BAYESIAN ECONOMETRICS

1. GENERAL BACKGROUND

THEMES:

(i) Unified set of principles

(ii) Flexible use of prior information

(iii) Explicit decision - theoretic basis for inference.

(iv) Estimators / tests which are "optimal"

PROBABILITY:

(i) Classical (A priori) probability.

(ii) Long - Run Relative Frequency.

(iii) Subjective (personalistic) probability.

Features:

(a) Once-and-for all events O.K.

(b) Results conditional on information set.

(c) Results revised with more information.

(d) "Subjective betting-odds"

BAYES' THEOREM:

Recall:

$$p(B|A) = p(B \cap A)/p(A)$$

$$p(A|B) = p(A \cap B)/p(B)$$

$$\Rightarrow \quad p(B|A) = p(A|B)p(B)/p(A)$$

More generally -

$$p(B_j|A) = \frac{p(A|B_j)p(B_j)}{\sum_i p(A|B_i)p(B_i)}$$

So:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

$$= \frac{p(y|\theta).p(\theta)}{\int p(y|\theta).p(\theta)\,d\theta}$$

So, $\int p(\theta|y)\,d\theta = 1$, & we can write:

$$p(\theta|y) \propto p(\theta).\,p(y|\theta) = p(\theta)\,L(\theta|y)$$

PRIOR $\rightarrow$ Bayes' Theorem $\leftarrow$ DATA

Bayes' Theorem

$\downarrow$

POSTERIOR

$\downarrow$ etc.

Posterior $\propto$ (Prior) $\times$ (Likelihood Fctn.)

DECISION THEORY

Loss fctn. :

$$L_1(\theta,\hat{\theta}) = c_1(\theta-\hat{\theta})^2$$

$$L_2(\theta,\hat{\theta}) = c_2|\theta-\hat{\theta}|$$

$$L_3(\theta,\hat{\theta}) = \begin{cases} 1 & ; |\theta-\hat{\theta}| > \varepsilon \\ 0 & ; |\theta-\hat{\theta}| \le \varepsilon \end{cases}$$

$(\varepsilon > 0)$

"MEL" Rule:

" Act so as to minimize posterior expected loss. "

(choose $\hat{\theta}$; select a model / hypothesis)

BAYES' RULE:

" Act so as to minimize average risk ".

(a) Posterior Expected Loss is :

$$\int_{\Omega} L(\theta,\hat{\theta})\,p(\theta|y)\,d\theta$$

(b) Risk = Expected Loss

$$R(\hat{\theta}) = \int_{y} L(\theta,\hat{\theta})\,p(y|\theta)\,dy$$

(c) Average risk

$$r(\hat{\theta}) = \int_{\Omega} R(\hat{\theta})\,p(\theta)\,d\theta$$

So, Bayes' Rule $\Rightarrow$

$$\min_{(\hat\theta)} [r(\hat\theta)] = \min_{(\hat\theta)} \int_\Omega \Big[\int_y L(\theta,\hat\theta)\, p(y|\theta)\, dy \Big] p(\theta)\, d\theta$$

$$= \min_{(\hat\theta)} \int_\Omega \Big[\int_y L(\theta,\hat\theta)\, p(\theta|y)\, p(y)\, dy \Big] d\theta$$

$$= \min_{(\hat\theta)} \int_y p(y) \Big[\int_\Omega L(\theta,\hat\theta)\, p(\theta|y)\, d\theta \Big] dy$$

(This requires convergence of $\int\int$ – Fubini)

Integrand is non-negative, so $\Leftrightarrow$

$$\min_{(\hat\theta)} \int_\Omega L(\theta,\hat\theta)\, p(\theta|y)\, d\theta$$

So, often, Bayes' Rule $\Rightarrow$ MEL Rule.

(If Bayes' Rule undefined, can still use MEL Rule.)

Results:

(a) L_1 : $\hat\theta$ = posterior mean

(b) L_2 : $\hat\theta$ = posterior median

(c) L_3 : $\hat\theta$ = posterior mode

If $p(\theta)$ is "proper", then Bayes' Rules are consistent + admissible, by construction. (Try to prove latter result.)

Example:

$$\min_{(\hat\theta)} \int_\Omega c_1 (\theta - \hat\theta)^2\, p(\theta|y)\, d\theta$$

$$\Rightarrow c_1 \int_\Omega \frac{d}{d\hat\theta} (\theta - \hat\theta)^2\, p(\theta|y)\, d\theta = 0$$

$$\Rightarrow \int_\Omega 2(\hat\theta - \theta)\, p(\theta|y)\, d\theta = 0$$

$$\Rightarrow \hat\theta \int_\Omega p(\theta|y)\, d\theta = \int_\Omega \theta\, p(\theta|y)\, d\theta$$

$$\Rightarrow \hat\theta = E(\theta|y).$$

(check 2nd. Deri

__Example__:

$X_i \sim N(\mu, \sigma_0^2)$; σ_0^2 known

Information about unknown μ represented by $N(\bar{\mu}, \bar{V})$.

$$x = (x_1, x_2, ----, x_n)$$

Now, $p(x \mid \mu) = L(\mu \mid x)$

$$= \left(\frac{1}{\sigma_0\sqrt{2\pi}}\right)^n . \exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

$$\propto \underbrace{\exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}}_{\text{kernel of density.}}$$

$$p(\mu) = \frac{1}{\sqrt{2\pi\bar{V}}}\exp\left[-\frac{1}{2\bar{V}}(\mu-\bar{\mu})^2\right]$$

$$\propto \exp\left\{-\frac{1}{2\bar{V}}(\mu-\bar{\mu})^2\right\}$$

Bayes' Theorem:

$$p(\mu \mid x) \propto p(\mu) . p(x \mid \mu)$$

So,

$$p(\mu \mid x) \propto \exp\left[-\frac{1}{2}\left\{\frac{1}{\bar{V}}(\mu-\bar{\mu})^2 + \frac{1}{\sigma_0^2}\sum_i (x_i-\mu)^2\right\}\right]$$

Now,

$$\sum_i (x_i-\mu)^2 = \sum_i (x_i-\bar{x})^2 + n(\bar{x}-\mu)^2$$

So,

$$p(\mu \mid x) \propto \exp\left\{-\tfrac{1}{2}\left[\frac{1}{\bar{V}}(\mu-\bar{\mu})^2 + \frac{1}{\sigma_0^2}\sum_i (x_i-\bar{x})^2 + n/\sigma_0^2(\bar{x}-\mu)^2\right]\right\}$$

$$\propto \exp\left\{-\tfrac{1}{2}\left[\mu^2\left(\frac{1}{\bar{V}} + n/\sigma_0^2\right) - 2\mu\left(\frac{\bar{\mu}}{\bar{V}} + \frac{n\bar{x}}{\sigma_0^2}\right) + \text{constant}\right]\right\}$$

$$\propto \exp\left\{-\tfrac{1}{2}\left(a\mu^2 + b\mu + c\right)\right\}$$

$$\propto \exp\left\{-\tfrac{1}{2}\left[a\left(\mu + b/2a\right)^2 + \left(c - b^2/4a\right)\right]\right\}$$

So:

$$p(\mu \mid x) \propto \exp\left\{-\tfrac{1}{2}\left(\frac{1}{\bar{V}} + n/\sigma_0^2\right)\left(\mu - \frac{\left(\frac{\bar{\mu}}{\bar{V}} + n\bar{x}/\sigma_0^2\right)}{\left(\frac{1}{\bar{V}} + n/\sigma_0^2\right)}\right)\right\}$$

$$\sim N[\bar{\bar{\mu}}, \bar{\bar{V}}]$$

where: $\frac{1}{\bar{\bar{V}}} = \left[\frac{1}{\bar{V}} + n/\sigma_0^2\right]$

$$\bar{\bar{\mu}} = \frac{\left[\frac{1}{v}\cdot\bar{\mu} + \frac{n}{\sigma_0^2}\cdot\bar{x}\right]}{\left[\frac{1}{v} + \frac{n}{\sigma_0^2}\right]}$$

(Recall : $\bar{x} \sim N(\mu, \sigma_0^2/n)$)

(a) Weighted-average interpretations.

(b) $\bar{\bar{\mu}}$ is BAYES estimator of μ.

(c) Example of "Natural Conjugacy".

MARGINAL DENSITIES:

Let's see how flexible the Bayesian approach is w.r.t. "nuisance parameters".

e.g., $\underset{\sim}{\theta} = (\underset{\sim}{\beta}, \sigma) = (\beta_1, \beta_2, \ldots, \beta_k, \sigma)$

$$p(\underset{\sim}{\theta}|y) = p(\beta_1, \ldots, \beta_k, \sigma|y)$$

$$p(\underset{\sim}{\beta}|y) = \int_0^\infty p(\underset{\sim}{\beta}, \sigma|y)\,d\sigma$$

$$p(\sigma|y) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} p(\underset{\sim}{\beta}, \sigma|y)\,d\beta_1\,d\beta_2\cdots d\beta_k$$

Then, $\hat{\beta}_i = E(\beta_i|y)$

$$= \int_{-\infty}^{\infty} \beta_i \cdot p(\beta_i|y)\,d\beta_i$$

where $p(\beta_i|y) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} p(\underset{\sim}{\beta}|y)\,d\beta_1\cdots d\beta_{i-1}\,d\beta_{i+1}\cdots d\beta_k$

& $\text{var.}(\beta_i|y) = \int_{-\infty}^{\infty} (\beta_i - \hat{\beta}_i)^2\, p(\beta_i|y)\,d\beta_i$

(Enables us to provide point & interval estimates of β_i, ..., etc.

Example:

$$C_t = bY_t + u_t$$

$$u_t \sim N(0, \sigma^2) \quad ; t = 1, \ldots, n.$$

Prior Information:

$0 \le b \le 1$

$E(b) = 0.8$

$\text{var.}(b) = 0.001$

$0 < \sigma < \infty$

Likelihood Function:

$$p(C \mid b, \sigma) \propto \sigma^{-n} \exp\left[-\frac{1}{2\sigma^2} \sum_t (C_t - bY_t)^2\right]$$

Prior pdf:

$$p(b, \sigma) = p(b) \cdot p(\sigma)$$

$$p(\sigma) \propto \sigma^{-1} \quad ; \quad 0 < \sigma < \infty$$

$$p(b) \propto b^{\alpha - 1} (1 - b)^{\gamma - 1} \quad ; \quad 0 \le b \le 1$$

$$(\alpha, \gamma > 0)$$

What are α and γ?

$$E(b) = \left(\frac{\alpha}{\alpha + \gamma}\right) = 0.8$$

$$\text{var.}(b) = \alpha\gamma / [(\alpha+\gamma)^2 (\alpha+\gamma+1)] = 0.001$$

Solve $\Rightarrow$ $\alpha = 127.2$; $\gamma = 31.8$

So: $p(b, \sigma \mid C) \propto p(b, \sigma)\, p(C \mid b, \sigma)$

$$\propto \sigma^{-(n+1)} \exp\left\{-\frac{1}{2\sigma^2} \sum_t (C_t - bY_t)^2\right] b^{126.2} (1-b)^{30.8}$$

(a) Proportionality constant ?

(b) What are $\hat{b}$, $\hat{\sigma}$?

(c) Practical issues ?

We can obtain the marginal posterior for 'b' by integrating out 'σ' analytically (see pp. 21-22 later on):

$$p(b \mid c) = \int_0^\infty p(b, \sigma \mid c)\, d\sigma$$

$$\propto \int_0^\infty \sigma^{-(n+1)} \exp\left\{-\frac{1}{2\sigma^2} \sum_t (C_t - bY_t)^2\right\} \cdot b^{126.2} (1 - b)^{30.8}\, d\sigma$$

$$\propto \left[\sum_t (C_t - bY_t)^2\right]^{-n/2} \cdot b^{126.2} \cdot (1 - b)^{30.8}$$

As $0 \le b \le 1$, we can easily normalize $p(b \mid c)$ using numerical integration, & then find the posterior mean, etc.

11(b)

```
Welcome to SHAZAM - Version 10.0 -  JUL 2004 SYSTEM=WIN-XP   PAR=300000
CURRENT WORKING DIRECTORY IS: C:\SHAZAM
 |_smpl 1950 1985
 |_dim post1 101 x 101 post2 401 xx 401
 |_read(C:\Documents and Settings\Discovery User\My
Documents\Teaching\546\2006\10_1.dat) year  y  c
 UNIT 88 IS NOW ASSIGNED TO: C:\Documents and Settings\Discovery User\My Docume
 nts\Teaching\546\2006\10_1.dat
    3 VARIABLES AND       36 OBSERVATIONS STARTING AT OBS    1950
 |_*
 |_* U.S. ANNUAL CONSUMPTION AND INCOME DATA
 |_*
 |_print c y
          C              Y
    733.2000        791.8000
    748.7000        819.0000
    771.4000        844.3000
    802.5000        880.0000
    822.7000        894.0000
    873.8000        944.5000
    899.8000        989.4000
    919.7000        1012.100
    932.9000        1028.800
    979.4000        1067.200
    1005.100        1091.100
    1025.200        1123.200
    1069.000        1170.200
    1108.400        1207.300
    1170.600        1291.000
    1236.400        1365.700
    1298.900        1431.300
    1337.700        1493.200
    1405.900        1551.300
    1456.700        1599.800
    1492.000        1688.100
    1535.800        1728.400
    1621.900        1797.400
    1689.600        1916.300
    1674.000        1896.600
    1711.900        1931.700
    1803.900        2001.000
    1883.800        2066.600
    1961.000        2167.400
    2004.400        2212.600
    2000.400        2214.300
    2024.200        2248.600
    2050.700        2261.500
    2145.900        2334.600
    2239.900        2468.400
    2312.600        2509.000
 |_*
 |_* SCALE THE DATA TO DEAL WITH PRECISION ISSUES BELOW
 |_* (THIS DOES NOT ALTER THE ANSWERS)
 |_*
 |_g y=y/2000
 |_g c=c/2000
 |_stat y c/means=m
NAME        N   MEAN        ST. DEV      VARIANCE      MINIMUM        MAXIMUM
Y          36   0.77830     0.27193      0.73943E-01   0.39590        1.2545
C          36   0.70486     0.24450      0.59780E-01   0.36660        1.1563
 |_g y=y-m:1
 |_g c=c-m:2
```

1.

```
 |_ols c y/ noanova

 REQUIRED MEMORY IS PAR=        56 CURRENT PAR=  300000
 OLS ESTIMATION
        36 OBSERVATIONS      DEPENDENT VARIABLE= C
 ...NOTE..SAMPLE RANGE SET TO:    1950,    1985

 R-SQUARE =    0.9985     R-SQUARE ADJUSTED =    0.9985
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   0.89918E-04
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   0.94825E-02
 SUM OF SQUARED ERRORS-SSE=   0.30572E-02
 MEAN OF DEPENDENT VARIABLE =  -0.49343E-16
 LOG OF THE LIKELIHOOD FUNCTION =   117.646

 VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED
ELASTICITY
   NAME    COEFFICIENT   ERROR      34 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
Y          0.89848     0.5894E-02   152.4     0.000 0.999     0.9993     2.5831
CONSTANT   0.71316E-16 0.1580E-02 0.4513E-13 1.000 0.000     0.0000    -1.4453

 DURBIN-WATSON = 0.8398     VON NEUMANN RATIO = 0.8638     RHO =  0.59226
 RESIDUAL SUM =  0.17347E-16  RESIDUAL VARIANCE =  0.89918E-04
 SUM OF ABSOLUTE ERRORS=  0.22134
 R-SQUARE BETWEEN OBSERVED AND PREDICTED = 0.9985
 RUNS TEST:    11 RUNS,    21 POS,      0 ZERO,    15 NEG   NORMAL STATISTIC = -
2.6112
 COEFFICIENT OF SKEWNESS =  -0.4322 WITH STANDARD DEVIATION OF 0.3925
 COEFFICIENT OF EXCESS KURTOSIS =     1.5286 WITH STANDARD DEVIATION OF 0.7681

 JARQUE-BERA NORMALITY TEST- CHI-SQUARE(2 DF)=    3.0545 P-VALUE= 0.217

    GOODNESS OF FIT TEST FOR NORMALITY OF RESIDUALS -  6 GROUPS
 OBSERVED   2.0   3.0 10.0 19.0   1.0   1.0
 EXPECTED   0.8   4.9 12.3 12.3   4.9   0.8
 CHI-SQUARE =     9.6555 WITH  2 DEGREES OF FREEDOM, P-VALUE= 0.008
 |_g c2=c**2
 |_g y2=y**2
 |_g cy=c*y
 |_*
 |_* FORM THE SUMS OF SQUARES AND CROSS PRODUCTS OF THE DATA
 |_*
 |_stat c2  y2  cy/sums=s
NAME        N   MEAN         ST. DEV      VARIANCE      MINIMUM       MAXIMUM
C2         36   0.58119E-01  0.50428E-01  0.25429E-02   0.36523E-05   0.20380
Y2         36   0.71889E-01  0.58902E-01  0.34694E-02   0.70299E-05   0.22677
CY         36   0.64592E-01  0.54471E-01  0.29671E-02   0.50671E-05   0.21497
 |_smpl 1 1
 |_gen1 lower=0
 |_gen1 upper=1
 |_*
 |_* OBTAIN THE NORMALIZING CONSTANT FOR THE MARGINAL POSTERIOR FOR "b".
 |_* NOTE THAT (N/2)=18. TAKING LOGS AND EXPONENTIATING IS TO CONTROL FOR
 |_* NUMERICAL OVERFLOW:
 |_*
 |_integ b lower upper answer=(b**126.2)*((1-b)**30.8)/ &
 | exp(18*log(s:1+b**2*s:2-2*b*s:3))
OBSERVATION  LOWERBOUND      UPPERBOUND       INTEGRAL
      1      0.0000000       1.0000000        13503422.
 |_g norm=1/answer
 |_p norm
NORM
   0.7405530E-07
```

11(d)

```
|_* NOW OBTAIN THE MEAN OF THE (NORMALIZED) MARGINAL POSTERIOR FOR "b":
|_*
|_integ b lower upper mean=norm*(b**127.2)*((1-b)**30.8)/ &
| exp(18*log(s:1+b**2*s:2-2*b*s:3))
OBSERVATION  LOWERBOUND    UPPERBOUND     INTEGRAL
      1      0.0000000      1.0000000     0.89310850
|_p mean
MEAN
  0.8931085
|_*
|_* THE BAYES ESTIMATOR OF "b" IS VERY SIMILAR TO THE MLE OF "b"
|_* AS THE DATA EVIDENCE IS VERY STRONG - SEE THE t-RATIO ON THE SLOPE
|_* COEFFICIENT IN THE OLS (MLE) REGRESSION EARLIER.
|_* ALSO NOTE THAT THE BAYES ESTIMATOR IS A LITTLE SMALLER THAN THE MLE,
|_* BECAUSE THE PRIOR MEAN WAS 0.8, WHICH WAS LESS THAN THE MLE.
|_* NOW PLOT THE MARGINAL POSTERIOR OVER THE FULL RANGE:
|_*
|_set nodoecho
|_do #=1,101
|_genl x:#=0+(#-1)*0.01
|_genl post1:#=norm*((x:#)**126.2)*((1-x:#)**30.8)/exp(18*log(s:1+(x:#)**2* &
| s:2-2*(x:#)*s:3))
|_endo
****** EXECUTION BEGINNING FOR DO LOOP  # =       1
****** EXECUTION FINISHED FOR DO LOOP  #=     101
|_smpl 1 101

|_plot post1 x/nopretty

REQUIRED MEMORY IS PAR=     104 CURRENT PAR=  300000
      101 OBSERVATIONS
                    *=POST1
                    M=MULTIPLE POINT
   57.451        |                                  *
   54.427        |
   51.404        |
   48.380        |
   45.356        |
   42.332        |
   39.309        |
   36.285        |
   33.261        |                                   *
   30.237        |
   27.214        |
   24.190        |
   21.166        |
   18.142        |
   15.119        |
   12.095        |
   9.0712        |
   6.0475        |
   3.0237        |                                  *
   0.0000        |MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM MMMM*
                  ________________________________________

              0.000     0.250     0.500     0.750     1.000

                               X
```

11(e)

```
|_*
|_* THIS PLOT IS NOT VERY INFORMATIVE, SO REFINE THE SCALE AND "ZOOM IN"
|_* ON THE RELEVANT RANGE OF VALUES FOR "b":
|_*
|_set nodoecho
|_do #=1,201
|_genl xx:#=0.874+(#-1)*0.0002
|_genl post2:#=norm*((xx:#)**126.2)*((1-xx:#)**30.8)/exp(18*log(s:1+(xx:#)**2* &
| s:2-2*(xx:#)*s:3))
|_endo
****** EXECUTION BEGINNING FOR DO LOOP  # =       1
****** EXECUTION FINISHED FOR DO LOOP  #=     201
|_smpl 1 201

|_plot post2 xx

REQUIRED MEMORY IS PAR=     105 CURRENT PAR=  300000
      201 OBSERVATIONS
                    *=POST2
                    M=MULTIPLE POINT
   80.000        |
   75.789        |
   71.579        |
   67.368        |             MM*
   63.158        |            MM M
   58.947        |            M  *M
   54.737        |           MM   M
   50.526        |           M    M*
   46.316        |          *M     M
   42.105        |          M      M
   37.895        |          M      *M
   33.684        |         M*       M
   29.474        |         M        M
   25.263        |        *M         M
   21.053        |        M          M
   16.842        |       MM          *M
   12.632        |       M            M
   8.4211        |      M*            *M
   4.2105        |    MMM              MM
   0.0000        | MMMM                MMMMMM
                  ________________________________________

           0.870     0.885     0.900     0.915     0.930

                             XX

|_graph post2 xx

REQUIRED MEMORY IS PAR=     105 CURRENT PAR=  300000
      201 OBSERVATIONS
 SHAZAM WILL NOW MAKE A PLOT FOR YOU
|_stop
```

PROS & CONS OF BAYESIAN ANALYSIS :

(a) Unity of approach to all types of inference.
Prior ; L.F. ; loss fctn. ; Bayes' Theorem;
MEL Rule.

(b) Prior information incorporated explicitly
and flexibly.

(c) Forces declaration of prior information.

(d) Explicit use of decision theory.

(e) Admissible estimators & tests.

(f) Nuisance parameters handled easily.

But :

(a) Difficulty / cost of specifying prior.

(b) Integration / computational issues.