

Financial Econometrics

Estimation and Inference

Gerald P. Dwyer

Trinity College, Dublin

January 2011

About Me

- Professor
- Federal Reserve Bank of Atlanta
- Research
- Brief summary
- jerry@jerrydwyer.com and <http://www.jerrydwyer.com>

What is Financial Econometrics?

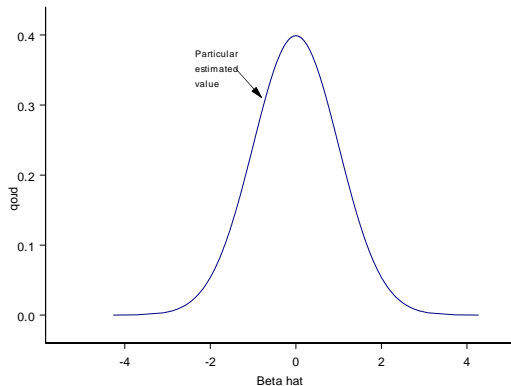
- Financial econometrics is the use of econometric procedures to answer financial questions using financial data
- What sorts of questions?
 - What factors affect stock returns and how much do they do so?
 - Interest rates on government debt have increased substantially. Are they likely to go down?
 - Statistical analysis to inform an economic analysis
 - How likely is it that a portfolio will lose 20 percent of its value in any given 12-month period?
 - Are stock prices mean reverting? Are stock returns mean reverting?
 - What is the effect of a recession on stock prices?
 - Is the value of the euro likely to go up or down? What does it depend on?
- Mostly time series data

Topics covered

- Estimation
- Summarizing data and behavior of returns
- Univariate time series
- Multivariate time series (Vector autoregressions)
- Multivariate time series (Error correction mechanisms)
- Volatility
- Multivariate volatility
- Nonlinear time series analysis
- Value at risk
- Simulation in financial econometrics

Purpose of inference

- What are plausible and implausible values of estimates of a particular parameter?
 - Point estimate



Criteria for estimators

- Classical statistics
 - Minimum Variance Unbiased Estimators (MVUE)
 - or Best Linear Unbiased Estimator (BLUE)
 - or Ordinary Least Squares (OLS)
 - Maximum likelihood
 - Conditional on the data, pick the most likely value

- Least squares with x fixed (nonstochastic)
- $y_i = x_i\beta + \varepsilon_i, \quad i = 1, \dots, N$
 - $E y_i = E x_i = 0$
 - $E \varepsilon_i = 0, E \varepsilon_i^2 = \sigma^2, E \varepsilon_i \varepsilon_j = 0 \forall i \neq j$
- $\hat{\beta} = \frac{\sum xy}{\sum x^2} = \frac{\sum xx\beta}{\sum x^2} + \frac{\sum x\varepsilon}{\sum x^2} = \beta + \frac{\sum x\varepsilon}{\sum x^2}$
- Suppose that x is not stochastic
 - x is deterministic, fixed in repeated samples
 - e.g. treatments of crops on plots
 - time trend
 - quarterly dummy variables
 - $E \hat{\beta} = E \beta + E \frac{\sum x\varepsilon}{\sum x^2} = \beta + \frac{\sum x E \varepsilon}{\sum x^2} = \beta$

Why x nonstochastic?

- $E \frac{\sum x\varepsilon}{\sum x^2}$
 - If x is not random, then $E \frac{\sum x\varepsilon}{\sum x^2} = \frac{\sum x E\varepsilon}{\sum x^2}$
 - If x is random, then $E \frac{\sum x\varepsilon}{\sum x^2} \neq \frac{E \sum x\varepsilon}{E \sum x^2}$ in general
 - Expectations operator is a linear operator
 - $E ax = a E x$ if a is a constant
 - If $\frac{x}{x^2}$ is a constant, then $E \frac{x\varepsilon}{\sum x^2} = \frac{x}{\sum x^2} E \varepsilon$
 - In general, $E \frac{x\varepsilon}{\sum x^2} \neq \frac{E(x\varepsilon)}{E \sum x^2}$ and $E \frac{x\varepsilon}{\sum x^2} \neq E \left[\frac{x}{\sum x^2} \right] E \varepsilon$

Right-hand side variable (x) stochastic and least squares works

- The case with x stochastic in which least squares works: x and ε are independent
 - $E \frac{x\varepsilon}{\sum x^2} = E [f(x) \varepsilon]$ with $f(x) = \frac{x}{\sum x^2}$
 - $E [f(x) \varepsilon] = E f(x) E \varepsilon = 0$ because x and ε are independent and $E \varepsilon = 0$

OLS is MVUE and BLUE

- $\text{Var} \left[\hat{\beta} \right]$ around true value a minimum among estimators that are unbiased
- $\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} = \sum w_i y_i, \quad w_i = \frac{x_i}{\sum x_i^2}$
 - $\hat{\beta}$ a linear function of the y_i and is unbiased
 - Best Linear Unbiased Estimator (BLUE)

Unbiasedness in a time series setting

- Unbiasedness will hardly come up in this class
- Why?
- Time series regression
 - $y_t = \beta y_{t-1} + \varepsilon_t, \quad t = 1, \dots, T$
 - Assume that y_{t-1} and ε_t are independent
 - Correlation of y_{t-1} and ε_t is zero
 - Implies that $E y_{t-1} \varepsilon_t = 0$
 - An ordered sequence of observations from 1 to T
 - This is called a first-order autoregression
 - y_0
 - \downarrow
 - $y_1 \leftarrow \varepsilon_1$
 - \downarrow
 - $y_2 \leftarrow \varepsilon_2$
 - ...

OLS an unbiased estimator?

- Ordinary Least Squares (OLS) estimator

- $$\hat{\beta} = \frac{\sum y_t y_{t-1}}{\sum y_{t-1}^2} = \frac{\sum \beta y_{t-1} y_{t-1}}{\sum y_{t-1}^2} + \frac{\sum \varepsilon_t y_{t-1}}{\sum y_{t-1}^2} = \beta + \frac{\sum \varepsilon_t y_{t-1}}{\sum y_{t-1}^2}$$

- y_{t-1} not fixed in repeated samples

- Can't have different ε_t and same $y_{t-1} \forall T$
- For example, a different ε_2 implies a different y_2 and a different y_2 implies a different y_3 , and so on

- So y_{t-1} must be stochastic

- Seems like y_{t-1} and ε_t are independent and they are

- By assumption, ε_t is independent of y_{t-1}

- But y_t depends on ε_t , and so does y_{t+1} , y_{t+2} , etc.

- $$E \hat{\beta} = \beta + E \frac{\sum \varepsilon_t y_{t-1}}{\sum y_{t-1}^2} = \beta + \sum E f \left(\frac{y_{t-1}}{\sum y_{t-1}^2} \right) \varepsilon_t$$

- ε_t and $\sum_{t=1}^T y_{t-1}^2$ cannot be independent and $\hat{\beta}$ is not an unbiased estimator in general

Unbiasedness in a time series context

- In general, estimators are not unbiased in a time series context because they're part of a sequence
- Will focus on consistency

Simple problem of estimating the mean of a normally distributed variable

- $y \sim N(\mu, \sigma^2)$ or can be written $y \sim NIND(\mu, \sigma^2)$
 - $\bar{y}_T = \frac{\sum y_t}{T}$ and $s_T^2 = \frac{\sum (y_t - \bar{y})^2}{T-1}$
 - It's well known that $E \bar{y}_T = \mu$ and $E s_T^2 = \sigma^2$
 - $\text{Var}[\bar{y}_T] = \frac{\text{Var}[\sum y_t]}{T^2}$
 - $= T^{-2} \sum \text{Var}[y_t]$ by independence
 - $= \frac{\sum \sigma^2}{T^2} = \frac{T\sigma^2}{T^2} = \frac{\sigma^2}{T}$
 - In general, estimators are not unbiased in a time series context because they're part of a sequence

Definition of convergence in probability

- Let θ_T be an estimator with sample size T and θ a parameter with some particular value
- Definition: θ_T **converges in probability** to a constant θ if
$$\lim_{T \rightarrow \infty} \Pr(|\theta_T - \theta| > \epsilon) = 0 \quad \forall \epsilon > 0$$
- Write $\text{plim } \theta_T = \theta$
- Example
 - θ_T is an estimator that takes on the values 0 and T
 - $\Pr(\theta_T = 0) = 1 - \frac{1}{T}$ and $\Pr(x_T = T) = \frac{1}{T}$
 - $\lim_{T \rightarrow \infty} \Pr(\theta_T = 1/T) = 0$
 - $\lim_{T \rightarrow \infty} \Pr(\theta_T = 0) = 1$
 - If $\theta = 0$, then $\lim_{T \rightarrow \infty} \Pr(|\theta_T - \theta| > \epsilon) = 0 \quad \forall \epsilon > 0$
 - and $\lim_{T \rightarrow \infty} \Pr(|\theta_T - \theta| < \epsilon) = 1 \quad \forall \epsilon > 0$
 - and $\text{plim } \theta_T = 0$

Consistency and convergence in mean square

- An estimator θ_T is a consistent estimator of θ if and only if $\text{plim } \theta_T = \theta$
- Often a convenient way to prove consistency is to prove mean square convergence
 - Convergence in mean square implies convergence in probability
 - Convergence in probability does not imply convergence in mean square
 - In other words, an estimator can converge in probability even though it does not converge in mean square
 - Definition: θ_T **converges in mean square** to a constant θ with $\text{Var } \theta_T = \sigma_T^2$ if
 - $\lim_{T \rightarrow \infty} E \theta_T = \theta$ and $\lim_{T \rightarrow \infty} \sigma_T^2 = 0$
 - then $\text{plim } \theta_T = \theta$
 - $\lim_{T \rightarrow \infty} \Pr(|\theta_T - \theta| > \epsilon) = 0 \quad \forall \epsilon > 0$
- This method of proof can be applied to show that the mean \bar{y}_T of a normally distributed variable is a consistent estimator of μ

Properties of probability limits

- Suppose have estimates a_T of a parameter α and b_T of a parameter β
- Suppose that $\text{plim } a_T = \alpha$ and $\text{plim } b_T = \beta$. Then
 - $\text{plim } (a_T + b_T) = \text{plim } a_T + \text{plim } b_T = \alpha + \beta$
 - $\text{plim } a_T \text{plim } b_T = \alpha\beta$
 - $\text{plim } a_T / \text{plim } b_T = \alpha/\beta$ if $\beta \neq 0$
- This can be contrasted with the expectation operator for which $E a_T b_T \neq \alpha\beta$ in general

Convergence in distribution

- Want nondegenerate distribution of estimator
 - If an estimator θ_T is a consistent estimator of θ , then estimator converges to a constant
 - We want some measure of the variability of the estimator
 - This is where the asymptotic distribution comes in
- The **asymptotic distribution** of an estimator is a distribution that is used to approximate the finite-sample distribution of the estimator
- Some function of the estimator converges to a distribution, the asymptotic distribution

Limiting Distribution

- Definition: If θ_T converges in distribution to the random variable θ , where $F(\theta)$ is the cumulative distribution function of θ , then $F(\theta)$ is the **limiting distribution** of θ_T
 - Often written $\theta_T \rightarrow^d F(\theta)$
 - If $F(\theta)$ is a common form such as $N(\mu, \sigma^2/T)$, this is often written as $\theta_T \rightarrow^d N(\mu, \sigma^2/T)$
 - (It is proved by showing, for example, that $\sqrt{T}\theta_T$ converges to $N(\mu, \sigma^2)$)

Bottom line on asymptotics and time series

- Consistency is more pertinent than unbiasedness
- The limiting distribution provides a way to estimate the variability of the estimator
 - Some algebra can show that the mean \bar{y}_T of a normally distributed variable has the asymptotic distribution $N(\mu, \sigma^2 / T)$
 - This is the same as the finite-sample distribution in this case, but the asymptotic distribution often is easier to find

Maximum likelihood estimation

- Maximum likelihood often is a convenient way to obtain a consistent estimator
- Maximum likelihood uses the distribution of the observations
- Maximum likelihood obtains point estimates of the parameters as the ones most likely to have generated the observations

Maximum likelihood estimation of the parameters of a normal distribution

- Have a sample of T observations, y_1, y_2, \dots, y_T
- Suppose they are generated independently from a normal distribution with mean μ and variance σ^2
 - Each observation has the distribution
$$\frac{1}{\sigma(2\pi)^{1/2}} \exp \left[-\frac{1}{2\sigma^2} (y_t - \mu)^2 \right]$$
 - The joint sample of T observations has the distribution
$$f(y_t | \mu, \sigma^2) = \frac{1}{\sigma^T (2\pi)^{T/2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - \mu)^2 \right]$$
 - The likelihood function of these data and parameters is
$$L(\mu, \sigma^2 | y_t) = \frac{1}{\sigma^T (2\pi)^{T/2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - \mu)^2 \right]$$
 - For exponential distributions, the log of the likelihood often is more convenient
- $\ln L(\mu, \sigma^2 | y_t) = -\frac{T}{2} 2\pi - T \ln \sigma - \frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - \mu)^2$

Maximum likelihood estimation of mean and variance

- $\ln L(\mu, \sigma^2 | y_t) = -\frac{T}{2} \ln 2\pi - T \ln \sigma - \frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - \mu)^2$
- Maximize likelihood as a function of parameters conditional on the data
 - Can do all at once or sequentially
- μ
 - Denote the estimator by a “hat” over it
 - $\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2} \sum_{t=1}^T (y_t - \hat{\mu}) = 0$
 - $\sum_{t=1}^T (y_t - \hat{\mu}) = 0$
 - $\sum_{t=1}^T y_t = T \hat{\mu}$
 - $\hat{\mu} = \frac{\sum_{t=1}^T y_t}{T} = \bar{y}$

Illustration

- Likelihood function

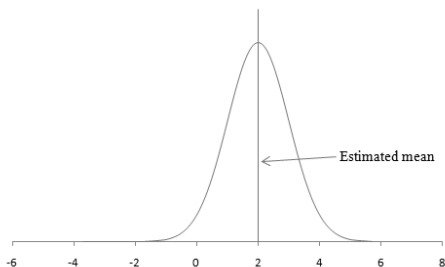


Figure: Likelihood function of normally distributed data with mean \bar{y} of 2 and variance $s_{ml}^2 = \sum (y - \bar{y})^2 / T$ of 1. The maximum likelihood estimator is the mean of the normal distribution.

Finish by estimating variance

- σ^2
 - Concentrate μ out of likelihood function by replacing it by \bar{y}
 - $\ln L(\sigma^2 | y_t) = -\frac{T}{2} \ln 2\pi - T \ln \sigma - \frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - \bar{y})^2$
 $= -\frac{T}{2} \ln 2\pi - \frac{T}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - \bar{y})^2$
 - $\frac{\partial \ln L}{\partial \sigma^2} = -\frac{T}{2} \frac{1}{\sigma^2} + \frac{1}{2} \frac{1}{\sigma^4} \sum_{t=1}^T (y_t - \bar{y})^2 = 0$
 - $-T + \frac{1}{\sigma^2} \sum_{t=1}^T (y_t - \bar{y})^2 = 0$
 - $\hat{\sigma}^2 = \frac{\sum_{t=1}^T (y_t - \bar{y})^2}{T}$

Consistency and unbiasedness

- $\hat{\mu} = \bar{y}$ and $\hat{\sigma}^2 = \frac{\sum_{t=1}^T (y_t - \bar{y})^2}{T}$ are consistent estimators of μ and σ
 - Not necessarily unbiased
 - $\hat{\sigma}^2 = \frac{\sum_{t=1}^T (y_t - \bar{y})^2}{T}$ is a biased estimator of σ
 - Not very important with enough observations

Properties of maximum likelihood estimation mentioned

- Maximum likelihood provides a couple of natural estimators of the variance of the estimator
 - Let θ be a parameter we have estimated
 - $\text{Var} [\hat{\theta}] \geq \left(\text{E} [(\partial \ln L(\theta|y) / \partial \theta)]^2 \right)^{-1}$
 - where the expectation with respect to the distribution of y is evaluated at the true parameter
 - Under regularity conditions
$$\text{E} [(\partial \ln L(\theta|y) / \partial \theta)]^2 = \text{E} \left[-\frac{\partial^2 \ln L(\theta|y)}{\partial \theta^2} \right]^{-1}$$
 - $-\text{E} \left[-\frac{\partial^2 \ln L(\theta|y)}{\partial \theta^2} \right]$ often is called the information matrix, I

Common properties of maximum likelihood estimators

- Let $\hat{\theta}_{ML}$ be the maximum likelihood of some estimator
 - Suppose that the likelihood function has a single peak and a unique maximum
- Fairly general properties
 - $\text{plim } \hat{\theta}_{ML} = \theta$
 - $\hat{\theta}_{ML} \rightarrow^d \text{N}(\theta, I^{-1})$
 - Asymptotic variance of $\hat{\theta}_{ML}$ is $\text{AVar}(\hat{\theta}_{ML}) = I^{-1}$
 - Asymptotic standard deviation of $\hat{\theta}_{ML}$ is $\text{ASD}(\hat{\theta}_{ML})$
 - t-ratio is $\frac{\hat{\theta}_{ML} - \theta}{\text{ASD}(\hat{\theta}_{ML})} \sim \text{N}(0, 1)$
 - Can do more complicated tests by likelihood ratio test
 - $-2 \ln \left(\frac{\text{max Likelihood Restricted}}{\text{max Likelihood Unrestricted}} \right) \sim \chi^2$ (degrees of freedom = number of restrictions)

Bayes rule

- Foundation is Bayes rule
- Combine likelihood function of parameters with prior information to get posterior distribution and conclusions
 - Prior – before the data
 - Posterior – after the data

Bayes rule is simple

- The application is the big jump
- Start from definition of conditional probability
 - $\text{pr}(A, B) = \text{pr}(A|B) \text{pr}(B)$
 - where $\text{pr}(A, B)$ is the joint probability of two events A and B
 - $\text{pr}(A|B)$ is the probability of the event A conditional on the event B
 - $\text{pr}(B)$ is the probability of B
 - This equation defines conditional probability
 $\text{pr}(A|B) = \text{pr}(A, B) / \text{pr}(B)$ if $\text{pr}(B) \neq 0$
 - Also can say $\text{pr}(A, B) = \text{pr}(B|A) \text{pr}(A)$
 - Two different ways of writing joint probability
 - Equate two definitions and get $\text{pr}(B|A) = \frac{\text{pr}(A|B) \text{pr}(B)}{\text{pr}(A)}$

Bayesian interpretation

$$\text{pr}(B|A) = \frac{\text{pr}(A|B) \text{pr}(B)}{\text{pr}(A)}$$

- Want to draw an inference about probability of observing event B
 - Observe some discrete event A
 - $\text{pr}(B|A)$ is the probability of B conditional on observing A
 - $\text{pr}(B)$ is prior probability that B is true
 - $\text{pr}(A|B)$ is probability of observing A if B is true
 - $\text{pr}(A)$ is the probability of observing A whether B is true or not
 - Note that $\text{pr}(A)$ is the unconditional probability of observing A
 - $\text{pr}(A) = \text{pr}(A|B) \text{pr}(B) + \text{pr}(A|\text{not } B) \text{pr}(\text{not } B)$

$$\text{pr}(B|A) = \frac{\text{pr}(A|B) \text{pr}(B)}{\text{pr}(A|B) \text{pr}(B) + \text{pr}(A|\text{not } B) \text{pr}(\text{not } B)}$$

Example of Bayesian analysis

$$\text{pr}(B|A) = \frac{\text{pr}(A|B) \text{pr}(B)}{\text{pr}(A|B) \text{pr}(B) + \text{pr}(A|\text{not } B) \text{pr}(\text{not } B)}$$

- Example: B is the result that have illness, say flu
 - A is some evaluation
 - Have a prior probability of having flu, $\text{pr}(B)$, say 50 percent
 - How informative is it if you go to doctor's office and he says you have the flu
 - Suppose that he says you have the flu
 - 80 percent of time when you do $\text{pr}(A|B)$
 - 20 percent when you don't $\text{pr}(A|\text{not } B)$
 - Then the probability of your having the flu given the doctor says you do is

$$\frac{.8 \cdot .5}{.8 \cdot .5 + .2 \cdot .5} = \frac{.40}{.50} = .80$$

- A lot of information in the doctor's evaluation

Second example of Bayesian analysis

$$\text{pr}(B|A) = \frac{\text{pr}(A|B) \text{pr}(B)}{\text{pr}(A|B) \text{pr}(B) + \text{pr}(A|\text{not } B) \text{pr}(\text{not } B)}$$

- Have a prior probability of having flu, $\text{pr}(B)$, say 50 percent
- Go to doctor's office and he says you have the flu
- Suppose that he says you have the flu
 - 60 percent of time when you do $\text{pr}(A|B)$
 - 40 percent when you don't $\text{pr}(A|\text{not } B)$
- Then the probability of your having the flu given the doctor says you do is

$$\frac{.6 \cdot .5}{.6 \cdot .5 + .4 \cdot .5} = \frac{.30}{.50} = .60$$

- If $\text{pr}(A|B)$ and $\text{pr}(A|\text{not } B)$ are both 0.5, then $\text{pr}(B|A) = .5$, the prior probability

- First example
 - prior probability of flu is .5
 - probability that doctor will say you have the flu is .8 if you do
 - posterior probability is .8
- Second example
 - prior probability of flu is .5
 - probability that doctor will say you have the flu is .6 if you do
 - posterior probability is .6
- You had a diffuse prior – equal probabilities of flu or not – and you learned what can be learned from doctor

Third example of Bayesian analysis

$$\text{pr}(B|A) = \frac{\text{pr}(A|B) \text{pr}(B)}{\text{pr}(A|B) \text{pr}(B) + \text{pr}(A|\text{not } B) \text{pr}(\text{not } B)}$$

- Have a prior probability of having flu, $\text{pr}(B)$, say 80 percent
- Go to doctor's office and he says you have the flu
- Suppose that he says you have the flu
 - 80 percent of time when you do $\text{pr}(A|B)$
 - 20 percent when you don't $\text{pr}(A|\text{not } B)$
- Then the probability of your having the flu given the doctor says you do is

$$\frac{.8 \cdot .8}{.8 \cdot .8 + .2 \cdot .2} = \frac{.64}{.68} = .94$$

Analysis in econometric context

- $\text{pr}(B|A) = \frac{\text{pr}(A|B)\text{pr}(B)}{\text{pr}(A)}$
- Can write this in terms of discrete or continuous probability distribution functions
 - Let B be a parameter β and $\text{pr}(B) \equiv \mathbb{P}(\beta)$
 - Might be CAPM parameter
 - Prior probability distribution of plausible values of β for some firm
 - Let A be some data we observe and $\text{pr}(A|B) = p(y|\beta)$, the probability of the data given β

Bayesian analysis of parameter values



$$p(\beta|y) = \frac{L(\beta|y) P(\beta)}{P(y)}$$

where $p(\beta|y)$ is the posterior probability distribution of values of β conditional on the data

- $P(y)$ is a normalizing constant independent of β so this can be analyzed using

$$p(\beta|y) \propto L(\beta|y) P(\beta)$$

where \propto means “proportional to”

- Purpose is to make inferences about the posterior distribution of parameter values
 - Very flexible
 - Coherent
 - Can be computationally demanding but computer time is cheap