EMPIRICAL BAYES METHODS IN
TIME SERIES ANALYSIS

by

Taghi M. Khoshgoftaar

Dissertation submitted to the Graduate faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

APPROVED:

Dr. Richard G. Krutchkoff, Chairman

Dr. Klaus H. Hinkelmann

Dr. Robert V. Foutz

Dr. Raymond H. Myers

Dr. Jerry E. Mann

May 1982

Blacksburg, Virginia

ACKNOWLEDGEMENTS

I would like to express my appreciation to the following:

Dr. Richard G. Krutchkoff for guidance, patience and help in preparing this dissertation.

Dr. Robert V. Foutz for co-reading a handwritten copy of the manuscript, making many helpful suggestions. I also benefitted earlier from some conversations with Dr. Foutz concerning the material covered in Chapter V.

Dr. Klaus H. Hinkelmann, Dr. Raymond H. Myers and Dr. Jerry E. Mann for serving on my thesis committee.

Miss Patsy Galliher for her typing this dissertation skillfully.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

1.1 Introduction to Time Series

Most statistical procedures are designed to be used with data originating from a series of independent experiments or survey interviews. The resulting data, or sample, x_i, $i = 1, 2, \ldots, T$ are taken to be representative of some population. The statistical analysis that then follows is largely concerned with making references about the properties of the population from the sample. With this type of data, the order in which the sample is presented to the statistician is irrelevant. With time series data, this is by no means the case.

A time series is a sequence of values ordered by a time parameter, such as hourly temperature readings. This dissertation concerns the analysis of time series which are similar to the previously observed ones. Specifically, one assumes the existence of a population of time series and an underlying mechanism which generates independent parameters for each member of the population. We further assume that we have previously observed several members of the population. Notice that we are not assuming the replication of a time series with fixed parameters.

1.2 Purpose

To illustrate the problem more clearly, consider a researcher in U.S. Department of Labor who is concerned with the relationship between layoff rates in manufacturing at time t and time t-k. Let the layoff rates at time t be represented by x_t. The researcher assumes that the relation is of the form $x_t = \rho x_{t-1} + \varepsilon_t$ where ρ is the unknown parameter, and ε_t is a random error. In order to use this equation for predicting layoff rates in a given year for different months, the parameter ρ must be estimated.

This is done by obtaining monthly data for 12 months on the layoff rates in previous years. For different years one might expect a similar relationship to hold between x_t and x_{t-1}.

It seems reasonable that for different years the parameter will not be same. In certain years, factors that have not been considered in the model, such as interest rate, consumer price index, inflation rate, etc., would effect a different relationship between x_t and x_{t-1}. Therefore, for different years, the parameter takes on different values.

The estimate of the parameter (ρ) for a given year is a function only of the data taken from that year. Intuitively, it seems that a better estimate could be obtained if we use the data taken from other years.

It should be pointed out that the parameters themselves are different from year to year, and thus all the data can not be pooled to estimate the parameter.

Let us assume that the values of the parameter vary from year to year in an unpredictable manner, therefore they are random variables. If the parameter for each year was known, then the histogram might provide evidence for some underlying distribution on the parameter.

The underlying distribution of the parameter will be referred to as a prior distribution. However, the exact form of the prior distribution will not be known to the researcher. In the Empirical Bayes approach, the researcher does not assume any specific form for the prior distribution.

The main purpose of this example is to show how additional estimates of the parameter can be used to improve the estimate in the present situation. Various estimators using past information will be proposed and comparisons will also be made to the maximum likelihood estimator.

1.3 Covariance, Weakly Stationary and Autocorrelations. Consider a process x_t, defined for all integer values of t. In general, the process will be generated by some scheme involving random inputs, and so x_t will be a random variable for each t and $(x_{t_1}, x_{t_2}, \ldots, x_{t_n})'$ will be an n x 1 vector random variable. To fully characterize such random variables, one needs to specify distribution functions. Let the mean of x_t be defined by μ_t, $\mu_t = E(x_t)$, and the covariance between x_t and x_s will be $\gamma_{t,s} = \text{cov}\ (x_t, x_s) = E[(x_t-\mu_t)(x_s-\mu_s)]$, so that $\gamma_{t,t}$ is the variance of x_t. The linear properties of the process can be described in terms of just these quantities. This dissertation concerns only the analysis of linear and weakly stationary processes.

A process x_t will be said to be weakly stationary if the mean of x_t is equal to μ and the variance of $x_t = \sigma_x^2 < \infty$ for all t, i.e. $E\ x_t = \mu$ and $\text{var}\ x_t = \sigma^2 < \infty$ for all t and $\text{cov}\ (x_t, x_s) = \gamma_{t-s} = \gamma_{s-t}$

so that, $\sigma_x^2 = \gamma_o$. Thus a weakly stationary process will have mean and variance that do not change through time, and the covariance between values of the process at two time points will depend only on the difference in time between these time points and not on the times themselves. Let $\rho_t = \frac{\gamma_t}{\gamma_o}$, then ρ_t will be called the autocorrelation of the process.

1.4 The Autoregressive Model

A process generated by the equation $x_t = ax_{t-1} + \varepsilon_t$ (1.4.1) where ε_t are uncorrelated (o, σ^2) random variables and $|a| < 1$, is one of the simplest and most often used models in time series analysis.

x_t in equation (1.4.1) is called first-order autoregressive process. In general, the p^{th} order autoregressive process is defined by the stochastic difference equation

$$x_t = a_1 x_{t-1} + \dots + a_p x_{t-p} + \varepsilon_t. \tag{1.4.2}$$

Under the weakly stationary condition, x_t can be written in the form

$$x_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}. \tag{1.4.3}$$

Mann and Wald (14) studied the estimation problem. Let $\varepsilon_t \sim N(0, \sigma^2)$, then the maximum likelihood estimates $\hat{a}_i$ satisfy the normal equations:

$$\sum_{t=1}^{T} x_t x_{t-j} - \sum_{i=1}^{p} \hat{a}_i \sum_{t=1}^{T} x_{t-i} x_{t-j} = 0 \text{ for } j = 1, \dots, p \tag{1.4.4}$$

For the first order A.R. process, $x_t = a x_{t-1} + \varepsilon_t$, the conditional M. L. E.* is:

*The computation of the maximum likelihood estimators is greatly simplified if we assume that x_1 is fixed and investigate the conditional likelihood.

$$\hat{a} = \frac{\sum_{t=2}^{T} x_t \, x_{t-1}}{\sum_{t=2}^{T} x_{t-1}^{2}} \,. \tag{1.4.5}$$

1.5 Regression with Time Series Errors

The model with first-order autoregressive errors can be written as

$$y_t = \underline{x}'_t \, \underline{\beta} + u_t \qquad t = 1, 2, \ldots, T$$

$$u_t = \rho u_{t-1} + \varepsilon_t$$

where

$$\underline{x}'_t = (x_{1,\,t}, x_{2,\,t}, \ldots, x_{k,\,t})$$

$$\underline{\beta}' = (\beta_1, \beta_2, \ldots, \beta_k)$$

$\varepsilon_t \sim N(0, \sigma^2)$ and $E(\varepsilon_t \, \varepsilon_s) = 0$ for $t \neq s$. Let us assume that $y_t = \beta \, x_t + u_t$ and $u_t = \rho u_{t-1} + \varepsilon_t$ then y_t can be written as

$$y_t = \rho y_{t-1} + \beta (x_t - \rho x_{t-1}) + \varepsilon_t$$

Note: x's are explanatory variables, that is, fixed and known numbers, and σ^2 is known.

The conditional M.L.E. for ρ and β are

$$\hat{\beta} = \frac{\sum_{t=2}^{T} (x_t - \hat{\rho} x_{t-1}) \, (y_t - \hat{\rho} x_{t-1})}{\sum_{t=2}^{T} (x_t - \hat{\rho} x_{t-1})^2} \tag{1.5.1}$$

and

$$\hat{\rho} = \frac{\sum_{t=2}^{T} (y_t - \hat{\beta}x_t)(y_{t-1} - \hat{\beta}x_{t-1})}{\sum_{t=2}^{T} (y_{t-1} - \hat{\beta}x_{t-1})^2} \qquad (1.5.2)$$

with

$$E[(y_t - \rho y_{t-1})|\rho,\beta] = \beta(x_t - \rho x_{t-1}).$$

Let, ρ be known, then

$$E(\hat{\beta}|\rho) = E\left[\frac{\Sigma (x_t - \rho x_{t-1})(y_t - \rho y_{t-1})}{\Sigma (x_t - \rho x_{t-1})^2}\right]$$

$$= \frac{\Sigma (x_t - \rho x_{t-1})(\beta)(x_t - \rho x_{t-1})}{\Sigma (x_t - \rho x_{t-1})^2}$$

$$= \beta$$

and

$$\text{var}(\hat{\beta}|\rho) = \frac{\sigma^2}{\Sigma (x_t - \rho x_{t-1})^2}$$

Therefore,

$$(\hat{\beta}|\rho) \sim N\left[\beta, \frac{\sigma^2}{\Sigma (x_t - \rho x_{t-1})^2}\right]$$

When β is known, then

$$\hat{\rho} = \frac{\Sigma u_t u_{t-1}}{\Sigma u_{t-1}^2}$$

Let $\beta = \hat{\beta}$, then we can write $\hat{u}_t = y_t - \hat{\beta}\, x_t$ and the M.L.E. of ρ is:

$$\hat{\rho} = \frac{\Sigma\, \hat{u}_t\, \hat{u}_{t-1}}{\Sigma\, \hat{u}_{t-1}^{\,2}}$$

Cochrance-Orcutt (6) have proposed a method for the solution of the normal equations (1.5.1), (1.5.2) known as the two step least squares procedure.

1.6 The Moving Average Model

We shall call the time series x_t, defined by $x_t = \varepsilon_t + \sum_{j=1}^{q} \alpha_j\, \varepsilon_{t-j}$, where q is an integer, α_j are real numbers (parameters) and ε_t are independent and normally distributed with mean 0 and variance σ^2, a moving average process of order q.

It is easily verified that

$$E\, x_t = 0$$

$$\gamma_s = \operatorname{cov}(x_t, x_{t-s}) = E\,(x_t\, x_{t-s}) = \qquad (1.6.1)$$

$$\sigma^2\, [\alpha_s + \alpha_1\, \alpha_{s+1} + \ldots + \alpha_{q-s}\, \alpha_q] \qquad \text{for } |s| \leq q$$

and $\gamma_s = 0$ for $|s| > q$.

Taking s = 0 gives

$$\operatorname{var} x_t = \gamma_0 = \sigma^2 \sum_{j=0}^{q} \alpha_j^{\,2}.$$

It is immediately seen that an MA(q) process, with $q < \infty$, is always weakly stationary.

It follows from 1.6.1, that, for a moving average process of order q, the autocorrelations ρ_s are all zero for $s > q$. Thus, the autocorrelation function takes a simple and easily recognized form. For example, consider the MA(1) process $x_t = \varepsilon_t + \alpha\varepsilon_{t-1}$ which has first autocorrelation $\rho_1 = \frac{\alpha}{1 + \alpha^2}$ and clearly $|\rho_1| \leq .5$. It can easily be proved that the largest possible first autocorrelation ρ_1 achievable from an MA(q) process is

$$\rho_1 \text{ (max)} = \cos [\pi/(q+2)]$$

A.1.6. The backward operator B, which is frequently employed for notational convenience, is an operator on a time sequence with the property $Bx_t = x_{t-1}$. Thus, on reapplication

$$B^k x_t = x_{t-k}$$

This operator will often be used in polynomial form, so that $d_0 x_t + d_1 x_{t-1} + d_2 x_{t-2} + \ldots + d_r x_{t-p}$ can be summarized as $d(B) x_t$ where

$$d(B) = d_0 + d_1B + d_2B^2 + \ldots d_pB^p$$

B.1.6

Under stationarity condition, it is easily verified that AR(p) process could always be written in an MA(∞) form. Suppose $p = 1$ then AR(1) (i.e. $x_t = ax_{t-1} + \varepsilon_t, |a| < 1$) can be written as $x_t = \sum_{j=0}^{\infty} a^j\varepsilon_{t-j} = \sum_{j=0}^{q} a^j \varepsilon_{t-j}$ since $|a| < 1$ and $a^j \to 0$ as $j \to \infty$.

1.7 The Mixed Autoregressive-Moving Average Model

An obvious generalization of the MA and AR models that includes them as special cases is the mixed model in which x_t is generated by $x_t + \sum_{j=1}^{p} \alpha_j x_{t-j} = \varepsilon_t + \sum_{k=0}^{q} \beta_j \varepsilon_{t-k}$ where ε_t are uncorrelated and normally distributed with mean zero and variance σ^2 (known).

For the ARMA(1,1) process $x_t + \alpha x_{t-1} = \varepsilon_t + \beta\varepsilon_{t-1}$ it is easily varified that

$$\text{var}\ (x_t) = \gamma_0 = \frac{1 - 2\alpha\beta + \beta^2}{1-\alpha^2}\ \sigma^2$$

$$\gamma_1 = \frac{(1+\alpha\beta)\ (\alpha + \beta)}{1-\alpha^2}\ \sigma^2$$

$$\gamma_s = \text{cov}\ (x_t, x_{t-s}) = (-\alpha)\ \gamma_{s-1} \qquad s \geq 2.$$

CHAPTER II

2.1 Spectral Analysis (Introduction)

The previous chapter was largely concerned with the description of time series in terms of models - known as the time domain approach. There is, however, an alternative approach, known as spectral or frequency-domain analysis, that often provides useful insights into the properties of a series.

Let x_t be a stationary process with autocovariance $\gamma_s =$ cov (x_t, x_{t-s}) and corresponding autocorrelations $\rho_s = \frac{\gamma_s}{\gamma_0}$. The autocovariance generating function $\gamma(Z)$ is defined by $\gamma(Z) = \sum_{\text{all } s} \gamma_s Z^s$ and it will be assumed that this function exists, at least for $|Z| = 1$. The function of particular interest in this chapter is defined by

$$f(w) = \frac{1}{2\pi} \gamma(Z) \text{ where } Z = e^{-iw}$$

$$= \frac{1}{2\pi} \sum_{\text{all } s} \gamma_s e^{-iws}$$

$$= \frac{\gamma_0}{2\pi} + \frac{1}{\pi} \sum_{s \geq 1} \gamma_s \cos (sw)$$

Note: $\gamma_s = \gamma_{-s}$.

The function f(w) is called the spectral density function.

From the definition it is seen that

(1) $f(w + 2\pi k) = f(w)$ for integer k so, $f(w)$ is periodic outside the interval $(-\pi, \pi)$ and thus one needs to consider the function only over this interval.

(2) $f(w) = f(-w)$ i.e. $f(w)$ is symmetric, so it is usual to plot $f(w)$ vs. w only for $0 \leq w \leq \pi$.

(3) $\gamma_s = \int_{-\pi}^{\pi} e^{isw} f(w)\, dw$, this important property follows from the fact that $\int_{-\pi}^{\pi} e^{i(k-s)w} dw = \begin{Bmatrix} 0 & \text{if } k \neq s \\ 2\pi & \text{if } k = s \end{Bmatrix}$

Thus, the sequence γ_s and the function $f(w)$ comprise a Fourier transform pair, with one being uniquely determined by the other, so that the time-domain approach, which concentrates on the γ_s sequence and the models derived from it, and the frequency domain approach, which is based on interpretation of $f(w)$ are theoretically equivalent to each other. The function $\frac{f(w)}{\gamma_0}$ has the properties of a probability density function when $w \in [-\pi, \pi]$ since $f(w) \geq 0$ and third property of $f(w)$, so $\int_{-\pi}^{\pi} \frac{f(w)}{\gamma_0} = 1$. Since the covariance function of a stationary time series is the correlation function multiplied by the variance of the process we have $\gamma_s = \int_{-\pi}^{\pi} e^{iws}\, dF(w)$. The function $F(w)$ have been called the spectral distribution function, and it is usual to write $F(w) = \int_{-\pi}^{w} \frac{f(w)}{\gamma_0} dw$. The other fundamental relation is that the time series itself can be expressed as follows:

$$x_t = \int_{-\pi}^{\pi} e^{itw}\, dZ(w)$$

where Z(w) is an orthogonal stochastic process, and

$$E\,\{dZ(w)\,\overline{dZ(\lambda)}\} = 0 \qquad \text{if } w \neq \lambda$$

$$= \lambda_0\, dF(w) \qquad \text{if } w = \lambda$$

Since f(w) and ρ_s are Fourier transforms of one another, it follows that x_t possesses a spectral density function of the form

$$f(w) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k\, e^{-iwk}$$

Recall that we have taken $\{\varepsilon_t\}$ to be a time series of uncorrelated $N(0,\, \sigma^2)$ random variables. The spectral density of $\{\varepsilon_t\}$ is $f_\varepsilon(w) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k \cos w\,k = \frac{\sigma^2}{2\pi}$ since

$$\gamma_k = \int_{-\pi}^{\pi} f(w)\, e^{iwk}\, d\,w = \int_{-\pi}^{\pi} \frac{\sigma^2}{2\pi}\, e^{iwk}\, d\,w$$

$$= \int_{-\pi}^{\pi} \frac{\sigma^2}{2\pi} \cos\,(wk)\, d\,w$$

$$= \begin{cases} \sigma^2 & \text{if } k = 0 \\ 0 & \text{otherwise.} \end{cases}$$

2.2 Estimates of the Spectral Density

A stationary Gaussian process is described completely by its mean $E\,x_t = \mu$ and its covariance sequence $E(x_t - \mu)(x_{t+k} - \mu) = \gamma_k$, $k = 0, 1, 2, \ldots$, or equivalently by its mean and spectral function $F(w)$ or equivalently by its mean and spectral density, $f(w)$.

Let us suppose that μ is equal to zero, and c_k is an estimate of γ_k, where c_k is

$$c_k = c_{-k} = \frac{1}{T-k} \sum_{t=1}^{T-k} x_t\, x_{t+k} \qquad t = 0, 1, 2, \ldots, T-1$$

It is easily verified that c_k is an unbiased estimate of γ_k. There is another useful estimate of γ_k. It is:

$$c'_k = \frac{1}{T} \sum_{t=1}^{T-k} x_t\, x_{t+k}$$

where

$$c_s = \frac{T}{T-s}\, c'_s$$

c'_s reduces the mean squared error of $\hat{f}(w)$.

The sample spectral density is

$$\hat{f}(w) = I(w) = \frac{1}{2\pi T} \left| \sum_{t=1}^{T} x_t\, e^{iwt} \right|^2 \qquad w \in [-\pi, \pi] \tag{2.2.1}$$

By rearranging (2.2.1), one can see that

$$I(w) = \frac{1}{2\pi} \sum_{j=-(T-1)}^{T-1} c'_j\, e^{iwj} = \frac{1}{2\pi} \sum_{j=-T+1}^{T-1} \left(\frac{1}{T} \sum_{t=1}^{T-|j|} x_t\, x_{t+k}\right) e^{iwj}$$

$$= \frac{1}{2\pi} \sum_{j=-T+1}^{T-1} c'_j \cos (wj) \qquad w \in [-\pi, \pi]$$

I(w) is called the periodogram of the time series.

2.3 Moments of the Estimate of Spectral Density

Let us consider the first and second order moments of the sample spectral density I(w) defined by (2.2.1). The first order moment is

$$E(I(\lambda)) = \frac{1}{2\pi} \sum_{j=-(T-1)}^{T-1} (1 - \frac{|j|}{T}) \gamma_j \cos (wj)$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{j=-T+1}^{T-1} (1 - \frac{|j|}{T}) \cos (vj) \cos (wj) f(v)\, dv$$

since $EC'_k = (T-|k|)\, \gamma_k/T$. This can also be written

$$E(I(w)) = \frac{1}{2\pi T} \int_{-\pi}^{\pi} \sum_{t,s=1}^{T} \cos v(t-s) \cos w(t-s) f(v)\, dv$$

$$= \frac{1}{2\pi T} \int_{-\pi}^{\pi} \left| \sum_{t=1}^{T} e^{i(v-w)t} \right|^2 f(v)\, dv$$

$$= \frac{1}{2\pi T} \int_{-\pi}^{\pi} \left[\frac{\sin \frac{1}{2} (v-w)T}{\sin \frac{1}{2} (v-w)} \right]^2 f(v)\, dv$$

Early investigators found that I(w) is approximately unbiased, but not consistent i.e.

$$E(I(w)) \doteq f(w)$$

$$\text{var } (I(w)) \doteq f^2(w) \text{ for any } T.$$

2.4 Smoothing the Periodogram

In order to obtain consistent estimates, Daniell (7) suggested that a time series of length $\underline{\underline{T}}$ be divided into n segments (subseries), each of length m, where $T = m \cdot n$. Let $I_j(w)$ be the periodogram of the j^{th} segment and $\{c'_{j,s}\}$ the corresponding autocovariances. Then, the average of the spectral estimates over all segments could be used as a smoother estimator.

$$\overline{I}(w) = \frac{1}{n} \sum_{j=1}^{n} I_j(w)$$

$$= \frac{1}{2\pi} \sum_{|s|<m} (n^{-1} \sum_{j=1}^{n} c_{j,s}) e^{isw}$$

$$= \frac{1}{2\pi} \sum_{|s|<m} (1 - \frac{|s|}{m}) \frac{1}{n(m-|s|)} \sum_{j} m\, c_{j,s}\, e^{isw} \qquad (2.4.1)$$

Now $\sum_j m\, c_{j,s}$ is,a sum of products of the $x_t\, x_{t+s}$, but not all such products, for the term $x_t\, x_{t+s}$ is included only if x_t and x_{t+s} fall in the same subseries of the series. There are $m-|s|$ terms in each $c_{j,s}$, and thus

$$\frac{1}{n(m - |s|)} \sum_{j} m\, c_{j,s}$$

is the average of these products. It seems reasonable to replace this part of (2.4.1) by $T\, c_s/(T-|s|) = c_s/(1 - \frac{|s|}{T})$, the average of all available products of the form $x_t\, x_{t+s}$. This leads to the modified function

$$\bar{I}(w) = \frac{1}{2\pi} \sum_{|s|<m} \frac{1 - \frac{|s|}{m}}{1 - \frac{|s|}{T}} c_s e^{isw} \tag{2.4.2}$$

$$= \frac{1}{2\pi} \sum_{|s|<m} w_s c_s e^{isw} = \frac{1}{2\pi} \sum_{|s|<m} w_s c_s \cos(sw)$$

where

$$w_s = \frac{1 - \frac{|s|}{m}}{1 - \frac{|s|}{T}}$$

The function $\bar{I}(w)$ defined in (2.4.2) is known as the Barlett spectral density estimate, and the numbers w_s are the corresponding lag weights.

Other Smoothing Coefficients

A. General Blackman-Tukey Estimates

Blackman and Tukey (1) have proposed two estimates (called hanning and hamming by them) which are special cases

$$\hat{f}(w) = I(w) = \frac{1}{2\pi} \sum_{|s|<m} (1 - 2a + 2a \cos \frac{\pi s}{m}) \cos (ws)\, c_s$$

for $0 \leq a \leq \frac{1}{4}$

B. Hanning. A case of the Blackman-Tukey estimate is

$a = \frac{1}{4}$, then

$$\hat{f}(w) = I(w) = \frac{1}{2\pi} \sum_{|s|<m} (1/2 + 1/2 \cos \frac{\pi s}{m}) \cos (ws)\, c_s$$

C. Parzen. Two smoothing coefficients which yield small mean squared error are due to Parzen.

$$\hat{f}(w) = I(w) = \frac{1}{2\pi} \sum_{|s|<m} \left(1 - \frac{s^2}{m^2}\right) \left(1 - \frac{|s|}{T}\right) \cos (ws)\, c_s$$

$$= \frac{1}{2\pi} \sum_{|s|<m} \left(1 - \frac{s^2}{m^2}\right) \cos (ws)\, c_s$$

Another suggestion by Parzen is for m even

$$\hat{f}(w) = I(w) = \frac{1}{2\pi} \sum_{s = -m/2}^{m/2} \left(1 - 6\frac{s^2}{m^2} + 6\frac{|s|^3}{m^3}\right) \cos (ws)\, c_s$$

$$+ \frac{2}{2\pi} \sum_{s = -m}^{-m/2 - 1} \left(1 - \frac{|s|}{m}\right)^3 \cos (ws)\, c_s$$

2.5 The Approximate Distribution of Smoothed Spectral Estimates

Suppose that $x_1, x_2, \ldots, x_T$ are observation from a stationary Gaussian time series. Recall,

$$I(w) = \hat{f}(w) = \frac{1}{2\pi} \sum_{j=-T+1}^{T-1} c_j\, e^{iwj} \quad -\pi \leq w \leq$$

Let $I^*(w)$ be the periodic extension of $I(w)$ to the entire real line, so that $I^*(w)$ has period 2π and $I(w) = I^*(w)$ for $-\pi \leq w \leq \pi$.

Let k be any fixed integer and define

$$A(w) = \frac{1}{2k+1} \sum_{j=-k}^{k} I^*\left[\frac{2\pi}{T}(s(w;\, T) + j)\right] \quad -\pi \leq w \leq \pi$$

where for each w, s (w; T) is the smallest integer minimizing

$$\left| w - \frac{2\pi}{T} s\ (w;\ T) \right|.$$

It can be shown, see Brillinger (3) that for $w \neq 0 \pmod{\pi}$ $A(w)$ is approximately distributed as

$$\frac{1}{4k+2}\ f(w)\ \chi^2\ (4k+2) \text{ or } \frac{(4k+2)A(w)}{f(w)} \sim \chi^2\ (4k+2)$$

Thus

$$f(A(w)\,|\,f(w)) \propto \frac{(A(w))^{2k}}{(f(w))^{2k+1}} \quad \exp\ \{-(2k+1)\ \frac{A(w)}{f(w)}\}$$

CHAPTER III

3.1 Empirical Bayes Estimation

Suppose there is an unobservable random parameter θ which takes on values in a set Θ, and that θ has the distribution function $G(\theta)$. $G(\theta)$ is called the prior distribution. Suppose there is an observable random variable x with the known family of conditional distribution functions $\{F(x|\theta);\ \theta \in \Theta\}$. We are interested in estimating θ by using the observation X. Let us suppose that this procedure occurs periodically with the same unknown $G(\theta)$, therefore, we have a sequence

$$(\theta_1,\ x_1),\ (\theta_2,\ x_2),\ \ldots,\ (\theta_n,\ x_n) \text{ where } \theta_i \sim G(\theta)$$

and

$$x_i \sim F(X|\theta_i) \quad i = 1,\ 2,\ \ldots,\ n\ .$$

Robbins (18) in 1955 introduced Empirical Bayes analysis and the theory and applications were developed by Robbins (18), Krutchkoff (12), Martz (15), Rutherford (19) and others. In the estimation problem, a decision space D is defined which coincides with the parameter space Θ, along with a decision function $\delta(x)$ which takes on values in D. For each $\delta(x)$ and θ a loss function is defined as $L(\delta(x),\ \theta) \geq 0$.

The loss function represents the loss incurred in choosing $\delta(x)$ as the estimate of θ. A Bayes estimator is the $\delta(X)$ which minimizes the overall expected loss.

That is

$$R(G) = E\ [L(\delta^*(x),\ \theta)]$$

$$= \min_{\delta(x)}\ E\ [L(\delta(x),\ \theta)]$$

$$= \min_{\delta(x)}\ R\ (\delta,\ \theta)$$

where $R(\delta(x),\ \theta)$ is called the risk of $\delta(x)$ with respect to G and $R(\delta(x),\ \theta) = \int_x L(\delta,\ \theta)\ dF(x,\ \theta)$. R(G) is called the Bayes risk and $\delta^*(x)$ is called a Bayes decision procedure with respect to G.

Consider a quadratic loss function $L(\delta(x),\ \theta) = (\delta(x)-\theta)^2$ then the Bayes estimator of θ when x is observed can be shown to be $E(\theta|x)$.

Proof: Let $R(\delta(x),\ x) = \int_\theta (\delta-\theta)^2\ dF(\theta|x)$

$$= E[(\delta-\theta)^2|x]$$

$$= E\ (\delta^2 - 2\delta\theta + \theta^2|x)$$

$$= \delta^2 - 2\delta E(\theta|x) + E(\theta^2|x)$$

$$= \delta^2 - 2\delta E(\theta|x) + E^2(\theta|x) + E(\theta^2|x) - E^2(\theta|x)$$

$$= (\delta-E(\theta|x))^2 + \mathrm{var}\ (\theta|x)$$

where var $(\theta|x) = E[(\theta-E(\theta|x))^2|x]$. Since var $(\theta|x)$ does not depend on δ it is obvious that $R(\delta, x)$ is minimized if $\delta(x) = E(\theta|x)$ this is a Bayes decision procedure.

Martz (15) has developed this result for the multivariate case. Recall, the Bayes decision procedure is a decision procedure δ^* such that $u(\delta^*) = \max_{\delta} u(\delta)$ where $u(\delta)$ is the utility of δ. Since the prior distribution of θ is unknown, we cannot obtain δ^*, but by using past experience, we should at least expect $u(\delta(x_1, x_2, \ldots, x_n)) \xrightarrow{P} u(\delta^*)$ as $n \to \infty$. Then we say that $\delta(x_1, x_2, \ldots, x_n)$ is asymptotically optimal. However, when the parameter space Θ is unbounded, then for certain loss functions, like, squared error loss function, it is often impossible to find such an estimator.

For squared error loss function Rutherford and Krutchkoff (20) have verified that if for $\gamma > 0$ and some real number B, the prior distribution is such that

$$E\{|\theta|^{2+\gamma}\} \leq B < \infty$$

then "ε asymptotically optimal" estimators can be found.

3.2 The Classes of Families of Distributions

Rutherford and Krutchkoff (19) have found general classes of families of distributions $\{F(x|\theta), \theta \in \Theta\}$ (called F_1, F_2, F_3) for which empirical Bayes estimators, $\tilde{\theta}$, can be found. In this dissertation, two types of these families of distributions, F_2 and F_3, have been used.

3.2.1 Families of the form F_2 if

(1) x is continuous for all $\theta \varepsilon \Theta$, and

(2) the density function $f(x|\theta)$ is such that

$$\frac{1}{f(x|\theta)} \frac{\partial f(x|\theta)}{\partial x} = a(x) + b(x)\,\theta$$

where $b(x) \neq 0$. So, we can write

$$\theta = \frac{a(x) + b(x)\theta - a(x)}{b(x)}$$

$$= \frac{1}{b(x)} \cdot \frac{1}{f(x|\theta)} \frac{\partial f(x|\theta)}{\partial x} - \frac{a(x)}{b(x)}$$

Using $f(x) = \int_\Theta f(x|\theta)\, dG(\theta)$ and $dF(\theta|x)\ \frac{f(x|\theta)\ dG(\theta)}{f(x)}$ we get

$$E(\theta|x) = \int_\Theta \frac{\partial f(x|\theta)}{\partial x} \cdot \frac{dF(\theta|x)}{b(x)f(x|\theta)} - \frac{a(x)}{b(x)}$$

$$= \int_\Theta \frac{\partial f(x|\theta)}{\partial x} \cdot \frac{f(x|\theta)}{b(x)f(x|\theta)} \cdot \frac{dG(\theta)}{f(x)} - \frac{a(x)}{b(x)}$$

$$= \frac{1}{b(x)f(x)} \int \frac{\partial f(x|\theta)}{\partial x}\, dG(\theta) - \frac{a(x)}{b(x)}$$

$$= \frac{1}{b(x)f(x)} \frac{\partial}{\partial x} \int f(x|\theta)\, dG(\theta) - \frac{a(x)}{b(x)}$$

$$= \frac{1}{b(x)f(x)} \cdot \frac{\partial f(x)}{\partial x} - \frac{a(x)}{b(x)}$$

(assuming the order of integration and differentiation can be interchanged). From $x_1, x_2, \ldots$, one can find a consistent estimator for the marginal density $f(x)$, see Parzen (16) and Rutherford (19).

3.2.2 Families of the Form F_3

A family of distributions $\{F(x|\theta);\ \theta \in \Theta\}$ belongs to F_3 if (1) $\underline{x}$ is a vector (size $k > 1$) of conditionally independent and identically distributed continuous random variables; (2) there is a statistic T, sufficient for θ, with a density of the form

$$f_k(t|\theta) = \begin{cases} h(k)\ \left(\frac{t}{\theta}\right)^{k/v}\ q(t,\ \theta) & \text{for } t \text{ in some interval} \\ 0 & \text{otherwise.} \end{cases}$$

where $v < k$ is an integer and $h(k)$ does not depend on t or θ. Since these are k independent and identically distributed random variables in $\underline{X}$, the density of the sufficient statistic based on (k-v) components is given by

$$f_{k-v}(t|\theta) = \begin{cases} h(k-v)\ \left(\frac{t}{\theta}\right)^{k-v/v}\ q(t,\ \theta); & t \text{ in some interval} \\ 0 & \text{otherwise.} \end{cases}$$

It follows that

$$\theta f_k(t|\theta) = h(k)\ \left(\frac{t}{\theta}\right)^{k-v/v}\ t\ q(t,\ \theta)$$

so we can write

$$\theta = \frac{th(k)}{h(k-v)}\ \frac{f_{k-v}\ (t|\theta)}{f_k(t|\theta)}$$

Therefore

$$E(\theta|t) = \int_{\Theta} \theta dF(\theta|t) = \int_{\Theta} \frac{\theta f_k(t|\theta)}{f_k(t)} dG(\theta)$$

$$= \frac{th(k)}{h(k-v)} \int_{\Theta} \frac{f_{k-v}(t|\theta)}{f_k(t|\theta)} \cdot \frac{f_k(t|\theta)\, dG(\theta)}{f_k(t)}$$

$$= \frac{th(k)}{h(k-v)} \cdot \frac{f_{k-v}(t)}{f_k(t)}$$

The empirical Bayes estimator is found by estimating the ratio

$$\frac{f_{k-v}(t)}{f_k(t)}$$

3.2.3 Estimation of Density Functions

Parzen (16) has found consistent estimators of densities. In order to estimate a density function f(t), as is necessary for class F_2 and F_3, the consistent estimators

$$f_N(x) = \frac{1}{2\pi Nh} \sum_{i=1}^{N} \left\{ \frac{\sin\left(\frac{x-x_i}{2h}\right)}{\left(\frac{x-x_i}{2h}\right)} \right\}^2$$

will be used, where $h = N^{-1/5} \sqrt{\sum_{i=1}^{N} \frac{(x_i-\bar{x})^2}{N}}$

In order to estimate $\frac{df(x)}{dx}$, the consistent estimators

$$f'_N(x) = \frac{f_N(x+h) - f_N(x)}{h}$$

will be used. In order to estimate the density of a sufficient (for θ) statistic $\underline{\underline{T}}$ based on (k-1) observation, the technique used by Rencher (17) will be used.

$$f_{n,k-1}(t) = \frac{1}{2\pi Nh} \sum_{i=1}^{N} \left\{ \frac{\sin \frac{t-t_i^*}{2h}}{\left[\frac{t-t_i^*}{2h}\right]} \right\}^2$$

where $t_i^* = \frac{k-1}{k} t_i$, i = 1, 2, ..., N. h is same as before, N is the number of past experiences and t_i, i = 1, 2, ..., N is the sum of past data.

3.3 Empirical Bayes Estimators for the Chi-Squared Density

Assume that x has the exponential density, $f(x|\theta) = \frac{1}{\theta} e^{-x/\theta}$. Suppose that $x_1, x_2, \ldots, x_n$ are independent observations of x, and let $t = \sum_{i=1}^{n} x_i$, then t is a sufficient statistic for θ and has gamma distribution.

$$f_n(t|\theta) = \frac{1}{(n-1)!\theta^n} t^{n-1} e^{-t/\theta}$$

$$= \frac{1}{(n-1)!} (\frac{t}{\theta})^n (e^{-t/\theta}/t) .$$

So, $f_n(t|\theta)$ belongs to class F_3, where

$$h(n) = \frac{1}{(n-1)!} , \; q(t, \theta) = (e^{-t/\theta}/t)$$

Therefore

$$E(\theta|t) = t \cdot \frac{h(n)}{h(n-v)} \cdot \frac{f_{n-v}(t)}{f_n(t)}$$

where $f_n(t)$ is the marginal density of the sufficient statistic.

Since

$$\frac{h(n)}{h(n-1)} = \frac{1}{n-1}, \text{ then}$$

$$E(\theta|t) = \frac{t}{n-1} \; \frac{f_{n-1}(t)}{f_n(t)}$$

Therefore, we obtain Empirical Bayes estimator, by substituting consistent estimators for $f_n(t)$ and $f_{n-1}(t)$.

CHAPTER IV

Empirical Bayes Estimates of Time Series Parameters

4.1 Prediction With the AR Model

Given T observations $x_1, x_2, \ldots, x_T$ of a AR(P), we would like to predict the $(T+k)^{th}$ observation, where k is a positive integer. (Note: The prediction is sometimes called the forecast of the $(T+k)^{th}$ observation.) Let us suppose that k = 1, and we are interested in forecasting x_{T+1}.

Let us assume that x_{T+1} is a parameter, since it is a quantity which we wish to estimate. Let

$$\underline{x}'_T = (x_1, x_2, \ldots, x_T);\ \underline{a}'_P = (a_1, a_2, \ldots, a_P).$$

Therefore, $E(x_{T+1}|\underline{a}_P, \underline{x}'_T) = a_1 x_T + a_2 x_{T-1} + \ldots + a_P\, x_{T-P+1}$ since $x_{T+1} = \varepsilon_{T+1} + a_1 x_T + \ldots + a_P\, x_{T-P+1}$.

In order to minimize the expected loss function $L(\hat{x}_{T+1}, x_{T+1}) = (\hat{x}_{T+1} - x_{T+1})^2$ it is necessary to use the Bayes estimate of $\underline{a}$ (with squared error loss function), $E(\underline{a}_P|\underline{x}_T)$, since the Bayes estimate of x_{T+1} is

$$E(x_{T+1}|\underline{x}_T) = \int_{x_{T+1}} x_{T+1}\, f(x_{T+1}|\underline{x}_T)\, dx_{T+1}$$

$$= \frac{\int_{\underline{a}_P} \int_{x_{T+1}} x_{T+1}\, f(x_{T+1}, \underline{x}_T, \underline{a}_P)\, dx_{T+1}\, d\underline{a}_P}{\int_{\underline{a}_P} \int_{x_{T+1}} f(x_{T+1}, \underline{x}_T, \underline{a}_P)\, dx_{T+1}\, d\underline{a}_P}$$

$$= \frac{\int_{\underline{a}_P} \int_{x_{T+1}} x_{T+1} f(x_{T+1}, \underline{x}_T, \underline{a}_P)\, dx_{T+1}\, d\underline{a}_P}{g(\underline{x}_T)}$$

$$= \frac{\int_{\underline{a}_P} \int_{x_{T+1}} x_{T+1} f(x_{T+1} | \underline{x}_T, \underline{a}_P)\, g(\underline{x}_T, \underline{a}_P)\, dx_{T+1}\, d\underline{a}_P}{g(\underline{x}_T)}$$

$$= \int_{\underline{a}_P} \left[\int_{x_{T+1}} x_{T+1} f(x_{T+1} | \underline{x}_T, \underline{a}_P)\, g(\underline{a}_P | \underline{x}_T)\, dx_{T+1} \right] d\underline{a}_P$$

$$= \int_{\underline{a}_P} g(\underline{a}_P | \underline{x}_T) \cdot E(x_{T+1} | \underline{x}_T, \underline{a}_P)\, d\underline{a}_P$$

$$= \int_{\underline{a}_P} g(\underline{a}_P | \underline{x}_T) \cdot \underline{a}'_P\, \underline{x}^*_P\, d\underline{a}_P \text{ where } \underline{x}^*_P = (x_T, x_{T-1}, \dots x_{T-P+1})'$$

Therefore $E(x_{T+1} | \underline{x}_T) = E(\underline{a}_P | \underline{x}_T)' \cdot \underline{x}^*_P$. Thus we need to find the empirical Bayes estimate of $\underline{a}_P$.

4.2 Empirical Bayes Estimates of the AR Parameters

For P = 1, first order AR model $x_t = ax_{t-1} + \varepsilon_t$, Durbin (8) showed that $\hat{a} \sim N(a, \frac{1-a^2}{n})$ approximately, where $\hat{a}$ is the M.L.E. of a.

Launer (13) showed that

$$E(a | \underline{x}_T) = E(a | \hat{a}, u) \neq E(a | \hat{a})$$

where $u = \sum_{j=2}^{T} x^2_{j-1}$ and is a random variable. He proposed using the marginal empirical Bayes estimator which was found by Clemmer and

Krutchkoff (5). In order to obtain the empirical Bayes estimate of a, we seek a form of the Bayes estimator for a that is useful in the empirical Bayes approach.

Recall,

$$\hat{a} = \frac{\sum_{t=2}^{T} x_t x_{t-1}}{\sum_{t=2}^{T} x^2_{t-1}} = \frac{\sum_{t=2}^{T-1} x_t x_{t-1}}{\sum_{t=2}^{T} x^2_{t-1}} + \frac{x_{T-1} x_T}{\sum_{t=2}^{T} x^2_{t-1}}$$

Note: This dissertation concerns the empirical Bayes procedure based on a time series generated by a stationary, Gaussian process i.e. $\varepsilon_t \sim N(0, \sigma^2)$ and σ^2 is known.

Thus

$$E(\hat{a} \mid a, \underline{x}_{T-1}) = \frac{c}{u} + \frac{x^2_{T-1}}{u} a$$

where

$$c = \sum_{t=2}^{T-1} x_t x_{t-1}, \quad u = \sum_{t=2}^{T} x^2_{t-1}$$

$$\underline{x}_{T-1} = (x_1, x_2, \ldots, x_{T-1})$$

Therefore,

$$f(\hat{a} \mid a, \underline{x}_{T-1}) = \frac{u}{\sqrt{2\pi}\sigma x_{T-1}} \exp \left\{- \frac{u^2}{2\sigma^2 x^2_{T-1}} \left(\hat{a} - \frac{c}{u} - \frac{x^2_{T-1}}{u} a\right)^2\right\}$$

or

$$(\hat{a}|a, \underline{x}_{T-1}) \sim N\left[(\frac{c}{u} + \frac{x^2_{T-1}}{u} a), \frac{x^2_{T-1}}{u^2} \sigma^2\right]$$

It is easily verified that

$$\frac{\frac{d}{d\hat{a}} f(\hat{a}|a, \underline{x}_{T-1})}{f(\hat{a}|a, \underline{x}_{T-1})} = -\frac{u^2}{\sigma^2 x^2_{T-1}} (\hat{a} - \frac{c}{u} - \frac{x^2_{T-1}}{u} a)$$

Thus

$$a = \frac{\sigma^2}{u} \cdot \frac{\frac{d}{d\hat{a}} f(\hat{a}|a, \underline{x}_{T-1})}{f(\hat{a}|a, \underline{x}_{T-1})} + \frac{u\hat{a}}{x^2_{T-1}} - \frac{c}{x^2_{T-1}}$$

By taking expectation, we get

$$E(a|\hat{a}, \underline{x}_{T-1}) = \frac{u\hat{a}}{x^2_{T-1}} - \frac{c}{x^2_{T-1}} +$$

$$\frac{\sigma^2}{u} \int_a \frac{\frac{d}{da} f(\hat{a}|a, \underline{x}_{T-1})}{f(\hat{a}|a, \underline{x}_{T-1})} dG\ (a|\hat{a}, \underline{x}_{T-1})$$

By using the law of conditional probability, we have

$$\int_a \frac{\frac{d}{da} f(\hat{a}|a, \underline{x}_{T-1})}{f(\hat{a}|a, \underline{x}_{T-1})} dG\ (a|\hat{a}, \underline{x}_{T-1}) =$$

$$\int_a \frac{\frac{d}{d\hat{a}} f(\hat{a}|a, \underline{x}_{T-1})}{f(\hat{a}|a, \underline{x}_{T-1})} \left[\frac{f(\hat{a}|a, \underline{x}_{T-1})\ g\ (a, \underline{x}_{T-1})\, da}{\int_a f(\hat{a}, \underline{x}_{T-1}\ |a)\ dG(a)}\right]$$

$$= \frac{\int_a \frac{d}{d\hat{a}} f(\hat{a} | a, \underline{x}_{T-1}) \; g \; (a, \underline{x}_{T-1}) da}{\int_a f(\hat{a}, \underline{x}_{T-1} | a) \; dG \; (a)}$$

By interchanging the order of integration and differentiation,

$$= \frac{\frac{d}{d\hat{a}} \int_a f(\hat{a} | a, \underline{x}_{T-1}) \; g \; (a, \underline{x}_{T-1}) \, da}{\int f(\hat{a}, \underline{x}_{T-1} | a) \; dG \; (a)}$$

$$= \frac{\frac{d}{d\hat{a}} \int_a f(\hat{a}, a, \underline{x}_{T-1}) \, da}{f(\hat{a}, \underline{x}_{T-1})} = \frac{\frac{d}{d\hat{a}} f(\hat{a}, \underline{x}_{T-1})}{f(\hat{a}, \underline{x}_{T-1})}$$

Therefore, we get

$$E(a | \hat{a}, \underline{x}_{T-1}) = \frac{u\hat{a}}{x^2_{T-1}} - \frac{c}{x^2_{T-1}} + \frac{\frac{d}{d\hat{a}} f(\hat{a}, \underline{x}_{T-1})}{f(\hat{a}, \underline{x}_{T-1})} \cdot \frac{\sigma^2}{u} \qquad (4.2.1)$$

We obtain empirical Bayes estimator corresponding to (4.2.1) by substituting consistent estimators for $f(\hat{a}, \underline{x}_{T-1})$ and $\frac{d}{d\hat{a}} f(\hat{a}, \underline{x}_{T-1})$.

Consistent estimators for multivariate densities were provided by Cacoullos (4).

4.3 Forecasting Several Steps Ahead from an Estimated Model

The forecasting theory developed in time series was based on the assumption that the coefficients in a model were given. In practice, of course, these quantities must be estimated from sample data and coefficient estimates substituted into the corresponding formulas.

To see the effect of this on forecasting more than one step ahead, consider the first-order AR model:

$$x_t = ax_{t-1} + \varepsilon_t$$

Given $x_1, \ldots, x_T$, then the optimal linear forecast of x_{T+k} is $a^k x_T$. The usual procedure is to substitute $\hat{a}$ for a in $(a^k x_T)$. Unfortunately, $\hat{a}^k$ is a biased estimator of a^k, for $k > 1$. For example $k = 2$, an asymptotically unbiased estimator is given by $\hat{a^2} = \hat{a}^2 + \hat{\sigma}^2_{\hat{a}}$ where $\hat{\sigma}^2_{\hat{a}}$ is the unbiased estimated variance of $\hat{a}$.

4.3.1 Forecasting Several Steps Ahead for a First Order AR Using Empirical Bayes Estimate of the Coefficient

Let $x_t = ax_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim N(0, \sigma^2)$. The problem is that, having observed $\underline{x}'_T = (x_1, x_2, \ldots, x_T)$ one wishes to predict x_{T+k}.

It is easily shown that the predictor $\hat{x}_{T+k}$ of x_{T+k} is obtained by appealing directly to the well-known result in Bayesian statistics that, for a quadratic loss function, the mean of x_{T+k} given x_T, i.e. $a^k x_T$, evaluated by the posterior density $g\,(a|\underline{x}_T)$ is optimal. Therefore $\hat{x}_{T+k} = [E(a^k|\underline{x}_T)]\; x_T$.

In order to obtain $\hat{x}_{T+k}$, one needs to obtain an empirical Bayes estimate of a^k. For $k = 2$, $\hat{x}_{T+2} = E(a^2)\; x_T$. To evaluate the second moments, we note that

$$\frac{d^2}{d\hat{a}^2} f(\hat{a}|a, \underline{x}_{T-1}) = \left(\frac{u}{\sigma^2}\right)^2 \left[a - \frac{u}{x^2_{T-1}}\hat{a} + \frac{c}{x^2_{T-1}}\right]^2 f(\hat{a}|a, \underline{x}_{T-1})$$

$$- \left(\frac{u}{\sigma^2}\right) \left(\frac{u}{x^2_{T-1}}\right) f(\hat{a}|a, \underline{x}_{T-1})$$

or

$$\frac{\frac{d^2}{d\hat{a}^2} f(\hat{a}|a, \underline{x}_{T-1})}{f(\hat{a}|a, \underline{x}_{T-1})} = \left(\frac{u}{\sigma^2}\right)^2 \left[a - \frac{u}{x^2_{T-1}} \hat{a} + \frac{c}{x^2_{T-1}}\right]^2$$

$$- \left(\frac{u}{\sigma^2}\right) \left(\frac{u}{x^2_{T-1}}\right)$$

where

$$u = \sum_{t=2}^{T} x^2_{t-1}, \quad c = \sum_{t=2}^{T-1} x_{t-1} x_t$$

Thus

$$\frac{\frac{d^2}{d\hat{a}^2} f(\hat{a}|a, \underline{x}_{T-1})}{f(\hat{a}|a, \underline{x}_{T-1})} = \left[a^2 + \left(\frac{c}{x^2_{T-1}}\right)^2 + \left(\frac{u}{x^2_{T-1}}\right)^2 \hat{a}^2 + 2 \frac{ac}{x^2_{T-1}}\right.$$

$$\left. - 2 \frac{au}{x^2_{T-1}} \hat{a} - 2 \left(\frac{u}{x^2_{T-1}}\right) \left(\frac{c}{x^2_{T-1}}\right) \hat{a}\right] - \left(\frac{u}{\sigma^2}\right) \left(\frac{u}{x^2_{T-1}}\right)$$

Therefore

$$E(a^2|\hat{a}, \underline{x}_{T-1}) = \left(\frac{\sigma^2}{u}\right) \left(\frac{u}{x^2_{T-1}}\right) +$$

$$2 \left(\frac{u}{x^2_{T-1}}\right) \left(\frac{c}{x^2_{T-1}}\right) \hat{a} + 2 \left(\frac{u}{x^2_{T-1}}\right) \hat{a}\, E(a|\hat{a}, \underline{x}_{T-1}) - \frac{2c}{x^2_{T-1}} E(a|\hat{a}, \underline{x}_{T-1})$$

$$- \left(\frac{u}{x^2_{T-1}}\right)^2 \hat{a}^2 - \left(\frac{c}{x^2_{T-1}}\right)^2 + \left(\frac{\sigma^2}{u}\right)^2 \int \frac{\frac{d^2}{d\hat{a}^2} f(\hat{a} | a, \underline{x}_{T-1})}{f(\hat{a} | a, \underline{x}_{T-1})} dG\,(a | \hat{a}, \underline{x}_{T-1}) \tag{4.3.1.1}$$

In order to obtain $E(a^2 | \hat{a}, \underline{x}_{T-1})$ one needs to evaluate

$$\int \frac{\frac{d^2}{d\hat{a}^2} f(\hat{a} | a, \underline{x}_{T-1})}{f(\hat{a} | a, \underline{x}_{T-1})} dG\,(a | \hat{a}, \underline{x}_{T-1}) =$$

$$\int \frac{\frac{d^2}{d\hat{a}^2} f(\hat{a} | a, \underline{x}_{T-1})}{f(\hat{a} | a, \underline{x}_{T-1})} \cdot \frac{f(\hat{a} | a, \underline{x}_{T-1})\, g(a, \underline{x}_{T-1})\, da}{\int f(\hat{a}, \underline{x}_{T-1} | a)\, dG\,(a)}$$

$$= \frac{\frac{d^2}{d\hat{a}^2} \int_a f(\hat{a} | a, \underline{x}_{T-1})\, g(a, \underline{x}_{T-1})\, da}{\int f(\hat{a}, \underline{x}_{T-1} | a)\, dG\,(a)} = \frac{\frac{d^2}{d\hat{a}^2} \int_a f(\hat{a}, a, \underline{x}_{T-1})\, da}{f(\hat{a}, \underline{x}_{T-1})}$$

$$= \frac{\frac{d^2}{d\hat{a}^2} f(\hat{a}, \underline{x}_{T-1})}{f(\hat{a}, \underline{x}_{T-1})}$$

where $f(\hat{a}, \underline{x}_{T-1}) = \int_a f(\hat{a}, \underline{x}_{T-1} | a)\, dG\,(a)$. It can also be written as

$$f(\hat{a}, \underline{x}_{T-1}) = f(\hat{a} | \underline{x}_{T-1}) \cdot h(\underline{x}_{T-1})$$

Therefore, we can obtain an empirical Bayes estimator corresponding to (4.3.1.1) by substituting consistent estimators for $f(\hat{a}, \underline{x}_{T-1})$ and $\frac{d^2}{d\hat{a}^2} f(\hat{a}, \underline{x}_{T-1})$. For k = 3

$$\frac{d^3}{d\hat{a}^3} f(\hat{a} | a, \underline{x}_{T-1}) = 2\left(\frac{u}{\sigma^2}\right)^2 \left(\frac{-u}{x^2_{T-1}}\right) \left(a - \frac{u}{x^2_{T-1}} \hat{a} + \frac{c}{x^2_{T-1}}\right)$$

$$\cdot\ f(\hat{a}|a, \underline{x}_{T-1}) + \left(\frac{u}{\sigma^2}\right)^3 \left(a - \frac{u}{x^2_{T-1}}\hat{a} + \frac{c}{x^2_{T-1}}\right)^3 f(\hat{a}|a, \underline{x}_{T-1})$$

$$- \left(\frac{u}{\sigma^2}\right) \left(\frac{u}{x^2_{T-1}}\right) \left(a - \frac{u}{x^2_{T-1}}\hat{a} + \frac{c}{x^2_{T-1}}\right)$$

$$= \left(\frac{u}{\sigma^2}\right)^3 T^{*3} f(\hat{a}|a, \underline{x}_{T-1}) - 3 \left(\frac{u}{\sigma^2}\right)^2 \left(\frac{u}{x^2_{T-1}}\right) T^* f(\hat{a}|a, \underline{x}_{T-1}) \tag{4.3.1.2}$$

where

$$T^* = \left(a - \frac{u}{x^2_{T-1}}\hat{a} + \frac{c}{x^2_{T-1}}\right)$$

Thus, we can get $E(a^3|\hat{a}, \underline{x}_{T-1})$, by using (4.3.1.2). For k = 4

$$\frac{d^4}{d\hat{a}^4} f(\hat{a}|a, \underline{x}_{T-1}) = \left(\frac{u}{\sigma^2}\right)^4 T^{*4} f(\hat{a}|a, \underline{x}_{T-1})$$

$$- 6 \left(\frac{u}{\sigma^2}\right)^3 \left(\frac{u}{x^2_{T-1}}\right) T^{*2} f(\hat{a}|a, \underline{x}_{T-1}) + 3 \left(\frac{u}{\sigma^2}\right) \left(\frac{u}{x^2_{T-1}}\right) f(\hat{a}|a, \underline{x}_{T-1}) .$$

For k = 5

$$\frac{d^5}{d\hat{a}^5} f(\hat{a}|a, \underline{x}_{T-1}) = \left(\frac{u}{\sigma^2}\right)^5 T^{*5} f(\hat{a}|a, \underline{x}_{T-1})$$

$$- 10 \left(\frac{u}{\sigma^2}\right)^4 \left(\frac{u}{x^2_{T-1}}\right) T^{*3} f(\hat{a}|a, \underline{x}_{T-1}) + 15 \left(\frac{u}{\sigma^2}\right) \left(\frac{u}{x^2_{T-1}}\right)^2$$

$$\cdot\ T^* f(\hat{a}|a, \underline{x}_{T-1})$$

4.4 Empirical Bayes Estimates of AR(2) Parameters

For second order AR model $x_t = \varepsilon_t + a_1 x_{t-1} + a_2 x_{t-2}$, Box and Jenkins (2) have found approximate maximum likelihood estimates of the AR parameters, and $\hat{a}_i$, satisfying the following equations:

$$D_{12} = \hat{a}_1 D_{22} + \hat{a}_2 D_{23} + \dots + \hat{a}_P D_{2,P+1}$$

$$D_{13} = \hat{a}_1 D_{23} + \hat{a}_2 D_{33} + \dots + \hat{a}_P D_{3,P+1}$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

$$D_{1,P+1} = \hat{a}_1 D_{2,P+1} + \hat{a}_2 D_{3,P+1} + \dots + \hat{a}_P D_{P+1,P+1}$$

which, in a matrix notation, can be written

$$\underline{d} = D_P \underline{\hat{a}}$$

where

$$\underline{\hat{a}}' = (\hat{a}_1, \hat{a}_2, \dots, \hat{a}_P); \; \underline{d}' = (D_{12}, D_{13}, \dots, D_{1,P+1})$$

$$D_P = \begin{pmatrix} D_{22} & D_{23} & \cdots & D_{2,P+1} \\ D_{23} & D_{33} & \cdots & D_{3,P+1} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ D_{2,P+1} & D_{3,P+1} & \cdots & D_{P+1,P+1} \end{pmatrix}$$

where

$$D_{ij} = D_{ji} = x_i\, x_j + x_{i+1}x_{j+1} + \ldots + x_{T+1-j}\, x_{T+1-i}$$

Therefore $\underline{\hat{a}} = D_P^{-1}\, \underline{d}$. For P = 2

$$\underline{\hat{a}} = D_2^{-1}\, \underline{d} = \begin{pmatrix} D_{22} & D_{23} \\ D_{23} & D_{33} \end{pmatrix}^{-1} \begin{pmatrix} D_{12} \\ D_{13} \end{pmatrix}$$

where

$$D_{22} = \sum_{t=3}^{T} x^2{}_{t-1} \qquad D_{33} = \sum_{t=5}^{T} x^2{}_{t-2}$$

$$D_{23} = x_2x_3 + x_3x_4 + \ldots + x_{T-2}\, x_{T-1}$$

$$D_{12} = x_1x_2 + x_2x_3 + \ldots + x_{T-1}\, x_T$$

$$D_{13} = x_1x_3 + x_2x_4 + \ldots + x_{T-2}\, x_T$$

Let $\underline{x}^*{}_{T-2} = (x_2, x_3, \ldots, x_{T-2}, x_{T-1})$. Therefore

$$E(\underline{\hat{a}}|\underline{a}, \underline{x}^*{}_{T-2}) = D_2^{-1}\, E(\underline{d}|\underline{a}, \underline{x}^*{}_{T-2})$$

$$E(\underline{d}|\underline{a}, \underline{x}^*{}_{T-2}) = \begin{pmatrix} c_1 + x_2\, E(x_1) + x_{T-1}\, E(x_T|\underline{a}, \underline{x}^*{}_{T-2}) \\ c_2 + x_3\, E(x_1) + x_{T-2}\, E(x_T|\underline{a}, \underline{x}^*{}_{T-2}) \end{pmatrix} \qquad (4.4.1)$$

where

$$c_1 = x_2x_3 + x_3x_4 + \ldots + x_{T-2}\, x_{T-1}$$

$$c_2 = x_2x_4 + x_3x_5 + \dots + x_{T-3}\, x_{T-1}$$

Let us suppose $E(x_1) = \mu_1$. Recall $x_T = a_1\, x_{T-1} + a_2\, x_{T-2} + \varepsilon_t$. Thus,

$$E(x_T | \underline{a}, \underline{x}^*_{T-2}) = a_1\, x_{T-1} + a_2\, x_{T-2}$$

By substituting Ex_1 and $E(x_T | \underline{a}, \underline{x}^*_{T-2})$ in (4.4.1) we get:

$$E(\underline{d} | \underline{a}, \underline{x}^*_{T-2}) = \begin{pmatrix} c_{11} + x_{T-1}\, (x_{T-1}, x_{T-2})\, \underline{a} \\ c_{22} + x_{T-2}\, (x_{T-1}, x_{T-2})\, \underline{a} \end{pmatrix}$$

where c_{11} and c_{22} are fixed

$$c_{11} = c_1 + x_2\, \mu_1$$

$$c_{22} = c_2 + x_3\, \mu_1$$

Thus

$$(\underline{\hat{a}} | \underline{a}, \underline{x}^*_{T-2}) \sim N\, (D_2^{-1} \cdot E(\underline{d} | \underline{a}, \underline{x}^*_{T-2}),\, V)$$

where

$$V = D_2^{-1} \begin{pmatrix} (x_2^{\,2} + x^2_{T-1})\, \sigma^2 & (x_2x_3 + x_{T-2}\, x_{T-1})\, \sigma^2 \\ (x_2x_3 + x_{T-2}\, x_{T-1})\, \sigma^2 & (x_3^{\,2} + x^2_{T-2})\, \sigma^2 \end{pmatrix} D_2^{-1}$$

$$= \sigma^2\, D_2^{-1} \begin{pmatrix} (x_2^{\,2} + x^2_{T-1}) & (x_2x_3 + x_{T-2}\, x_{T-1}) \\ (x_2x_3 + x_{T-2}\, x_{T-1}) & (x_3^{\,2} + x^2_{T-2}) \end{pmatrix} D_2^{-1}$$

$$= \sigma^2\, D_2^{-1}\, F\, D_2^{-1}$$

$$f(\underline{\hat{a}}|\underline{a}, \underline{x}^*_{T-2}) \propto \exp - 1/2\ [\underline{\hat{a}} - E(\underline{\hat{a}}|\underline{a}, \underline{x}^*_{T-2})]'\ V^{-1}\ [\underline{\hat{a}} - E(\underline{\hat{a}}|\underline{a}, \underline{x}^*_{T-2})]$$

Therefore

$$\log f(\underline{\hat{a}}|\underline{a}, \underline{x}^*_{T-2}) \propto - 1/2\ [\underline{\hat{a}} - E(\underline{\hat{a}}|\underline{a}, \underline{x}^*_{T-2})]'\ V^{-1}\ [\underline{\hat{a}} - E(\underline{\hat{a}}|\underline{a}, \underline{x}^*_{T-2})]$$

$$\frac{\partial \log f(\underline{\hat{a}}|\underline{a}, \underline{x}^*_{T-2})}{\partial \underline{\hat{a}}} = -V^{-1}\ [\underline{\hat{a}} - E(\underline{\hat{a}}|\underline{a}, \underline{x}^*_{T-2})]$$

$$\underline{\hat{a}} + V\ \frac{\partial \log f(\underline{\hat{a}}|\underline{a}, \underline{x}^*_{T-2})}{\partial \underline{\hat{a}}} = D_2^{-1}\ E(\underline{d}|\underline{a}, \underline{x}^*_{T-2})$$

$$D_2\underline{\hat{a}} + D_2 V\ \frac{\partial \log f(\underline{\hat{a}}|\underline{a}, \underline{x}^*_{T-2})}{\partial \underline{\hat{a}}} = E(\underline{d}|\underline{a}, \underline{x}^*_{T-2})$$

$$= \begin{pmatrix} c_{11} + (x^2_{T-1} & x_{T-1}\ x_{T-2})\ \underline{a} \\ c_{22} + (x_{T-1}\ x_{T-2} & x^2_{T-2})\ \underline{a} \end{pmatrix}$$

$$= \begin{pmatrix} c_{11} \\ c_{22} \end{pmatrix} + \begin{pmatrix} x^2_{T-1} & x_{T-1}\ x_{T-2} \\ x_{T-1}\ x_{T-2} & x^2_{T-2} \end{pmatrix} \underline{a}$$

Thus

$$B\underline{a} = D_2 \,\hat{\underline{a}} - \begin{pmatrix} c_{11} \\ c_{22} \end{pmatrix} + \sigma^2 D_2 D_2^{-1} F D_2^{-1} \frac{\partial \log f(\hat{\underline{a}} | \underline{a}, \underline{x}^*_{T-2})}{\partial \hat{\underline{a}}}$$

$$= D_2 \,\hat{\underline{a}} - \begin{pmatrix} c_{11} \\ c_{22} \end{pmatrix} + \sigma^2 F D_2^{-1} \frac{\partial \log f(\hat{\underline{a}} | \underline{a}, \underline{x}^*_{T-2})}{\partial \hat{\underline{a}}}$$

where

$$B = \begin{pmatrix} x^2_{T-1} & x_{T-1}\, x_{T-2} \\ x_{T-1}\, x_{T-2} & x^2_{T-2} \end{pmatrix} \qquad \text{is singular.}$$

$$F = \begin{pmatrix} (x^2_2 + x^2_{T-1}) & (x_2 x_3 + x_{T-2}\, x_{T-1}) \\ (x_2 x_3 + x_{T-2}\, x_{T-1}) & (x^2_3 + x^2_{T-2}) \end{pmatrix}$$

Therefore

$$E\ [(B\underline{a}) | \hat{\underline{a}}, \underline{x}^*_{T-2}] = D_2 \,\hat{\underline{a}} - \begin{pmatrix} c_{11} \\ c_{22} \end{pmatrix} + \sigma^2 F D_2^{-1}$$

$$\cdot \int \frac{\frac{\partial}{\partial \hat{\underline{a}}} f(\hat{\underline{a}} | \underline{a}, \underline{x}^*_{T-2})}{f(\hat{\underline{a}} | \underline{a}, \underline{x}^*_{T-2})}\, dG\ (\underline{a} | \hat{\underline{a}}, \underline{x}^*_{T-2})$$

$$= D_2 \hat{\underline{a}} - \begin{pmatrix} c_{11} \\ c_{22} \end{pmatrix} + \sigma^2 F D_2^{-1} \int \frac{\frac{\partial}{\partial \hat{\underline{a}}} f(\hat{\underline{a}} | \underline{a}, \underline{x}^*_{T-2})}{f(\hat{\underline{a}} | \underline{a}, \underline{x}^*_{T-2})}$$

$$\cdot \frac{f(\underline{\hat{a}} | \underline{a}, \underline{x}^*_{T-2})\ g(\underline{a}, \underline{x}^*_{T-2})\ d\underline{a}}{\int f(\underline{\hat{a}}, \underline{x}^*_{T-2} | a)\ dG\ (a)}$$

$$= D_2\underline{\hat{a}} - \begin{pmatrix} c_{11} \\ \\ c_{22} \end{pmatrix} + \sigma^2\ F\ D_2^{-1}\ \frac{\frac{\partial}{\partial \underline{\hat{a}}} \int f(\underline{\hat{a}} | \underline{a}, \underline{x}^*_{T-2})\ g(\underline{a}, \underline{x}^*_{T-2})\ d\underline{a}}{\int f(\underline{\hat{a}}, \underline{x}^*_{T-2} | \underline{a})\ dG\ (\underline{a})}$$

$$= D_2\underline{\hat{a}} - \begin{pmatrix} c_{11} \\ \\ c_{22} \end{pmatrix} + \sigma^2\ F\ D_2^{-1}\ \frac{\frac{\partial}{\partial \underline{\hat{a}}} \int f(\underline{\hat{a}}, \underline{a}, \underline{x}^*_{T-2})\ d\underline{a}}{\int f(\underline{\hat{a}}, \underline{x}^*_{T-2} | \underline{a})\ dG\ (\underline{a})}$$

$$= D_2\underline{\hat{a}} - \begin{pmatrix} c_{11} \\ \\ c_{22} \end{pmatrix} + \sigma^2\ F\ D_2^{-1}\ \frac{\frac{\partial}{\partial \underline{\hat{a}}}\ f(\underline{\hat{a}}, \underline{x}^*_{T-2})}{f(\underline{\hat{a}}, \underline{x}^*_{T-2})}$$

where

$$f(\underline{\hat{a}}, \underline{x}^*_{T-2}) = \int f(\underline{\hat{a}}, \underline{x}^*_{T-2} | \underline{a})\ dG(\underline{a})$$

By substituting consistent estimators for $f(\underline{\hat{a}}, \underline{x}^*_{T-2})$ and $\frac{\partial}{\partial \underline{\hat{a}}} f(\underline{\hat{a}}, \underline{x}^*_{T-2})$, we can get empirical Bayes estimator for $B\underline{a}$ i.e. $E(B\underline{a} | \underline{\hat{a}}, \underline{x}^*_{T-2})$.

CHAPTER V

5.1 Empirical Bayes Estimation of the Spectral Density

In section 2.5, we have found that the estimate of f(w) is approximately distributed as gamma.

Let us suppose that the estimation problem has occurred N-1 times previously.

Let us define $A_j(w)$ to be the estimate of spectral density for the j^{th} experiment (time series), $1 \leq j \leq N$. If $A_{j,i}(w)$ represents the estimate in the i^{th} subinterval of j^{th} time series and if $A_j(w) = \sum_{i=-(k+1)}^{k} A_{j,i}(w)$ then $A_{j,i}(w)$ belongs to F_3 (see section 3.2.2), $i = -(k+1), -k, -k+1, \ldots, k-1, k$; $j = 1, 2, \ldots, N$. Since (1) $\underline{A}_j = (A_{j,-(k+1)}(w), A_{j,-k}(w), \ldots, A_{j,k}(w))$ is a vector of size (2k+1) of independent and identically distributed continuous random variables, (2) There is a statistic $A_N(w)$ ($= \sum_{i=-(k+1}^{k} A_{N,i}(w)$) sufficient for f(w), with a density of the form

$$f_{(2k+1)}(A_N(w)\,|\,f(w)) \doteq \frac{1}{(2k)!2^{2k+1}} \left(\frac{A_N(w)(4k+2)}{f(w)}\right)^{2k}$$

$$\cdot \exp\left[-1/2 \left(\frac{A_N(w)}{f(w)}\right) (4k+2)\right]$$

Therefore

$$h(2k+1) = \frac{(2k+1)^{2k}}{(2k!)2}$$

$$q(A_N(w),\ f(w)) = \exp\left[- \frac{A_N(w)}{f(w)} (2k+1)\right]$$

$$h(2k) = \frac{(2k)}{(2k-1)!2}$$

Thus

$$E(f(w)\,|A_N(w)) = A_N(w) \cdot \frac{h(2k+1)}{h(2k)} \cdot \frac{f_{2k}(A_N(w))}{f_{2k+1}(A_N(w))}$$

where $f_{2k+1}(A_N(w))$ is the marginal density of the sufficient statistic since

$$\frac{h(2k+1)}{h(2k)} = (1 + \frac{1}{2k})^{2k}$$

then

$$E(f(w)\,|A_N(w)) = A_N(w)\,(1 + \frac{1}{2k})^{2k} \cdot \frac{f_{2k}(A_N(w))}{f_{2k+1}(A_N(w))}$$

Therefore, we obtain empirical Bayes estimator, by substituting consistent estimators for $f_{2k}(A_N(w))$ and $f_{2k+1}(A_N(w))$.

Let $f_{N,(2k+1)}(A_N(w)) = f_{2k+1}(A_N(w))$ then

$$f_{N,(2k+1)}(A_N(w)) = \frac{1}{2\pi Nh} \sum_{i=1}^{N} [\sin(f_i)/f_i]^2$$

where

$$\dot{f}_i = [A_N(w) - A_i(w)]/2h$$

$$h = N^{-1/5} [\sum_{i=1}^{N} (A_i(w) - \overline{A}(w))^2/N]^{1/2}$$

$$\overline{A}(w) = \frac{1}{N} \sum_{i=1}^{N} A_i(w)$$; see Rencher (17).

5.2 The ARMA Parameterization

The following assumption will be made about all time series x_t mentioned in this dissertation.

Assumption. The joint distribution of $x_1, \ldots, x_T$ is a multivariate normal distribution with mean zero and variance-covariance matrix Σ.

One way to analyze time series data is to assume a parameterization known as the autoregressive-moving average (ARMA) parameterization. Recall, a time series process is an ARMA (p,q) if it satisfies the following equation.

$$x_t + \sum_{j=1}^{P} \alpha_j x_{t-j} = \varepsilon_t + \sum_{j=1}^{q} \beta_j \varepsilon_{t-j} \qquad (5.2.1)$$

where ε_t is a sequence of uncorrelated and normally distributed random variables with mean zero and variance σ^2. Special cases of the ARMA (p,q) are moving average mA(q) = ARMA (0,q) and the autoregressive AR(P) = ARMA (P,0). The parameters in (5.2.1) are $\alpha_1, \alpha_2, \ldots, \alpha_P$, $\beta_1, \ldots, \beta_q$ since we are assuming σ^2 is known.

Let us denote the parameter vector by

$$\underline{m} = (\alpha_1, \alpha_2, \ldots, \alpha_P, \beta_1, \beta_2, \ldots, \beta_q)'$$

Therefore, we can write the likelihood function as follows:

$$L(\underline{x}_T | \underline{m}) = (2\pi)^{-T/2} \; |\Sigma|^{-1/2} \exp\{-\tfrac{1}{2} \underline{x}'_T \Sigma^{-1} \underline{x}_T\} \qquad (5.2.2)$$

One can use the approximate likelihood which is due to Whittle (21), however we are using a different but equivalent approach.

Let $f(w)$ be a spectral density, which for convenience will now be assumed to be defined over $(0, 2\pi)$ rather than $(-\pi, \pi)$. Now, $f(w)$ is symmetric about π. Define

$$\sigma(t) = \int_{-\pi}^{\pi} f(w)\, e^{iwt}\, dw \tag{5.2.3}$$

The function σ is the covariance function corresponding to the spectral distribution function

$$F(w) = \int_{0}^{w} f(\lambda)\, d\lambda$$

When T is large, the integral (5.2.3) could be approximated by the following formula:

$$\sigma(t) \doteq \frac{2\pi}{T} \sum_{k=1}^{T} f(w_k) e^{iw_k t} \tag{5.2.4}$$

where

$$w_k = \frac{2\pi k}{T}$$

Let us suppose that x_t is a stationary Gaussian process with covariance function σ given by (5.2.3).

Note: In the following paragraph the likelihood for the $\underline{x}_T = (x_1, x_2, \ldots, x_T)'$ is computed approximately by using the covariance function σ given by (5.2.4).

Let A be a T x T matrix:

$$A = \begin{bmatrix} 0 & 1 & 0 & . & . & . & 0 \\ 0 & 0 & 1 & . & . & . & 0 \\ 0 & 0 & 0 & . & . & . & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & . & 0 \\ 0 & 0 & 0 & . & . & . & 1 \\ 1 & 0 & 0 & . & . & . & 0 \end{bmatrix}$$

Define the vectors

$$u_k = [1, e^{iw_k}, e^{2iw_k}, \ldots, e^{Tiw_k}]' \qquad k = 1, 2, \ldots, T$$

and the matrix

$$M = \frac{1}{\sqrt{T}} \; [u_1, u_2, \ldots, u_T]$$

Let D be a diagonal matrix.

$$D = \text{diag}\,(e^{iw_1}, e^{iw_2}, \ldots, e^{iw_T})$$

Then AM = MD and MM* = M*M = I where M* is the conjugate transpose of M. Thus A = MDM*. Now, the variance-covariance matrix of $\underline{x}_T$ = $(x_1, x_2, \ldots x_T)'$ is

$$\Sigma = \sum_{j=1}^{T} \sigma(j) A^j = \sum_{j=1}^{T} \sigma(j)\, MD^j M^*$$

$$= M \left[\sum_{j=1}^{T} \sigma(j)\, D^{j} \right] M^*$$

The matrix $\sum_{j=1}^{T} \sigma(j)\, D^{j}$ is diagonal, and the k^{th} diagonal element is

$$\sum_{j=1}^{T} \sigma(j)\, e^{ijw_k} = \sum_{j=1}^{T} \sum_{n=1}^{T} \frac{2\pi}{T} f(w_n) e^{iw_n j} \cdot e^{iw_k j}$$

$$= \frac{2\pi}{T} \sum_{n=1}^{T} f(w_n) \sum_{j=1}^{T} e^{ij(w_n + w_k)} = \begin{cases} 2\pi f(w_{T-k}) & \text{if } k \neq T \\ 2\pi f(w_0) & \text{if } k = T \end{cases}$$

$$= 2\pi f(w_k)$$

since f is symmetric about π and exponential function sums to zero unless $n = T - k$ or $n = -k$. Let

$$E = 2\pi \text{ diag } (f(w_1), \ldots, f(w_T))$$

one can find that

$$\underline{x}'_T \Sigma^{-1} \underline{x}_T = \underline{x}'_T ME^{-1} M^* \underline{x}_T$$

$$= \sum_{k=1}^{T} \{\frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t e^{iw_k t}\} \{2\pi f(w_k)\}^{-1} \cdot$$

$$\{\frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t e^{-iw_k t}\} = \sum_{k=1}^{T} \frac{I(w_k)}{f(w_k)}$$

where the periodogram $I(w_k)$ is defined

$$I(w_k) = \frac{1}{2\pi} \left| \sum_{t=1}^{T} x_t \, e^{iw_k t} \right|^2$$

In similar way, one can find that

$$|\Sigma| = \prod_{k=1}^{T} (2\pi f(w_k))$$

Since, $\Sigma = M \left[\sum_{j=1}^{T} \sigma(j) D^j \right] M^*$, the matrix $\sum_{j=1}^{T} \sigma(j) D^j$ is diagonal, and the k^{th} diagonal element is $2\pi f(w_k)$. Therefore,

$$\hat{L}(\underline{x}_T | \underline{m}) = (2\pi)^{-T/2} \exp \left\{-1/2 \left[\sum_{k=1}^{T} \log (2\pi f(w_k) + \sum_{k=1}^{T} I(w_k)/f(w_k) \right]\right\}$$

5.3 The Empirical Bayes Estimate of Moving Average and Autoregressive and Mixed Models Using Spectral Density

The techniques we shall use are based on estimates of the spectral density, and approximate likelihood.

5.3.1 Autoregressive With P Parameters

The spectral density of the Pth order AR time series x_t defined by

$$x_t + \sum_{j=1}^{P} a_j \, x_{t-j} = \varepsilon_t \tag{5.3.1.1}$$

where $a_P \neq 0$, and ε_t are uncorrelated and normally distributed, i.e. $\varepsilon_t \sim N(0, \sigma^2)$, is given by

$$f_x(w) = \frac{\sigma^2}{2\pi} \left[\left(\sum_{j=0}^{P} a_j \, e^{-iwj} \right) \left(\sum_{j=0}^{P} a_j \, e^{iwj} \right) \right]^{-1} \tag{5.3.1.2}$$

where $a_0 = 1$, the proof of this theorem can be seen in Fuller (9). We have shown in section 2.3, that $I(w_k)$ is approximately unbiased but not consistent, therefore, we use better estimator for $f(w_k)$, say, $\hat{f}(w_k)$ where

$$\hat{f}(w_k) = \frac{\hat{\sigma}^2}{2\pi} [(\sum_{j=0}^{P} \hat{a}_j e^{-iw_k j}) (\sum_{j=0}^{P} \hat{a}_j e^{iw_k j})]^{-1} \tag{5.3.1.3}$$

where $\hat{a}_j$, $j = 1, 2, \ldots, P$ are maximum likelihood estimates of a_j, $a_0 = 1$. Recall

$$f(\underline{x}_T|\underline{a}) \doteq (2\pi)^{-T/2} \exp \{-\frac{1}{2} [\sum_{k=1}^{T} \log (2\pi f(w_k)) + \sum_{k=1}^{T} \frac{I(w_k)}{f(w_k)}]\} \tag{5.3.1.4}$$

Thus, by substituting $\hat{f}(w_k)$ for $I(w_k)$ in (5.3.1.4), then we get

$$\hat{f}(\underline{x}_T|\underline{a}) = (2\pi)^{-T/2} \exp \{-\frac{1}{2} [\sum_{k=1}^{T} \log (2\pi f(w_k)) + \sum_{k=1}^{T} \frac{\hat{f}(w_k)}{f(w_k)}]\}$$

Then

$$\log \hat{f}(\underline{x}_T|\underline{a}) = - \frac{T}{2} \log 2\pi - \frac{1}{2} \sum_{k=1}^{T} \log 2\pi + \frac{1}{2} \sum_{k=1}^{T} \log f^{-1}(w_k)$$

$$- \frac{1}{2} \sum_{k=1}^{T} \hat{f}(w_k)/f(w_k)$$

where

$$\underline{a} = (a_1, a_2, \ldots, a_P)'$$

Thus

$$\frac{\partial \log \hat{f}(\underline{x}_T|\underline{a})}{\partial \hat{a}_j} = -\frac{1}{2} \sum_{k=1}^{T} \frac{1}{f(w_k)} \cdot \frac{\partial \hat{f}(w_k)}{\partial \hat{a}_j}$$

$$\frac{\partial \hat{f}(w_k)}{\partial \hat{a}_j} = \frac{\sigma^2}{2\pi} [(\sum_{j=0}^{P} \hat{a}_j e^{-iw_k j}) (\sum_{j=0}^{P} \hat{a}_j e^{iw_k j})]^{-2} \cdot$$

$$[(e^{-iw_k j}) (\sum_{j=0}^{P} \hat{a}_j e^{iw_k j}) + (e^{iw_k j}) (\sum_{j=0}^{P} a_j e^{-iw_k j})]$$

$$= (\frac{2\pi}{\sigma^2}) [\hat{f}(w_k)]^2 (2 \cos (w_k j)) + (\frac{2\pi}{\sigma^2}) [\hat{f}(w_k)]^2 \underline{c}'_j \underline{\hat{a}}$$

where

$$\underline{c}'_j = (2 \cos (w_k(j-1)), \ldots, 2 \cos (w_k(P-j)))$$

$$\underline{\hat{a}} = (\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_P)'; \; 1 \leq j \leq P$$

Let us define C to be a matrix P x P, where the j^{th} row of C is $\underline{c}'_j$, then

$$\frac{\partial \hat{f}(w_k)}{\partial \underline{\hat{a}}} = (\frac{2\pi}{\sigma^2}) [\hat{f}(w_k)]^2 \underline{d} + (\frac{2\pi}{\sigma^2}) [\hat{f}(w_k)]^2 C \underline{\hat{a}} \qquad (5.3.1.5)$$

where

$$\underline{d}' = (2 \cos w_k, \ldots, 2 \cos p\, w_k)$$

$$\frac{\partial \log f(\underline{x}_T | \underline{a})}{\partial \hat{\underline{a}}} = -\frac{1}{2} \sum_{k=1}^{T} \frac{\partial \hat{f}(w_k)}{\partial \hat{\underline{a}}} \cdot \frac{1}{f(w_k)} \tag{5.3.1.6}$$

therefore

$$\frac{\partial \log f(\underline{x}_T | \underline{a})}{\partial \hat{\underline{a}}} = -\frac{1}{2} \sum_{k=1}^{T} \left(\frac{2\pi}{\sigma^2}\right) [\hat{f}(w_k)]^2 \; C \hat{\underline{a}} \left(\frac{1}{f(w_k)}\right)$$

$$-\frac{1}{2} \sum_{k=1}^{T} \left(\frac{2\pi}{\sigma^2}\right) [\hat{f}(w_k)]^2 \; \underline{d} \left(\frac{1}{f(w_k)}\right) \tag{5.3.1.7}$$

The particular parameterization of the spectral density which will be considered is

$$\frac{1}{f(w_k)} = \sum_{j=1}^{n} e_j \cdot y_j(w_k)$$

where $\underline{e} = (e_1, e_2, \ldots, e_n)'$ is the parameter vector. The integer n will be assumed to be fixed and the set $\{y_j;\ j = 1, 2, \ldots, n\}$ will be assumed to consist of known functions.

It can be shown that the reciprocal spectral density of an AR(P) parameterization may be written as

$$\frac{1}{f(w_k)} = \sum_{j=1}^{P+1} e_j \cdot \cos((j-1)w_k) \tag{5.3.1.8}$$

where the relationship between the parameter vector $\underline{e}$ and $\underline{a}$ of the AR(P) is

$$e_j = \begin{cases} \frac{2\pi}{\sigma^2} (1 + \sum_{k=1}^{P} a^2_k); \; j = 1 \\ \frac{4\pi}{\sigma^2} \left(\sum_{k=1}^{P-j+1} a_{k+j-1} a_k - a_{j-1} \right); \; j = 2, 3, \ldots, P \\ - \frac{4\pi}{\sigma^2} a_P; \; j = P + 1 \end{cases} \quad (5.3.1.9)$$

and $y_j(w_k) = \cos[(j-1)w_k]$; $j = 1, \ldots, P+1$

Thus, by substituting $\frac{1}{f(w_k)}$ in (5.3.1.7), we get

$$\frac{\partial \log f(\underline{x}_T|\underline{a})}{\partial \underline{\hat{a}}} = - \frac{1}{2} \sum_{k=1}^{T} (\frac{2\pi}{\sigma^2}) (\hat{f}(w_k))^2 C \, \underline{\hat{a}} \, \underline{h}' \, \underline{e}$$

$$- \frac{1}{2} \sum_{k=1}^{T} (\frac{2\pi}{\sigma^2}) (\hat{f}(w_k))^2 \, \underline{d} \, \underline{h}' \, \underline{e}$$

where $\underline{h}' = (\cos(0), \cos(w_k), \ldots, \cos(Pw_k))$

$$\frac{\partial \log \hat{f}(\underline{x}_T|\underline{a})}{\partial \underline{\hat{a}}} = [- \frac{1}{2} \sum_{k=1}^{T} (\frac{2\pi}{\sigma^2}) (\hat{f}(w_k))^2 \, C \, \underline{\hat{a}} \, \underline{h}'] \, \underline{e}$$

$$- [\frac{1}{2} \sum_{k=1}^{T} (\frac{2\pi}{\sigma^2}) (\hat{f}(w_k))^2 \, \underline{d} \, \underline{h}'] \, \underline{e} = B \, \underline{e}$$

where

$$B = \left\{ \left[-\frac{1}{2} \sum_{k=1}^{T} \left(\frac{2\pi}{\sigma^2}\right) (\hat{f}(w_k))^2 \; C \; \underline{\hat{a}} \; \underline{h}'\right] + \left[-\frac{1}{2} \sum_{k=1}^{T} \left(\frac{2\pi}{\sigma^2}\right) (\hat{f}(w_k))^2 \; \underline{d} \; \underline{h}'\right] \right\}$$

B is P x (P+1) matrix. In addition, we need to find

$$\frac{\partial \log \hat{f}(\underline{x}_T | \underline{a}, \sigma^2)}{\partial \hat{\sigma}^2}$$

$$\frac{\partial \log \hat{f}(\underline{x}_T | \underline{a}, \sigma^2)}{\partial \hat{\sigma}^2} = -\frac{1}{2} \sum_{k=1}^{T} \frac{\partial \hat{f}(w_k)}{\partial \hat{\sigma}^2} \cdot \frac{1}{f(w_k)}$$

where

$$\frac{\partial \hat{f}(w_k)}{\partial \hat{\sigma}^2} = \frac{\hat{f}(w_k)}{\hat{\sigma}^2}$$

Therefore

$$\frac{\partial \log \hat{f}(\underline{x}_T | \underline{a}, \sigma^2)}{\partial \hat{\sigma}^2} = -\frac{1}{2} \sum_{k=1}^{T} \frac{\hat{f}(w_k)}{\hat{\sigma}^2} \cdot \frac{1}{f(w_k)} \qquad (5.3.1.10)$$

Thus, by substituting $\frac{1}{f(w_k)}$ in (5.3.1.10), we get

$$\frac{\partial \log \hat{f}(\underline{x}_T | \underline{a}, \sigma^2)}{\partial \hat{\sigma}^2} = -\frac{1}{2} \sum_{k=1}^{T} \frac{\hat{f}(w_k)}{\hat{\sigma}^2} \cdot \underline{h}' \; \underline{e}$$

$$= \underline{1}' \; \underline{e} \quad \text{where } \underline{1}' = -\frac{1}{2} \sum_{k=1}^{T} \frac{\hat{f}(w_k)}{\hat{\sigma}^2} \underline{h}'$$

$$\underline{h}' = (\cos(0), \cos(w_k), \ldots, \cos(Pw_k))$$

Let us define $\underline{m} = (a_1, \ldots, a_P, \sigma^2) = (\underline{a}, \sigma^2)$. Then, it is easily shown that

$$\frac{\partial \log \hat{f}(\underline{x}_T|\underline{m})}{\partial \underline{\hat{m}}} = \begin{bmatrix} \dfrac{\partial \log \hat{f}(\underline{x}_T|\underline{m})}{\partial \underline{\hat{a}}} \\ \\ \dfrac{\partial \log \hat{f}(\underline{x}_T|\underline{m})}{\partial \hat{\sigma}^2} \end{bmatrix}$$

$$= \begin{bmatrix} B\,\underline{e} \\ \\ \underline{1}'\,\underline{e} \end{bmatrix} = \begin{bmatrix} B \\ \\ \underline{1}' \end{bmatrix} \underline{e} = A\,\underline{e}$$

where A is a nonsingular matrix (P+1) x (P+1). Thus,

$$\underline{e} = A^{-1} \frac{\partial \log \hat{f}(\underline{x}_T|\underline{m})}{\partial \underline{\hat{m}}}$$

$$= A^{-1} \frac{1}{\hat{f}(\underline{x}_T|\underline{m})} \cdot \frac{\partial \hat{f}(\underline{x}_T|\underline{m})}{\partial \underline{\hat{m}}}$$

Then

$$E(\underline{e}|\underline{\hat{m}}) = A^{-1} \int_{\underline{e}} \frac{1}{\hat{f}(\underline{x}_T|\underline{m})} \cdot \frac{\partial \hat{f}(\underline{x}_T|\underline{m})}{\partial \underline{\hat{m}}}\, dG(\underline{e}|\underline{\hat{m}})$$

$$= A^{-1} \int_{\underline{e}} \frac{1}{\hat{f}(\underline{x}_T|\underline{m})} \cdot \frac{\partial \hat{f}(\underline{x}_T|\underline{m})}{\partial \underline{\hat{m}}} \cdot \frac{g(\underline{\hat{m}}|\underline{e})\, dG(\underline{e})}{f(\underline{\hat{m}})}$$

$$= A^{-1} \int_{\underline{e}} \frac{\partial(\underline{x}_T)}{\partial(\underline{x}_T)} \cdot \frac{\partial(\hat{\underline{m}})}{\partial(\hat{\underline{m}})} \cdot \frac{1}{\hat{f}(\underline{x}_T|\underline{m})} \cdot \frac{g(\hat{\underline{m}}|\underline{e})}{f(\hat{\underline{m}})} \frac{\partial \hat{f}(\underline{x}_T|\underline{m})}{\partial \hat{\underline{m}}} \cdot dG(\underline{e})$$

$$= A^{-1} \int_{\underline{e}} \frac{\partial(\underline{x}_T)}{\partial(\hat{\underline{m}})} \cdot \frac{dG(\hat{\underline{m}}|\underline{e})}{dF(\underline{x}_T|\underline{m})} \cdot \frac{1}{f(\hat{\underline{m}})} \frac{\partial \hat{f}(\underline{x}_T|\underline{m})}{\partial \hat{\underline{m}}} dG(\underline{e})$$

where,

$$dF(\underline{x}_T|\underline{m}) = \hat{f}(\underline{x}_T|\underline{m}) \ \partial(\underline{x}_T)$$

and

$$dG(\hat{\underline{m}}|\underline{e}) = g(\hat{\underline{m}}|\underline{e}) \ \partial(\hat{\underline{m}})$$

since $\hat{\underline{m}} = f(\underline{x}_T)$, then

$$dG(\hat{\underline{m}}|\underline{e}) = |J| \ dF(\underline{x}_T|\underline{m})$$

where

$$J = \frac{\partial(\underline{x}_T)}{\partial(\hat{\underline{m}})}$$

Thus

$$E(\underline{e}|\hat{\underline{m}}) = A^{-1} \int_{\underline{e}} J^2 \cdot \frac{1}{f(\hat{\underline{m}})} \frac{\partial \hat{f}(\underline{x}_T|\underline{m})}{\partial \hat{\underline{m}}} dG(\underline{e})$$

$$= A^{-1} \int_{\underline{e}} J^2 \frac{1}{f(\hat{\underline{m}})} \frac{\partial \hat{f}(\underline{x}_T|\underline{e})}{\partial \hat{\underline{m}}} dG(\underline{e})$$

since, there is a 1 - 1 relationship between $\underline{e}$ and $\underline{m}$, therefore

$$\hat{f}(\underline{x}_T | \underline{e}) = \hat{f}(\underline{x}_T | \underline{m})$$

$$= A^{-1} \frac{J^2}{f(\hat{\underline{m}})} \int_{\underline{e}} \frac{\partial \hat{f}(\underline{x}_T | \underline{e})}{\partial \hat{m}} \, dG\,(\underline{e}) =$$

By interchanging the order of integration and differentiation, we obtain

$$E(\underline{e} | \hat{\underline{m}}) \doteq A^{-1} \frac{J^2}{f(\hat{\underline{m}})} \cdot \frac{\partial \hat{f}(\underline{x}_T)}{\partial \hat{\underline{m}}} \qquad (5.3.1.11)$$

We obtain empirical Bayes estimator corresponding to (5.3.1.11) by substituting consistent estimators for J, $f(\hat{\underline{m}})$ and $\frac{\partial \hat{f}(\underline{x}_T)}{\partial \hat{\underline{m}}}$.

CHAPTER VI

6.1 Analysis of Multiple Series

The spectral representation of vector time series follows in a straightforward manner from that of univariate time series. Let us suppose that

$$x(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \cdot \\ \cdot \\ \cdot \\ x_p(t) \end{pmatrix}$$

is a multivariate Gaussian process with zero mean and continuous spectrum.

If we let f(w) denote the matrix with typical element $f_{jm}(w)$, we have the matrix representation

$$\Gamma(h) = \int_{-\pi}^{\pi} e^{iwh} f(w)\, dw$$

$$f(w) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-iwh} \Gamma(h)$$

Therefore, the estimate of f(w) can be represented in matrix form, by the expression

$$\hat{f}(w) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-iwh} W_M(h)\, C(h)$$

which is called a weighted covariance estimator. The weight function $W_M(h)$ is called a lag window, and integer M is the lag number, and

$C(h)$ denote the matrix with elements $c_{j,m}(h)$, where

$$c_{j,m}(h) = \frac{1}{N} \sum_{t=1}^{N-h} x_j(t+h)\, x_m(t); \; h = 0, 1, \ldots, N-1$$

Thus, the vector finite Fourier transform can be written as

$$Z(w) = \frac{1}{(2\pi N)^{1/2}} \sum_{t=1}^{N} x(t)\, e^{-iwt}$$

and the periodogram matrix is

$$I_N(w) = Z(w)\, Z^*(w)$$

where asterisk designates conjugate transpose. Then, the smoothed periodogram estimator is

$$\hat{f}(w) = \sum_{k=-\left[\frac{N-1}{2}\right]}^{[N/2]} K(w-w_k)\, I_N(w_k)$$

where $w_k = \frac{2\pi k}{N}$ and $K(w-w_k)$ is called weight function. For the Daniell estimator $K(w - w_k) = \frac{1}{N}$.

6.2 The Complex Wishart Distribution

Let $x_1, x_2, \ldots, x_P$ be independent N x 1 vectors, each complex normal with $\underline{0}$ means and covariance matrix Σ, then the N x N matrix

$$W = \sum_{j=1}^{P} x_j\, x^*_j,$$

has a complex Wishart distribution with P degrees of freedom and covariance matrix Σ, denoted $W^c(P, N, \Sigma)$. The density is

$$\frac{1}{\tilde{\Gamma}_N(P)\ |\Sigma|^P}\ |W|^{P-N}\ e^{-tr(\Sigma^{-1}W)}$$

where $P \geq N$, $W \geq 0$ and

$$\tilde{\Gamma}_N(P) = (\pi)^{(1/2)(N-1)N} \prod_{j=1}^{N} \Gamma(P-j+1)$$

6.3 The Approximate Distribution of Smoothed Spectral Estimates

Recall, the Daniell estimator

$$\hat{f}(w_k) = \frac{1}{n} \sum_{r=-\left[\frac{n-1}{2}\right]}^{[n/2]} I_N(w_k - w_v) \tag{6.3.1}$$

where $n = \frac{r}{2}$ and r is given by

$$r = N/\pi \int_{-\pi}^{\pi} W_M^{\ 2}(w)\ dw$$

Then, it can be shown that $n\hat{f}(w_k)$ is approximately the sum of n independent, identically distributed random matrices of the form

$$I_N(w_k) = Z(w_k)\ Z(w_k)^* \tag{6.3.2}$$

where $Z(w_k)$ has multivariate complex normal distribution with mean $\underline{0}$ and covariance matrix $f(w_k)$. Goodman (10) has found that the sum of n independent identically distributed matrices of the form (6.3.2) has the complex P-dimensional Wishart distribution.

6.4 Empirical Bayes Estimate of a 2 x 2 Spectral Density Matrix

In section 6.3, we have shown that S has the complex Wishart distribution $W^c(2, n, f(w))$, where

$$S = \sum_{v=-\left[\frac{n-1}{2}\right]}^{[n/2]} I_N(w_k - w_v)$$

$$f(w_k) = \begin{pmatrix} f_{xx}(w_k) & f_{xy}(w_k) \\ f_{xy}(w_k) & f_{yy}(w_k) \end{pmatrix}$$

and

$$f_n(S|f) = \frac{|S|^{n-2} \exp[-\mathrm{tr}\ f^{-1}S]}{\pi[\prod_{j=1}^{2}(n-j)!]\ |f|^n} \tag{6.4.1}$$

Using (6.4.1) we can write

$$|f| f_n(S|f) = \frac{|S|}{(n-2)(n-1)} \cdot f_{n-1}(S|f) \tag{6.4.2}$$

Thus

$$\log f_n(S|f) \ \alpha \ \log |S|^{n-2} - \mathrm{tr}(f^{-1}S) \tag{6.4.3}$$

Recall,

$$S = n\hat{f} = n \begin{pmatrix} \hat{f}_{xx} & \hat{f}_{xy} \\ \hat{f}_{xy} & \hat{f}_{yy} \end{pmatrix}$$

then

$$|S| = n^2(\hat{f}_{xx}\hat{f}_{yy} - \hat{f}_{xy}{}^2)$$

with

$$P = \begin{pmatrix} \hat{f}_{yy} & -2\hat{f}_{xy} \\ -2\hat{f}_{xy} & \hat{f}_{xx} \end{pmatrix}$$

$$\frac{\partial \log |S|^{n-2}}{\partial P} = (n-2)\ (n) \frac{S}{|S|}$$

$$\operatorname{tr}\ (f^{-1}S) = \frac{n}{|f|}\ (\hat{f}_{xx}f_{yy} + f_{xx}\hat{f}_{yy} - 2f_{xy}\hat{f}_{xy})$$

$$\frac{\partial \operatorname{tr}\ (f^{-1}S)}{\partial P} = n \frac{f}{|f|}$$

Then it is easily verified that

$$\frac{f}{|f|} = (n-2) \frac{S}{|S|} - \frac{1}{n} \frac{\partial \log f_n(s|f)}{\partial P} \tag{6.4.4}$$

Therefore, we can obtain empirical Bayes estimate for θ, from (6.4.4) where $\theta = \frac{f}{|f|}$

$$E(\frac{f}{|f|}|\hat{f}) = \frac{n-2}{n} \frac{\hat{f}}{|\hat{f}|} - \frac{1}{n} \int \frac{\partial \log f_n(\hat{f}|f)}{\partial P}\, dG(f|\hat{f})$$

$$= \frac{n-2}{n} \frac{\hat{f}}{|\hat{f}|} - \frac{1}{n} \int \frac{\frac{\partial}{\partial P} f_n(\hat{f}|f)}{f_n(\hat{f}|f)}\, dG(f|\hat{f})$$

$$= \frac{n-2}{n} \frac{\hat{f}}{|\hat{f}|} - \frac{1}{n} \int \frac{\frac{\partial}{\partial P} f_n(\hat{f}|f)}{f_n(\hat{f}|f)} \cdot \frac{f_n(\hat{f}|f)\ dG(f)}{f_n(\hat{f})}$$

$$= \frac{n-2}{n} \frac{\hat{f}}{|\hat{f}|} - \frac{1}{n} \frac{\frac{\partial}{\partial P} f_n(\hat{f})}{f_n(\hat{f})} \tag{6.4.5}$$

where $f_n(\hat{f}) = \int f_n(\hat{f}|f)\, dG(f)$.

Thus, we can get empirical Bayes estimator corresponding to (6.4.5), by substituting consistent estimators for $f_n(\hat{f})$ and $\frac{\partial}{\partial P} f_n(\hat{f})$. Since $Z(w_v)$ has multivariate complex normal distribution with mean zero and covariance matrix $f(w_k)$, therefore $\overline{Z}$ has multivariate complex normal distribution with mean $\underline{0}$ and covariance matrix $\frac{1}{n} f(w_k)$, since $\overline{Z}$ and S are (conditionally) independent. Hence the joint conditional distribution of $\overline{Z}$ and S can be expressed as

$$f_n(\overline{Z}, S|f) = f(\overline{Z}|f) \cdot f_n(S|f)$$

It follows from (6.4.4) that

$$f = (n-2) \frac{S}{|S|} |f| - \frac{1}{n} |f| \frac{\partial \log f_n(S|f)}{\partial P}$$

Let us write f as

$$f = A - B$$

where $A = (n-2) \frac{S}{|S|} |f|$

$$B = \frac{1}{n} |f| \frac{\partial \log f_n(S|f)}{\partial P}$$

then

$$\int A\, dF(f|\overline{Z}, S) = \int (n-2) \frac{S}{|S|} |f|\, dF(f|\overline{Z}, S)$$

$$= (n-2)\ \frac{S}{|S|}\int |f|\ \frac{f(\overline{Z}|f)f_n(S|f)}{f_n(\overline{Z},S)}\ dG(f)$$

Using (6.4.2) we can write

$$= (n-2)\ \frac{S}{|S|}\int f(\overline{Z}|f)\ \frac{|S|}{(n-2)(n-1)}\ \frac{f_{n-1}(S|f)}{f_n(\overline{Z},S)}\ dG(f)$$

$$= \frac{S}{m-1} \cdot \frac{f_{n-1}(\overline{Z},S)}{f_n(\overline{Z},S)}$$

$$\int B\ dF(f|\overline{Z},S) = \int \frac{1}{n}\ |f|\ \frac{\partial \log\ f_n(S|f)}{\partial P}\ \frac{f(\overline{Z}|f)f_n(S|f)}{f_n(\overline{Z},S)}\ dG(f)$$

$$= \int \frac{1}{n}\ |f|\ \frac{\partial f_n(S|f)}{\partial P} \cdot \frac{f(\overline{Z}|f)}{f_n(\overline{Z},S)}\ dG(f)$$

$$= \frac{1}{n}\ \frac{1}{f_n(\overline{Z},S)}\ \frac{\partial}{\partial P}\ \int |f|\ f_n(S|f)f(\overline{Z}|f)\ dG(f)$$

(assuming we can interchange the order of integration and differentiation). Therefore, by using (6.4.2) we can write

$$= \frac{1}{n} \cdot \frac{1}{f_n(\overline{Z},S)} \cdot \frac{\partial}{\partial P} \int \frac{|S|}{(n-2)(n-1)}\ f_{n-1}\ (S|f)f(\overline{Z}|f)\ dG(f)$$

$$= \frac{1}{n(n-1)(n-2)} \cdot \frac{1}{f_n(\overline{Z},S)} \cdot \frac{\partial}{\partial P}\ [|S| f_{n-1}(\overline{Z},S)]$$

$$= \frac{1}{n(n-1)(n-2)} \cdot \frac{1}{f_n(\overline{Z},S)}\ \left[|S|\ \frac{\partial}{\partial P}\ f_{n-1}(\overline{Z},S) + \frac{S}{n} \cdot f_{n-1}(\overline{Z},S)\right]$$

Thus, the empirical Bayes estimator of f is

$$E(f\,|\,\overline{Z},S) \doteq \frac{S}{n-1} \cdot \frac{f_{n-1}(\overline{Z},S)}{f_n(Z,S)} +$$

$$\frac{1}{n(n-1)(n-2)} \cdot \frac{1}{f_n(\overline{Z},S)} \quad [|S|\ \frac{\partial}{\partial P}\ f_{n-1}(\overline{Z},S) + \frac{S}{n}\, f_{n-1}(\overline{Z},S)]$$

CHAPTER VII

The Estimation of Moving Average and Autoregressive and Mixed Models, Using Spectral Methods

7.1 Moving Average

The likelihood function of MA process, under Gaussian assumption, can be written as

$$L(\underline{x}_T|\underline{m}) = (2\pi)^{-T/2} |\Sigma|^{-1/2} \exp\{-1/2\ \underline{x}'_T\ \Sigma^{-1}\ \underline{x}_T\} \tag{7.1.1}$$

Therefore, it is a function of quadratic form $\underline{x}'_T\ \Sigma^{-1}\ \underline{x}_T$. In section 5.2, we have shown that

$$\underline{x}'_T\ \Sigma^{-1}\ \underline{x}_T \doteq \sum_{k=1}^{T} \frac{I(w_k)}{f(w_k)} \tag{7.1.2}$$

Thus, in order to get an approximate maximum likelihood estimator, we have to minimize $\underline{x}'_T\ \Sigma^{-1}\ \underline{x}_T$. This leads to the equations

$$\sum_{l=0}^{q} \alpha_l \sum_{j=1}^{T} \frac{I(w_j)\cos(k-l)w_j}{f^2(w_j)} = 0;\ k = 1, 2, \ldots, q;\ \alpha_0 = 1 \tag{7.1.3}$$

where

$$f(w_j) = \frac{\sigma^2}{2\pi} \left| \sum_{k=0}^{q} \alpha_k e^{ikw_j} \right|^2$$

These equations are very difficult to solve. If we replace $f(w_j)$ by $\hat{f}(w_j)$ then our estimation equations become

$$\sum_{l=0}^{q} \hat{\alpha}_l^{(1)} \sum_{j=1}^{T} \frac{I(w_j)\cos(k-l)w_j}{\hat{f}^2(w_j)} = 0;\ k = 1, 2, \ldots, q;$$

$$\hat{\alpha}_0 = \hat{\alpha}_0^{(1)} = 1 \qquad (7.1.4)$$

The superscript (1) on $\hat{\alpha}_l^{(1)}$ is meant to differentiate it from the other estimator, $\hat{\alpha}_l$.

Hannan (11) has shown (7.1.4) does not given an improved estimator and he found that $\tilde{\alpha}^{(1)}$ gives us a better estimator, i.e. an asymptotically efficient estimator, where

$$\tilde{\alpha}^{(1)} = 2\hat{\alpha} - \hat{\alpha}^{(1)} \qquad (7.1.5)$$

We can write (7.1.4) as

$$\hat{\alpha}^{(1)} = -\hat{A}^{-1}\hat{a} \qquad (7.1.6)$$

where $\hat{A}$ is a q x q matrix with entries $(\hat{A})_{ij} = \hat{a}_{i-j}$; i,j = 1, 2, ..., q and

$$\hat{a}_k = T^{-1} \sum_{j=1}^{T} I(w_j)\cos(kw_j)/\hat{f}^2(w_j);\ k = 0, 1, 2, \ldots, q$$

Theorem 1

Let x_t be generated by a moving average of order q, i.e.

$$x_t = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \ldots + \alpha_q \varepsilon_{t-q} = \sum_0^q \alpha_k \varepsilon_{t-k},\ \alpha_0 = 1$$

for which the zeroes of h(m) must lie outside the unit circle, i.e. x_t is invertible, and ε_t is iid $N(0, \sigma^2)$ then $\tilde{\alpha}^{(1)}$ converges to α and

$\sqrt{T}\,(\tilde{\alpha}^{(1)} - \alpha)$ is asymptotically normal with zero mean and covariance matrix Φ^{-1}, where Φ has $\phi(i-j)$ in row i, column j, and

$$\phi(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{|h(e^{i\lambda})|^2}\, e^{i\lambda k}\, d\lambda$$

Note: $h(m) = 1 + \alpha_1 m + \alpha_2 m^2 + \ldots + \alpha_q m^q$. See Hannan (11) for proof.

7.2 Autoregressive Process

Let us assume that we have a pth order AR model, i.e.

$$x_t + \sum_{j=1}^{P} \beta_j\, x_{t-j} = \varepsilon_t$$

It can be verified that approximate maximum likelihood estimators satisfy in the following equation.

$$\sum_{l=0}^{P} \hat{\beta}_l^{(1)} \sum_{j=1}^{T} I(w_j)\cos(k-l)\,w_j = 0;\ k = 1, 2, \ldots, p$$

$$\hat{\beta}_0^{(1)} = 1 \qquad (7.2.1)$$

and

$$f(w_j) = \frac{\sigma^2}{2\pi}\left|\sum_{k=0}^{P} \beta_k\, e^{ikw_j}\right|^{-2}$$

We may write (7.2.1) as $\hat{\beta}^{(1)} = -\hat{B}^{-1}\,\hat{b}$ where $\hat{B}$ is a p x p matrix with entries $(\hat{B})_{ij} = \hat{b}_{i-j}$; i,j = 1, 2, ..., p and $\hat{b}_k = T^{-1}\sum_{j=1}^{T} I(w_j)\cos(kw_j)$; k = 0, 1, 2, ..., p.

Theorem 2

Let x_t be generated by a autoregressive of order P for which the zeroes of g(m) must lie outside the unit circle, i.e. x_t is stationary, and ε_t is i.i.e. $N(0, \sigma^2)$ then $\tilde{\beta}^{(1)}$ converges to β and

$\sqrt{T}\,(\tilde{\beta}^{(1)} - \beta)$ is asymptotically normal with zero mean and covariance matrix Ψ^{-1}, where Ψ has $\psi(i-j)$ in row i and column j, $\psi(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{|g(e^{i\lambda})|^2} e^{i\lambda k}\, d\lambda$ and $\tilde{\beta}^{(1)} = 2\hat{\beta} - \hat{\beta}^{(1)}$.

Note: $g(m) = 1 + \beta_1 m + \beta_2 m^2 + \ldots + \beta_p m^p$. See Hannan (11) for proof.

7.3 Empirical Bayes Estimate of the Moving Average Parameters

In Section 7.1, we found that $\underline{\tilde{\alpha}}^{(1)} \sim N(\frac{\alpha}{\sqrt{T}}, \frac{\Phi^{-1}}{T})$. Thus,

$$f(\underline{\tilde{\alpha}}^{(1)}|\underline{\alpha}) \propto \exp\{-\frac{1}{2T}(\underline{\tilde{\alpha}}^{(1)} - \frac{\underline{\underline{\alpha}}}{\sqrt{T}})\,\Phi^{-1}\,(\underline{\tilde{\alpha}}^{(1)} - \frac{\underline{\underline{\alpha}}}{\sqrt{T}})\}$$

$$\log f(\underline{\tilde{\alpha}}^{(1)}|\underline{\alpha}) \propto -\frac{1}{2T}(\underline{\tilde{\alpha}}^{(1)} - \frac{\underline{\underline{\alpha}}}{\sqrt{T}})\,\Phi^{-1}\,(\underline{\tilde{\alpha}}^{(1)} - \frac{\underline{\underline{\alpha}}}{\sqrt{T}})$$

Now

$$\frac{\partial \log f(\underline{\tilde{\alpha}}^{(1)}|\underline{\alpha})}{\partial \underline{\tilde{\alpha}}^{(1)}} = -\frac{1}{T}\,\Phi^{-1}\,(\underline{\tilde{\alpha}}^{(1)} - \frac{\alpha}{\sqrt{T}})$$

for which

$$\underline{\alpha} = T^{3/2}\,\Phi\,\frac{\partial \log f(\underline{\tilde{\alpha}}^{(1)}|\underline{\alpha})}{\partial \underline{\tilde{\alpha}}^{(1)}} + T^{1/2}\,\underline{\underline{\tilde{\alpha}}}^{(1)}$$

In order to determine the empirical Bayes estimators, we substitute a consistent estimator for Φ, then we have

$$\underline{\alpha} = T^{1/2}\,\underline{\tilde{\alpha}}^{(1)} + T^{3/2}\hat{\Phi}\,\frac{\partial \log f(\underline{\tilde{\alpha}}^{(1)}|\underline{\alpha})}{\partial \underline{\tilde{\alpha}}^{(1)}}$$

Therefore

$$E(\underline{\alpha}|\underline{\tilde{\alpha}}^{(1)}) = T^{1/2}\,\underline{\tilde{\alpha}}^{(1)} + T^{3/2}\,\hat{\Phi}\,\frac{\frac{\partial}{\partial\underline{\tilde{\alpha}}^{(1)}} f(\underline{\tilde{\alpha}}^{(1)})}{f(\underline{\tilde{\alpha}}^{(1)})}$$

By estimating $f(\underline{\tilde{\alpha}}^{(1)})$ and $f'(\underline{\tilde{\alpha}}^{(1)}) = \frac{\partial}{\partial\underline{\tilde{\alpha}}^{(1)}} f(\underline{\tilde{\alpha}}^{(1)})$ consistently; however, we obtain an empirical Bayes estimator for $\underline{\alpha}$.

7.4 Empirical Bayes Estimate of the Autoregressive Parameters

We noted previously that $\tilde{\beta}^{(1)} \sim N\left(\frac{\beta}{\sqrt{T}}, \psi^{-1}/T\right)$. Therefore,

$$f(\underline{\tilde{\beta}}^{(1)}|\underline{\beta})\alpha \quad \exp\left\{-\frac{1}{2T}\left(\underline{\tilde{\beta}}^{(1)} - \frac{\beta}{\sqrt{T}}\right)\Psi^{-1}\left(\underline{\tilde{\beta}}^{(1)} - \frac{\beta}{\sqrt{T}}\right)\right\}$$

$$\log f(\underline{\tilde{\beta}}^{(1)}|\underline{\beta})\ \alpha\ -\frac{1}{2T}\left(\underline{\tilde{\beta}}^{(1)} - \frac{\beta}{\sqrt{T}}\right)\Psi^{-1}\left(\underline{\tilde{\beta}}^{(1)} - \frac{\beta}{\sqrt{T}}\right)$$

In order to get the empirical Bayes estimators, we substitute a consistent estimator for Ψ^{-1}, denoted by $\hat{\Psi}^{-1}$, then we have

$$\log \hat{f}(\underline{\tilde{\beta}}^{(1)}|\underline{\beta})\alpha\ -\frac{1}{2T}\left(\underline{\tilde{\beta}}^{(1)} - \frac{\beta}{\sqrt{T}}\right)\hat{\Psi}^{-1}\left(\underline{\tilde{\beta}}^{(1)} - \frac{\beta}{\sqrt{T}}\right)$$

where

$$\hat{\Psi} = \left\{\sum_{l=0}^{P} \hat{\beta}_l\, \hat{b}_l\right\}^{-1} \hat{B}$$

Now

$$\frac{\partial \log \hat{f}(\underline{\tilde{\beta}}^{(1)}|\underline{\beta})}{\partial\underline{\beta}^{(1)}} = -\frac{1}{T}\,\hat{\Psi}^{-1}\left(\underline{\tilde{\beta}}^{(1)} - \frac{\beta}{\sqrt{T}}\right)$$

for which

$$\underline{\beta} = T^{3/2}\,\hat{\Psi}\,\frac{\partial \log \hat{f}(\underline{\tilde{\beta}}^{(1)}|\underline{\beta})}{\partial\underline{\tilde{\beta}}^{(1)}} + T^{1/2}\,\underline{\tilde{\beta}}^{(1)}$$

Thus,

$$E(\underline{\beta}|\underline{\tilde{\beta}}^{(1)}) = T^{1/2}\ \underline{\tilde{\beta}}^{(1)} + T^{3/2}\ \hat{\Psi} \int_{\underline{\beta}} \frac{\partial \log \hat{f}(\underline{\tilde{\beta}}^{(1)}|\underline{\beta})}{\partial \underline{\tilde{\beta}}^{(1)}}\, dG\ (\underline{\beta}|\underline{\tilde{\beta}}^{(1)})$$

$$= T^{1/2}\ \underline{\tilde{\beta}}^{(1)} + T^{3/2}\ \hat{\Psi} \int_{\underline{\beta}} \frac{\frac{\partial}{\partial \underline{\tilde{\beta}}^{(1)}} f(\underline{\tilde{\beta}}^{(1)}|\underline{\beta})}{f(\underline{\tilde{\beta}}^{(1)}|\beta)}\ \frac{f(\underline{\tilde{\beta}}^{(1)}|\underline{\beta})}{f(\underline{\tilde{\beta}}^{(1)})}\, dG(\beta)$$

$$= T^{1/2}\ \underline{\tilde{\beta}}^{(1)} + T^{3/2}\ \hat{\Psi}\ \frac{\frac{\partial}{\partial \underline{\tilde{\beta}}^{(1)}} f(\underline{\tilde{\beta}}^{(1)})}{f(\underline{\tilde{\beta}}^{(1)})} \qquad (7.4.1)$$

We obtain empirical Bayes estimator corresponding to (7.4.1) by substituting consistent estimators for $f(\underline{\tilde{\beta}}^{(1)})$ and

$\frac{\partial}{\partial \underline{\tilde{\beta}}^{(1)}} f(\underline{\tilde{\beta}}^{(1)})$.

CHAPTER VIII

Estimation of the Autocorrelation (Serial Correlation)

In regression analysis of economic time series a situation often arises in which a certain observed quantity represents a dependent variable at one time and an independent variable at a later time, for instance, the following relation may exist between x_t (price at time t) and x_{t-k} (price at time t-k).

$$x_t = \rho_k x_{t-k} + \varepsilon_t \text{ where } k \geq 1$$

A k^{th} order serial correlation, ρ_k, is a measurement of a serial dependence between x_t and x_{t-k}.

8.1 The Estimation of Serial Correlation

In the process $x_t = \rho_k x_{t-k} + \varepsilon_t$, $k > 1$ where the ε_t are independent drawing from a normal distribution with mean zero and variance σ^2 (known), the parameter ρ_k may have any positive or negative values. The process will only be stationary one if $|\rho_k| < 1$. It can be shown that, the conditional maximum likelihood estimate of ρ_k is

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{T} x_{t-k} x_t}{\sum_{t=k+2}^{T} x^2_{t-k}}$$

In most cases, $\hat{\rho}_k$ is denoted by r_k, where

$$r_k = \frac{\sum_{t=k+1}^{T} x_t\, x_{t-k}}{\sum_{t=1}^{T} x^2_t}$$

since both $\sum_{t=k+1}^{T} x^2_t$ and $\sum_{t=k+1}^{T} x^2_{t-k} = \sum_{t=1}^{T-k} x^2_t$ estimate $(T-k)\sigma^2$ in the case of stationary process with mean zero these may be replaced by $\sum_{t=1}^{T} x^2_t$ (possibly multiplied by $(T-k)/T$).

8.2 Empirical Bayes Estimate of ρ_k

Recall,

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{T} x_t\, x_{t-k}}{\sum_{t=k+1}^{T} x^2_{t-k}}$$

where $\hat{\rho}_k$ is the conditional M.L.E. of ρ_k.

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{T-1} x_t\, x_{t-k}}{\sum_{t=k+1}^{T} x^2_{t-k}} + \frac{x_T\, x_{T-k}}{\sum_{t=k+1}^{T} x^2_{t-k}}$$

Thus,

$$E(\hat{\rho}_k \mid \rho_k, \cdot \underline{x}_{T-1}) = \frac{c_1}{u_1} + \frac{x^2_{T-k}}{u_1} \quad \rho_k$$

where

$$C_1 = \sum_{t=k+1}^{T-1} x_t x_{t-k}, \; u_1 = \sum_{t=k+1}^{T} x^2_{t-k}$$

$$\underline{x}_{T-1} = (x_1, x_2, \ldots, x_{T-1})$$

Therefore

$$f(\hat{\rho}_k | \rho_k, \underline{x}_{T-1}) = \frac{u_1}{\sqrt{2\pi}\sigma\, x_{T-k}} \exp\left\{- \frac{u_1^{\,2}}{2\sigma^2\, x^2_{T-k}} (\hat{\rho}_k - \frac{c_1}{u_1} - \frac{x^2_{T-k}}{u_1} \rho_k)^2\right\}$$

or

$$(\hat{\rho}_k | \rho_k, \underline{x}_{T-1}) \sim N\left[(\frac{c_1}{u_1} + \frac{x^2_{T-k}}{u_1} \rho_k), \frac{x^2_{T-k}}{u^2_1} \sigma^2\right]$$

It is easily verified

$$\frac{\frac{d}{d\hat{\rho}_k} f(\hat{\rho}_k | \rho_k, \underline{x}_{T-1})}{f(\hat{\rho}_k | \rho_k, \underline{x}_{T-1})} = - \frac{u_1^{\,2}}{c_1^{\,2}\, x^2_{T-k}} (\hat{\rho}_k - \frac{c_1}{u_1} - \frac{x^2_{T-k}}{u_1} \rho_k)$$

Thus

$$\rho = \frac{\sigma^2}{u_1} \cdot \frac{\frac{d}{d\hat{\rho}_k} f(\hat{\rho}_k | \rho_k, \underline{x}_{T-1})}{f(\hat{\rho}_k | \rho_k, \underline{x}_{T-1})} + \frac{u_1 \hat{\rho}_k}{x^2_{T-k}} - \frac{c_1}{x^2_{T-k}}$$

By taking expectation, we get

$$E(\rho_k | \hat{\rho}_k, \underline{x}_{T-1}) = \frac{u_1 \hat{\rho}_k}{x^2_{T-k}} - \frac{c}{x^2_{T-k}}$$

$$+ \frac{\sigma^2}{u_1} \int_{\rho_k} \frac{\frac{d}{d\hat{\rho}_k} f(\hat{\rho}_k | \rho_k, \underline{x}_{T-1})}{f(\hat{\rho}_k | \rho_k, \underline{x}_{T-1})} dG(\rho_k | \hat{\rho}_k, \underline{x}_{T-1})$$

By using the law of conditional probability, we have

$$\int_{\rho_k} \frac{\frac{d}{d\hat{\rho}_k} f(\hat{\rho}_k | \rho, \underline{x}_{T-1})}{f(\hat{\rho}_k | \rho, \underline{x}_{T-1})} dG(\rho_k | \hat{\rho}_k, \underline{x}_{T-1}) =$$

$$\cdot \int_{\rho_k} \frac{\frac{d}{d\hat{\rho}_k} f(\hat{\rho}_k | \rho, \underline{x}_{T-1})}{f(\hat{\rho}_k | \rho, \underline{x}_{T-1})} \cdot \frac{f(\hat{\rho}_k | \rho_k, \underline{x}_{T-1}) \, g(\rho_k, \underline{x}_{T-1}) \, d\rho_k}{\int f(\hat{\rho}_k, \underline{x}_{T-1} | \rho_k) \, dG(\rho_k)}$$

$$= \frac{\int_{\rho_k} \frac{d}{d\hat{\rho}_k} f(\hat{\rho}_k | \rho_k, \underline{x}_{T-1}) \, g(\rho_k, \underline{x}_{T-1}) \, d\rho_k}{\int_{\rho_k} f(\hat{\rho}_k, \underline{x}_{T-1} | \rho_k) \, dG(\rho_k)}$$

By interchanging the order of integration and differentiation

$$= \frac{\frac{d}{d\hat{\rho}_k} \int_{\rho_k} f(\hat{\rho}_k | \rho_k, \underline{x}_{T-1}) \, g(\rho_k, \underline{x}_{T-1}) \, d\rho_k}{\int_{\rho_k} f(\hat{\rho}_k, \underline{x}_{T-1} | \rho_k) \, dG(\rho_k)}$$

$$= \frac{\frac{d}{d\hat{\rho}_k} \int_{\rho_k} f(\hat{\rho}_k, \rho_k, \underline{x}_{T-1}) \, d\rho_k}{f(\hat{\rho}_k, \underline{x}_{T-1})}$$

$$= \frac{\frac{d}{d\hat{\rho}_k} f(\hat{\rho}_k, \underline{x}_{T-1})}{f(\hat{\rho}_k, \underline{x}_{T-1})}$$

Therefore, we get

$$E(\rho_k | \rho_k, \underline{x}_{T-1}) = \frac{u_1 \hat{\rho}_k}{x^2_{T-k}} - \frac{c_1}{x^2_{T-k}} + \frac{\frac{d}{d\hat{\rho}_k} f(\hat{\rho}_k, \underline{x}_{T-1})}{f(\hat{\rho}_k, \underline{x}_{T-1})} \cdot \frac{\sigma^2}{u_1}$$

We obtain empirical Bayes estimator, by substituting consistent estimators for $f(\hat{\rho}_k, \underline{x}_{T-1})$ and $\frac{d}{d\hat{\rho}_k} f(\hat{\rho}_k, \underline{x}_{T-1})$.

CHAPTER IX

Unobservable Variables in Regression

9.1 Unobservable Variables and Errors in Variables

In econometric analysis, we often find that there are variables which affect the observable variables but which are not themselves directly observable. Unobservable variables may be due to measurement error in the observed magnitudes, and variables such as "Permanent Income" or "Expected Price" are examples.

9.2 Statistical Consequences of Errors in Variables

When errors and inaccuracies exist in explanatory variables x, what effect will this measurement error have on the sampling properties of the conventional or least squares estimator?

Let us assume

$$y_n = \sum_{j=-s}^{t} \beta_j x_{n-j} + \varepsilon_n$$

Suppose that ε_n and x_m are independent for all m, n where ε_i's are i.i.d., normally distributed with mean zero and variance σ^2 (σ^2 is known).

First, we discuss the case j = 1, that is,

$$y_t = \beta x_t + \varepsilon_t \tag{9.2.1}$$

where

$$x^*_t = x_t + u_t \tag{9.2.2}$$

x^*_t is the observed or measured value of the true value x_t and u_t denote errors of observation. Substitution of (9.2.2) into (9.2.1) results in

$$y_t = \beta(x^*_t - u_t) + \varepsilon_t$$

or

$$y_t = \beta x^*_t + e_t \text{ where } e_t = \varepsilon_t - \beta u_t$$

In this case, even if we make the usual favorable assumptions that u_t and ε_t are mutually and serially independent with constant variances (σ^2_u and σ^2), independent of the true value x_t and $E(u_t) = E(\varepsilon_t) = 0$, so that $x^*_t - E(x^*_t) = u_t$, the covariance between x^*_t and e_t does not vanish because

$$E[x^*_t - E(x^*_t)]e_t = E\, u_t(\varepsilon_t - \beta u_t) = -\beta\sigma^2_u$$

Thus the classical assumption that the stochastic term is independent of the regressor is not met and the least squares estimate of the parameters will be biased. The bias of

$$\hat{\beta} = \sum_{i=1}^{T} (x_t - \bar{x})\ (y_t - \bar{y}) / \sum_{i=1}^{T} (x_t - \bar{x})^2$$

$$E(\hat{\beta}) = \beta + E\left[\frac{\sum_{1}^{T} (x_t - \bar{x})e_t}{\sum_{t=1}^{T} (x_t - \bar{x})^2}\right]$$

The bias will not disappear as the sample size becomes infinity large, and the inconsistency is reflected in terms of the probability limit as

$$P \lim (\hat{\beta}-\beta) = P \lim \frac{\sum_{1}^{T} (x_t - \bar{x})\, e_t/T}{\sum_{1}^{T} (x_t-\bar{x})^2/T}$$

$$\lim_{T} \frac{\sum_{1}^{T}(x_t-\bar{x})\, e_t/T}{\sum_{1}^{T} (x_t-\bar{x})^2/T} \xrightarrow{P} -\frac{\beta\sigma_u^{\,2}}{\sigma_{x*}^{\,2}} = -\frac{\beta\sigma_u^{\,2}}{\sigma_x^{\,2} + \sigma_u^{\,2}}$$

where

$$\sigma_{x*}^{\,2} = P \lim \sum_{1}^{T} (x_t-\bar{x})^2/T \text{ and}$$

$$\sigma_u^{\,2} = P \lim \sum_{1}^{T} u_t^{\,2}/T \text{ are assumed to exist.}$$

The asymptotic bias indicates that $\hat{\beta}$ will always underestimate β by $\frac{\beta\sigma_u^{\,2}}{\sigma_x^{\,2}}$ which can be small only if $\sigma_x^{\,2}$ or $\sigma_{x*}^{\,2}$ is relatively larger than $\sigma_u^{\,2}$. In other words, asymptotic bias can be small only if the variation of the true variable is large relative to that of the measurement errors.

9.3 Estimating the Coefficient (β), Using Spectral Method

Recall, we have the relation between spectral

$$f_{yx}(\lambda) = \beta e^{i\lambda} f_x(\lambda)$$

If we assume, that x_t is stationary with absolutely continuous spectrum, then we can estimate β from the approximate relation.

$$\hat{f}_{yx}(\lambda_k) = \hat{\beta}\, e^{i\lambda_k}\, \hat{f}_x(\lambda_k) \text{ where } \lambda_k = \frac{2\pi k}{2M}$$

$\hat{f}_{yx}$, $\hat{f}_x$ can be computed from mean corrected data. Thus

$$\hat{\beta} = \frac{1}{2M} \sum_{k=-M+1}^{M} \frac{\hat{f}_{yx}(\lambda_k)}{\hat{f}_x(\lambda_k)}\, e^{-i\lambda_k}$$

It can be shown that for large T, the distribution of the $\sqrt{T}\,(\hat{\beta}-\beta)$ is normal with mean zero and variance $\frac{1}{2\pi} \int_{-\pi} \frac{f_{\varepsilon}(\lambda)}{f_y(\lambda)}\, d\lambda$, see Hannan (11).

9.3.1 Empirical Bayes Estimate of β

We noted previously that

$$\hat{\beta} \sim N\left(\frac{\beta}{\sqrt{T}}, V\right) \text{ where } V = \frac{1}{2\pi T} \int_{-\pi}^{\pi} \frac{f\varepsilon(\lambda)}{f_x(\lambda)}\, d\lambda$$

Therefore

$$f(\hat{\beta}|\beta) \propto \exp\left\{-\frac{1}{2V}\left(\hat{\beta} - \frac{\beta}{\sqrt{T}}\right)^2\right\}$$

$$\log f(\hat{\beta}|\beta) \propto -\frac{1}{2V}\left(\hat{\beta} - \frac{\beta}{\sqrt{T}}\right)^2$$

In order to get the empirical Bayes estimators, we substitute a consistent estimator for V, and is denoted by $\hat{V}$, then we have

$$\log \hat{f}(\hat{\beta}|\beta) \propto -\frac{1}{2\hat{V}}\left(\hat{\beta} - \frac{\beta}{\sqrt{T}}\right)^2$$

where

$$\hat{V} = \frac{1}{2\pi T} \int_{-\pi}^{\pi} \frac{\hat{f}_{\varepsilon}(\lambda)}{\hat{f}_x(\lambda)}\, d\lambda$$

$$\frac{d \log f(\hat{\beta}|\beta)}{d\hat{\beta}} = -\frac{1}{\hat{V}} (\hat{\beta} - \frac{\beta}{\sqrt{T}})$$

$$\beta = \sqrt{T}\ \hat{\beta} + \hat{V}\ T^{1/2}\ \frac{d \log f(\hat{\beta}|\beta)}{d\hat{\beta}}$$

Thus,

$$E(\beta|\hat{\beta}) = \sqrt{T}\ \hat{\beta} + \hat{V}\ T^{1/2} \int_{\beta} \frac{d \log f(\hat{\beta}|\beta)}{d\hat{\beta}}\ dG\ (\beta|\hat{\beta})$$

$$= \sqrt{T}\ \hat{\beta} + \hat{V}\ T^{1/2} \int_{\beta} \frac{\frac{d}{d\hat{\beta}} f(\hat{\beta}|\beta)}{f(\hat{\beta}|\beta)} \cdot \frac{f(\hat{\beta}|\beta)}{f(\hat{\beta})}\ dG\ (\beta)$$

$$= \sqrt{T}\ \hat{\beta} + \hat{V}\ T^{1/2}\ \frac{\frac{d}{d\hat{\beta}} f(\hat{\beta})}{f(\hat{\beta})} \qquad (9.4.1)$$

We obtain empirical Bayes estimator corresponding to (9.4.1) by substituting consistent estimators for $f(\hat{\beta})$ and $\frac{d}{d\hat{\beta}} f(\hat{\beta})$.

9.4 Unobservable Variables in Linear Regression With MA(q) Errors

In this section we are concerned with estimating β in the linear model with moving average errors of order q. This model is given by

$$y_t = \beta\ x_t + u_t;\ t = 1, 2, \ldots, T$$

and

$$u_t = \varepsilon_t + \alpha_1\ \varepsilon_{t-1} + \alpha_2\ \varepsilon_{t-2} + \ldots + \alpha_q\ \varepsilon_{t-q}$$

For $q = 1$; $u_t = \varepsilon_t + \alpha\ \varepsilon_{t-1}$ where $E(\varepsilon_t) = 0$, $E({\varepsilon^2}_t) = {\sigma_\varepsilon}^2$, $E(\varepsilon_t\varepsilon_s) = 0$ for $t \neq s$ and $|\alpha| < 1$. The covariance matrix for $\underline{u}_t$ is

$E[\underline{u}_t \; \underline{u}'_t] = \sigma_u^{\;2} W$ where $\sigma_u^{\;2} = \sigma_\varepsilon^{\;2}\,(1 + \alpha^2)$.

$$W = \begin{bmatrix} 1+\alpha^2 & \alpha & 0 & \cdots & 0 & 0 \\ \alpha & 1+\alpha^2 & \alpha & \cdots & 0 & 0 \\ 0 & \alpha & 1+\alpha^2 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & 1+\alpha^2 & \alpha \\ 0 & 0 & 0 & \cdots & \alpha & 1+\alpha^2 \end{bmatrix}$$

For known α the GLS estimator for the β is given by

$$\hat{\beta} = \frac{\underline{x}' \; W^{-1} \; \underline{y}}{\underline{x}' \; W^{-1} \; \underline{x}}$$

Estimating the Coefficient (β), Using Spectral Method

Recall, we have the relation between spectral

$$f_{yx}(\lambda) = \beta e^{i\lambda} \; f_x(\lambda)$$

If we assume, that x_t is stationary with continuous spectrum, then we can estimate β from the approximate relation.

$$\hat{f}_{yx}(\lambda_k) = \hat{\beta} \; e^{i\lambda_k} \; \hat{f}_x(\lambda_k); \; \lambda_k \doteq \frac{2\pi k}{2M}$$

where $\hat{f}_{yx}$, $\hat{f}$ can be computed from mean corrected data. Therefore

$$\hat{\beta} = \frac{1}{2M} \sum_{k=-M}^{M} \frac{\hat{f}_{yx}(\lambda_k)}{\hat{f}_x(\lambda_k)} e^{-i\lambda_k}$$

It can be shown that for large T, the distribution of $\sqrt{T}(\hat{\beta}-\beta)$ is normal with mean zero and variance

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_u(\lambda)}{f_x(\lambda)} \, d\lambda, \text{ see Hannan (11).}$$

9.5 Unobservable Variables in Linear Regression With AR(P) Errors

In this section we are concerned with estimation of the linear model

$$y_t = \beta x_t + u_t;\ t = 1, 2, 3, \ldots, T$$

where the u'_ts are autocorrelated, and

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots + \rho_P u_{t-P} + \varepsilon_t$$

For P = 1, $u_t = \rho u_{t-1} + \varepsilon_t$ where $E(\varepsilon_t) = 0$, $E(\varepsilon^2{}_t) = \sigma_\varepsilon{}^2$, $E(\varepsilon_t \varepsilon_s) = 0$ for $t \neq s$, $|\rho| < 1$. The covariance matrix for u_t is

$$E[\underline{u}_t\ \underline{u}_t'] = \frac{\sigma_\varepsilon{}^2}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{T-1} \\ \rho & 1 & \rho & \ldots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \ldots & \rho^{T-3} \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \ldots & 1 \end{bmatrix}$$

We can use 2 step least square method for estimating β and ρ.

9.6 Empirical Bayes Estimate of β With AR(P) or MA(q) or ARMA(p,q) Errors

We noted previously that

$$\hat{\beta} \sim N\left(\frac{\beta}{\sqrt{T}}, V\right) \text{ where } V = \frac{1}{2\pi T} \int_{-\pi}^{\pi} \frac{f_u(\lambda)}{f_x(\lambda)} \, d\lambda$$

Therefore $f(\hat{\beta}|\beta) \propto \exp\{-\frac{1}{2V}(\hat{\beta} - \frac{\beta}{\sqrt{T}})^2\}$

$$\log f(\hat{\beta}|\beta) \propto -\frac{1}{2V}\left(\hat{\beta} - \frac{\beta}{\sqrt{T}}\right)^2$$

In order to get the empirical Bayes estimators, we substitute a consistent estimator for V and is denoted by $\hat{V}$, then we have

$$\log \hat{f}(\hat{\beta}|\beta) \propto -\frac{1}{2\hat{V}}\left(\hat{\beta} - \frac{\beta}{T}\right)^2$$

where

$$\hat{V} = \frac{1}{2\pi T} \int_{-\pi}^{\pi} \frac{\hat{f}_u(\lambda)}{\hat{f}_x(\lambda)} \, d\lambda$$

$$\frac{d \log f(\hat{\beta}|\beta)}{d\hat{\beta}} = -\frac{1}{\hat{V}}\left(\hat{\beta} - \frac{\beta}{\sqrt{T}}\right)$$

$$\beta = \sqrt{T}\,\hat{\beta} + T^{1/2}\,\hat{V}\,\frac{d \log \hat{f}(\hat{\beta}|\beta)}{d\hat{\beta}}$$

Therefore

$$E(\beta|\hat{\beta}) = \sqrt{T}\,\hat{\beta} + \hat{V}\,T^{1/2} \int_{\beta} \frac{d \log f(\hat{\beta}|\beta)}{d\hat{\beta}} \, dG(\beta|\hat{\beta})$$

$$= \sqrt{T}\ \hat{\beta} + \sqrt{T}\ \hat{V} \int_{\beta} \frac{\frac{d}{d\beta} f(\hat{\beta}|\beta)}{f(\hat{\beta}|\beta)} \cdot \frac{f(\hat{\beta}|\beta)}{f(\hat{\beta})}\, dG(\beta)$$

$$= \sqrt{T}\ \hat{\beta} + \sqrt{T}\ \hat{V} \frac{\frac{d}{d\hat{\beta}} f(\hat{\beta})}{f(\hat{\beta})} \qquad (9.6.1)$$

We can obtain empirical Bayes estimator corresponding to (9.6.1) by substituting consistent estimators for $f(\hat{\beta})$ and $\frac{d}{d\hat{\beta}} f(\hat{\beta})$. We can generalize the simple linear regression case to a multiple regression, in which we have $\underline{\underline{P}}$ explanatory variables ($P > 1$), with AR or MA or ARMA errors.

9.7 Regression With Time Series Structure in u_t and x_t

For the linear model $x_t = \beta x_t + u_t$, we have considered various time series model about u_t, where the independent variable(s) x_t does not follow any "time series model". However, when u_t follows an ARMA process, it is possible that x_t also follows an ARMA model; therefore, we can write the model as

$$y_t = \beta x_t + u_t \text{ where}$$

$$u_t + \alpha_1 u_{t-1} + \alpha_2 u_{t-2} + \dots + \alpha_P u_{t-P} = \varepsilon_t + \lambda_1 \varepsilon_{t-1} + \dots + \lambda_q \varepsilon_{t-q}$$

$$x_t + \gamma_1 x_{t-1} + \gamma_2 x_{t-2} + \dots + \gamma_P x_{t-P} = e_t + \rho_1 e_{t-1} + \dots + \rho_q e_{t-q}$$

Let's consider the simple model.

$$\begin{cases} y_t = \beta x_t + u_t \\ u_t = \rho u_{t-1} + \varepsilon_t \\ x_t = \lambda x_{t-1} + e_t \end{cases}$$

Where the e_t are normal independent $(0, \sigma_e^2)$ random variables, the ε_t are normal independent $(0, \sigma_\varepsilon^2)$ random variables, e_t is independent of ε_j for all t,j and $|\rho| < 1$, $|\lambda| < 1$. For known ρ the GLS estimator for β is given by

$$\hat{\beta} = \frac{(1-\rho^2)\, x_1 y_1 + \sum_{t=2}^{T} (x_t - \rho x_{t-1})(y_t - \rho y_{t-1})}{(1-\rho^2)\, x_1^2 + \sum_{t=2}^{T} (x_t - \rho x_{t-1})^2}$$

Estimating the Coefficient (β), Using Spectral Method

Recall, we have the relation between spectral

$$f_{yx}(\lambda) = \beta\, e^{i\lambda} f_x(\lambda)$$

If we assume that x_t is stationary with continuous spectrum. Then we can estimate β from the approximate relation

$$\hat{f}_{yx}(\lambda_k) = \hat{\beta}\, e^{i\lambda_k} \hat{f}_x(\lambda_k); \quad \lambda_k = \frac{2\pi k}{2M}$$

where $\hat{f}_{yx}$, $\hat{f}_x$ can be computed from mean corrected data. Therefore,

$$\hat{\beta} = \frac{1}{2M} \sum_{k=-M}^{M} \frac{\hat{f}_{yx}(\lambda_k)}{\hat{f}_x(\lambda_k)}\, e^{-i\lambda_k}$$

CHAPTER X

10.1 An Example

The monthly seasonally adjusted United States layoff rates and total accession rates in manufacturing from 1966 to 1977 is given in Table I. (See Amini (1).)

Let the layoff (or total accession) rates at time t be represented by x_t. The usual relationship assumed between layoff rates (or total accession rates) at time t and time t-1 is of the form

$$(x_t - \mu) = \rho(x_{t-1} - \mu) + \varepsilon_t \qquad (10.1.1)$$

where μ and ρ are the unknown parameters, and ε_t is a random error. This assumption will be made here and tested. If this is the true model, then the residuals constitute a white noise process. Thus, if the true values of the ε_t were known, then the sample autocorrelations of the residuals would give a very clear indication of any departures from white noise behavior. Unfortunately, in practice it is necessary to estimate the coefficient of equation (10.1.1), and so, denoting the estimate of ρ by $\hat{\rho}$. What one has available are estimated residuals.

$$\hat{\varepsilon}_t = (x_t - \hat{\mu}) - \hat{\rho}(x_{t-1} - \hat{\mu})$$

It seems reasonable to expect the autocorrelations

$$r_k(\hat{\varepsilon}) = \sum_{t=k+1}^{T} \hat{\varepsilon}_t \, \hat{\varepsilon}_{t-k} / \sum_{t=1}^{T} \hat{\varepsilon}_t^{\,2}$$

TABLE I.

The Monthly Seasonally Adjusted United States Layoff Rates in Manufacturing (per 100 Employees).

						Layoffs						
Year	Jan.	Feb.	Mar.	April	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.
1966	1.2	1.1	1.1	1.1	1.1	1.2	1.5	1.2	1.0	1.1	1.2	1.3
1967	1.4	1.5	1.6	1.5	1.4	1.3	1.4	1.3	1.3	1.3	1.2	1.2
1968	1.4	1.3	1.2	1.2	1.2	1.2	1.3	1.3	1.2	1.2	1.1	1.1
1969	1.0	1.1	1.1	1.0	1.0	1.3	1.1	1.2	1.2	1.3	1.3	1.4
1970	1.5	1.7	1.8	1.9	1.9	1.9	1.6	1.9	1.9	2.2	2.0	1.7
1971	1.7	1.5	1.5	1.5	1.6	1.5	1.5	2.0	1.7	1.5	1.4	1.3
1972	1.2	1.2	1.1	1.2	1.1	1.4	1.3	1.1	1.0	1.0	.9	.9
1973	.8	.7	.8	.8	.8	.8	1.1	.9	.8	.8	.9	1.1
1974	1.3	1.2	1.1	1.1	1.1	1.2	1.2	1.4	1.4	1.9	2.4	2.5
1975	2.9	2.9	2.6	2.5	2.5	2.2	1.7	1.7	1.7	1.6	1.5	1.3
1976	1.1	1.0	1.2	1.3	1.3	1.4	1.4	1.5	1.5	1.5	1.3	1.2
1977	1.2	1.4	1.1	1.1	1.1	1.2	1.3	1.3	1.3	1.1	.9	1.0
1978	.9	.9	.9	.9	p1.0							

p = preliminary.

$\tilde{\rho}$ = .60150599 $\hat{\rho}$ = .46388790

TABLE I. (Cont'd.)

The Monthly Seasonally Adjusted United States Total Accession Rates in Manufacturing (per 100 Employees).

					Total Accession							
Year	Jan.	Feb.	Mar.	April	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.
1966	4.9	5.0	5.3	5.1	5.0	4.9	4.9	5.0	5.0	4.9	4.7	4.7
1967	4.5	4.4	4.3	4.3	4.4	4.5	4.4	4.3	4.4	4.4	4.5	4.6
1968	4.4	4.4	4.6	4.6	4.6	4.5	4.5	4.7	4.6	4.8	4.9	4.9
1969	4.9	4.8	4.9	4.9	4.7	5.0	4.7	4.5	4.7	4.6	4.5	4.6
1970	4.4	4.4	4.0	4.0	4.1	4.1	4.3	3.9	3.9	3.8	3.7	3.8
1971	3.8	3.7	3.7	3.8	3.8	3.8	3.8	4.0	4.0	3.9	4.0	4.2
1972	4.3	4.3	4.4	4.3	4.4	4.2	4.3	4.4	4.4	4.6	4.6	4.9
1973	5.0	5.1	5.0	4.8	4.7	4.6	4.6	4.6	4.7	4.8	4.9	4.7
1974	4.7	4.6	4.5	4.5	4.6	4.3	4.2	4.1	3.9	3.6	3.1	3.1
1975	3.1	3.2	3.2	3.7	3.6	3.7	4.0	4.0	3.8	3.7	3.8	3.9
1976	4.1	4.2	4.3	4.1	4.0	3.8	3.8	3.8	3.7	3.6	3.9	4.1
1977	4.0	4.6	4.2	4.0	4.1	3.9	3.8	3.8	3.9	3.8	3.9	4.5
1978	4.0	4.0	4.0	4.2	4.1	P3.9						

$\tilde{\rho}$ = .50531292 $\hat{\rho}$ = .30782837

TABLE II.

MSE and Standard Error of MSE for the Estimates of the First Order AR Parameter.

$\rho \sim u(.455, .555)$, $\varepsilon \sim N(0, .002)$

Classical

$MSE(\hat{\rho}) = .0578743$

Standard Error = .003081243

Empirical Bayes

Experiences	$MSE(\tilde{\rho})$	S.E.	R
12	.0319950	.000996953	.5528

$\rho \sim \exp(.71)$, $\varepsilon \sim N(0, .002)$

Classical

$MSE(\hat{\rho}) = .0588847$

Standard Error = .002955245

Empirical Bayes

Experiences	$MSE(\tilde{\rho})$	S.E.	R
12	.0398491	.001249441	.6767

to yield valuable information about possible model inadequacies. Therefore, the comparison of $r_k(\hat{\varepsilon})$ with bounds $\pm$ 2 S.D. $(r_k(\hat{\varepsilon}))$ should provide a general indication of possible departure from white noise behavior in the ε_t. The results of diagnostic checks are shown in Table III. This indicates that a first order autoregressive model is appropriate.

In order to use equation (10.1.1) for forecasting layoff (or total accession) rates in a given year, for different months, the parameter ρ must be estimated. Table III shows that the values of the $\hat{\rho}$ vary from year to year in an unpredictable way. Therefore, let us assume that the values of the parameter (ρ) is a random variable from year to year. At the time we wish to estimate ρ_N (the present value of ρ, i.e., ρ_{77}), we have the sequence of observed values $\underline{X}_{66}$, $\underline{X}_{69}$, $\underline{X}_{70}$, ..., $\underline{X}_{76}$, $\underline{X}_{77}$. Let $z_{12} = (\underline{X}_{66}, \underline{X}_{67}, \ldots, \underline{X}_{77})$ with Z_{12} the corresponding random variable. In the empirical Bayes method we used z_{12} to estimate ρ_{77}. In the classical method we used $\underline{X}_{77}$ to estimate ρ_{77}.

The results of these estimation methods are shown in Table I. Note that the estimates do indeed differ. Which one is better remains to be shown.

10.2 The Simulation Study and its Results

In order to demonstrate the superiority of the empirical Bayes procedure, we decided to simulate data similar to the data of the previous example. ρ was constrained to be near .5, so that the variation in $\hat{\rho}$ would be similar to that of Table III. See Table IV.

Table III

$\hat{\rho}$	= classical estimate of ρ
$\tilde{\rho}$	= empirical Bayes estimate of ρ
$S.D.(\hat{\rho}_i)$	= standard deviation of $\hat{\rho}$

Table III (Cont'd.)

Layoffs

Year	$\hat{\rho}_i$	S.D.($\hat{\rho}_i$)	Test of Model	$\tilde{\rho}_i$
1966	.2393	.2603	yes*	.5194
1967	.8143	.1676	yes	.9341
1968	.9357	.1020	yes	.7814
1969	.7098	.2032	yes	.6779
1970	.3798	.2670	yes*	-.5619
1971	.2899	.2762	yes*	.6749
1972	.6668	.2152	yes	.8669
1973	.3170	.2737	yes*	.3942
1974	.7030	.2052	yes	.7215
1975	.7520	.1903	yes	.7813
1976	.8335	.1594	yes	.9154
1977	.4639	.2550	yes	.6015

*Indicates that AR(1) is an appropriate model but H_0: $\rho = 0$ is not rejected.

Table III (Cont'd.)

Total Accessions

Year	$\hat{\rho}_i$	S.D.$(\hat{\rho}_i)$	Test of Model	$\tilde{\rho}_i$
1966	.6660	.2154	yes	.80140
1967	.5992	.2311	yes	.51020
1968	.9401	.0985	yes	.97260
1969	.4172	.2623	yes*	.64470
1970	.9269	.1086	yes	.63611
1971	.9255	.1091	yes	.8447
1972	.9616	.0794	yes	.9851
1973	.8221	.1643	yes	.8754
1974	.7330	.1962	yes	.9980
1975	.9331	.1039	yes	.8261
1976	.7705	.1841	yes	.7913
1977	.3078	.2738	yes*	.5053

*Indicates that AR(1) is an appropriate model but H_o: $\rho = 0$ is not rejected.

Table IV

Sample of Simulated Data

													$\hat{\rho}$
1	1.3	1.2	1.4	1.4	1.4	1.4	1.6	1.5	1.8	1.5	1.4	1.6	.6392
2	1.6	1.5	1.4	1.2	1.2	1.3	1.2	1.6	1.4	1.4	1.5	1.5	.3818
3	1.5	1.2	1.1	1.3	1.6	1.5	1.3	1.1	1.0	1.6	1.7	1.7	.4632
4	1.7	1.6	1.5	1.4	1.3	1.5	1.5	1.6	1.4	1.2	1.2	1.4	.5231
5	1.4	1.4	1.5	1.6	1.7	1.6	1.5	1.3	1.2	1.3	1.3	1.0	.8782
6	1.0	1.3	1.6	1.4	1.3	1.4	1.6	1.6	1.3	1.1	1.2	1.7	.1934
7	1.7	1.6	1.6	1.9	1.4	1.3	1.5	1.2	1.0	1.0	1.0	.9	.8115
8	.9	1.2	1.2	1.3	1.0	1.3	1.3	1.3	1.3	1.4	1.3	1.3	.3581
9	1.3	1.2	1.1	1.1	1.1	1.0	1.3	1.5	1.3	1.2	1.0	1.2	.7484
10	1.2	1.4	1.3	1.5	1.6	1.5	1.5	1.5	1.2	1.5	1.5	1.1	.1204
11	1.1	1.3	1.5	1.5	1.4	1.4	1.5	1.5	1.4	1.5	1.6	1.3	.4724
12	1.3	1.1	1.4	1.4	1.5	1.3	1.4	1.4	1.4	1.5	1.5	1.4	.7643

The empirical Bayes estimator (4.2.1) was investigated for $N = 12$ (number of past experiences) so that the variation in $\hat{\rho}$ would be similar to that of Table III. The observation x_{it} were constructed according to the model. (See Table IV.)

$$(x_{i,t} - \mu) = \rho_i(x_{i,t-1} - \mu) + \varepsilon_{i,t}$$

where $i = 1, 2, \ldots, N$ denotes the particular experience and $t = 1, 2, \ldots, 12$ denotes the observation within the i^{th} experiment. The values for ρ_i, ε_{ij} were randomly generated using uniform and exponential density functions for ρ_i and normal density function for ε_{ij}.

For each experiment, the estimate $\hat{\rho}_i$ was found by (1.4.5) and stored. The empirical Bayes estimate was found using past (and present) values of $\hat{\rho}_i$, i.e., $\hat{\rho}_i$ $(i = 1, 2, \ldots, 12)$ were found and then $\hat{\rho}_i$. $i = 1, 2, \ldots, 12$ were used to obtain $\tilde{\rho}_{12}$ (empirical Bayes estimate of ρ_{12}).

In this manner a run of 12 experiences was simulated and empirical Bayes estimate was calculated. Since the true parameter values were known, the squared error for estimate could be found, and by averaging the results of 5000 repetitions of the entire run of 12 experiences, the mean squared errors were estimated. For example, $\tilde{\rho}_{12}$ was given independently 5000 times and the average of the squared error $(\tilde{\rho}_{12} - \rho_{12})^2$ was found.

The object of the Monte Carlo study is to compare the empirical Bayes estimator $\tilde{\rho}_N$ given by (4.2.1) with the conditional maximum likelihood estimator $\hat{\rho}_{12}$ given by (1.4.5).

Our criterion for comparison is mean squared error, and we therefore observe the ratio

$$R = \frac{\text{empirical Bayes mean squared error}}{\text{maximum likelihood mean squared error}}$$

Table II shows the value of R, MSE($\tilde{\rho}$), MSE($\hat{\rho}$), obtained for the time series of 12 observations for $\rho \sim U(.455, .555), \varepsilon \sim N(0, .002)$ and $\rho \sim$ exponential with parameter $\beta = .71$, $f(\rho) = \frac{1}{\beta} e^{-\rho/\beta}$, where $.455 < \rho < .530$, $\varepsilon \sim N(0, .002)$. It was found that the empirical Bayes estimators produced smaller mean squared errors than the maximum likelihood estimator (see Table II).

Note that the empirical Bayes estimators have about half the mean squared error of the maximum likelihood estimators.

10.3 Summary

In this dissertation empirical Bayes estimators were obtained for:

(1) The autoregressive model, $x_t = \sum_{j=1}^{P} \alpha_j x_{t-j} + \varepsilon_t$,

(2) The moving average model, $x_t = \sum_{j=1}^{q} \beta_j \varepsilon_{t-j} + \varepsilon_t$,

(3) The mixed autoregressive-moving average model,

$$x_t + \sum_{j=1}^{P} \alpha_j x_{t-j} = \varepsilon_t + \sum_{j=1}^{q} \beta_j \varepsilon_{t-j},$$

(4) The regression with time series errors, $y_t = \beta x_t + u_t$;

$u_t = \rho u_{t-1} + \varepsilon_t$,

(5) The spectral density function, $f(w)$,

(6) Multiple time series,

(7) Serial correlation,

(8) Unobservable variables in regression.

The monthly seasonally adjusted United States labor turnover rates in manufacturing were analyzed. It was found that the AR(1) is an appropriate model for layoff rates and total accession rates.

Empirical Bayes and classical estimates of AR(1) parameter, ρ, were computed.

By Monte Carlo simulation the empirical Bayes estimator of first order autoregressive parameter, ρ, was shown to have smaller mean squared errors than the conditional maximum likelihood estimator for 11 past experiences.

BIBLIOGRAPHY

1. Amini, S. B., and Khoshgoftaar, T. M., (1978). Term project concerning the analysis of the monthly seasonally adjusted United States labor turnover rates in manufacturing. M.I.T. Cambridge, Massachusetts. (Unpublished)

2. Blackman, R. B., and J. W. Tukey (1959). The measurement of power spectra from the point of view of communications engineering, Dover Publications, Inc., New York (printed originally (1958), Bell System Tech. J., 37, 185-282, 485-569).

3. Box, G. P., and Jenkins, G. M. (1976). Time series analysis forecasting and control. Holden-Day, Inc., 274-284.

4. Brillinger, D. R. (1969). Asymptotic properties of spectral estimates of second order, Biometrika, 56, 375.

5. Cacoullos, T. (1966). Estimation of a multivariate density. Ann. Inst. Statist. Math. (Tokyo) 18, 174-183.

6. Clemmer, B. A., and Krutchkoff, R. G. (1968). The use of Empirical Bayes approach in a linear regression model. Biometrika 55, 3, 525-534.

7. Cochrance, D., and Orcutt, G. H. (1949). Application of least squares regression to relationships containing auto-correlated error terms, Journ. Amer. Stat. Asso., 44, p. 32.

8. Daniell, P. J. (1946). Discussion on the symposium on autocorrelation in time series, J. R. Stat. Soc. (Suppl.) 8, 88-90.

9. Durbin, J. (1960). Estimation of parameters in time series regression models. J.R.S.S., B, 22, 139.

10. Fuller, W. A. (1976). Introduction to statistical time series, John Wiley & Sons, Inc., 139-171.

11. Goodman, N. R. (1957). On the joint estimation of the spectra, cospectrum and quadrature spectrum of a two-dimensional stationary Gaussian process. Sci. Paper No. 10, Engrng. Statist. Lab., New York Univ., New York.

12. Hannan, E. J. (1970). Multiple time series, John Wiley & Sons, Inc., 369-398, 475-485.

13. Krutchkoff, R. G. (1972). Empirical Bayes estimation, The American Statistician 26, No. 5.

14. Launer, R. L. (1970). Empirical Bayes Estimation in time series analysis, Ph.D. Thesis. Virginia Polytechnic Institute, Blacksburg, Virginia, 47-5.

15. Mann, H. B., and Wald, A. (1943). On the statistical treatment of linear stochastic difference equations. Econometrica, 11, 173.

16. Martz, H. F., Jr., and Krutchkoff, R. G. (1969). Empirical Bayes estimators in a multiple linear regression model. Biometrika 56, 367-373.

17. Parzen, E. (1962). On estimation of a probability density function and mode. Ann. Math. Statist. 33, 1065-76.

18. Rencher, A. C., and Krutchkoff, R. G. (1975). Some empirical Bayes estimators allowing for varying error variances. Biometrika 62, 643-650.

19. Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. Ann. Math. Stat. 35, 1-20.

20. Rutherford, J. R., and Krutchkoff, R. G. (1969). Some empirical Bayes technique in point estimation. Biometrika 56, 133-137.

21. Rutherford, J. R., and Krutchkoff, R. G. (1969). ε-asymptotic optimality of empirical Bayes estimators. Biometrika 56, 220-223.

22. Whittle, P. (1951). Hypothesis testing in time series analysis (Uppsala: Almqvist & Wiksells).

The vita has been removed from the scanned document

EMPIRICAL BAYES METHODS IN TIME SERIES ANALYSIS

by

Taghi M. Khoshgoftaar

Abstract

In the case of repetitive experiments of a similar type, where the parameters vary randomly from experiment to experiment, the Empirical Bayes method often leads to estimators which have smaller mean squared errors than the classical estimators.

Suppose there is an unobservable random variable θ, where $\theta \sim G(\theta)$, usually called a prior distribution. The Bayes estimator of θ cannot be obtained in general unless $G(\theta)$ is known. In the empirical Bayes method we do not assume that $G(\theta)$ is known, but the sequence of past estimates is used to estimate θ.

This dissertation involves the empirical Bayes estimates of various time series parameters: The autoregressive model, moving average model, mixed autoregressive-moving average, regression with time series errors, regression with unobservable variables, serial correlation, multiple time series and spectral density function. In each case, empirical Bayes estimators are obtained using the asymptotic distributions of the usual estimators.

By Monte Carlo simulation the empirical Bayes estimator of first order autoregressive parameter, ρ, was shown to have smaller mean squared errors than the conditional maximum likelihood estimator for 11 past experiences.