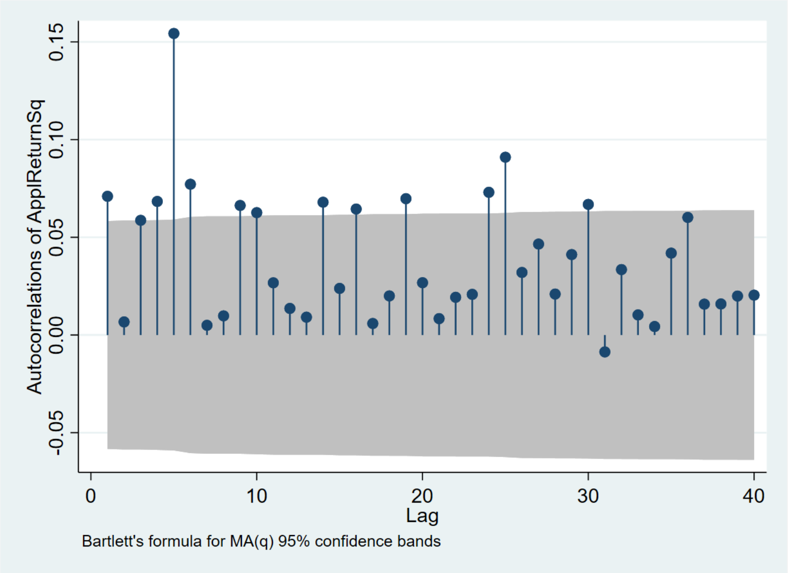

Figure 3: Weekly Squared Apple Price Return Correlogram

Maximum Likelihood

Maximum Likelihood (ML):Idea

Have data set $\{y_i\}_{i=1}^N$ would like to estimate the unknown PDF $g(y_i)$

Maximum Likelihood (ML):Idea

Have data set $\{y_i\}_{i=1}^{N}$ would like to estimate the unknown PDF $g(y_i)$

ML assumes $g(y_i) = f(y_i, \beta_0)$ for some $p \times 1$ parameter β_0

Maximum Likelihood (ML):Idea

Have data set $\{y_i\}_{i=1}^{N}$ would like to estimate the unknown PDF $g(y_i)$

ML assumes $g(y_i) = f(y_i, \beta_0)$ for some $p \times 1$ parameter β_0

E.g if $y_i \sim Bern(\beta_0)$ when $g(y_i) = \beta_0^{y_i}(1-\beta_0)^{1-y_i}$ then $g(y_i) = f(y_i, \beta)$ at $\beta = \beta_0$

Maximum Likelihood (ML):Idea

Have data set $\{y_i\}_{i=1}^N$ would like to estimate the unknown PDF $g(y_i)$

ML assumes $g(y_i) = f(y_i, \beta_0)$ for some $p \times 1$ parameter β_0

E.g if $y_i \sim Bern(\beta_0)$ when $g(y_i) = \beta_0^{y_i}(1 - \beta_0)^{1-y_i}$ then $g(y_i) = f(y_i, \beta)$ at $\beta = \beta_0$

ML is a general method to estimate the distribution of some variable, which involves estimating a set of parameters.

Maximum Likelihood (ML)

ML assumes we know the function $f(\cdot, \beta)$ is known

Once we know β_0 we know the distribution of y_i

Need to estimate β_0.

Need to construct an estimator that has "nice" properties like the OLS estimator of a linear model (under the linear model assumptions)

ML estimator achieves this (under assumptions)

Maximum Likelihood (ML)

OLS chooses β to minimise the sum of squared residuals

Many estimators in Econometrics can be expressed as the minimiser/maximiser of some function

For ML want to choose β such that $f(y_i, \beta)$ is as near as possible to the unknown $g(y_i)$

$$K(\beta) = E\left[\log\left(\frac{g(y_i)}{f(y_i,\beta)}\right)\right]$$

Known as Kullback Leibler Info. Criterion

Can show that $K(\beta) \geq 0$ and $K(\beta_0) = 0$

Maximum Likelihood (ML)

If there is a unique β_0 such that $f(y_i, \beta) = g(y_i)$

Then $K(\beta) = 0$ only at β_0

Hence β_0 is the minimiser of $K(\beta)$

For OLS β_0 minimises $E[(y_i - x_i'\beta)^2]$

OLS estimator solves sample version of this

ML minimises sample version of $K(\theta)$

Maximum Likelihood (ML) Idea

$K(\beta) = E[\log(g(y_i))] - E[\log(f(y_i, \beta)]$

Hence minimiser of $K(\theta)$ is the minimiser of $-E[\log(f(y_i, \beta)]$

Minimiser of $-E[\log(f(y_i, \beta)]$ is maximiser of $E[\log(f(y_i, \beta)]$

$E[\log(f(y_i, \beta)]$ known as Likelihood Function

Hence the name- Maximum Likelihood Estimation

Maximum Likelihood (ML) Idea

Need to estimate $E[\log(f(y_i, \beta)]$

Use $\frac{1}{N}\sum_{i=1}^{N} \log f(y_i, \beta)$

$f(y_1, .., y_n)$ is the joint likelihood function

$\frac{1}{N}\sum_{i=1}^{N} \log f(y_i, \beta) = \frac{1}{N}\log(f(y_1, .., y_n)$ when $y_1, ..y_n$ are i.i.d.

Intuition often given is that β select to maximise probability of observing the sample $(y_1, .., y_n)$

Parameter Inference and Testing

Distribution of ML estimators

- To perform inference on any of the estimated parameters we need a distribution.
- The asymptotic distribution of the $(k \times 1)$ dimensional $\widehat{\beta}_{ML}$ is (without any derivation):

$$\widehat{\beta}_{ML} \overset{a}{\sim} N\left(\beta_0, \mathcal{I}_n^{-1}\left(\beta_0\right)\right)$$

centered around the true but unknown β_0, where

$$\mathcal{I}_n\left(\beta_0\right) = -E\left(\frac{\partial^2 l\left(\beta_0\right)}{\partial \beta_0 \partial \beta_0'}\right)$$

is called the information matrix, which is unobserved.

Distribution of ML estimators

$\hookrightarrow$ In practice we use the estimated version

$$\widehat{\beta}_{ML} \overset{a}{\sim} N\left(\beta_0, \mathcal{I}_n^{-1}\left(\widehat{\beta}_{ML}\right)\right)$$

where

$$\underset{(k\times k)}{\mathcal{I}_n^{-1}\left(\widehat{\beta}_{ML}\right)} = -\left(\frac{\partial^2 l\left(\widehat{\beta}_{ML}\right)}{\partial\widehat{\beta}_{ML}\partial\widehat{\beta}'_{ML}}\right)^{-1} \tag{1}$$

and $l\left(\widehat{\beta}_{ML}\right)$ is the log-likelihood function evaluated at the ML estimate.

$\hookrightarrow$ For some $l\left(\widehat{\beta}_{ML}\right)$ its derivatives can be written down analytically. Often we require numerical estimates.

Inference

- *Single restrictions*: t- are feasible and asymptotically $N(0,1)$ distributed.if you use $se_{\widehat{\beta}_i}$ from (1)
- *Multiple restrictions* are tested using a Likelihood-Ratio test

$$LR = 2\left(\log L\left(\widehat{\beta}_u\right) - \log L\left(\widehat{\beta}_r\right)\right)$$

where $\widehat{\beta}_u$ is the unrestricted parameter estimate and $\widehat{\beta}_r$ the parameters estimate with restrictions imposed (e.g. variable exclusions)
- If restrictions are valid then imposing them should cost very little (in terms of likelihood)
- Asymptotically $LR \sim \chi^2_p$ where (as usual) p is the number of restrictions imposed

Parameter Estimation

Parameter Estimation / Nonlinear Optimisation

- How do we find the ML estimates**?**
- Here only minimal details! The procedure is a sophisticated trial and error strategy which follows the following steps:

1. Propose an intitial value for β, $\widehat{\beta}_0$ and calculate the log-likelihood function $l\left(\widehat{\beta}_0\right)$ .Declare iteration $0 = best$
2. Iterate over $i = 1, 2, 3, \ldots$
   - 2.1 Find an alternative $\widehat{\beta}_i = \widehat{\beta}_{best} + d$
   - 2.2 calculate the log-likelihood function $l\left(\widehat{\beta}_i\right)$
   - 2.3 `IF` $l\left(\widehat{\beta}_i\right) > l\left(\widehat{\beta}_{best}\right)$ declare $i = best$ and go to next i
     `ELSE` go to 2a and try a few different d
3. Stop the iterations in 2 if no improvements can be found and declare $\widehat{\beta}_{ML} = \widehat{\beta}_{best}$

- The actual implementation of this algorithm differs in how to determine d and when to stop the iterations.