Financial Econometrics

Estimation and Inference

Gerald P. Dwyer

Trinity College, Dublin

January 2016

Outline

Who am I?

- Professor and BB&T Scholar at Clemson University
- 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?
- Statistical analysis to inform an economic analysis
  - What factors affect stock returns and how much do they do so?
  - Interest rates on Irish government debt have fallen substantially. Are they likely to go up or down?
  - 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?
  - Is the value of the euro likely to go up or down? What does it depend on?
  - The Swiss franc has risen a lot in the last week and there are widespread losses. Who is losing and why did the have the trades on that they had?
- Mostly time series data

Topics covered

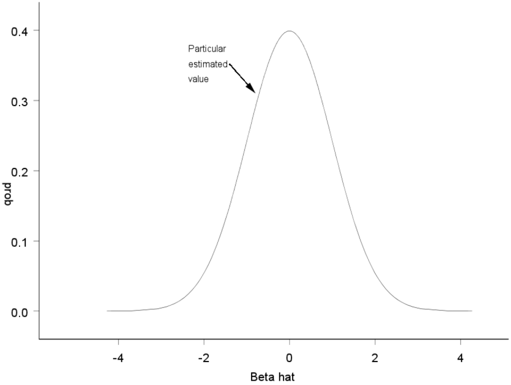
- Estimation
- Summarizing data and behavior of returns
- Event studies
- Univariate time series
- Multivariate time series (Vector autoregressions)
- Multivariate time series (Error correction mechanisms)
- Volatility
- Multivariate volatility
- Nonlinear time series analysis
- Value at risk

Topics covered and text

- Brooks, Chs. 1 and 2 - Estimation and Summarizing data and behavior of returns
- Class slides and Campbell, Lo and MacKinley Ch. 4, Event studies
- Brooks, Ch. 5 - Univariate time series
- Brooks, Ch. 6 - Multivariate time series (Vector autoregressions)
- Brooks, Ch. 7 - Multivariate time series (Error correction mechanisms)
- Brooks, Ch. 8 - Volatility and Multivariate volatility
- Brooks, Ch. 9 - Nonlinear time series analysis
- Riskmetrics Brochure - Value at risk

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

OLS

- Ordinary least squares with x fixed (nonstochastic)
- 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

$$y_i = x_i\beta + \varepsilon_i, \; i = 1, ..., N$$
$$\mathrm{E}\, y_i = \mathrm{E}\, x_i = 0$$
$$\mathrm{E}\, \varepsilon_i = 0, \; \mathrm{E}\, \varepsilon_i^2 = \sigma^2, \; \mathrm{E}\, \varepsilon_i \varepsilon_j = 0 \; \forall \; i \neq j$$

OLS with nonstochastic regressors is unbiased

- Properties of equation

$$y_i = x_i\beta + \varepsilon_i, \;\; i = 1, ..., N$$
$$\mathrm{E}\, y_i = \mathrm{E}\, x_i = 0$$
$$\mathrm{E}\, \varepsilon_i = 0, \;\; \mathrm{E}\, \varepsilon_i^2 = \sigma^2, \;\; \mathrm{E}\, \varepsilon_i \varepsilon_j = 0 \; \forall \; i \neq j$$

OLS with nonstochastic regressors is unbiased

- Properties of equation

$$y_i = x_i\beta + \varepsilon_i, \;\; i = 1, ..., N$$
$$\mathrm{E}\, y_i = \mathrm{E}\, x_i = 0$$
$$\mathrm{E}\, \varepsilon_i = 0, \;\; \mathrm{E}\, \varepsilon_i^2 = \sigma^2, \;\; \mathrm{E}\, \varepsilon_i \varepsilon_j = 0 \; \forall \; i \neq j$$

- $\widehat{\beta}$ can be written

$$\widehat{\beta} = \frac{\sum xy}{\sum x^2} = \frac{\sum xx\beta}{\sum x^2} + \frac{\sum x\varepsilon}{\sum x^2} = \beta + \frac{\sum x\varepsilon}{\sum x^2}$$

OLS with nonstochastic regressors is unbiased

- And the expected value of $\widehat{\beta}$ is

$$\begin{aligned}
\mathrm{E}\,\widehat{\beta} = & \mathrm{E}\,\beta + \mathrm{E}\,\frac{\sum x\varepsilon}{\sum x^2} \\
= & \beta + \frac{\sum x\,\mathrm{E}\,\varepsilon}{\sum x^2} \\
= & \beta
\end{aligned}$$

Why x nonstochastic?

- Consider the term $\mathrm{E}\,\frac{\sum x\varepsilon}{\sum x^2}$ in $\mathrm{E}\,\widehat{\beta} = \beta + \mathrm{E}\,\frac{\sum x\varepsilon}{\sum x^2}$
- If x is not random, then

$$\mathrm{E}\,\frac{\sum x\varepsilon}{\sum x^2} = \frac{\sum x\,\mathrm{E}\,\varepsilon}{\sum x^2}$$

- If x is random, then in general

$$\mathrm{E}\,\frac{\sum x\varepsilon}{\sum x^2} \neq \frac{\mathrm{E}\sum x\varepsilon}{\mathrm{E}\sum x^2}$$

Why unbiased if x nonstochastic?

- Expectations operator is a linear operator
- If a is a constant, then

$$\mathrm{E}\, ax = a\, \mathrm{E}\, x$$

- If $\frac{x}{\sum x^2}$ is a constant, then

$$\mathrm{E}\, \frac{x\varepsilon}{\sum x^2} = \frac{x}{\sum x^2}\, \mathrm{E}\, \varepsilon$$

- In general,

$$\mathrm{E}\, \frac{x\varepsilon}{\sum x^2} \neq \frac{\mathrm{E}\,(x\varepsilon)}{\mathrm{E} \sum x^2} \text{ and } \mathrm{E}\, \frac{x\varepsilon}{\sum x^2} \neq \mathrm{E}\left[\frac{x}{\sum x^2}\right] \mathrm{E}\, \varepsilon$$

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

$$\mathrm{E}\,\frac{x\varepsilon}{\sum x^2} = \mathrm{E}\left[f\left(x\right)\varepsilon\right] \text{ with } f\left(x\right) = \frac{x}{\sum x^2}$$

$$\mathrm{E}\left[f\left(x\right)\varepsilon\right] = \mathrm{E}\,f\left(x\right)\mathrm{E}\,\varepsilon \text{ if } x \text{ and } \varepsilon \text{ are independent}$$

$$\mathrm{E}\,f\left(x\right)\mathrm{E}\,\varepsilon = 0 \text{ because } \mathrm{E}\,\varepsilon = 0$$

- If x and ε are normally distributed and uncorrelated, then least squares is unbiased
  - Sufficient but not necessary

OLS is MVUE and BLUE

- MVUE: Var $\left[\widehat{\beta}\right]$ around true value is a minimum among estimators that are unbiased
- BLUE: $\widehat{\beta}$ is a linear function of the y_i, is unbiased and has minimum variance among unbiased estimators
  - Estimator is a linear function of the y_i because

$$\widehat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} = \sum w_i y_i, \quad w_i = \frac{x_i}{\sum x_i^2}$$

Unbiasedness in a time series setting

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

Unbiasedness in a time series setting

- Unbiasedness will hardly come up in this class
- Why?
- Time series regression with dependence on past values
  - $y_t = \beta y_{t-1} + \varepsilon_t, \quad t = 1, ..., T$
    - Assume that y_{t-1} and ε_t are independent
    - Correlation of y_{t-1} and ε_t is zero
    - Implies that $\mathrm{E}\, y_{t-1}\varepsilon_t = 0$
- 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 for an autoregression?

$$\begin{aligned}\widehat{\beta} =& \frac{\sum y_t y_{t-1}}{\sum y_{t-1}^2} \\ =& \frac{\sum \beta y_{t-1} y_{t-1}}{\sum y_{t-1}^2} + \frac{\sum \varepsilon_t y_{t-1}}{\sum y_{t-1}^2} \\ =& \beta + \frac{\sum \varepsilon_t y_{t-1}}{\sum y_{t-1}^2}\end{aligned}$$

- y_{t-1} not fixed in repeated samples
  - Can't have different ε_t and same set of y's $\forall\ t = 1, ..., T$
  - 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

OLS an unbiased estimator?

- Just because y_{t-1} is stochastic doesn't mean that OLS is not unbiased

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

- 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.

$$\begin{aligned} \mathrm{E}\,\widehat{\beta} =& \beta + \mathrm{E}\,\frac{\sum \varepsilon_t y_{t-1}}{\sum y_{t-1}^2} \\ =& \beta + \sum \mathrm{E}\left[\left(\frac{y_{t-1}}{\sum y_{t-1}^2}\right)\varepsilon_t\right] \end{aligned}$$

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

Digression: Least squares and error term

- One way to see why $\mathrm{E} \sum x\varepsilon = 0$ is required for unbiasedness:
- What is the correlation of the error term and right-hand-side variables in a computed regression? The covariance of x and the computed error term is identically zero (for any correct program)

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

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 $\overline{y}_T$ of a normally distributed variable has the asymptotic distribution $\mathrm{N}\left(\mu, \sigma^2 / T\right)$
  - This is the same as the finite-sample distribution in this case, but the asymptotic distribution often is easier to find

Simple problem of estimating the mean of a normally distributed variable

- In general, estimators are not unbiased in a time series context because they're part of a sequence but they can be unbiased if dependence over time is unimportant
- Suppose y is normally distributed

$$y \sim \mathrm{N}\left(\mu, \sigma^2\right) \text{ or can be written } y \sim \mathrm{NIND}\left(\mu, \sigma^2\right)$$

- By definition

$$\overline{y}_T = \frac{\sum y_t}{T} \text{ and } s_T^2 = \frac{\sum \left(y_t - \overline{y}\right)^2}{T-1}$$

- It's shown in basic statistics that

$$\mathrm{E}\,\overline{y}_T = \mu \text{ and } \mathrm{E}\,s_T^2 = \sigma^2$$

- Also

$$\mathrm{Var}\left[\overline{y}_T\right] = \frac{\mathrm{Var}\left[\sum y_t\right]}{T^2} = T^{-2} \sum \mathrm{Var}\left[y_t\right]$$

$$= \frac{\sum \sigma^2}{T^2} = \frac{T\sigma^2}{T^2} = \frac{\sigma^2}{T}$$

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 $\lim_{T\to\infty} \Pr\left(|\theta_T - \theta| > \epsilon\right) = 0 \ \forall \ \epsilon > 0$
  - ϵ is "some constant value", not an error term
- Write $\operatorname{plim} \theta_T = \theta$

Example of consistent estimator

- Suppose that $\theta = 0$ and θ_T is an estimator that takes on the values 0 and T

$$\Pr(\theta_T = 0) = 1 - \frac{1}{T} \text{ and } \Pr(\theta_T = T) = \frac{1}{T}$$

- Therefore

$$\lim_{T \to \infty} \Pr(\theta_T = T) = 0$$
$$\lim_{T \to \infty} \Pr(\theta_T = 0) = 1$$

- Because $\theta = 0$,

$$\lim_{T \to \infty} \Pr(|\theta_T - \theta| > \epsilon) = 0 \ \forall \ \epsilon > 0$$

- and

$$\lim_{T \to \infty} \Pr(|\theta_T - \theta| < \epsilon) = 1 \ \forall \ \epsilon > 0$$

- and therefore

$$\operatorname{plim} \theta_T = \theta$$

- If θ equalled something other than zero, then θ_T is an inconsistent estimator of θ

Properties of probability limits

- Suppose have estimates a_T of a parameter α and b_T of a parameter β
- Suppose that $\operatorname{plim} a_T = \alpha$ and $\operatorname{plim} b_T = \beta$
- Then

$$\begin{aligned}
\operatorname{plim}(a_T + b_T) &= \operatorname{plim} a_T + \operatorname{plim} b_T = \alpha + \beta \\
\operatorname{plim}(a_T b_T) &= \operatorname{plim} a_T \operatorname{plim} b_T = \alpha\beta \\
\operatorname{plim}(a_T / b_T) &= \operatorname{plim} a_T / \operatorname{plim} b_T = \alpha/\beta \text{ if } \beta \neq 0
\end{aligned}$$

- This can be contrasted with the expectation operator for which, in general,

$$\begin{aligned}
\mathrm{E}\, a_T b_T &\neq \alpha\beta \\
\mathrm{E}\, a_T / b_T &\neq \alpha/\beta
\end{aligned}$$

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 $\mathrm{F}(\theta)$ is the cumulative distribution function of θ, then $\mathrm{F}(\theta)$ is the **limiting distribution** of θ_T
- Often written
$$\theta_T \to^d \mathrm{F}(\theta)$$
- If $\mathrm{F}(\theta)$ is a common form such as $\mathrm{N}\left(\mu, \sigma^2 / T\right)$, this is often written as
$$\theta_T \to^d \mathrm{N}\left(\mu, \sigma^2 / T\right)$$
  - Proved by showing, for example, that $\sqrt{T}\theta_T$ converges to $\mathrm{N}\left(\mu, \sigma^2\right)$

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 $\mathrm{N}\left(\mu, \sigma^2 / T\right)$
  - This is the same as the finite-sample distribution in this case, but the asymptotic distribution often is easier to find

Maximum likelihood estimation is commonly invoked to justify an estimator

- 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 provides a relatively straightforward way of estimating the variance of parameters

Maximum likelihood estimation of the parameters of a normal distribution

- Have a sample of T observations, $y_1, y_2, ..., y_T$
- 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

$$\mathrm{f}\left(y_t|\mu, \sigma^2\right) = \frac{1}{\sigma^T (2\pi)^{T/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{t=1}^{T} (y_t - \mu)^2\right]$$

- The likelihood function of these data and parameters is

$$\mathrm{L}\left(\mu, \sigma^2|y_t,\right) = \frac{1}{\sigma^T (2\pi)^{T/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{t=1}^{T} (y_t - \mu)^2\right]$$

The log of the likelihood function

- The likelihood function of the parameters for a normal distribution is

$$\mathrm{L}\left(\mu,\sigma^2|y_t,\right)=\frac{1}{\sigma^T\left(2\pi\right)^{T/2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(y_t-\mu\right)^2\right]$$

- The log of the likelihood often is more convenient for exponential distributions such as the normal distribution

$$\ln\mathrm{L}\left(\mu,\sigma^2|y_t,\right)=-\frac{T}{2}2\pi-T\ln\sigma-\frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(y_t-\mu\right)^2$$

Maximum likelihood estimation of mean and variance

- The log of the likelihood function

$$\ln \mathrm{L}\left(\mu, \sigma^2 | y_t,\right) = -\frac{T}{2} \ln 2\pi - T \ln \sigma - \frac{1}{2\sigma^2} \sum_{t=1}^{T} (y_t - \mu)^2$$

- Maximize likelihood as a function of parameters conditional on the data
  - Can do all at once or sequentially
- Want to estimate μ
- Denote the estimator by a "hat" over it
- Maximize by solving

$$\frac{\partial \ln \mathrm{L}}{\partial \mu} = \frac{1}{\sigma^2} \sum_{t=1}^{T} (y_t - \widehat{\mu}) = 0$$

Maximum likelihood estimation of mean and variance

- Maximize by solving

$$\frac{\partial \ln \mathrm{L}}{\partial \mu} = \frac{1}{\sigma^2} \sum_{t=1}^{T} (y_t - \widehat{\mu}) = 0$$

$$\sum_{t=1}^{T} (y_t - \widehat{\mu}) = 0$$

$$\sum_{t=1}^{T} y_t = T\widehat{\mu}$$

$$\widehat{\mu} = \frac{\sum_{t=1}^{T} y_t}{T} = \overline{y}$$

Illustration

- Likelihood function

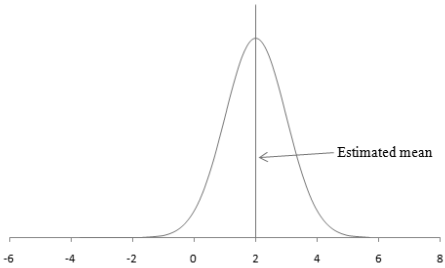


Figure: Likelihood function of normally distributed data with mean $\overline{y}$ of 2 and variance $s_{ml}^2 = \sum (y - \overline{y})^2 / T$ of 1. The maximum likelihood estimator is the mean of the normal distribution.

Finish by finding estimator of variance

- Estimator of σ^2
- Concentrate μ out of likelihood function by replacing it by $\overline{y}$

$$\begin{aligned} \ln \mathrm{L}\left(\sigma^2 | y_t\right) &= -\frac{T}{2}\ln 2\pi - T\ln\sigma - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(y_t - \widehat{\mu}\right)^2 \\ &= -\frac{T}{2}\ln 2\pi - \frac{T}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(y_t - \overline{y}\right)^2 \end{aligned}$$

- Maximize the concentrated likelihood function with respect to σ^2

Finish by finding estimator of variance

- Maximize likelihood function with respect to σ^2

$$\ln \mathrm{L}\left(\sigma^2 | y_t\right) = -\frac{T}{2}\ln 2\pi - \frac{T}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(y_t - \overline{y}\right)^2$$

- Solve

$$\frac{\partial \ln \mathrm{L}}{\partial \sigma^2} = -\frac{T}{2}\frac{1}{\widehat{\sigma}^2} + \frac{1}{2}\frac{1}{\widehat{\sigma}^4}\sum_{t=1}^{T}\left(y_t - \overline{y}\right)^2 = 0$$

$$-T + \frac{1}{\widehat{\sigma}^2}\sum_{t=1}^{T}\left(y_t - \overline{y}\right)^2 = 0$$

$$\widehat{\sigma}^2 = \frac{\sum_{t=1}^{T}\left(y_t - \overline{y}\right)^2}{T}$$

Consistency and unbiasedness

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

Properties of maximum likelihood estimators commonly mentioned

- Maximum likelihood provides a couple of natural estimators of the variance of the estimator
- Let θ be a parameter we have estimated

$$\mathrm{Var}\left[\widehat{\theta}\right] \geq \left(\mathrm{E}\left[\left(\partial \ln \mathrm{L}\left(\theta|y\right)/\partial\theta\right)\right]^2\right)^{-1}$$

  where the expectation with respect to the distribution of y's is evaluated at the true parameter
- Under regularity conditions

$$\mathrm{E}\left[\left(\partial \ln \mathrm{L}\left(\theta|y\right)/\partial\theta\right)\right]^2 = -\mathrm{E}\left[\frac{\partial^2 \ln \mathrm{L}\left(\theta|y\right)}{\partial\theta^2}\right]$$

- The term information matrix denotes

$$I = -\mathrm{E}\left[\frac{\partial^2 \ln \mathrm{L}\left(\theta|y\right)}{\partial\theta^2}\right]$$

- Therefore, under regularity conditions,

$$\mathrm{Var}\left[\widehat{\theta}\right] = I^{-1}$$

Typical properties of maximum likelihood estimators

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

$$\text{pr}\left(A, B\right) = \text{pr}\left(A|B\right)\text{pr}\left(B\right)$$

  - where $\text{pr}\left(A, B\right)$ is the joint probability of two events A and B
  - $\text{pr}\left(A|B\right)$ is the probability of the event A conditional on the event B
  - $\text{pr}\left(B\right)$ is the probability of B
- This equation defines conditional probability

$$\text{pr}\left(A|B\right) = \text{pr}\left(A, B\right) / \text{pr}\left(B\right) \text{ if } \text{pr}\left(B\right) \neq 0$$

- Also can say

$$\text{pr}\left(A, B\right) = \text{pr}\left(B|A\right) \; pr\left(A\right)$$

- Equate two definitions and get

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

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 doctor says you have the flu
    - 80 percent of time when you do $\text{pr}(A|B)$
    - 20 percent when you don't $\text{pr}(A|\text{not } B)$
  - If the doctor says you have the flu, then the probability of your having the flu 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

Diffuse prior

- 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

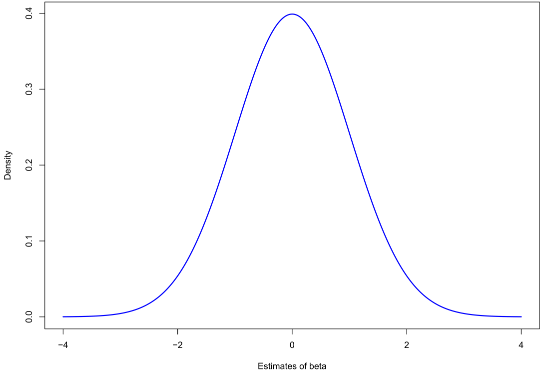
- $\mathrm{pr}\left(B|A\right) = \frac{\mathrm{pr}(A|B)\,\mathrm{pr}(B)}{\mathrm{pr}(A)}$
- Can write this in terms of discrete or continuous probability distribution functions
  - Let B be a parameter β and $\mathrm{pr}\left(B\right) \equiv p\left(\beta\right)$
    - ★ Might be CAPM parameter
  - Prior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

Bayesian analysis of parameter values

- $$p\left(\beta|y\right) = \frac{\mathrm{L}\left(\beta|y\right)p\left(\beta\right)}{p\left(y\right)}$$
  where $p\left(\beta|y\right)$ is the posterior probability distribution of values of β conditional on the data
- $p\left(y\right)$ is a normalizing constant independent of β so this can be analyzed using
  $$p\left(\beta|y\right) \propto \mathrm{L}\left(\beta|y\right)p\left(\beta\right)$$
  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

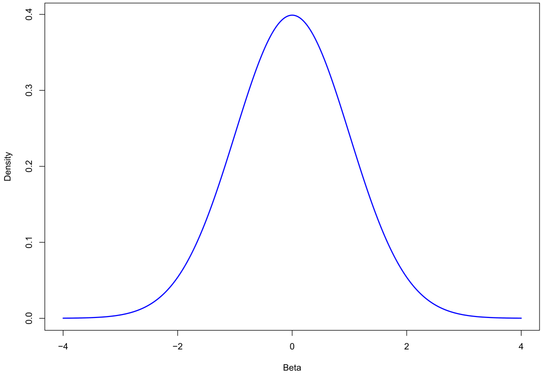
Comparison of classical and Bayesian analysis

- Classical: Probability distribution of estimator $\widehat{\beta}$
  - True value is a number, zero in this case if the estimator is unbiased



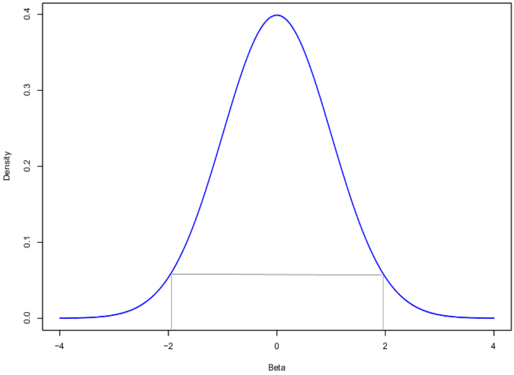
Comparison of classical and Bayesian analysis

- Bayesian: Posterior probability distribution of various possible values of β
  - True value is one of these possible values, with some more probable than others



Interpretration of Estimate of Variability

- Estimate of five percent confidence interval for a normal distribution



Summary

- Estimation issues
  - Unbiased
  - Estimate of variability
  - Consistency
  - Maximium likelihood estimator
- Bayesian statistics
  - Plausibility of posterior value after seeing data
  - Natural interpretation of variability