Bayesian Analysis

Advanced Economic Theory
Lecture 12

1

Classical Econometrics

Classical econometric methods assume that there is true parameter underlying the data generating process such as θ and its true value is unknown.

The objective of sample statistics $\hat{\theta}$ is to represent this unknown parameter as best as possible.

The estimated parameter $\hat{\theta}$ is a random variable and has its own distribution where as the true parameter θ is a fixed number but unknown.

The estimator should be unbiased, [$E(\hat{\theta}-\theta)=0$], and

Efficiency of an estimator is judged by the minimum square error [$E(\hat{\theta}-\theta)(\hat{\theta}-\theta)$]

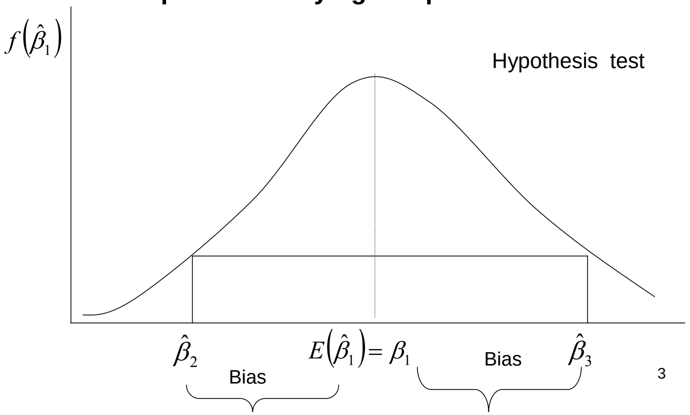
d the data generating process is given by

$$f(y;\theta)=\prod_{i=1}^{T}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[\frac{-(y-\mu)^2}{2\sigma^2}\right]$$

2

There is a TRUE parameter?

Purpose of studying sample is to find it.



3

Bayesian Econometrics

In Bayesian analysis the value of true parameter θ unknown like in the classical approach but it is not fixed.

Instead θ has a probability distribution and it is updated continuously based on sample information – priors.

The prior density is given by $f(\theta)$ and this may represent all available information up to that point.

$$f(\theta_1)=\int_{-\infty}^{\infty} f(\theta)\,d\theta_2 d\theta_3 \ldots d\theta_n$$

The sample density of variable y is treated as conditional on the random variable θ given by $f(y/\theta)$.

The joint density of y and θ is result of the product of prior density and the sample likelihood function.

$$f(y,\theta)=f(y/\theta)f(\theta)$$

4

Bayesian Law: Priors, Data and Posterior

$$P(A/B) = \frac{P(B \cap A)}{P(B)} \qquad (1)$$

Reverse position of A and B and get $P(B \cap A) = P(B/A)P(A)$ and replace in (1)

$$P(A/B) = \frac{P(B/A)P(A)}{P(B)} \qquad (2)$$

$P(A)$ measures the strength of subjective belief in event A

and $P(A/B)$ is the manipulation of that believe after data (event B) has been breveted

.

(Ref: Zellner (1985))

Bayesian Method: A Numerical Example from Lancaster (2004);

In the context of economics suppose there are two contending theories:

theory 1: $C = 10 + 0.9Y$; $\theta_1 = (10, 0.9)$ and prior $P(\theta_1) = 0.5$

. theory 2: $C = Y$; $\theta_2 = 1$ prior $P(\theta_2) = 0.5$

Now you can form a belief on data conditional on $P(E/\theta_1)$ and $P(E/\theta_2)$ probabilities of observing certain data given beliefs on θ_1 and θ_2.

If probability of E is less likely with θ_1 than θ_2, then $P(E/\theta_1) = 0.1$ and $P(E/\theta_2) = 0.6$ then the probability of event is given by

$$P(E) = P(E/\theta_1)P(\theta_1) + P(E/\theta_2)P(\theta_2) = (0.1 \times 0.5) + (0.6 \times 0.5) = 0.35$$

Now from the Bayes' theorem it is possible to update priors on θ_1 and θ_2

$$P(\theta_1/E) = \frac{P(E/\theta_1)P(\theta_1)}{P(E)} = \frac{0.1 \times 0.5}{0.35} = \frac{1}{7}$$

$$P(\theta_2/E) = \frac{P(E/\theta_2)P(\theta_2)}{P(E)} = \frac{0.6 \times 0.5}{0.35} = \frac{6}{7}$$

Components of a Bayesian Model

A Bayesian econometric model $f(\theta/y) = \frac{f(y/\theta)f(\theta)}{f(y)}$ has two components;

$f(y/\theta)$ is called the **likelihood function,**
it describes what you see for every particular values of parameter set $\theta \in \Theta$ as the $\theta_1 = (10, 0.9)$ in the consumption function above.

The second component $f(\theta)$ is the distribution over the parameter space Θ; **prior distribution** denoting the beliefs about particular parameter values $\theta \in \Theta$.

Parameter $\theta \in \Theta$ is unknown both before and after the data has been observed

but data are unknown before the information has been gathered but known afterwards.

$f(\theta/y)$ is the posterior formed after taking account of prior and data.

Bayesian algorithm (Lancaster (2004))

1. Formulate an economic model with conditional probability distribution over parameter space $\theta \in \Theta$; such as $f(\theta/y) = \frac{f(y/\theta)f(\theta)}{f(y)}$
2. Organise beliefs about θ into a prior probability distribution over Θ.
3. Collect data and insert into the model as given in step 1.
4. Criticise the model.

Numerical Example for Bayesian Inference (prior from a pilot survey)

Suppose one is interested in obtaining mean expenditure on food. Its variance is believed to be 57.6 but the mean is unknown.

Before conducting a large sample survey one small pilot survey of six people is implemented with data

[30, 23.69, 29.04, 11.48, 30.83, 27.81]

with $\bar{y}_0 = \frac{1}{6}\sum y_i = 25.475$. Prior distribution of the mean food expenditure $\beta \sim N\left(\bar{y}_0, \frac{\sigma^2}{T_0}\right) = N(25.475, 9.6)$. With

$$P\left[\hat{\beta}_1 - 1.96\left(\mathrm{var}\left(\hat{\beta}_1\right)\right)^{\frac{1}{2}} < \beta_1 < \hat{\beta}_1 + 1.96\left(\mathrm{var}\left(\hat{\beta}_1\right)\right)^{\frac{1}{2}}\right] = 0.95$$

$$P\left[25.475 - 1.96(9.6)^{\frac{1}{2}} < \beta_1 < 25.475 + 1.96(9.6)^{\frac{1}{2}}\right] = 0.95$$

$$P[19.40 < \beta_1 < 31.55] = 0.95$$

Numerical Example for Bayesian Inference (data from a larger survey)

Sample of size 40 is conducted, the distribution of sample with new information is

$$\beta \sim N\left(\bar{y}_1, \frac{\sigma^2}{T_1}\right) = N(23.5945, 1.44)$$

Large Sample of 40

9.46	17.83	22.87	17.87
10.56	32.81	26.52	25.54
14.81	22.13	21	39
21.71	23.46	37.52	20.44
22.79	16.81	21.69	30.1
18.19	21.35	27.4	20.9
22	14.87	30.69	48.71
18.12	33	19.56	
23.13	25.19	30.58	
19	17.77	41.12	
19.46	22.44	15.38	

$$P\left[\hat{\beta}_1 - 1.96\left(\mathrm{var}\left(\hat{\beta}_1\right)\right)^{\frac{1}{2}} < \beta_1 < \hat{\beta}_1 + 1.96\left(\mathrm{var}\left(\hat{\beta}_1\right)\right)^{\frac{1}{2}}\right] = 0.95$$

$$P\left[23.5945 - 1.96(1.44)^{\frac{1}{2}} < \beta_1 < 23.5945 + 1.96(1.44)^{\frac{1}{2}}\right] = 0.95$$

$$P[21.24 < \beta_1 < 25.95] = 0.95$$

Posterior from Prior and Data

The Bayesian rule $f(\beta/y) = f(\beta/y_0, y_1) = kf(y/\beta)f(\beta)$ links these pre-sample prior to the post sample distribution and new distribution is given by $\beta \sim N\left(\bar{\bar{\beta}}, \frac{\sigma^2}{T_0 + T_1}\right)$

$$\bar{\bar{\beta}} = \frac{h_0\bar{y}_0 + h_1\bar{y}_1}{h_0 + h_1} = \frac{T_0\bar{y}_0 + T_1\bar{y}_1}{T_0 + T_1} = \frac{6 \times 25.475 + 40 \times 23.5945}{6 + 40} = 23.84$$

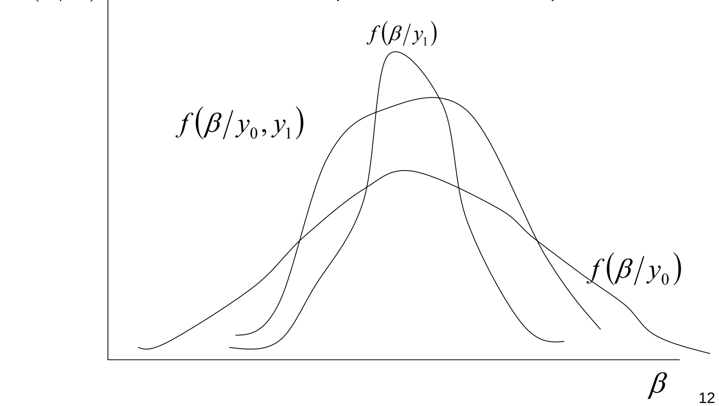
$h_0 = \left(\frac{\sigma^2}{T_0}\right)^{-1} = \left(\frac{57.6}{9}\right)^{-1} = 0.104$ and $h_1 = \left(\frac{\sigma^2}{T_1}\right)^{-1} = \left(\frac{57.6}{40}\right)^{-1} = 0.694$ are inverse of the variance from the small and large samples and indicate the precision of the estimate.

$$\bar{\bar{\sigma}}_\beta^2 = \frac{1}{\bar{\bar{h}}} = \frac{1}{h_0 + h_1} = \frac{1}{\frac{T_0}{\sigma^2} + \frac{T_1}{\sigma^2}} = \frac{\sigma^2}{T_0 + T_1} = \frac{57.6}{40 + 6} = 1.25$$

Posterior distribution combines information from both pilot and large surveys. It is closer to the mean and variance of the large sample survey as it is more precise because it has more observations.

Prior, Likelihood and Posterior Mean Expenditure In Bayesian Analysis

Bayesian procedure involves continuous updating about the underlying distribution of a parameter based on sample information.



Classical Estimation

Take a simple regression Model $Y_i = \beta_1 + \beta_2 X_{1,i} + e_i$, with following data.

Observations	Y	X_1	X_2	YX_1	YX_2	X_1X_2	X_1^2	X_2^2	Y^2
Total	**28**	**35**	**28**	**97**	**126**	**117**	**245**	**126**	**140**

The easiest way to estimate regression with one explanatory variable is by using the equation in deviation form. $y_i = \beta_2 x_{1,i} + e_i$. This implies $\hat{\beta}_2 = \dfrac{\sum_i x_{1i} y_i}{\sum_i x_{1i}^2}$

where using the deviation formula and the information given in the table

$\sum_i x_{1i} y_i = \sum_i \left(X_{1i} - \overline{X}_{1i}\right)\left(Y_i - \overline{Y}_i\right) = \sum_i X_{1i} Y_i - N\overline{X}_{1i}\overline{Y}_i$ = 97-7(4) (5)= 97-140 = -43

and $\sum_i x_{1i}^2 = \sum_i \left(X_{1,i} - \overline{X}_1\right)^2 = \sum_i X_{1,i}^2 - N\overline{X}_1^2$ = 245 – 7(5)^2 = 245-175 = 70;

$$\hat{\beta}_2 = \frac{-43}{70} = -0.6143$$

$$\beta_1 = \overline{Y} - \hat{\beta}_2 \overline{X}_1 = 4 - (-0.6143)5 = 4 + 3.0715 = 7.07$$

Classical Hypothesis Test

$y_i = \beta_2 x_{1,i} + e_i \quad y_i = \hat{y}_i + e_i \Rightarrow \sum_i y_i^2 = \sum_i \hat{y}_i^2 + \sum_i e_i^2$

$\sum_i y_i^2 = \hat{\beta}_2^2 \sum_i x_{1,i}^2 + \sum_i \hat{e}_i^2 \Rightarrow 28 = (\text{-}0.6143)2 * 70 + \sum_i \hat{e}_i^2$

$\sum_i \hat{y}_i^2 = (\text{-}0.6143)^2 * 70 = 26.41$ and $\sum_i \hat{e}_i^2 = 28 - 26.41 = 1.584$

$$R^2 = \frac{\sum_i \hat{y}_i^2}{\sum_i y_i^2} = \frac{26.41}{28} = 94.32\%;$$

$$\overline{R}^2 = 1 - \frac{\sum_i \hat{e}_i^2 \big/ (N-K)}{\sum_i y_i^2 \big/ (N-1)} = 1 - \frac{1.584/(7-2)}{28/(7-1)} = 1 - \frac{0.3168}{4.667} = 93.21\%$$

$$\text{var}(e_i) = \hat{\sigma}^2 = \frac{\sum_i \hat{e}_i^2}{N-K} = \frac{1.584}{5} = 0.3168;$$

The variance of the slope parameter is given by:

$$\text{var}\left(\hat{\beta}_2\right) = \frac{\hat{\sigma}^2}{\sum_i x_{1,i}^2} = \frac{0.3168}{70} = 0.00453$$

Calculate t value for the given estimates to test $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 \neq 0$.

$t_{calc} = \dfrac{\beta_2 - \hat{\beta}_2}{\sqrt{\text{var}\left(\hat{\beta}_2\right)}} = \dfrac{-0.6143}{\sqrt{0.0043}} = -9.367$; The theoretical t value for $t_{5,0.05} = 2.571$

Distribution of Bayesian Parameter in a Single Variable Regression

$$z_1 = \frac{\beta_2 - \hat{\beta}_2}{\sqrt{\text{var}\left(\hat{\beta}_2\right)}} \text{ and } \beta_2 = \hat{\beta}_2 - z_1\sqrt{\text{var}\left(\hat{\beta}_2\right)}$$

β_2 is now a random variable

$$E[\beta_2] = \hat{\beta}_2 - E(z_1)\sqrt{\text{var}\left(\hat{\beta}_2\right)} = \hat{\beta}_2$$

$$\text{var}[\beta_2] = \text{var}(z_1)\,\text{var}\left(\sqrt{\text{var}\left(\hat{\beta}_2\right)}\right)^2 \qquad \text{var}\left(\hat{\beta}_2\right) = \frac{\sigma^2}{\sum_i x_{1,i}^2} = \sigma_{\beta_2}^2 = \left(\sqrt{0.0043}\right)^2 = 0.0043$$

$$E[\beta_2] = -0.6143 - 9.367\sqrt{0.004} = 1.207$$

$$\beta_2 \sim N\left(\hat{\beta}_2, \sigma_{\beta_2}^2\right)$$

Similarly for $\beta_1 \sim N\left(\hat{\beta}_1, \sigma_{\beta_1}^2\right)$; $\sigma_{\beta_1}^2 = \dfrac{\sigma^2 \sum_i X_{1,i}^2}{T\sum_i x_{1,i}^2}$

Bayesian estimation of Consumption Function

Errors are normally distributed with mean zero and known variance $\sigma^2 = 6000$.

The least square estimates of consumption on income gives:

$$C_t = \beta_1 + \beta_2 Y_t + e_t$$

$$\hat{c}_t = -1475.2 + 1.0591 y_t$$

$$\text{var}\left(\hat{\beta}_2\right) = \frac{\sigma^2}{\sum_i x_{1,i}^2} = \sigma_{\beta_2}^2 = 0.0033439$$

$$\text{var}\left(\hat{\beta}_1\right) = \frac{\sigma^2 \sum X_t^t}{T\sum_i x_{1,i}^2} = \sigma_{\beta_1}^2 = 334022.4$$

Consumption Function: Data and Shazam Programme

```
file 33 table25.5
read (33) y c
sample 18 27
ols c y/pcov cov=covb coef=b
matrix covb=covb*(6000/$sig2)
print covb
dim seb 2
matrix seb=sqrt(diag(covb))
sample  1 2
genr lower=b-1.96*seb
genr upper=b+1.96*seb
print lower upper
sample 1 1
gen1 mean=b:2
gen1 var=(seb:2)**2
read b1
0
distrib b1/mean=mean var=var
sample 1 2
read b2
1
0.9
gen1 mean=b:1
gen1 var=(seb:1)**2
distrib b2/mean=mean var=var
stop
```

	Y	C
1960	6036	5561
1961	6113	5579
1962	6271	5729
1963	6378	5855
1964	6727	6099
1965	7027	6362
1966	7280	6607
1967	7513	6730
1968	7728	7003
1969	7891	7185
1970	8134	7275
1971	8322	7409
1972	8562	7726
1973	9042	7972
1974	8867	7826
1975	8944	7926
1976	9175	8272
1977	9381	8551
1978	9735	8808
1979	9829	8904
1980	9722	8783
1981	9769	8794
1982	9725	8818
1983	9930	9139
1984	10421	9475
1985	10563	9713
1986	10780	10014

Bayesian Inference on Parameters of Consumption Function

Using normal distribution of 95 percent confidence interval

$$P\left[\hat{\beta}_1 - 1.96\left(\mathrm{var}\left(\hat{\beta}_1\right)\right)^{\frac{1}{2}} < \beta_1 < \hat{\beta}_1 + 1.96\left(\mathrm{var}\left(\hat{\beta}_1\right)\right)^{\frac{1}{2}}\right] = 0.95$$

$$P\left[-1475.2 - 1.96\sqrt{334022.4} < \beta_1 < -1475.2 + 1.96\sqrt{334022.4}\right] = 0.95$$

$$P\left[-2607.97 < \beta_1 < -342.47\right] = 0.95$$

Similarly

$$P\left[\hat{\beta}_2 - 1.96\left(\mathrm{var}\left(\hat{\beta}_2\right)\right)^{\frac{1}{2}} < \beta_2 < \hat{\beta}_2 + 1.96\left(\mathrm{var}\left(\hat{\beta}_2\right)\right)^{\frac{1}{2}}\right] = 0.95$$

$$P\left[1.0591 - 1.96\sqrt{0.0033439} < \beta_2 < 1.0591 + 1.96\sqrt{0.0033439}\right] = 0.95$$

$$P\left[0.946 < \beta_1 < 1.172\right] = 0.95$$

Both of above parameters are counter intuitive and infeasible considering the economic theory as one would expect $\beta_1 > 0$ and $\beta_2 < 1$.

Post Sample Distribution and Odds Ratio

Post sample density functions for β_1 and β_2 implied by above sample information is

$$\beta_1 \sim N(-1475.2, 334022.4)$$

$$\beta_2 \sim N(1.0591, 0.0033439)$$

$$P(\beta_1 < 0) = P\left[Z < \frac{0 + 1475.2}{\sqrt{334022.4}}\right] = P[Z < 2.5525] = 0.995$$

$$P(\beta_2 > 1) = P\left[Z > \frac{1 - 1.0591}{\sqrt{0.0033439}}\right] = P[Z > -1.032] = 0.846$$

The odd ratio from the post sample density function in favour of hypothesis $H_0 : \beta_2 \geq 0.9$ compared to alternative $H_1 : \beta_2 < 0.9$

$$odd\_ratio = \frac{P[\beta_2 \geq 0.9]}{P[\beta_2 < 0.9]} = \frac{P\left[Z \geq \frac{0.9 - 1.0591}{\sqrt{0.0033439}}\right]}{P\left[Z < \frac{0.9 - 1.0591}{\sqrt{0.0033439}}\right]} = \frac{P(Z \geq -2.7513)}{P(Z < -2.7513)} = \frac{0.99702}{0.00298} = 335.927$$

Thus from the sample information, $\beta_2 \geq 0.9$ is likely to happen 336 times more than not happening like this.

Post Sample Distribution of Random Parameters

Post sample bivariate random variables β_1 and β_2 can be generated by a

$$\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \sim N\left[\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}\right]$$

From the sample estimates these values are $\hat{\beta}_1$= -1475.2; $\hat{\beta}_1$ = 1.0591;
$\sigma_{22} = \mathrm{var}(\beta_2) = 0.0033439$; $\sigma_{11} = \mathrm{var}(\beta_1) = 334022.4$;
$\sigma_{12} = \mathrm{cov}(\beta_1, \beta_2) = -33.39065$

A bivariate distribution of $\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ can be generated from standard normally distribute variables Z_1 and Z_2 as following.

$$\beta_1 = \hat{\beta}_1 + \sqrt{\sigma_{11}} Z_1 \quad \text{and} \quad \beta_2 = \hat{\beta}_2 + \frac{\sigma_{11}}{\sqrt{\sigma_{11}}} Z_1 + \sqrt{\sigma_{11} - \frac{\sigma_{11}}{\sqrt{\sigma_{11}}}} Z_2$$

Shazam Program for Bayesian Estimation

Bayesian Estimation 1	Bayesian Estimation 2
`file 33 table25.5`	`file 33 table25.5`
`read (33) y c`	`read (33) y c`
`sample 18 27`	`sample 1 10`
`ols c y/pcov cov=covb coef=b`	`ols c y/pcov cov=covb coef=b`
`matrix covb=covb*(6000/$sig2)`	`matrix covb=covb*(6000/$sig2)`
`print covb`	`print covb`
`dim seb 2`	`dim seb 2`
`matrix seb=sqrt(diag(covb))`	`matrix seb=sqrt(diag(covb))`
`sample  1 2`	`sample  1 2`
`genr lower=b-1.96*seb`	`genr lower=b-1.96*seb`
`genr upper=b+1.96*seb`	`genr upper=b+1.96*seb`
`print lower upper`	`print lower upper`
`sample 1 1`	`sample 1 1`
`gen1 mean=b:2`	`gen1 mean=b:2`
`gen1 var=(seb:2)**2`	`gen1 var=(seb:2)**2`
`read b1`	`read b1`
`0`	`0`
`distrib b1/mean=mean var=var`	`distrib b1/mean=mean var=var`
`sample 1 2`	`sample 1 2`
`read b2`	`read b2`
`1`	`1`
`0.9`	`0.9`
`gen1 mean=b:1`	`gen1 mean=b:1`
`gen1 var=(seb:1)**2`	`gen1 var=(seb:1)**2`
`distrib b2/mean=mean var=var`	`distrib b2/mean=mean var=var`
`stop`	`stop`

Classical Structure of a VAR Model

$$y_{1,t} = a_{10} + \sum_{j=1}^{P} a_{11j} y_{1,t-j} + .. + \sum_{j=1}^{P} a_{1nj} y_{n,t-j} + \sum_{j=1}^{r} b_{11j} x_{1,t-j} + .. + \sum_{j=1}^{P} b_{1mj} x_{m,t-j} + e_{1t}$$

$$\vdots$$

$$y_{n,t} = a_{n0} + \sum_{j=1}^{P} a_{n1j} y_{1,t-j} + .. + \sum_{j=1}^{P} a_{nnj} y_{n,t-j} + \sum_{j=1}^{r} b_{11j} x_{1,t-j} + .. + \sum_{j=1}^{P} b_{1mj} x_{m,t-j} + e_{nt}$$

Simple Example

$$y_t = a_{10} + a_{11} y_{1t-1} + a_{12} y_{2,t-2} + b_{11} x_{t-1} + b_{12} x_{t-2} + e_{1t}$$

$$x_t = a_{20} + a_{21} y_{1t-1} + a_{22} y_{2,t-2} + b_{21} x_{1,t-1} + b_{22} x_{t-2} + e_{2t}$$

Impulse Response Analysis in Classical VAR Model

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-2} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{t-2} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & b_{11} & b_{12} \\ a_{21} & a_{22} & b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-2} \\ x_{1t-1} \\ x_{t-2} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$

$$\overline{Y} = (I - A)^{-1} BX + (I - A)^{-1} U$$

$$\overline{X} = (I - B)^{-1} AY + (I - B)^{-1} U$$

$$Y_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad Y_1 = (I - A)^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad Y_2 = (I - A)^{-1} Y_1 = (I - A)^{-1} (I - A)^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

VAR estimation using the Classical Model

SYS(9) Estimating the system by OLS (using YMPEEXM_UK.xls)
The estimation sample is: 1963 (3) to 2003 (2)

Results from unrestricted VAR

Y = + 1.044*Y_1 - 0.008679*M4_1 + 31.1*RPI_1 - 19.99*INTREST_1
(SE) (0.00739) (0.00202) (13.5) (25.7)
- 7.521
(203)

M4 = + 0.1555*Y_1 + 0.975*M4_1 - 26.58*RPI_1 + 117*INTREST_1
(SE) (0.0501) (0.0137) (91.5) (174)
- 1526
(1.37e+003)

RPI = - 3.419e-005*Y_1 + 8.877e-006*M4_1 + 0.933*RPI_1 + 0.1266*INTREST_1
(SE) (1.31e-005) (3.59e-006) (0.0239) (0.0456)
+ 0.1257
(0.359)

INTREST = + 3.34e-005*Y_1 - 1.001e-005*M4_1 + 0.01489*RPI_1 + 0.793*INTREST_1
(SE) (1.42e-005) (3.9e-006) (0.026) (0.0495)
+ 1.224
(0.39)

Estimating VAR with Restrictions

Results from a restricted VAR (Y -0INT, M - 0Y, RPI -0INT, INT-0Y)

Y = + 1.039*Y_1 - 0.007245*M4_1 + 28.31*RPI_1 + 0*INTREST_1
(SE) (0.00602) (0.00162) (12.7) (0)
- 73.9
(164)

M4 = + 0*Y_1 + 1.018*M4_1 - 109.8*RPI_1 + 715.3*INTREST_1 - 3500
(SE) (0) (0.00163) (94.5) (145) (1.4e+003)

RPI = - 1.398e-005*Y_1 + 3.125e-006*M4_1 + 0.9549*RPI_1 + 0*INTREST_1
(SE) (1.11e-005) (2.99e-006) (0.0231) (0)
+ 0.7069
(0.298)

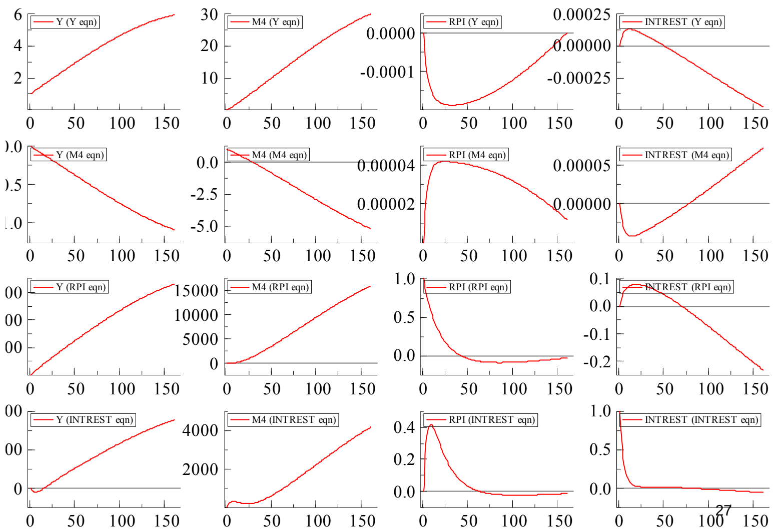
INTREST = + 0.0001661*Y_1 - 4.766e-005*M4_1 + 0.1502*RPI_1 +
0*INTREST_1
(SE) (1.88e-005) (5.07e-006) (0.0398) (0)
+ 4.788
(0.512)

PcGive Commands for Structural VAR and for Restricted VAR

PcGive Commands for unrestricted VAR	PcGive Commands for restricted VAR

module("PcGive");
package("PcGive");
usedata("YMPEEXM_UK.xls");
system
{
Y = Y, M4, RPI, INTREST;
Z = Y_1, M4_1, RPI_1, INTREST_1;
U = Constant;
}
estimate("OLS", 1963, 3, 2003, 2);
	module("PcGive");
	package("PcGive");
	usedata("YMPEEXM_UK.xls");
	system
	{
	Y = Y, M4, RPI, INTREST;
	Z = Y_1, M4_1, RPI_1, INTREST_1;
	U = Constant;
	}
	model
	{
	Y = Y_1, M4_1, RPI_1, INTREST_1;
	M4 = Y_1, M4_1, RPI_1, INTREST_1;
	RPI = Y_1, M4_1, RPI_1, INTREST_1;
	INTREST = Y_1, M4_1, RPI_1, INTREST_1;
	}
	
	constraints
	{
	&3 = 0;
	&4 =0;
	&11 =0;
	&15 =0;
	}
	estimate("CFIML", 1963, 3, 2003, 2);

Impulse Response Analysis in Classical VAR



OLS procedure to minimise the sum of squared error terms.

$$S = e'e = (Y - X\beta)'(Y - X\beta)$$
$$= Y'Y - 2\beta' X'Y + \beta' X' X\beta$$

Imposing a restriction involves constrained optimisation with a Lagrange multiplier.

$$L = e'e + 2\lambda(r' - \beta' R')$$

$$L = (Y - X\beta)'(Y - X\beta) + 2\lambda(r' - \beta' R')$$

$$L = Y'Y - 2\beta' X'Y + \beta' X' XX\beta + 2\lambda(r' - \beta' R')$$

Partial derivation of this constrained minimisation function (Lagrangian function) wrt β and λ yields

(i) $\frac{\partial L}{\partial \beta} = -2X'Y + 2X'Xb - 2\lambda R' = 0$

(ii) $\frac{\partial L}{\partial \lambda} = 2(r - Rb) = 0$

$$X'Xb = X'Y + \lambda R'$$
$$(X'X)^{-1}X'Xb = (X'X)^{-1}(X'Y + \lambda R')$$
$$b = (X'X)^{-1}X'Y + (X'X)^{-1}R'\lambda$$
$$b = \hat{\beta} + (X'X)^{-1}R'\lambda$$

This is the restricted least square estimator but need still to be solved for λ. For that multiply the above equation both sides by R

$$Rb = R\hat{\beta} + R(X'X)^{-1}R'\lambda = r$$
$$\lambda = \left[R(X'X)^{-1}R'\right]^{-1}\left[Rb - R\hat{\beta}\right]$$
$$\lambda = \left[R(X'X)^{-1}R'\right]^{-1}\left[r - Rb\right]$$
$$b = \hat{\beta} + (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}\left[r - Rb\right]$$

Thus the restricted least square estimator is a linear function of the restriction $Rb - r = 0$.

$$E(b) = E(\hat{\beta}) + (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}\left[r - RE(b)\right]$$
$$E(b) = \beta$$

For variance we need to use property of an idempotent matrix AA=A.

Such as $A = \begin{bmatrix} 0.4 & 0.8 \\ 0.3 & 0.6 \end{bmatrix}$

Recall in unrestricted case

$$\hat{\beta} = (X'X)^{-1}X'Y = \beta + (X'X)^{-1}X'e$$
$$E(b) - \beta = (X'X)^{-1}X'e + (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}\left[r - RE(b) - R(X'X)^{-1}X'e\right]$$

Since $Rb - r = 0$

$$E(b) - \beta = M(X'X)^{-1}X'e$$

Where M is the idempotent matrix:

$$M = I - (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R$$

The variance covariance matrix of

$$\mathrm{cov}(b) = [E(b) - \beta][E(b) - \beta]' = E\left[M(X'X)^{-1}X'ee'X(X'X)^{-1}M'\right]$$
$$\mathrm{cov}(b) = \sigma^2 M(X'X)^{-1}M$$
$$\mathrm{cov}(b) = \sigma^2 M(X'X)^{-1}$$
$$M = \sigma^2(X'X)^{-1}\left[I - (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}\right]R$$

Thus the variance of the restricted least square estimator is smaller than the variance of the unrestricted least square estimator.

$$M = \sigma^2(X'X)^{-1}\left[I - (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}\right]R$$

Model $y_t = x_{t1}\beta_1 + x_{t2}\beta_2 + x_{t3}\beta_3 + e_t$

One general restriction may be

Null Hypothesis: $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$

A simple restriction $Rb - r = 0$

$$Rb = r \Rightarrow \begin{array}{ccc} R & b & r \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} & \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = & \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \end{array}$$

where b are the estimates obtained from the sample estimation.

F-test for Restrictions in Multiple Regression

$$y_t = x_{t,1}\beta_1 + x_{t,2}\beta_2 + x_{t,3}\beta_3 + e_t$$

Estimate β_1 β_2 and β_3, and test the following Test for Restrictions

a. $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 \neq 0$

b. $H_0 : \beta_1 = \beta_2 = 0$ against $H_1 : \beta_1$ or β_2 or both are nonzero

c. $H_0 : \beta_2 = \beta_3 = 0$ against $H_1 : \beta_2$ or β_3 or both are nonzero

d. $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ against $H_1 : \beta_1$ or β_2 or β_3 or all are nonzero

$$F = \frac{(Rb - r)'\left[R \operatorname{cov}(b) R'\right]^{-1}(Rb - r)}{J}$$

Estimates of the Model

y_t	x_{t1}	x_{t2}	x_{t3}
1	1	0	-1
-1	-1	1	0
2	1	0	0
0	0	1	0
4	1	2	0
2	0	3	0
2	0	0	1
0	1	-1	1
2	0	0	1

$$\begin{bmatrix} \hat\beta_1 \\ \hat\beta_2 \\ \hat\beta_3 \end{bmatrix} = \begin{bmatrix} \sum_i x_{1,i}^2 & \sum_i x_{1,i}x_{2,i} & \sum_i x_{1,i}x_{3,i} \\ \sum_i x_{1,i}x_{2,i} & \sum_i x_{2,i}^2 & \sum_i x_{2,i}x_{3,i} \\ \sum_i x_{1,i}x_{3,i} & \sum_i x_{2,i}x_{3,i} & \sum_i x_{3,i}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_i y_i x_{1,i} \\ \sum_i y_i x_{2,i} \\ \sum_i y_i x_{3,i} \end{bmatrix}$$

$$\begin{bmatrix} \hat\beta_1 \\ \hat\beta_2 \\ \hat\beta_3 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 16 & -1 \\ 0 & -1 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 8 \\ 13 \\ 3 \end{bmatrix} \text{ or } \begin{bmatrix} \hat\beta_1 \\ \hat\beta_2 \\ \hat\beta_3 \end{bmatrix} = \begin{bmatrix} 1.6 \\ 0.873 \\ 0.968 \end{bmatrix}$$

Variance of Error and the Parameters

$$e_t = y_t - x_{t1}\beta_1 - x_{t2}\beta_2 - x_{t3}\beta_3$$
$$= y - 1.6x1 - 0.873x2 - 0.968x3$$

$$\sum_i \hat e_t^2 = 6.946 \; ; \; \hat\sigma^2 = \frac{6.946}{9-3} = 1.158$$

Computed covariance matrix

$$\operatorname{cov}(b_1, b_2) = \hat\sigma^2 (X'X)^{-1} = 1.158 \begin{bmatrix} 0.2 & 0 & 0 \\ 0 & 0.063 & 0.0186 \\ 0 & 0.016 & 0.254 \end{bmatrix} = \begin{bmatrix} 0.2315 & 0 & 0 \\ 0 & 0.0735 & 0.0184 \\ 0 & 0.0184 & 0.294 \end{bmatrix}$$

$$\begin{bmatrix} SE(\hat\beta_1) \\ SE(\hat\beta_2) \\ SE(\hat\beta_3) \end{bmatrix} = \begin{bmatrix} \sqrt{0.2315} \\ \sqrt{0.0735} \\ \sqrt{0.294} \end{bmatrix} = \begin{bmatrix} 0.48118 \\ 0.2711 \\ 0.5422 \end{bmatrix}$$

Test of Restrictions -1

ase of (d) $Rb = r \Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

$$F = \frac{\left[\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \right]' \begin{bmatrix} 0.2315 & 0 & 0 \\ 0 & 0.0735 & 0.0184 \\ 0 & 0.0184 & 0.294 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}}{J}$$

$$F = \frac{\begin{bmatrix} \beta_1 & \beta_2 & \beta_3 \end{bmatrix} \begin{bmatrix} 0.2315 & 0 & 0 \\ 0 & 0.0735 & 0.0184 \\ 0 & 0.0184 & 0.294 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}}{J}$$

Test of Restrictions -2

$$F = \frac{\begin{bmatrix} 1.6 & 0.873 & 0.968 \end{bmatrix} \begin{bmatrix} 0.2315 & 0 & 0 \\ 0 & 0.0735 & 0.0184 \\ 0 & 0.0184 & 0.294 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1.6 \\ 0.873 \\ 0.968 \end{bmatrix}}{3}$$

= 7.79 .F critical value for d.f.= (3,6) at 5% confidence interval is 4.76.

F calculated is bigger than F critical => Reject null hypothesis, which says $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$.

At least one of these parameters is significant and explains variation in y.

Shazam Program to Test of Restrictions- from HGJ 10.2

Single and Multiple Restrictions

```
OLS y x1 x2 x3/noconstant pcov
test
    test x1 =0
end
OLS y x1 x2 x3/noconstant pcov
test
    test x1 =0
    test x2=0
end
OLS y x1 x2 x3/noconstant pcov
test
    test x1 =0
    test x2=0
end
read pcov / rows=3 cols=3
0.23153 0.0000    0.000
0.0000  0.073503  0.0184
0.0000  0.018376  0.29401
```

Calculation of F-statistics

```
read bigr/ rows=3 cols=3
1  0  0
0  1  0
0  0  1
read sr/ rows=3 cols=1
0
0
0
matrix invcov=inv(pcov)
print invcov
matrix F=(beta'*invcov*beta)/3
matrix inv2 =inv((bigr*pcov*bigr'))
matrix FF = ((bigr*beta-sr)'*inv2*(bigr*beta-sr))/3
print F ff
```


Shazam Program to Test of Restrictions- from HGJ 10.2

```
Estimates:

VARIABLE    ESTIMATED   STANDARD   T-RATIO          PARTIAL STANDARDIZED ELASTICITY
   NAME     COEFFICIENT    ERROR       6 DF    P-VALUE CORR.  COEFFICIENT  AT MEANS
 X1          1.6000     0.48118      3.3252     0.0159 0.8051  0.75425      0.40000
 X2         0.87302     0.27111      3.2201     0.0181 0.7959  0.71281      0.43651
 X3         0.96825     0.54223      1.7857     0.1244 0.5891  0.43034      0.16138

Computed variance-covariance matrix

VARIANCE-COVARIANCE MATRIX OF COEFFICIENTS
 X1         0.23153
 X2          0.0000      0.73503E-01
 X3          0.0000      0.18376E-01  0.29401
               X1            X2            X3
*F-statistics for the first restriction
|_test
 |_ test x1 =0
 |_end
F STATISTIC =   11.056673     WITH    1 AND    6 D.F.  P-VALUE= 0.01590
WALD CHI-SQUARE STATISTIC =   11.056673     WITH    1 D.F.  P-VALUE= 0.00088
=> restriction is not valid because the computed F is greater than the critical F(1,6) = 13.75

*F-statistics for the second restriction
F STATISTIC =   10.712866     WITH    2 AND    6 D.F.  P-VALUE= 0.01047
 WALD CHI-SQUARE STATISTIC =   21.425731     WITH    2 D.F.  P-VALUE= 0.00002
=> restriction is not valid because the computed F is greater than the critical F(2,6) = 10.92
*F-statistics for the third restriction
F STATISTIC =   7.7897623     WITH    3 AND    6 D.F.  P-VALUE= 0.01716
 WALD CHI-SQUARE STATISTIC =   23.369287     WITH    3 D.F.  P-VALUE= 0.00003
=> restriction is not valid because the computed F is greater than the critical F(1,6) = 9.15
```

```
read pcov / rows=3 cols=3
0.23153 0.0000    0.000
0.0000  0.073503  0.0184
0.0000  0.018376  0.29401

read bigr/ rows=3 cols=3
1  0  0
0  1  0
0  0  1

read sr/ rows=3 cols=1
0
0
0

matrix invcov=inv(pcov)
print invcov
matrix F=(beta'*invcov*beta)/3
matrix inv2 =inv((bigr*pcov*bigr'))
matrix FF = ((bigr*beta-sr)'*inv2*(bigr*beta-sr))/3
print F ff

|_read pcov / rows=3 cols=3
       3 ROWS AND           3 COLUMNS, BEGINNING AT ROW       1
 |_read bigr/ rows=3 cols=3
       3 ROWS AND           3 COLUMNS, BEGINNING AT ROW       1
 |_read sr/ rows=3 cols=1
    1 VARIABLES AND        3 OBSERVATIONS STARTING AT OBS       1


 |_matrix invcov=inv(pcov)
 |_print invcov
     INVCOV
     3 BY     3 MATRIX
    4.319095       0.000000       0.000000
    0.000000       13.82113     -0.8649666
    0.000000     -0.8638383       3.455306
 |_matrix F=(beta'*invcov*beta)/3
 |_matrix inv2 =inv((bigr*pcov*bigr'))
 |_matrix FF = ((bigr*beta-sr)'*inv2*(bigr*beta-sr))/3
 |_print F ff
     F
    7.789598
     FF
    7.789598
```

41

Theil's Mixed Estimation Method for Bayesian VAR

Data generating process

$$\underset{T\times 1}{Y_j} = \underset{T\times k}{X}\ \underset{k\times 1}{\beta_j} + \underset{T\times 1}{u_j} \quad \textbf{where } u_j \sim N\left(0, \sigma^2_{j,j} I_T\right)$$

Prior distributions are

$$\underset{k\times 1}{r_j} = \underset{k\times k}{R}\ \underset{k\times 1}{\beta_j} + \underset{k\times 1}{v_j}$$

Mixed estimator is obtained by combining above two

$$\underset{(T+k)\times 1}{\begin{bmatrix} Y_j \\ r_j \end{bmatrix}} = \underset{(T+k)\times k}{\begin{bmatrix} X_j \\ R_j \end{bmatrix}} \left[\beta_j\right] + \underset{(T+k)\times 1}{\begin{bmatrix} u_j \\ v_j \end{bmatrix}}$$

$$E\begin{bmatrix} u_j \\ v_j \end{bmatrix}\begin{bmatrix} u_j & v_j \end{bmatrix} = \begin{bmatrix} \sigma^2_{jj} I_T & 0 \\ 0 & I_k \end{bmatrix}$$

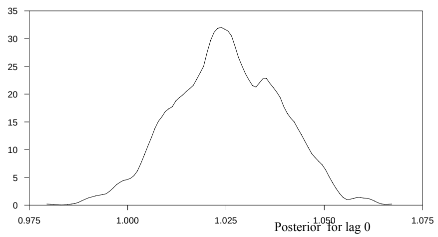
Point estimation.

$$\beta_j = \left[\begin{bmatrix} X_j & R_j \end{bmatrix}\begin{bmatrix} \sigma^2_{jj} I_T & 0 \\ 0 & I_k \end{bmatrix}^{-1}\begin{bmatrix} X_j \\ R_j \end{bmatrix}\right]^{-1}\left[\begin{bmatrix} X_j & R_j \end{bmatrix}\begin{bmatrix} \sigma^2_{jj} I_T & 0 \\ 0 & I_k \end{bmatrix}^{-1}\begin{bmatrix} Y_j \\ r_j \end{bmatrix}\right]$$

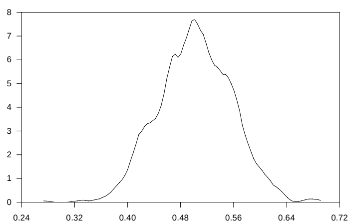
42

GIBB's Sappling; Litterman's Prior, Sim's Lag test

Posterior for Sum

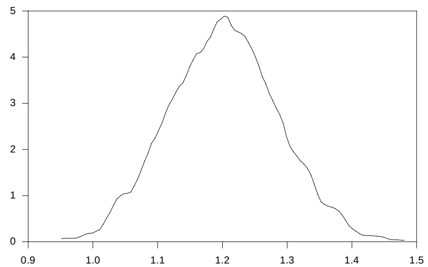


Posterior for lag 0

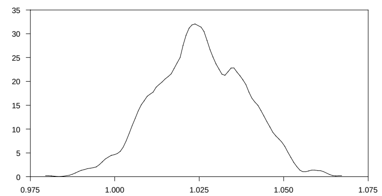


43

Posterior for intercept

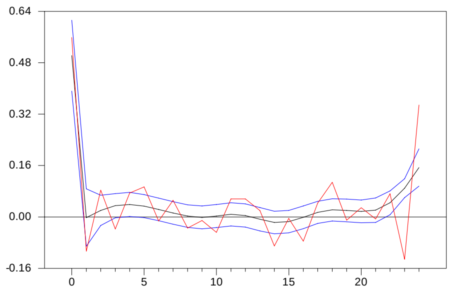


Posterior for residual sum square



44

Graph of lag distributions



45

Linear Regression - Estimation by Least Squares
Dependent Variable LONGRATE
Monthly Data From 1957:01 To 2000:12
Usable Observations 528 Degrees of Freedom 502
Centered R**2 0.917863 R Bar **2 0.913772
Uncentered R**2 0.990240 T x R**2 522.847
Mean of Dependent Variable 7.0826325758
Std Error of Dependent Variable 2.6033106352
Standard Error of Estimate 0.7644503877
Sum of Squared Residuals 293.36096641
Regression F(25,502) 224.3891
Significance Level of F 0.00000000
Log Likelihood -594.04873
Durbin-Watson Statistic 0.136490

Variable	Coeff	Std Error	T-Stat	Signif
1. Constant	1.184722143	0.085793634	13.80897	0.00000000
2. SHORTRATE	0.558763042	0.078151332	7.14976	0.00000000
3. SHORTRATE{1}	-0.105999176	0.130705524	-0.81098	0.41776318
4. SHORTRATE{2}	0.083775337	0.137269641	0.61030	0.54194084
5. SHORTRATE{3}	-0.037398107	0.138159675	-0.27069	0.78674260
6. SHORTRATE{4}	0.074312980	0.138335545	0.53719	0.59137186
7. SHORTRATE{5}	0.093572926	0.139123064	0.67259	0.50151697
8. SHORTRATE{6}	-0.012797235	0.140861303	-0.09085	0.92764810
9. SHORTRATE{7}	0.051411002	0.141386702	0.36362	0.71629504
10. SHORTRATE{8}	-0.034320781	0.141467576	-0.24261	0.80841028
11. SHORTRATE{9}	-0.011066640	0.139005898	-0.07961	0.93657699
12. SHORTRATE{10}	-0.047991318	0.135507520	-0.35416	0.72336783
13. SHORTRATE{11}	0.056020011	0.134388454	0.41685	0.67696532
14. SHORTRATE{12}	0.056027656	0.135061974	0.41483	0.67844413
15. SHORTRATE{13}	0.019883194	0.134291180	0.14806	0.88235462
16. SHORTRATE{14}	-0.090273444	0.135308929	-0.66717	0.50497300
17. SHORTRATE{15}	-0.004664073	0.138607622	-0.03365	0.97317005
18. SHORTRATE{16}	-0.075216464	0.140978647	-0.53353	0.59390227
19. SHORTRATE{17}	0.043352086	0.141004266	0.30745	0.75862663
20. SHORTRATE{18}	0.107845780	0.140544452	0.76734	0.44323845
21. SHORTRATE{19}	-0.009978597	0.138845595	-0.07187	0.94273534
22. SHORTRATE{20}	0.028710596	0.138164809	0.20780	0.83546970
23. SHORTRATE{21}	-0.006156344	0.138036298	-0.04460	0.96444431
24. SHORTRATE{22}	0.072220410	0.137201967	0.52638	0.59885650
25. SHORTRATE{23}	-0.131475075	0.130730251	-1.00570	0.31504566
26. SHORTRATE{24}	0.348639932	0.078024887	4.46832	0.00000975

46

RATS Routine

```
* GIBBS.PRG
* Manual example 13.4
*
cal 1955 1 12
allocate 2000:12
open data rates.rat
data(format=rats) / shortrate longrate
*
* Compute OLS estimates to start
*
linreg longrate
# constant shortrate{0 to 24}
*
* Save the OLS coefficients, residual sum of squares, X'X and
* X'y
*
compute rssols =%rss
compute betaols=%beta
compute xxols =inv(%xx)
compute xyols =xxols*%beta
*
* Set the number of draws
* Set the prior degrees of freedom (NU) and mean (S2) for
* sigma**2
* Set the prior precision for the second difference in the
* coefficients
*
compute ndraws=1000
compute s2=.50**2.0
compute nu=4.0
compute hb=1.0/(.03**2)
*
* Generate the precision matrix for the prior. For this example,
* the prior is flat on the constant and zero lag coefficient,
* and the second differences on the remainder of the lag
* polynomial are given independent Normal priors with common
* precision. In general, PRIORH will be the inverse of your
* prior variance.
```

47

Components of a Bayesian Model

Probability statements regarding θ after the sample on y has been observed given by a posterior density

$$f(\theta/y) = \frac{f(y,\theta)}{f(y)} = \frac{f(y/\theta)f(\theta)}{\int_{-\infty}^{\infty} f(y/\theta)f(\theta)d\theta} \text{ or simply}$$

$$f(y,\theta) = f(\theta/y)f(y)$$

This is the Bayes's law named after the eighteenth century mathematician Thomas Bayes.

48

Prior density function of β **is** $f(\beta)$

Sample information $f(y_t/\beta) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}(y_t-\beta)^2\right\}$

$$f(y/\beta) = f(y_1, y_2, y_3, \ldots.., y_T/\beta) = f(y_1/\beta).f(y_2/\beta)\ldots f(y_T/\beta)$$

Joint density $= \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t-\beta)^2\right\}$

Post sample density that represents revised prior

$$f(\beta/y) = \frac{f(y/\beta)f(\beta)}{f(y)} = \boldsymbol{k} \times \text{sample information} \times \text{prior information.}$$

Bayesian Updating Procedure

$f(\beta)=1$ **for** $-\infty < \beta < \infty$ assume the variance of distribution σ^2 is known.

$$f(\beta/y) = kf(y/\beta)f(\beta) = k\left(2\pi\sigma^2\right)^{\frac{-T}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t-\beta)^2\right\} \times 1$$

Convert this to a density function of β

$$\sum_{t=1}^{T}(y_t-\beta)^2 = \sum_{t=1}^{T}\left[(y_t-\bar{y})-(\beta-\bar{y})\right]^2 = \sum_{t=1}^{T}(y_t-\bar{y})^2 + \sum_{t=1}^{T}(\beta-\bar{y})^2 - 2(\beta-\bar{y})\sum_{t=1}^{T}(y_t-\bar{y}$$

$$= \sum_{t=1}^{T}(y_t-\bar{y})^2 + T(\beta-\bar{y})^2$$

substituting this into the posterior density function

$$f(\beta/y) = kf(y/\beta)f(\beta) = k\left(2\pi\sigma^2\right)^{\frac{-T}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{t=1}^{T}(y_t-\bar{y})^2 + T(\beta-\bar{y})^2\right]\right\} \times 1$$

$$f(\beta/y) = k\left(2\pi\sigma^2\right)^{\frac{-T}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t-\bar{y})^2\right\} \exp\left\{-\frac{T}{2\sigma^2}T(\beta-\bar{y})^2\right\}$$

$$f(\beta/y) = c_1 \exp\left\{-\frac{T}{2\sigma^2}(\beta-\bar{y})^2\right\} \text{ where } c_1 = k\left(2\pi\sigma^2\right)^{\frac{-T}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t-\bar{y})^2\right\}$$

Distribution of Bayesian Parameter in a Single Variable Regression

$$z_1 = \frac{\beta_2 - \hat{\beta}_2}{\sqrt{\operatorname{var}(\hat{\beta}_2)}} \text{ and } \beta_2 = \hat{\beta}_2 - z_1\sqrt{\operatorname{var}(\hat{\beta}_2)}$$

β_2 is now a random variable

$$E[\beta_2] = \hat{\beta}_2 - E(z_1)\sqrt{\operatorname{var}(\hat{\beta}_2)} = \hat{\beta}_2$$

$$\operatorname{var}[\beta_2] = \operatorname{var}(z_1)\operatorname{var}\left(\sqrt{\operatorname{var}(\hat{\beta}_2)}\right)^2 = \operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum_i x_{1,i}^2} = \sigma_{\beta_2}^2$$

$$\beta_2 \sim N\left(\hat{\beta}_2, \sigma_{\beta_2}^2\right)$$

Similarly for $\beta_1 \sim N\left(\hat{\beta}_1, \sigma_{\beta_1}^2\right)$; $\sigma_{\beta_1}^2 = \dfrac{\sigma^2\sum_i X_{1,i}^2}{T\sum_i x_{1,i}^2}$

References

Bauwens L , M. Lubrano and J. F. Richard (1999) Bayesian Inference in Dynamic Econometric Models, Oxford University Press.

ESTIMA (2004) RATS User Guide, Version 6,Evaston IL 60201 USA.

Griffiths W. E. and R. C. Hill and G. G. Judge (1993) Learning and Practicing Econometrics, John Willey.

Fu Q. (2000) Bayesian Multivariate Time Series Models for Forecasting European Macroeconomic Series, PhD Dissertation, University of Hull.

Hamilton J. D. (1994) Time Series Analysis, Princeton University Press.

Judge G.G., W.E. Griffiths, R. C. Hill, H. Lutkepohl and T.C. Lee(1990) Theory and Practice of Econometrics, Wiley.

Lancaster T (2004) An Introduction to Modern Bayesian Econometrics, Blackwell

Zellner A (1985) Bayesian Econometrics, Econometrica, 53:2:253-269.

Asymptotic Theory and GMM

Advanced Economic Theory

Lecture 13

Asymptotic Theory

$$\underset{t \to \infty}{Lim} \; P\left(\left|\hat{\theta}-\theta\right| \le \varepsilon\right)=1 \quad t \to \infty; \; P\lim\left(\hat{\theta}\right)=\theta$$

Chebycheff inequality $P\left\{\left|z-E(z)\right| > \delta\right\} < \dfrac{V\{z\}}{\delta^2}$ for all $\delta > 0$.

For the OLS estimator this implies that

$P\left\{\left|b_k-\beta_k\right| > \delta\right\} < \dfrac{V\{\beta_k\}}{\delta^2} = \dfrac{\sigma^2 (X'X)^{-1}}{\delta^2}$ as the sample size increases

$(X'X)^{-1} = \left(\sum_i^n x_i x_i'\right)^{-1}$ decreases as $\left(\sum_i^n x_i x_i'\right)$ becomes larger. When N become infinitely larger then

$\lim_{N\to\infty} P\left\{\left|b_k-\beta_k\right| > \delta\right\} < \dfrac{V\{\beta_k\}}{\delta^2} = \dfrac{\sigma^2 (X'X)^{-1}}{\delta^2} = 0$ for all $\delta > 0$.

This is the same thing as saying that the b converges in probability in β.

$P\lim(b)=\beta$ when the estimator of β converges to its true value it is a **consistent estimator** with asymptotic distribution as: $b \overset{a}{\sim} N\left(\beta, \left(\sum_i^n x_i x_i'\right)^{-1}\right)$

Sargan's General Instrumental Variable Estimator

If $E(X'u) \neq 0$ assume a set of instruments Z which as same dimension as X but uncorrelated with the error term, $E(Z'u)=0$ and variance $Var(Z'u)=(Z'Z)\sigma^2$.

The instrumental variable estimator is obtained by setting $(Z'u)=0$

or $Z'(Y-X\beta)=0$ or $\hat{\beta}_{IV} = (Z'X)^{-1} Z'Y$

Sargan (1958) had proposed generalised instrumental variable estimator (GIVE)

$Min(Y-X\beta)'Z'(Z'Z)^{-1}Z'(Y-X\beta)$

$E(u'u)=\Omega$

Then GIVE minimizes $Min(Y-X\beta)'Z'(Z'\Omega Z)^{-1}Z'(Y-X\beta)$.

The GMM estimator proposed by Hansen (1982) is similar to GIVE estimator

It minimises $uZWZ'u'$

$Min\,(Y-X\beta)'Z'WZ'(Y-X\beta) \Rightarrow (X'Z)W(Z'X)\hat{\beta} = (X'Z)WZ'Y$

$\hat{\beta}_{GMM} = (X'ZWZ'X)^{-1}(X'ZWZ'Y)$

Hansen's weighting matrix $W = (Z'Z)^{-1}/\sigma^2 = Var(Z'u)^{-1}$ when $E(u'u)=\Omega$

Changing this weighting matrix

$\hat{\beta}_{GMM} = \left(X'Z(Z'\Omega Z)^{-1}Z'X\right)^{-1}\left(X'Z(Z'\Omega Z)^{-1}Z'Y\right)$

$Cov\left(\hat{\beta}_{GMM}\right) = \left(X'Z(Z'\Omega Z)^{-1}Z'X\right)^{-1}$

The popularity of GMM is due to its generality as it can be applied to linear and non-linear estimations.

Dynamic Panel data models

$$y_{i,t} = \gamma y_{i,t}\beta + \alpha_i + \varepsilon_{i,t} \quad \text{with} \quad \gamma < 1$$

$$\gamma_{FE} = \frac{\sum_t^T \sum_i^N \left(y_{i,t} - \bar{y}_i\right)\left(y_{i,t} - \bar{y}_{i,t-1}\right)}{\sum_t^T \sum_i^N \left(y_{i,t-1} - \bar{y}_{i,-1},\right)^2} \;;\; \bar{y}_i = T^{-1}\sum_i y_{i,t} \text{ and } \bar{y}_{i,-1} = T^{-1}\sum_i y_{i,t-1}$$

This is not asymptotically unbiased estimator:

$$\gamma_{FE} = \gamma + \frac{(1/NT)\sum_t^T \sum_i^N \left(\varepsilon_{i,t} - \bar{\varepsilon}_i\right)\left(y_{i,t} - \bar{y}_{i,t-1}\right)}{(1/NT)\sum_t^T \sum_i^N \left(y_{i,t-1} - \bar{y}_{i,-1},\right)^2}$$

$$\underset{N\to\infty}{P\lim} \quad (1/NT)\sum_t^T \sum_i^N \left(\varepsilon_{i,t} - \bar{\varepsilon}_i\right)\left(y_{i,t} - \bar{y}_{i,t-1}\right) = -\frac{\sigma_\varepsilon^2}{T^2}\frac{(T-1) - T\gamma + \gamma^T}{(1-\gamma)^2} \neq 0$$

Dynamic Panel data models

Instrumental variable methods have been suggested to solve this inconsistency

$$\hat{\gamma}_{IV} = \frac{\sum_t^T \sum_i^N y_{i,t-2}\left(y_{i,t-1} - \bar{y}_{i,t-2}\right)}{\sum_t^T \sum_i^N y_{i,t-2}\left(y_{i,t-1} - y_{i,t-2},\right)} \quad \text{and}$$

where y_{i-2} is used as instrument of $\left(y_{i,t-1} - y_{i-2}\right)$

It is asymptotically

$$\underset{N\to\infty}{P\lim} \quad (1/NT)\sum_t^T \sum_i^N \left(\varepsilon_{i,t} - \bar{\varepsilon}_i\right) y_{i,t-2} = 0$$

Moment conditions with vector of transformed error terms

$$\Delta\varepsilon_i = \begin{pmatrix} \varepsilon_{i,2} - \varepsilon_{i,1} \\ \varepsilon_{i,3} - \varepsilon_{i,2} \\ . \\ - \\ \varepsilon_{i,T} - \varepsilon_{i,T-1} \end{pmatrix}$$

Dynamic Panel data models

$$Z_i = \begin{bmatrix} [y_{i0}] & 0 & 0 & . & . & 0 \\ 0 & [y_{i0}, y_{i1}] & . & . & . & . \\ 0 & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & [y_{i0}, y_{i,T-2}] \end{bmatrix}$$

$$E\{Z_i' \Delta\varepsilon_i\} = 0$$

Or for moment estimator write the transformed errors as
$E\{Z_i'(\Delta y_{i,t} - \gamma\Delta y_{i,t})\} = 0$

$$\min_\gamma \left(\frac{1}{N}\sum_{i=1}^N Z_i'(\Delta y_{i,t} - \gamma\Delta y_{i,t})\right)' W_N \left(\sum_i^n Z_i'(\Delta y_{i,t} - \gamma\Delta y_{i,t})'\right)$$

GMM method includes the most efficient instrument

$$\hat{\gamma}_{GMM} = \left(\left(\sum_{i=1}^N \Delta y_{i,-1}' Z_i\right) W_N \left(\sum_{i=1}^N Z_i' \Delta y_{i,-1}'\right)\right)^{-1} \times \left(\sum_{i=1}^N \Delta y_{i,-1}' Z_i\right) W_N \left(\sum_{i=1}^N Z_i' \Delta y_i'\right)$$

The optimal weighting matrix W_N can guarantee the most efficient estimator, this should be inversely proportional to transformed covariance matrix.

$$\hat{W}_N^{opt} = \left(\frac{1}{N}\sum_{i=1}^N Z_i' \Delta\varepsilon_i \Delta\varepsilon_i' Z_i\right)^{-1}$$

Computer Demonstration: Generating Panel data in Excel, Growth batch, Unemployment batch,

References

Blundell R and Bond S. (1998) Initial conditions and moment restrictions in dynamic panel data models, Journal of Econometrics, 87: 115-143.

Davidson J. (2002) Econometric Theory, Blackwell.

Davidson R and MacKinnon J. G. (2004) Econometric Theory and Methods, Oxford.

Judge G.G., W.E. Griffiths, R. C. Hill, H. Lutkepohl and T.C. Lee(1990) Theory and Practice of Econometrics, Wiley.

Hansen L.P. (1982) Large sample properties of generalized method of moment estimators, Econometrica, 50:4:1029-1054.

Hamilton (1994) Time Series Analysis, Princeton.

K. Kuthberson , S. Hall and M Taylor (1992) Applied time series analysis , Michigan University Press.

Newey W.K and McFadden D. (1994) Large Sample Estimation and Hypothesis Testing, Handbook of Econometrics, IV, (ed.) R. F. Engle and D. L. McFadden, 1994.

Wooldridge J. M. (2001) Applications of Generalised Methods of Moments Estimation, Journal of Economic Perspectives, 15:4:2001:87-100.