Empirical Bayes procedures in Time Series Regression Models

by

Ying-keh Wu

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Statistics

in

Department of Statistics

APPROVED:

Dr. Richard G. Krutchkoff, Chairman

Dr. Jesse C. Arnold

Dr. Yashaswini Mittal

Dr. Robert V. Foutz

Dr. Oral Capps, Jr.

May, 1986

Blacksburg, Virginia

Empirical Bayes procedures in Time Series Regression Models

by

Ying-keh Wu

Dr. Richard G. Krutchkoff, Chairman

Department of Statistics

(ABSTRACT)

Abstract

In this dissertation empirical Bayes estimators for the coefficients in time series regression models are presented. Due to the uncontrollability of time series observations, explanatory variables in each stage do not remain unchanged. A generalization of the reults of O'Bryan and Susarla is established and shown to be an extension of the results of Martz and Krutchkoff .

Alternatively, as the distribution function of sample observations is hard to obtain except asymptotically, the results of Griffin and Krutchkoff on empirical linear Bayes estimation are extended and then applied to estimating the coefficients in time series regression models. Comparisons between the performance of these two approaches are also made.

Finally, predictions in time series regression models using empirical Bayes estimators and empirical linear Bayes estimators are discussed.

Acknowledgements

Acknowledgments

Most of the work done in this dissertation is carried out while the student is supervised by Professor R. G. Krutchkoff. The student wishes to express his earnest appreciation to Dr Krutchkoff for his invaluable guidance, encouragement and critical advice throughout the investigation. Dr Krutchkoff is readily available to provide assistance when needed. The student would also like to acknowledge the stimulating comments made by the members of his committee: Professors J. Arnold, Y. Mittal, R. Foutz and O. Capps.

The student would also like to thank and give special appreciation to Professor
Head of the Department of Statistics, VPI & SU for the financial support and his encouragement during the student's tenure.

Finally, in dedicating this study to my wife and family, I am acknowledging their unfailing support, without which the study would never have been materilized.

Table of Contents

TABLE OF CONTENTS

Chapter 1 Introduction

1.1 General Description

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,...,T, are taken to be representative of some population. The statistical analysis that then follows is largely concerned with making inferences about the features of the population from which the sample comes. 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.

In econometrics, as in many other scientific disciplines, a research worker is constantly faced with the fundamental problem of arriving at an appropriate model for the explanation and hence the prediction of various economic phenomena. One main source of data used for such purposes is time series observations which have become increasingly available at various time intervals such as annually, quarterly or even monthly periods. The main feature of time series observations which distinguishes them from other sources is their time dependence property. In order to make efficient use of such data, it is essential that we should consider dynamic specifications in both the deterministic part(or mean part; systematic part) and the stochastic part(or disturbance part) of a model. We assume that a model is fully specified by the joint probability density of its dependent variables. The mean of such a distribution, which specifies only a portion of the model, will be referred to as the deterministic part, and when we are dealing with normally distributed variates we only need to consider their variance-covariance matrix which completely defines the stochastic part. Models which have a dynamic specification in the deterministic part are usually referred to in the literature as 'distributed lag' models; and those which have a dynamic specification in the stochastic part are called 'autocorrelated disturbance' models.

To illustrate the problem more clearly, consider a research worker in the U. S. Department of Interior who is concerned with the relationship among household consumption y_t and other influences x_t (like income, asset holdings, etc.) and also previous consumption levels y_{t-1}--- with this latter relation reflecting how past consumption habits and patterns may influence present decisions. The research worker assumes that the relation is of the form $y_t = \alpha y_{t-1} + \beta x_t + \varepsilon_t$, where α and β are the unknown parameters, and ε_t is a random disturbance. In order to use this equation for predicting household consumption level in a given year for different months, the parameters α and β need be estimated. This is done by obtaining monthly data for twelve months based on the consumption in the previous years. For different years, one might expect a similar relationship to hold among y_t, y_{t-1} and x_t .

It seems reasonable that for different years the parameters 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 y_t, y_{t-1} . Therefore, for different years, the parameters take on different values. Although these values probably change slowly, it is easier to consider the values constant for a particular run, considered here to be a year. Here we assume that the observations in distinct year are independent.

The estimate of the parameters α and β for a given year, therefore, 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 as well as the present year.

It should be recalled, however, that the parameters themselves are different from year to year, and thus it is obviously not proper to estimate the parameters by pooling all the data.

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

The main purpose of this example is to show how additional estimates of the parameters can be used to improve the estimate in the present situation.

The underlying distribution of the parameters will be referred to as a prior distribution. However, the exact form of the prior distribution will not be known to the research worker. In the empirical Bayes and empirical linear Bayes approaches, the research worker does not assume any known form for the prior distribution of the parameters, but obtains the estimators by means of supplementary information from similar independent experiments.

1.2 The Purpose of The Dissertation

In this study we try to the employ empirical Bayes(EB) approach to time series regression models (TSRM) which will be described in chapter 4. Notice that applying the EB approach to stochastic processes is quite different from applying it to classical statistics. Ordinarily, there is only one realization (sample function) of observations available in a stochastic process, and the distribution of that realization is usually hard to find except asymptotically. Firstly, we need to justify the use of EB estimation theory for the TSRM which is a special kind of stochastic processes.

Alternatively, we introduce another approach called the empirical linear Bayes (ELB) approach complementary to the EB approach. The ELB approach needs only the first two moments of the coefficients (treated as random variables) to be estimated. The purpose of this study is to give the EB and ELB approaches for estimating coefficients in a TSRM, assuming the availability of a sufficiently large number of independent past observations.

In chapter 2 we describe briefly the EB and ELB approaches and show how they may be exploited in a linear stochastic difference equation. We then extend the EB estimation theory for non-identical but independent component problems in the multivariate normal case in chapter 3. In chapter 4 we introduce the time series regression model and use it as an example, estimating the coefficients by invoking the results developed in chapter 3. We investigate ELB estimation theory and its applications in the next two chapters. In chapter 7 we discuss predictions for those models by utilizing the EBE and the ELBE we established in previous chapters. Finally, some conclusions and comparisons are given in the last chapter.

1.3 A Review of The Literature

In order to apply the EB approach to stochastic models, which was this studies' initial goal, I read Launer's dissertation (1970): " Empirical Bayes Estimation in Time Series" and Khoshgoftaar's (1982): " Empirical Bayes Methods in Time Series Analysis." Both of these have obtained some results in estimating coefficients in time series models by utilizing the EB approach similar to that of Martz and Krutchkoff (1969). However, Launer didn't investigate TSRM, while Khoshgoftaar concentrated his study on the frequency domain, rather than on the time domain. In this respect, Li and Hui published two papers (1983 a,b) using EB approach similar to that of Martz and Krutchkoff (1969) to estimate random coefficients of an ARMA model. The methods they used were simple and their estimates were of considerable improvement over the LSE or MLE.

In 1979 two German researchers, Bunke and Gladitz, applied the EB approach to identify (estimate) random parameters in linear stochastic difference equations. However, they restricted the EB approach to a parametric version, (that is, assuming that random coefficients have the joint normal distribution function but with distributional parameters, such as mean and variance unknown). Their method, strictly speaking, is not an EB approach. (This approach is called semi-empirical Bayes in Martz's dissertation (1967) and type II maximum likelihood in Good's book (Good, 1965). Bunke and Gladitz gave a complete investigation in convergence properties such as consistency and asymptotic optimality for the estimators. In this study we shall generalize their approach by relaxing the known (normal) distribution to the wide sense distribution, leaving the first two moments to be estimated.

In another paper " Empirical linear Bayes decision rules for a sequence of linear models with different regressor matrices" Bunke and Gladitz (1974)try to avoid the use of the consistent estimators for functionals (which include a relatively complicated density estimation) in the EB approach, but suggest instead the ELB approach. The ELB approach originates from Griffin and Krutchkoff (1971). It turns out that ELBE is also asymptotically optimal.

Since time series observations are different from the usual repeated sampled statistics, the classical empirical Bayes estimation theory cannot be applied exactly to the time series model. Thus, the modification of the theory such that it can be applied correctly must be the first issue of this study. To this end, the literature concerning empirical Bayes methods was surveyed. Some results developed by O'Bryan and Susarla (1975) can be adapted to those models. They weaken the restriction that the component problems need to be identically distributed. As one knows, it is not reasonable in time series regression models to assume that X (matrix of regressors) remains unchanged, since X is also a kind of stochastic process itself. By extending the results developed by O'Bryan and Susarla to the multivariate case, we are able to use the EB approach correctly in estimating the coefficients in a TSRM.

Appendix :

Important results used in later chapters

(1) C_δ - Inequality

$$E\,\|X + Y\|^\delta \le 2^{(\delta-1)^+}(E\,\|X\|^\delta + E\,\|Y\|^\delta) \quad \text{for } \delta > 0$$

where X and Y are random vectors. For details see Loeve (1963).

(2) Theorem 2.9 (Lehmann, 1959)

Let φ be any bounded measurable function on $(\mathcal{X}, \mathcal{A})$. Then

(i) the integral

$$\int \varphi(x)\ \exp[\sum_{j=1}^{k} \theta_j T_j(x)]d\mu(x) \qquad (*)$$

considered as a function of the complex variables $\theta_j = \xi_j + \eta_j$ ($j = 1, \dots, k$) is an analytic function in each of these variables in the region R of parameter points for which $(\xi_1, \dots, \xi_k)$ is an interior point of the natural parameter space Ω ;

(ii) the derivative of all orders with respect to the θ's of the integral (*) can be computed under the integral sign.

(3) Chebyshev Weak Law of Large Numbers (WLLN) :

Let $\xi_1, \xi_2, \dots$ be random variables , and let m_n and σ_n denote the mean and the s.d. of ξ_n . If $\sigma_n \to 0$ as $n \to \infty$, then $\xi_n - m_n$ converges in probability to zero .

Corollary

If $\xi_1, \dots, \xi_n, \dots$ are independent and we define

$$\xi_n = \frac{1}{n}\sum_{\nu=1}^{n} \xi_\nu, \quad \overline{m}_n = \frac{1}{n}\sum_{\nu=1}^{n} m_\nu.$$

then $\sum_{\nu=1}^{n} \sigma_\nu^2 = o(n^2)$ implies that $\xi_n - \overline{m}_n$ converges in probability to zero .

(4) Lemma (Okamoto 1973 Ann. Stat. 763-765)

Let ($y_1, \dots, y_T$) possess for all $T \in N$ an absolutely continuous distribution w.r.t. the T - dim. Lebesgue measure. Further if there is a $t_1 \in N \cup \{0\}$ such that rank [$(\mathbf{x}_1, \dots, \mathbf{x}_{t_1})$] = q, then it holds that rank [Z_T] = 1 + q a.s. for $T \geq \max(t_1, q+2)$. Here $\mathbf{x}_t$ is q-dim nonrandom vectors, and

$$Z_T = \begin{bmatrix} y'_0 & \mathbf{x}'_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ y'_{T-1} & \mathbf{x}'_T \end{bmatrix}$$

From the Lemma we can assume that $Z'_T Z_T > \mathbf{0}$.

Chapter 2 Survey of Empirical Bayes Estimator and Empirical Linear Bayes Estimator

2.1 The Parameters Are Randomly Varying

Essentially, a statistical model is an approximation of reality and, therefore, is subject to a number of misspecifications such as the exclusion of relevant variables, wrong choice of functional forms, etc. These misspecifications often lead to random parameters. Consider, for example, the regression of the quantity demanded of a certain commodity on its price. If time series data are used on these variables, it is quite likely that the price elasticity of demand will not remain the same over the sample period. Consider, also, the example of a regression relationship between savings and incomes of a sample of households. The marginal propensity to save is likely to differ for each household because of different average ages and wealth holdings. Thus, the assumption that the parameters are fixed for different periods of observation is restrictive and usually incorrect.

Problems in system identification often lead, in their mathematical description, to a sequence of models with randomly varying parameters. For instance, in chemical plants the process parameters must be estimated periodically. In different, approximately independent, periods the corresponding parameters may be different for several reasons, such as fret, repair, and changing outside conditions. Sometimes, it is reasonable to assume that the changing parameters are generated by a random source.

2.2 Bayes Estimation and Linear Bayes Estimation

As was showed in the last section, parameters may be at random in time series situations. The question now is how to estimate those random parameters when observations are available. The approaches we shall employ are the empirical Bayes estimation procedure and the empirical linear Bayes estimation procedure.

Let $(\underline{\theta}, \mathrm{X})$ be a random vector such that

(1) $\underline{\theta}$ has a distribution function G (called the prior distribution)

and

(2) Conditionally on $\underline{\theta}$, X has a probability function $f(\, \mathbf{x} \mid \underline{\theta}\,)$ of known form with respect to a σ-finite positive product measure $\mathbf{m}$.

We want to estimate $\underline{\theta}$ by some function $\mathbf{t} = \mathbf{t}(\mathbf{x})$ under the squared loss function

$$L(\mathbf{t}, \underline{\theta}) = (\mathbf{t} - \underline{\theta})'(\mathbf{t} - \underline{\theta})$$

To minimize the expected squared loss function of the estimate $\mathbf{t}$:

$$E((\mathbf{t} - \underline{\theta})'(\mathbf{t} - \underline{\theta})) = E[\ \mathrm{tr}(\mathbf{t} - \underline{\theta})(\mathbf{t} - \underline{\theta})']$$

[1] within the class of linear functions $\mathbf{a} + B\mathrm{X}$ of X for given G, we have

$$\bar{\mathbf{t}}(\mathrm{X}) = E(\underline{\theta}) + \mathrm{cov}(\underline{\theta}, \mathrm{X})(\mathrm{var}\mathrm{X})^{-1}(\mathrm{X} - E\mathrm{X})$$

and

$$E[(\tilde{\mathbf{t}} - \underline{\theta})'(\tilde{\mathbf{t}} - \underline{\theta})] = \text{tr}\ [\text{var}(\underline{\theta}) - \text{cov}(\underline{\theta}, \mathbf{X})(\text{var}\ \mathbf{X})^{-1}\ \text{cov}(\underline{\theta}, \mathbf{X})']$$

[2] within the class of all Borel functions of X for given G, we have

$$\mathbf{t}^*(\mathbf{X}) = E(\underline{\theta} \mid \mathbf{X}) = \frac{\int \underline{\theta} f(\mathbf{X} \mid \underline{\theta}) dG(\underline{\theta})}{\int f(\mathbf{X} \mid \underline{\theta}) dG(\underline{\theta})}$$

and

$$E(\mathbf{t}^* - \underline{\theta})'(\mathbf{t}^* - \underline{\theta}) = E\ [\text{tr}\ \text{Var}(\underline{\theta} \mid \mathbf{X})].$$

Here, $\tilde{\mathbf{t}}(\mathbf{x})$ is called the linear Bayes estimator for $\underline{\theta}$, while $\mathbf{t}^*(\mathbf{x})$ is called the Bayes estimator for $\underline{\theta}$

.

2.3 Empirical Bayes Estimator and Empirical Linear Bayes Estimator

Notice that if G is unknown (in practice this is always the case), neither $\tilde{\mathbf{t}}$ nor $\mathbf{t}^*$ can be used directly to estimate $\underline{\theta}$. In such case, we need to consider the empirical Bayes and empirical linear Bayes approaches.

Suppose that there are a large number N of independent versions of the component problem : ($\underline{\theta}_i$, $\mathbf{X}_i$) for i = 1,...N are i i d random vectors such that

(i) the $\underline{\theta}_i$ have the distribution function G

and

(ii) the $\mathbf{X}_i$ have the conditional probability density function $f(\mathbf{x} \mid \underline{\theta}_i)$.

By observing $\mathbf{X}_1, \dots, \mathbf{X}_N$ which are i i d random variables with marginal pdf

$$f(\mathrm{x}) = \int f(\mathrm{x} \mid \underline{\theta})dG(\underline{\theta}),$$

we can try to gather enough information about G to approximate $\tilde{\mathbf{t}}(\mathrm{X}) = E(\underline{\theta}) + \operatorname{cov}(\underline{\theta}, \mathrm{X})(\operatorname{Var} \mathrm{X})^{-1}(\mathrm{X} - E\mathrm{X})$ (a modest aim), or $\mathbf{t}^{*}(\mathrm{X}) = E(\underline{\theta} \mid \mathrm{X})$ (a more grandiose one).

The methods that seek to approximate $\tilde{\mathbf{t}}(\mathrm{X})$ are called *the empirical linear Bayes* (ELB) approach, and to those that seek to approximate $\mathbf{t}^{*}(\mathrm{X})$ are called *the empirical Bayes* (EB) approach . For further details see chapter 3 and chapter 5 .

2.4 Different Procedures For Obtaining Observations

In the literature of stochastic process there are two types of observational procedures that are commonly discussed. One either observes a single realization $\{X(u), 0 \le u \le t\}$ completely over a fixed interval [0 , t] of time, or observes n (> 1) independent and identical copies of the process, each over the fixed duration [0 , t].

It is almost impossible, as we shall soon discover, to obtain precise statistical properties of the usual estimators and to devise tests of hypotheses for a finite number N of independent copies of the process. Thus, our attention in this area has been mainly concentrated on obtaining the asymptotic properties of the estimators and on devising a test-statistics whose asymptotic distribution can be obtained.

Corresponding to the two procedures of observation mentioned above, there are two methods of studying the asymptotic properties of the different procedures in statistical inference. When a single realization of the process is observed, the asymptotic properties are obtained as the period t of observation becomes large. In the second situation, t is kept fixed and the number N of observed realization is allowed to grow to infinity. In this second situation, we do not face any major new problems since the classical theory of statistics for i i d random variables is still applicable. We shall, therefore, concentrate on studying the asymptotic properties for a single realization of a stochastic process as the period t of observation tends to infinity. It is for this reason that we need establish new empirical Bayes estimation techniques different from those contributed by Robbins (1956 , 1964).

There is still one more thing to be noted. Let us suppose that the observations are regarded as a single realization from some underlying data generation process. Unlike the second situation, only one set of observations can be obtained in practice. However, it is possible to regard the underlying data generation process as being capable, in principle, of producing an infinite number of realizations over that same period. It is on this basis that the properties of the stochastic process are derived and statistical inferences made. Furthermore, we assume the existence of a population of stochastic processes and an underlying mechanism which generate independent parameters for each member of the population. Notice that we are not assuming the replication of a stochastic process with fixed parameters.

There are many situations in which the previous assumptions might be valid. Consider, for example, the problem of estimating the parameters of an economic time series. If the time series represents the value of stock of an electronic corporation, then we might as well assume that this corporation is a member of the population of all electronics firms.

We often find situations which assume that one single realization of a stochastic process is composed of short pieces of several realizations, each with parameters different but similar to the others. If we need an estimate of a parameter over the whole range, the usual approach is to pool all the data together. However, since the pieces are not identical to each other (due to the randomly varying parameters), the value of the parameter for a given piece will not be same as that of the other pieces. Intuitively, it seems that a ' better ' estimate could be obtained by using all the individual pieces, than by straightforward pooling.

Here we assume that the values of the parameter vary in an unpredictable manner from piece to piece of the single realization, that means they behave like random variables. If the estimate of the parameter for each pieces was known, then the frequency of each piece might provide evidence for some underlying distribution on the parameter. In doing this we must assume that the situations for each piece are independent. There are many phenomena in which the previous assumptions might be valid. For instance, a single realization can only be observed over certain periods of time, and the time gaps between successive periods are large so that each observable periods may be considered as an independent realization. One such example of the above would be the amount of dissolved oxygen in a lake system which freezes in a long winter. Under this situation, the process parameters may also fluctuate for certain mean values from year to year. The entire history of the system should then be used in process model estimation.

In view of this description, we see that it may be reasonable to assume that the parameters behave as random variables. We can, therefore, consider some unspecified underlying joint distribution on the parameters for which the unknown set of parameters for each year is an observation from this unknown distribution referred to as a 'prior' distribution.

It is usually the case, however, that the exact form of the prior distribution will be unknown. In the empirical Bayes approach, presented in the following chapter, we do not apply the form of this distribution assumed to be unknown. Such an empirical Bayes approach, in the words of Robbins, " offers certain advantages over any approach which ignores the fact that the parameters are themselves random variables as well as over any approach which assumes a personal or conventional distribution of the parameters not just subject to change with experience" (Robbins, 1964).

On the other hand , in the empirical linear Bayes approach the researchers try to estimate the first two moments of the prior distribution using past independent pieces of observations by usual statistical moment estimation methods.

Chapter 3 An Empirical Bayes Estimation Problem

WITH NONIDENTICAL COMPONENTS INVOLVING NORMAL DISTRIBUTIONS ----

THE MULTIVARIATE CASE

1. Introduction

O'Bryan and Susarla (1975) discussed the empirical Bayes decision problem, where the component problems are non-identical in a wide sense. Specifically they exhibited the empirical Bayes procedures when the component problem is the squared-error loss estimation of θ, where θ, a random variable, has an unknown distribution G on some subset of the real line, based on an observation X from $f(.\mid\theta)$, the univariate normal density with mean $a\theta + b$ and variance σ^2, where $a \neq 0$ and b and σ^2 may vary with the component problem. In this chapter we attempt to extend their result to the multivariate case, namely empirical Bayes procedures are exhibited when the component problem is the squared error loss estimation of $\underline{\theta}$, where $\underline{\theta}$, a k-vector random variable, has an unknown distribution G on some subset of the R^k, based on an observation X from $f(.\mid\underline{\theta})$, the multivariate normal density with mean $A\underline{\theta} + \mathbf{b}$ and variance Σ where matrix $A_{k\times k}$ (invertible), and the k-vector $\mathbf{b}$ and the covariance matrix $\Sigma_{k \times k}$ may vary with the component problem. (hereafter referred to as the $A - \mathbf{b} - \Sigma$ component problem).

In the empirical Bayes estimation problem there are sequences $\{A_j\}$,$\{\mathbf{b}_j\}$ and $\{\Sigma_j\}$ of known quantities where A_j is invertible and Σ_j positive definite for j = 1 , 2 ,... , and a sequence $\{(\underline{\theta}_j, \mathbf{X}_j)\}$ of independent random vectors where the unobservable $\underline{\theta}_j$ are i i d G and, for $\underline{\theta}_j = \underline{\theta}$, $\mathbf{X}_j$ has density $f(. \mid \underline{\theta})$. With $\mathbf{t}_n(\mathbf{X}_{n+1}) = \mathbf{t}_n(\mathbf{X}_1, \dots, \mathbf{X}_n; \mathbf{X}_{n+1})$ a non-randomized estimator for use in the (n+ 1)st problem, its risk is given by

$$R_n(\mathbf{t}_n, G) \equiv E\|\mathbf{t}_n - \underline{\theta}_{n+1}\|^2 = E(\mathbf{t}_n - \underline{\theta}_{n+1})'(\mathbf{t}_n - \underline{\theta}_{n+1}) \tag{3.1}$$

where E denotes expectation with respect to all random variables involved. With $R_{n+1}(G)$ denoting the infimum Bayes risk in the $A_{n+1} - \mathbf{b}_{n+1} - \Sigma_{n+1}$ component problem, the fact that

$$R_n(\mathbf{t}_n, G) \geq R_{n+1}(G) \tag{3.2}$$

motivates the following definition.

Definition A sequence of rules $\{\mathbf{t}_n\}$ is said to be *asymptotically optimal with order $g(n)$* (denoted hereafter by a.o. (g(n))) for some function $g(n) \to 0$ as $n \to \infty$ if for some constant $c > 0$,

$$0 \leq R_n(\mathbf{t}_n, G) - R_{n+1}(G) \leq cg(n) \tag{3.3}$$

In the $A_{n+1} - \mathbf{b}_{n+1} - \Sigma_{n+1}$ component problem, a non-randomized estimator based on $\mathbf{X}_{n+1} = \mathbf{x}$ which is Bayes with respect to G is given similar to Miyasawa (1960) by

$$\mathbf{t}_{G,n}(\mathbf{x}) = A_{n+1}^{-1}(\mathbf{x} - \mathbf{b}_{n+1}) + A_{n+1}^{-1}\Sigma_{n+1}\frac{p'_{n+1}(\mathbf{x})}{p_{n+1}(\mathbf{x})} \tag{3.4}$$

where for $j = 1, 2, \dots$

$$p_j(\mathbf{x}) = \frac{1}{(2\pi)^{k/2}|\Sigma_j|^{1/2}}\int_{R^k} \exp\ \{\frac{-1}{2}(\mathbf{x} - A_j\underline{\theta} - \mathbf{b}_j)^T\Sigma_j^{-1}(\mathbf{x} - A_j\underline{\theta} - \mathbf{b}_j)\}dG(\underline{\theta}) \tag{3.5}$$

It is well-known that if $\underline{\theta}$ and $\mathbf{t}_n$ are L_2 vector random variables, then

$$R_n(\mathbf{t}_n,G) - R_{n+1}(G) = E\|\mathbf{t}_n - \mathbf{t}_{G,n}\|^2 \tag{3.6}$$

so that in order to define $\{\mathbf{t}_n\}$ satisfying (3.3), it seems reasonable to define $\mathbf{t}_n$ to be a natural estimator of $\mathbf{t}_{G,n}$. In view of (3.4), such a $\mathbf{t}_n$ can be defined by taking the ratio of estimators of p'_{n+1} and p_{n+1}. The definition of these estimators and their expected mean squared errors are the subjects of the next section which discusses the more general problem of estimating p_{n+1} and its kth derivative for k = 1,2,.... Section 3.3 exhibits $\{\bar{\mathbf{t}}_n\}$ satisfying (3.3) when G has support in $[T_1,T_2]_k$ while section 3.4 exhibits $\{\mathbf{t}_n^*\}$ satisfying (3.3) when G has a $(2 + \gamma)$th moment for some $\gamma > 0$.

3.2 Estimation of a Density and its lth Derivative

By Fubini's theorem for j = 1,2,..., the characteristic function of k-variate $\mathbf{X}_j$ is given by

$$\varphi_j(\mathbf{t}) = E[\exp\{i\mathbf{t}^T\mathbf{X}_j\}] = \exp\{-\mathbf{t}^T\Sigma_j\mathbf{t}/2 + i\mathbf{t}^T\mathbf{b}_j\}\int_{R^k} \exp\{i\mathbf{t}^T A_j\underline{\theta}\}dG(\underline{\theta}). \tag{3.7}$$

Abbreviating A_{n+1}, $\mathbf{b}_{n+1}$, Σ_{n+1} and φ_{n+1} by A, $\mathbf{b}$, Σ, and φ respectively for simplicity, consider the problem of estimation of p_{n+1} and its lth derivative $p_{n+1}^{(l)}$ using $\mathbf{X}_1, \ldots, \mathbf{X}_n$ for $l = 1, 2, \ldots\ldots$ Since $\int_{R^k}|\varphi(\mathbf{t})|dt < \infty$, it follows (see p. 333 Billingsley (1979)) that

$$p_{n+1}(\mathbf{x}) = (2\pi)^{-k}\int_{R^k} \exp\{-i\mathbf{t}^T\mathbf{x}\}\varphi(\mathbf{t})d\mathbf{t}. \tag{3.8}$$

It immediately follows from (3.7) and (3.8) that, for any $0 < M_i < \infty$, i = 1, ..., k

$$
\begin{aligned}
&|(2\pi)^k p_{n+1}(\mathbf{x}) - \int_{-\mathbf{M}}^{\mathbf{M}} \exp\{-i\mathbf{t}^T\mathbf{x}\}\varphi(\mathbf{t})d\mathbf{t}| \\
&\leq (2\pi)^k[\int_{-\infty^k}^{-\mathbf{M}} + \int_{\mathbf{M}}^{\infty^k} \exp\{-\lambda_{0,n+1}\mathbf{t}^T\mathbf{t}/2\}d\mathbf{t}] \quad \text{where } \lambda_{0,n+1} = \text{ min root of } \Sigma_{n+1} \\
&\leq 2(2\pi)^k \int_{\mathbf{M}}^{\infty^k} \exp\{-\lambda_{0,n+1}\mathbf{t}^T\mathbf{t}/2\}d\mathbf{t} \\
&\leq 2(2\pi)^k \exp\{-\lambda_{0,n+1}\mathbf{M}^T\mathbf{M}/2\}
\end{aligned}
\tag{3.9}
$$

where $\mathbf{M} = (M_1, \ldots, M_k)$, $\infty^k = (\infty, \ldots, \infty)$ (k components) . (For details see the appendix at the end of this chapter) .

Since for $j = 1, \ldots, n$

$$
\varphi(\mathbf{t}) = \exp\{\mathbf{t}^T(AA_j^{-1}\Sigma_j A_j^{-1^T}A^T - \Sigma)\mathbf{t}/2 + i\mathbf{t}^T(\mathbf{b} - AA_j^{-1}\mathbf{b}_j)\}\varphi_j(A_j^{-1^T}A^T\mathbf{t}) \tag{3.10}
$$

an unbiased estimator of $\int_{-\mathbf{M}}^{\mathbf{M}}\exp\{-i\mathbf{t}^T\mathbf{x}\}\varphi(\mathbf{t})d\mathbf{t}$ based on $\mathbf{X}_1, \ldots, \mathbf{X}_n$ is $\hat{p}_{n+1,\mathbf{M}}(\mathbf{x})$ where

$$
\begin{aligned}
(2\pi)^k\hat{p}_{n+1,\mathbf{M}}(\mathbf{x}) = \frac{1}{n}\sum_{j=1}^{n} \text{Re} \int_{-\mathbf{M}}^{\mathbf{M}} &\exp\{\mathbf{t}^T(AA_j^{-1}\Sigma_j A_j^{-1^T}A^T - \Sigma)\mathbf{t}/2\} \\
&\exp\, i\mathbf{t}^T\{\mathbf{b} - \mathbf{x} - (AA_j^{-1}(\mathbf{b}_j - \mathbf{X}_j))\}d\mathbf{t}.
\end{aligned}
\tag{3.11}
$$

($\varphi(\mathbf{t}) = E_{\mathbf{X}}(\exp\, i\mathbf{t}^T\mathbf{X})$ implies that $\exp\,\{i\mathbf{t}^T\mathbf{X}\}$ is an unbiased estimator of $\varphi(\mathbf{t})$.) Since each summand in (3.11) is bounded in absolute value by $c_2(\mathbf{M}^T\mathbf{M})^{k/2}\ \exp\{\mathbf{M}^T\mathbf{M}\rho_n/2\}$ where

$$
\rho_n = \max_{1 \leq j \leq n} |\text{ roots of } (AA_j^{-1}\Sigma_j A_j^{-1^T}A^T - \Sigma)| \tag{3.12}
$$

and c_2 is a constant depending on k and $\mathbf{M}$, the independence of $\mathbf{X}_1, \ldots, \mathbf{X}_n$ yields

$$
\text{var}\ (\hat{p}_{n+1,\mathbf{M}}(\mathbf{x})) \leq n^{-1}c_3(\mathbf{M}^T\mathbf{M})^k \exp\{\mathbf{M}^T\mathbf{M}\rho_n\} \tag{3.13}
$$

where $c_3 = (2\pi)^{-2k}c_2^2$. Combining (3.9) and (3.13) yields for $M_i > 1$, i = 1, ..., k

$$
\begin{aligned}
&\operatorname*{lub}_{\mathbf{x}}\ E(\hat{p}_{n+1,\mathbf{M}}(\mathbf{x}) - p_{n+1}(\mathbf{x}))^2 \\
&\leq 2c_1^2 \exp\{-\lambda_{0,n+1}\mathbf{M}^T\mathbf{M}\} + 2n^{-1}c_3(\mathbf{M}^T\mathbf{M})^k \exp\{\mathbf{M}^T\mathbf{M}\rho_n\}
\end{aligned}
\tag{3.14}
$$

Let $l \in \{1, 2, \dots\}$. Since $\int_{R^k} | t_s^l \; \varphi(\mathbf{t}) | dt < \infty$,

$$p_{n+1,s}^{(l)}(\mathbf{x}) = (2\pi)^{-k} \int (it_s)^l \; \exp\{ - i\mathbf{t}^T\mathbf{x}\} \varphi(\mathbf{t}) d\mathbf{t} \tag{3.15}$$

where

$$p_{n+1,s}^{(l)}(\mathbf{x}) = \frac{\partial^l p_{n+1}(\mathbf{x})}{\partial x_s^l}, \quad \begin{aligned} s &= 1, \dots, k \\ \mathbf{t} &= (t_1, \dots, t_k)^T, \\ \mathbf{x} &= (x_1, \dots, x_k)^T. \end{aligned}$$

and

$$p_{n+1}^{(l)}(\mathbf{x}) = (p_{n+1,1}^{(l)}(\mathbf{x}), \dots, p_{n+1,k}^{(l)}(\mathbf{x})).$$

Following a similar procedure it can be shown that for $0 < M_i < \infty, i = 1, \dots, k$.

$$\begin{aligned} (2\pi)^k \hat{p}_{n+1,s,\mathbf{M}}^{(l)}(\mathbf{x}) = \frac{1}{n} \sum_{j=1}^{n} \mathrm{Re} \int_{-\mathbf{M}}^{\mathbf{M}} (it_s)^l \; &\exp\{\mathbf{t}^T(AA_j^{-1}\Sigma_j A_j^{-1^T} A^T - \Sigma)\mathbf{t}/2\} \\ &\exp\{i\mathbf{t}^T(\mathbf{b} - \mathbf{x} - AA_j^{-1}(\mathbf{b}_j - \mathbf{X}_j))\} d\mathbf{t} \end{aligned} \tag{3.16}$$

for s = 1, ..., k. This yields an unbiased estimator of $p_{n+1,s}^{(l)}(\mathbf{x})$ with the integral truncated to the k dimensional hypercube $(- \mathbf{M}, \mathbf{M})_k$ and that for $M_i > 1, i = 1, \dots, k$

$$\begin{aligned} &\operatorname{lub}_{\mathbf{x}} E\|\hat{p}_{n+1,\mathbf{M}}^{(l)}(\mathbf{x}) - P_{n+1}^{(l)}(\mathbf{x})\|^2 \\ &\le 2 \sum_{s=1}^{k} \operatorname{lub}_{\mathbf{x}} E(\hat{p}_{n+1,s,\mathbf{M}}^{(l)}(\mathbf{x}) - p_{n+1,s}^{(l)}(\mathbf{x}))^2 \\ &\le c_4 n^{-1} (\frac{\mathbf{M}^T\mathbf{M}}{k})^{k+l} \exp(\rho_n \mathbf{M}^T\mathbf{M}) + 2^{2k+1} c_5 (\mathbf{M}^T\mathbf{M})^{l-1} \; \exp(- \lambda_{\min}\mathbf{M}^T\mathbf{M}) \end{aligned} \tag{3.17}$$

Remark 1 In case of simple identical component problems, that is , $A_j = I, \mathbf{b}_j = 0$, $\Sigma_j = \Sigma, j = 1, 2, \dots$, (3.11) becomes:

$$\hat{p}_{n+1,\mathbf{M}}(\mathbf{x}) = \frac{2}{n(2\pi)^k} \sum_{j=1}^{n} \prod_{h=1}^{k} \frac{\sin M_h(X_{hj} - x_h)}{(X_{hj} - x_h)} \tag{3.18}$$

where $\mathbf{x} = (x_1, \ldots, x_k)'$ and $\mathbf{X}_j = (X_{1j}, \ldots, X_{kj})'$, and for $l = 1$ (3.16) becomes

$$\hat{p}'_{n+1,s,\mathbf{M}}(\mathbf{x}) = \frac{2}{n(2\pi)^k}\sum_{j=1}^{n}$$

$$\{\prod_{\substack{h=1 \\ h\neq s}}^{k} \frac{\sin M_h(X_{hj} - x_h)}{(X_{hj} - x_h)}[\frac{\sin M_s(X_{sj} - x_s) - (X_{sj} - x_s)M_s \cos M_s(X_{sj} - x_s)}{(X_{sj} - x_s)^2}]\}. \tag{3.19}$$

Notice that the estimators of components of $p'_{n+1}(\mathbf{x})$ are all closed forms in contrast to what occurs in kernel density estimation.

Remark 2 In case of the simple component problem --- i i d but different sample sizes: $A_n = m_n I, \mathbf{b}_n = \mathbf{0}, \Sigma_n = m_n\Sigma, n = 1, 2, \ldots$. When $\{m_n\}$ is a sequence of positive integers such that there exists some positive integer $\overline{m}$ such that $m_n < \overline{m} < \infty$ and $\Sigma > 0$. This case would result when $\mathbf{X}$ is the sufficient statistic from a sample size m in the $I - 0 - \Sigma$ component problem.

The following theorem establishes the consistency of $\hat{p}_{n+1,\mathbf{M}}(\mathbf{x})$, and $\hat{p}^{(l)}_{n+1,\mathbf{M}}(\mathbf{x})$.

Theorem 1 If there exist constants λ and ρ such that

$$\lambda_{0,j} \geq \lambda > 0 \quad \text{for} \quad j = 1, 2, \ldots \tag{A1}$$

where $\lambda_{0,j} =$ min root of Σ_j, and

$$\rho_n \leq \rho\lambda_{0,n} \quad \text{for} \quad n = 1, 2, \ldots \tag{A2}$$

and $\mathbf{M}^T(n)\mathbf{M}(n) = (\rho_n + \lambda_{0,n})^{-1} \log n$, then $\hat{p}_{n+1,\mathbf{M}(n)}$ and $\hat{p}^{(l)}_{n+1,\mathbf{M}(n)}$ for $l = 1, 2, \ldots$, are such that

$$\operatorname*{lub}_{\mathbf{x}} E\,(\hat{p}_{n+1,\mathbf{M}(n)}(\mathbf{x}) - p_{n+1}(\mathbf{x}))^2 = O(n^{-\frac{1}{1+\rho}}(\log n)^k) \tag{3.20}$$

and

$$\operatorname*{lub}_{\mathbf{x}} E\|\hat{p}^{(l)}_{n+1,\mathbf{M}(n)}(\mathbf{x}) - p^{(l)}_{n+1}(\mathbf{x})\|^2 = O(n^{-\frac{1}{1+\rho}}(\log(n))^{k+l}) \tag{3.21}$$

Proof See the appendix at the end of this chapter.

When $l = 1$, (3.16) becomes

$$(2\pi)^k \hat{p}'_{n+1,s,\mathbf{M}}(\mathbf{x}) = \frac{1}{n}\sum_{j=1}^{n} \int_0^{\mathbf{M}} t_s \ \exp\{\mathbf{t}^T(AA_j^{-1}\Sigma_j A_j^{-1^T} A^T - \Sigma)\mathbf{t}/2\} \\ \times \sin \mathbf{t}^T(\mathbf{b} - \mathbf{x} - AA_j^{-1}(\mathbf{b} - \mathrm{X}_j))d\mathbf{t}. \tag{3.22}$$

Then with $\mathbf{M}(n)$ given in Theorem 1, $\hat{p}_{n+1,\mathbf{M}(n)}$ by (3.11) and $\hat{p}'_{n+1,s,\mathbf{M}(n)}$ by (3.22), defining

$$\hat{p}_{n+1}(\mathbf{x}) = \mathrm{tr}^*(\hat{p}_{n+1,\mathbf{M}(n)}(\mathbf{x})) \tag{3.23}$$

$$\hat{p}'_{n+1,s}(\mathbf{x}) = \mathrm{tr}'(\hat{p}'_{n+1,s,\mathbf{M}(n)}(\mathbf{x})) \tag{3.24}$$

where tr^* and tr' stand for retractions (i.e. truncations) to $[0,((2\pi)^k|\Sigma_{n+1}|)^{-1/2}]$ and $[-((2\pi)^k e)^{-1/2}\sigma_{ss}^{-1}|\Sigma_{n+1}|^{-1/2}, +((2\pi)^k e)^{-1/2}\sigma_{ss}^{-1}|\Sigma_{n+1}|^{-1/2}]$ respectively yields the following corollary to Theorem 1

Corollary 1 Under the conditions of Theorem 1

$$\operatorname*{lub}_{\mathbf{x}} E(\hat{p}_{n+1}(\underline{x}) - p_{n+1}(\mathbf{x}))^2 = O(n^{-\frac{1}{1+\rho}}(\log n)^k) \tag{3.25}$$

and

$$\operatorname*{lub}_{\mathbf{x}} E\|\hat{p}'_{n+1}(\mathbf{x}) - p'_{n+1}(\mathbf{x})\|^2 = O(n^{-\frac{1}{1+\rho}}(\log n)^{k+1}). \tag{3.26}$$

Remark 3 The rates mentioned above are better than those which has been obtained by using the kernel estimates of Cauculos (1966) (in estimating p and p'). The reason that better rates are attained is that we have exploited the fact that the density p is a mixture of normal densities.

3.3 An a. o. Sequence $\{t_n\}$ with G in a Compact Subset of R^k

In this section assume that G has support in

$$[T_1,T_2]_k \qquad [A3]$$

and, as in Section 3.2, abbreviate A_{n+1}, $\mathbf{b}_{n+1}$, and Σ_{n+1} by A, **b** and Σ respectively. Recalling the form of $\mathbf{t}_{G,n}$ from (3.4), it is immediate that

$$\|p'_{n+1}(\mathbf{x})/p_{n+1}(\mathbf{x})\| \le \|\Sigma^{-1}\|(\|A\| \;\max\; \{|T_1|,|T_2|\} + \|\mathbf{x}\| + \|\mathbf{b}\|) \qquad (3.27)$$

where $\|A\|$ denotes $(\mathrm{tr} \;\; AA^T)^{1/2}$. In view of the discussion following (3.6), define

$$\mathbf{t}_n(\mathbf{x}) = \;\; \mathrm{tr}[A_{n+1}^{-1}(\mathbf{x} - \mathbf{b}_{n+1}) + A_{n+1}^{-1}\Sigma_{n+1}(\hat{p}'_{n+1}(\mathbf{x})/\hat{p}_{n+1}(\mathbf{x}))] \qquad (3.28)$$

where the ratio **0/0** is to be interpreted as **0** , tr stands for retraction to $[T_1,T_2]_k$, and $\hat{p}_{n+1}$ and $\hat{p}'_{n+1}$ are defined in (3.23) and (3.24) respectively.

In light of (3.6), $E\|\mathbf{t}_{G,n} - \mathbf{t}_n\|^2$ is of interest. From (3.4) and (3.28), it is bounded for $0 < \gamma < 1$ by

$$\sum_{s=1}^{k} |T_2 - T_1|^{2-\gamma}E[\min\{|\frac{q_{n+1,s}(\mathbf{x})}{p_{n+1}(\mathbf{x})} - \frac{\hat{q}_{n+1,s}(\mathbf{x})}{\hat{p}_{n+1}(\mathbf{x})}|, |T_2 - T_1|\}]^{\gamma} \qquad (3.29)$$

where $q_{n+1,s}(\mathbf{x})$ = the sth component of $A_{n+1}^{-1}\Sigma_{n+1}p'_{n+1}(\mathbf{x})$, and $\hat{q}_{n+1,s}(\mathbf{x})$ = the sth component of $A_{n+1}^{-1}\Sigma_{n+1}\hat{p}'_{n+1}(\mathbf{x})$.

Theorem 2 Let (A1), (A2) and (A3) attain and for some constants $0 < a_* < a^*$ and $b^* > 0$, assume that

$$\mathrm{lub}_n\{\|\mathbf{b}_n\|, \lambda_{0,n}\} \le b^* \qquad (A4)$$

and

$$a_* \le \|A_n\| \le a^* \quad \text{for} \quad n = 1, 2, \ldots \tag{A5}$$

hold. Then for $0 < \gamma < 1$, $\{t_n\}$ with t_n defined by (3.28) is a.o. $(n^{-\frac{\gamma}{2(1+\rho)}}(\log n)^{\frac{(k+1)}{2}\gamma})$.

In the proof of Theorem 2 (see the appendix in the end of this chapter), the following lemma from the appendix of Singh (1977) will be exploited in bounding (3.29).

Lemma 1 Let y, z and L be real numbers with $z \neq 0$ and $0 < L < \infty$. If Y and Z are random variables, then for all $\gamma > 0$,

$$\begin{aligned} &E(\min\{|\frac{y}{z} - \frac{Y}{Z}|, L\})^\gamma \\ &\le 2^{\gamma + (\gamma-1)^+}|z|^{-\gamma}\{E|y - Y|^\gamma + (|\frac{y}{z}|^\gamma + z^{-(\gamma-1)^+}L^\gamma)E|z - Z|^\gamma\} \end{aligned}$$

where $(a)^+ = \max\{a, 0\}$ (For the proof of Lemma 1 see Singh, 1977)

3.4 An a.o. Sequence $\{t_n\}$ with E $(\|\underline{\theta}\|^{2+\gamma})$ bounded

Since the results on the squared error loss estimation of $p_{n+1}^{(i)}$ did not depend on (A3), it is possible to obtain an a.o. $\{t_n^*\}$ sequence with a weaker rate for risk convergence under the assumption that

$$B_\gamma = E(\|\underline{\theta}\|^{2+\gamma}) < \infty \tag{A6}$$

for some $\gamma > 0$ in place of (A3). To obtain this rate of risk convergence result, the following lemma is required.

Lemma 2 (Susarla 1974)

For any $a \in (0, 1)$ and $\beta > 0, p_{n+1}(p(\mathbf{y}) < \beta) \leq c^{*}\beta^{a}$ where c^{*} is a suitable constant, and $p(\mathbf{y}) = \int p_{\underline{\theta}}(\mathbf{y})dG(\underline{\theta})$, with $p_{\underline{\theta}}$ a normal density.

Proof See the appendix in the end of this chapter.

Now define the empirical Bayes rule $\{\mathbf{t}_n^*\}$ by

$$\mathbf{t}_n^*(\mathbf{x}) = A_{n+1}^{-1}(\mathbf{x} - \mathbf{b}_{n+1}) + A_{n+1}^{-1}\Sigma_{n+1}(\hat{p}'_{n+1}(\mathbf{x})/\max\{\hat{p}_{n+1}(\mathbf{x}), \delta_{n+1}\}) \tag{3.30}$$

where $\delta_{n+1}(\downarrow 0)$ will be specified in the following theorem.

Theorem 3 Let (A1), (A2), (A4), (A5) and (A6) hold. Define $(2\pi)^k|\Sigma_{n+1}|\delta_{n+1}^2 = n^{-\beta}$ for some β such that $0 < 2\beta(1 + \rho) < 1$ where ρ is a constant appearing in (A2). Then the sequence $\{\mathbf{t}_n^*\}$ defined by (3.30) is a.o. $((\log n)^{\frac{-\gamma}{2+\gamma}})$.

Proof: See the appendix

Meanwhile the following lemma will also be exploited.

Lemma 3 If (A4), (A5), (A6) hold, for a given δ in (0, $((2\pi)^k|\Sigma_{n+1}|)^{-1/2}$), then

$$\begin{aligned} &E\{\|p'_{n+1}(X_{n+1})/p_{n+1}(X_{n+1})\|^2[p_{n+1}(X_{n+1}) \leq \delta]\} \\ &\leq c\{\log(2\pi|\Sigma_{n+1}|\delta^2)^{-1}\}^{-\frac{\gamma}{2+\gamma}} \end{aligned}$$

for some constant c depending on the constants appearing in (A4), (A5) and (A6).

Proof See the appendix which follows.

Appendix

(3.5)

$$p_{n+1}(\mathbf{x}) = \frac{1}{(2\pi)^{k/2}|\Sigma_{n+1}|^{1/2}}\int_{R^k} \exp\{-\frac{1}{2}(\mathbf{x} - A_{n+1}\underline{\theta} - \mathbf{b}_{n+1})^T\Sigma_{n+1}^{-1}(\mathbf{x} - A_{n+1}\underline{\theta} - \mathbf{b}_{n+1})\}dG(\underline{\theta})$$

$$p'_{n+1}(\mathbf{x}) = \frac{1}{(2\pi)^{k/2}|\Sigma_{n+1}|^{1/2}}\int_{R^k} \exp\{-\frac{1}{2}(\mathbf{x} - A_{n+1}\underline{\theta} - \mathbf{b}_{n+1})^T\Sigma_{n+1}^{-1}(\text{----})\}dG(\underline{\theta})$$
$$\times\, (-\Sigma_{n+1}^{-1}(\mathbf{x} - A_{n+1}\underline{\theta} - \mathbf{b}_{n+1}))$$

Thus

$$\frac{p'_{n+1}(\mathbf{x})}{p_{n+1}(\mathbf{x})} = -\Sigma_{n+1}^{-1}(\mathbf{x} - A_{n+1}\underline{\theta} - \mathbf{b}_{n+1})$$

i.e. $E(\underline{\theta} \mid \mathbf{x}) = A_{n+1}^{-1}[\mathbf{x} - \mathbf{b}_{n+1} + \Sigma_{n+1}\frac{p'_{n+1}(\mathbf{x})}{p_{n+1}(\mathbf{x})}]$

(3.6)

$$R_n(\mathbf{t}_n, G) - R_{n+1}(G)$$
$$= E\|\mathbf{t}_n - \underline{\theta}_{n+1}\|^2 - E\|\mathbf{t}_{G,n} - \underline{\theta}_{n+1}\|^2$$
$$= E\|\mathbf{t}_n - \mathbf{t}_{G,n} + \mathbf{t}_{G,n} - \underline{\theta}_{n+1}\|^2 - E\|t_{G,n} - \underline{\theta}_{n+1}\|^2$$
$$= E\|\mathbf{t}_n - \mathbf{t}_{G,n}\|^2 + E\|\mathbf{t}_{G,n} - \underline{\theta}_{n+1}\|^2 + 2E(\mathbf{t}_n - \mathbf{t}_{G,n})^T(\text{----}) - E\|\mathbf{t}_{G,n} - \underline{\theta}_{n+1}\|^2$$
$$= E\|\mathbf{t}_n - \mathbf{t}_{G,n}\|^2 + 2E_{\mathbf{X}_1,\ldots,\mathbf{X}_{n+1}}\{(\mathbf{t}_n - \mathbf{t}_{G,n})E_{\underline{\theta}}[(\mathbf{t}_{G,n} - \underline{\theta}_{n+1}) \mid \mathbf{X}_1, \ldots, \mathbf{X}_{n+1}]\}$$
$$= E\|\mathbf{t}_n - \mathbf{t}_{G,n}\|^2$$

(3.9)

$$
\begin{aligned}
&(2\pi)^k | p_{n+1}(\mathbf{x}) - \frac{1}{(2\pi)^k}\int_{-\mathbf{M}}^{\mathbf{M}} \exp(- i\mathbf{t}^T\mathbf{x})\varphi(\mathbf{t})d\mathbf{t} | \\
&= (2\pi)^k | \int_{-\infty^k}^{\infty^k} \exp(- i\mathbf{t}^T\mathbf{x})\varphi(\mathbf{t})d\mathbf{t} - \int_{-\mathbf{M}}^{\mathbf{M}} \exp(- i\mathbf{t}^T\mathbf{x})\varphi(\mathbf{t})d\mathbf{t} | \\
&= (2\pi)^k | \int_{-\infty^k}^{-\mathbf{M}} \exp(- i\mathbf{t}^T\mathbf{x})\varphi(\mathbf{t})d\mathbf{t} + \int_{\mathbf{M}}^{\infty^k} \exp(- i\mathbf{t}^T\mathbf{x})\varphi(\mathbf{t})d\mathbf{t} | \\
&\le (2\pi)^k [\int_{-\infty^k}^{-\mathbf{M}} | \exp(- i\mathbf{t}^T\mathbf{x})\varphi(\mathbf{t}) | d\mathbf{t} + \int_{\mathbf{M}}^{\infty^k} | \exp(- i\mathbf{t}^T\mathbf{x})\varphi(\mathbf{t}) | d\mathbf{t}] \\
&\le (2\pi)^k [\int_{-\infty^k}^{-\mathbf{M}} \exp(- \mathbf{t}^T\Sigma\mathbf{t}/2)d\mathbf{t} + \int_{\mathbf{M}}^{\infty^k} \exp(- \mathbf{t}^T\Sigma\mathbf{t}/2)d\mathbf{t}] \\
&= (2\pi)^k 2\int_{\mathbf{M}}^{\infty^k} \exp(- \mathbf{t}^T\Sigma\mathbf{t}/2)d\mathbf{t} \\
&\le 2(2\pi)^k \int_{\mathbf{M}}^{\infty^k} \exp(- \lambda_{\min}\mathbf{t}^T\mathbf{t}/2)d\mathbf{t}
\end{aligned}
$$

(where $\lambda_{\min} > 0$. (since Σ : positive definite .) The last inequality follows from the following relations :

$$\lambda_{\min} X^T X \le X^T A X \le \lambda_{\max} X^T X,$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the maximum and minimum eigenvalue of the matrix A respectively . (See Rao, C. R. Intro. to Linear Statistics and Applications 1973 p. 69))

$$
\begin{aligned}
&= 2(2\pi)^k \int_{M_1}^{\infty} \cdots \int_{M_k}^{\infty} \exp\{ - \frac{\lambda_{\min}}{2}(t_1^2 + \cdots + t_k^2)\} dt_k \ldots dt_1 \\
&= 2(2\pi)^k \int_{M_1}^{\infty} \exp\{ - \frac{\lambda_{\min}}{2} t_1^2\} dt_1 \ldots \int_{M_k}^{\infty} \exp\{ - \frac{\lambda_{\min}}{2} t_k^2\} dt_k \\
&= 2(2\pi)^k (2\pi\lambda_{\min})^{1/2} \Phi(- \sqrt{\lambda_{\min}}\, M_1) \ldots (2\pi\lambda_{\min})^{1/2} \Phi(- \sqrt{\lambda_{\min}}\, M_k) \\
&\le c_1 \exp\{ - \lambda_{0,n+1} \mathbf{M}^T\mathbf{M}/2\}
\end{aligned}
$$

where c_1 is a constant depending on $\lambda_{\min}$ and k and Φ is the standard normal distribution function. The last inequality follows from

$$
\begin{aligned}
\Phi(x) &\le ((2\pi)^{1/2}x)^{-1} \exp\{ - \frac{x^2}{2}\} \quad \text{for} \quad x > 0 \\
&\le (2\pi)^{-1/2} \exp\{ - x^2/2\} \quad \text{for} \quad x > \mathrm{I}.
\end{aligned}
$$

(3.10)

$$\varphi(\mathbf{t}) = \exp\{ - \mathbf{t}^T\Sigma\mathbf{t}/2 + i\mathbf{t}^T\mathbf{b}\} \int \exp\{i\mathbf{t}^T A\underline{\theta}) dG(\underline{\theta}) \qquad (i)$$

and

$$\varphi(A_j^{-1^T}A^T\mathbf{t}) = \exp\{-\mathbf{t}^T AA_j^{-1}\Sigma_j A_j^{-1^T}A^T\mathbf{t}/2 + i\mathbf{t}^T AA_j^{-1}\mathbf{b}_j\} \int \exp\{i\mathbf{t}^T AA_j^{-1}A_j\underline{\theta})dG(\underline{\theta}) \qquad (ii)$$

From (i) and (ii),

$$\frac{\varphi(\mathbf{t})}{\varphi(A_j^{-1^T}A^T\mathbf{t})} = \frac{\exp\{-\mathbf{t}^T\Sigma\mathbf{t}/2 + i\mathbf{t}^T\mathbf{b}\}\int \exp\{i\mathbf{t}^T A\underline{\theta})dG(\underline{\theta})}{\exp\{-\mathbf{t}^T AA_j^{-1}\Sigma_j A_j^{-1^T}A^T\mathbf{t}/2 + i\mathbf{t}^T AA_j^{-1}\mathbf{b}_j\}\int \exp\{i\mathbf{t}^T A\underline{\theta}\}dG(\underline{\theta})}$$

Hence

$$\varphi(\mathbf{t}) = \exp\{\mathbf{t}^T(AA_j^{-1}\Sigma_j A_j^{-1^T}A^T - \Sigma)\mathbf{t}/2 + i\mathbf{t}^T(\mathbf{b} - AA_j^{-1}\mathbf{b}_j)\}\varphi(A_j^{-1^T}A^T\mathbf{t})$$

Proof of boundedness of each summand of (3.11) :

The absolute value of each summand of (3.11) is

$$\begin{aligned}
&\left| \mathrm{Re}\int_{-M_1}^{M_1}\cdots\int_{-M_k}^{M_k} \exp\{\mathbf{t}^T(AA_j^{-1}\Sigma_j A_j^{-1^T}A^T - \Sigma)\mathbf{t}/2\} \ \exp i\mathbf{t}^T(\mathbf{b} - \mathbf{x} - AA_j^{-1}(\mathbf{b}_j - X_j))dt_k \ldots dt_1 \right| \\
&\le \int_{-M_1}^{M_1}\cdots\int_{-M_k}^{M_k} \exp\{\mathbf{t}^T(AA_j^{-1}\Sigma_j A_j^{-1^T}A^T - \Sigma)/2\}dt_k \ldots dt_1 \\
&\le \int_{-M_1}^{M_1}\cdots\int_{-M_k}^{M_k} \exp\{\rho_n\mathbf{t}^T\mathbf{t}/2\}dt_k \ldots dt_1 \\
&= \int_{-M_1}^{M_1} \exp(\rho_n t_1^2/2)dt_1 \ldots \int_{-M_k}^{M_k} \exp(\rho_n t_k^2/2)dt_k \\
&\le 2^k(M_1 \ldots M_k) \ \exp\{\rho_n\mathbf{M}^T\mathbf{M}/2\} \\
&\le 2^k\left(\frac{\mathbf{M}^T\mathbf{M}}{k}\right)^{k/2} \exp\{\rho_n\mathbf{M}^T\mathbf{M}/2\} \\
&= c_2(\mathbf{M}^T\mathbf{M})^{k/2} \ \exp\{\rho_n\mathbf{M}^T\mathbf{M}/2\}
\end{aligned}$$

where c_2 is a constant depending on k and $\mathbf{M}$

(3.13)

$$\mathrm{var}(\hat{p}_{n+1,\mathbf{M}}(\mathbf{x})) \le E(\hat{p}_{n+1,\mathbf{M}}(\mathbf{x}))^2$$
$$= (n(2\pi)^k)^{-2} \sum_{j=1}^{n} E(\,\mathrm{Re}[\int_{-M_1}^{M_1} \cdots \int_{-M_k}^{M_k} \exp\{\mathbf{t}^T(AA_j^{-1}\Sigma_j A_j^{-1^T} A^T - \Sigma)\mathbf{t}/2\} \exp i\mathbf{t}^T(\mathbf{b} - \mathbf{x} - (AA_j^{-1}(\mathbf{b}_j - X_j))d\mathbf{t}])^2$$
$$\le (n(2\pi)^k)^{-2} \sum_{j=1}^{n} E(c_2(\mathbf{M}^T\mathbf{M})^{k/2} \; \exp\{\rho_n \mathbf{M}^T\mathbf{M}/2\})^2$$
$$= (n(2\pi)^k)^{-2} n c_2^2 (\mathbf{M}^T\mathbf{M})^k \; \exp\{\rho_n \mathbf{M}^T\mathbf{M}\}$$
$$= n^{-1}(2\pi)^{-2k} c_2^2 (\mathbf{M}^t\mathbf{M})^k \; \exp\{\rho_n \mathbf{M}^T\mathbf{M}\}$$

(3.14)

$$\operatorname{lub}_{\mathbf{x}} E(\hat{p}_{n+1,\mathbf{M}}(\mathbf{X}) - p_{n+1}(\mathbf{x}))^2$$
$$= \operatorname{lub}_{\mathbf{x}} E(\hat{p}_{n+1,\mathbf{M}}(\mathbf{x}) - \int_{-\mathbf{M}}^{\mathbf{M}} + \int_{-\mathbf{M}}^{\mathbf{M}} - p_{n+1}(\mathbf{x}))^2$$
$$\le 2 \operatorname{lub}_{\mathbf{x}} E(\hat{p}_{n+1,\mathbf{M}}(\mathbf{x}) - \int_{-\mathbf{M}}^{\mathbf{M}})^2 + 2 \operatorname{lub}_{\mathbf{x}} E(\int_{-\mathbf{M}}^{\mathbf{M}} - p_{n+1}(\mathbf{x}))^2 \quad (\text{by} C_\delta - \text{ineq.})$$
$$= 2 \operatorname{lub}_{\mathbf{x}} \; \mathrm{var}(\hat{p}_{n+1,\mathbf{M}}(\mathbf{x})) + 2 \operatorname{lub}_{\mathbf{x}} E(\int_{-\mathbf{M}}^{\mathbf{M}} - p_{n+1}(\mathbf{x}))^2$$
$$\le 2n^{-1} c_3 \, (\mathbf{M}^T\mathbf{M})^k \; \exp\{\rho_n \mathbf{M}^T\mathbf{M}\} + 2c_1^2 \; \exp\{-\lambda_{\min}\mathbf{M}^T\mathbf{M}\}$$

(3.17)

$$\operatorname{lub}_{\mathbf{x}} E\|\hat{p}^{(l)}_{n+1,\mathbf{M}}(\mathbf{x}) - p^{(l)}_{n+1}(\mathbf{x})\|^2$$
$$= \operatorname{lub}_{\mathbf{x}} E\|((\hat{p}^{(l)}_{n+1,1,\mathbf{M}}(\mathbf{x}) - p^{(l)}_{n+1,1}), \cdots, (\hat{p}^{(l)}_{n+1,k,\mathbf{M}}(\mathbf{x}) - p^{(l)}_{n+1,k}(\mathbf{x})))\|^2$$
$$= \operatorname{lub}_{\mathbf{x}} E[\sum_{s=1}^{k} (\hat{p}^{(l)}_{n+1,s,\mathbf{M}}(\mathbf{x}) - p^{(l)}_{n+1,s}(\mathbf{x}))^2]$$
$$\le \sum_{s=1}^{k} \operatorname{lub}_{\mathbf{x}} E(\hat{p}^{(l)}_{n+1,s,\mathbf{M}}(\mathbf{x}) - p^{(l)}_{n+1,s}(\mathbf{x}))^2$$
$$= \sum_{s=1}^{k} \operatorname{lub}_{\mathbf{x}} E(\hat{p}^{(l)}_{n+1,s,\mathbf{M}}(\mathbf{x}) - \int_{-\mathbf{M}}^{\mathbf{M}} + \int_{-\mathbf{M}}^{\mathbf{M}} - p^{(l)}_{n+1,s}(\mathbf{x}))^2$$
$$\le 2\sum_{s=1}^{k} \operatorname{lub}_{\mathbf{x}} E(\hat{p}^{(l)}_{n+1,s,\mathbf{M}}(\mathbf{x}) - \int_{-\mathbf{M}}^{\mathbf{M}})^2 + 2\sum_{s=1}^{k} \operatorname{lub}_{\mathbf{x}} E(\int_{-\mathbf{M}}^{\mathbf{M}} - p^{(l)}_{n+1,s}(\mathbf{x}))^2$$
$$= 2\sum_{s=1}^{k} \operatorname{lub}_{\mathbf{x}} \; \mathrm{var}\hat{p}^{(l)}_{n+1,s,\mathbf{M}}(\mathbf{x}) + 2\sum_{s=1}^{k} \operatorname{lub}_{\mathbf{x}} E(\int_{-\mathbf{M}}^{\mathbf{M}} - p^{(l)}_{n+1,s}(\mathbf{x}))^2$$
$$\le 2\sum_{s=1}^{k} (n(2\pi)^k)^{-2} \sum_{j=1}^{n} 2^k [\int_0^{M_1} \exp\{\rho_n t_1^2/2\} dt_1 \cdots \int_0^{M_s} t_s^l \; \exp\{\rho_n t_s^2/2\} dt_s \cdots \int_0^{M_k} \exp\{\rho_n t_k^2/2\} dt_k]^2$$
$$+ 2\sum_{j=1}^{k} [2^k \int_{M_1}^{\infty} \cdots \int_{M_k}^{\infty} t_s^l \; \exp\{-\lambda_{\min}\mathbf{t}^T\mathbf{t}/2\} d\mathbf{t}]^2$$

$$\leq 2(n(2\pi)^k)^{-2} \sum_{l=1}^{k} n[2^k(M_1 \dots M_k)\frac{M_s^l}{l+1} \exp(\rho_n \mathbf{M}^T\mathbf{M}/2)]^2$$
$$+ 22^{2k} c_4 \sum_{s=1}^{k} M_s^{2(l-1)} \exp\{-\lambda_{\min}\mathbf{M}^T\mathbf{M}\}$$
$$\leq c_5 n^{-1}(\frac{\mathbf{M}^T\mathbf{M}}{K})^{k+l} \exp(\rho_n \mathbf{M}^T\mathbf{M}) + 2^{2k+1} c_6(\mathbf{M}^T\mathbf{M})^{l-1} \exp\{-\lambda_{\min}\mathbf{M}^T\mathbf{M}\}$$

where c_5 and c_6 are constants depending on k , $\lambda_{\min}$, l and $\mathbf{M}$. Note that $l = 1, 2, \dots$

(3.18)

$$\begin{aligned}
\hat{p}_{n+1,\mathbf{M}}(\mathbf{x}) &= \frac{1}{n(2\pi)^k} \sum_{j=1}^{n} \int_{-\mathbf{M}}^{\mathbf{M}} \exp it^T(X_j - \mathbf{x})dt \\
&= \frac{1}{n(2\pi)^k} \sum_{j=1}^{n} \int_{-M_1}^{M_1} \exp it_1(X_{1j} - x_1)dt_1 \dots \int_{-M_k}^{M_k} \exp it_k(X_{kj} - x_k)dt_k \\
&= \frac{1}{n(2\pi)^k} \sum_{j=1}^{n} \prod_{h=1}^{k} \frac{\exp iM_h(X_{hj} - x_h) - \exp - iM_h(X_{hj} - x_h)}{i(X_{hj} - x_h)} \\
&= \frac{1}{n(2\pi)^k} \sum_{j=1}^{n} \prod_{h=1}^{k} \frac{2 \sin M_h(X_{hj} - x_h)}{X_{hj} - x_h}
\end{aligned}$$

Proof of Theorem 1 :

(i) By (3.14)

$$\begin{aligned}
&\operatorname{lub}_{\mathbf{x}} E(\hat{p}_{n+1,\mathbf{M}(n)}(\mathbf{x}) - p_{n+1}(\mathbf{x}))^2 \\
&\leq 2c_1^2 \exp\{-\lambda_{0,n}\mathbf{M}^T\mathbf{M}\} + 2n^{-1}c_3(\mathbf{M}^T\mathbf{M})^k \exp\{\rho_n\mathbf{M}^T\mathbf{M}\} \\
&= 2c_1^2 \exp\{-\lambda_{0,n}(\rho_n + \lambda_{0,n})^{-1}\log n\} + 2n^{-1}c_3[(\rho_n + \lambda_{0,n})^{-1}\log n]^k \exp\{\rho_n(\rho_n + \lambda_{0,n})^{-1}\log n\} \\
&= O(n^{\frac{-\lambda_{0,n}}{\rho_n + \lambda_{0,n}}}) + O(n^{-1}(\log n)^k n^{\frac{\rho_n}{\rho_n + \lambda_{0,n}}}) \\
&= O(n^{-\frac{\lambda_{0,n}}{\rho_n + \lambda_{0,n}}}) + O(n^{-\frac{\lambda_{0,n}}{\rho_n + \lambda_{0,n}}}(\log n)^k) \\
&= O(n^{-\frac{\lambda_{0,n}}{\rho_n + \lambda_{0,n}}}(\log n)^k) \\
&\leq O(n^{-\frac{1}{1+\rho}}(\log n)^k)
\end{aligned}$$

(ii) By (3.17)

$$
\begin{aligned}
&\operatorname*{lub}_{\mathbf{x}} E\|\hat{p}^{(l)}_{n+1,\mathbf{M}(n)}(\mathbf{x}) - p^{(l)}_{n+1}(\mathbf{x})\|^2 \\
&\le c'_5 n^{-1}(\mathbf{M}^T\mathbf{M})^{k+l}\exp(\rho_n\mathbf{M}^T\mathbf{M}) + c'_6(\mathbf{M}^T\mathbf{M})^{l-1}\exp(-\lambda_{0,n}\mathbf{M}^T\mathbf{M}) \\
&= c'_5 n^{-1}(\rho_n + \lambda_{0,n})^{-(k+l)}(\log n)^{k+l}\exp\{\rho_n(\rho_n+\lambda_{0,n})^{-1}\log n\} \\
&\qquad + c'_6(\rho_n+\lambda_{0,n})^{-l+1}(\log n)^{l-1}\exp(-\frac{\lambda_{0,n}}{\rho_n+\lambda_{0,n}}\log n) \\
&= O(n^{-1}(\log n)^{k+l} n^{\rho_n/(\rho_n+\lambda_{0,n})}) + O(n^{-\frac{\lambda_{0,n}}{\rho_n+\lambda_{0,n}}}(\log n)^{l-1}) \\
&= O(n^{-\frac{\lambda_{0,n}}{\rho_n+\lambda_{0,n}}}(\log n)^{k+l}) \\
&\le O(n^{-\frac{1}{1+\rho}}(\log n)^{k+l})
\end{aligned}
$$

Proof of (3.27)

From (3.4)

$$
\frac{p'_{n+1}(\mathbf{x})}{p_{n+1}(\mathbf{x})} = \Sigma^{-1}_{n+1}A_{n+1}[\mathbf{t}_{G,n}(\mathbf{x}) - A^{-1}_{n+1}(\mathbf{x} - \mathbf{b}_{n+1})]
$$

Thus

$$
\begin{aligned}
\|\frac{p'_{n+1}(\mathbf{x})}{p_{n+1}(\mathbf{x})}\| &\le \|\Sigma^{-1}_{n+1}\|[\|A_{n+1}\|\|\mathbf{t}_{G,n}(\mathbf{x})\| + \|\mathbf{x}\| + \|\mathbf{b}_{n+1}\|] \\
&\le \|\Sigma^{-1}_{n+1}\|[\|A_{n+1}\|\max\{|T_1|, |T_2|\} + \|\mathbf{x}\| + \|\mathbf{b}_{n+1}\|]
\end{aligned}
$$

where $\|A\| = (\operatorname{tr} AA^T)^{1/2}$.

Proof of (3.29)

$$
\begin{aligned}
&E\|\mathbf{t}_{G,n} - \mathbf{t}_n\|^2 \\
&= \sum_{s=1}^{k} E(\mathbf{t}_{G,n,s} - \mathbf{t}_{n,s})^2 \quad \text{where } \mathbf{t}_{G,n} = (\mathbf{t}_{G,n,1}, \dots, \mathbf{t}_{G,n,k})^T \text{ and similarly for } \mathbf{t}_n \\
&\le \sum_{s=1}^{k} E\,[\min\{(\mathbf{t}_{G,n,s} - \mathbf{t}^*_{n,s})^2, |T_2 - T_1|^2\}] \quad (\text{since } \mathbf{t}_n = \operatorname{tr}(\mathbf{t}^*_n)) \\
&\le \sum_{s=1}^{k} |T_2 - T_1|^{2-\gamma} E\,[\min\{|\frac{q_{n+1,s}(\mathbf{x})}{p_{n+1}(\mathbf{x})} - \frac{\hat{q}_{n+1,s}(\mathbf{x})}{\hat{p}_{n+1}(\mathbf{x})}|, |T_2 - T_1|\}]^\gamma
\end{aligned}
$$

where

$q_{n+1,s}(\mathbf{x}) =$ the sth component of $A_{n+1}^{-1}\Sigma_{n+1}p'_{n+1}(\mathbf{x})$,

and

$\hat{q}_{n+1,s}(\mathbf{x}) =$ the sth component of $A_{n+1}^{-1}\Sigma_{n+1}\hat{p}'_{n+1}(\mathbf{x})$.

Proof of Theorem 2 :

$$R_n(\mathbf{t}_n, G) - R_{n+1}(G) = E\|\mathbf{t}_{G,n} - \mathbf{t}_n\|_2^2$$
$$\leq \sum_{s=1}^{k} |T_2 - T_1|^{2-\gamma} E\, [\min\{|\frac{q_{n+1,s}(\mathbf{X}_{n+1})}{p_{n+1}(\mathbf{X}_{n+1})} - \frac{\hat{q}_{n+1,s}(\mathbf{X}_{n+1})}{\hat{p}_{n+1}(\mathbf{X}_{n+1})}|, |T_2 - T_1|\}]^{\gamma} \quad \text{(from (3.29))}$$
$$= \sum_{s=1}^{k} |T_2 - T_1|^{2-\gamma} E_{\mathbf{X}_{n+1}}(E_{\mathbf{X}_1,\ldots,\mathbf{X}_n}[\min\{|\frac{q_{n+1,s}(\mathbf{x})}{p_{n+1}(\mathbf{x})} - \frac{\hat{q}_{n+1,s}(\mathbf{x})}{\hat{p}_{n+1}(\mathbf{x})}|, |T_2 - T_1|\} \mid \mathbf{X}_{n+1} = \mathbf{x}])^{\gamma}$$
$$\leq \sum_{s=1}^{k} |T_2 - T_1|^{2-\gamma} 2^{\gamma} E_{\mathbf{X}_{n+1}}[|p_{n+1}(\mathbf{X}_{n+1})|^{-\gamma}\{E|q_{n+1,s} - \hat{q}_{n+1,s}|^{\gamma}\}]$$
$$+ \sum_{s=1}^{k} |T_2 - T_1|^{2-\gamma} 2^{\gamma} E_{\mathbf{X}_{n+1}}\{(|\frac{q_{n+1,s}}{p_{n+1}}|^{\gamma} + |T_2 - T_1|^{\gamma}) E|p_{n+1} - \hat{p}_{n+1}|^{\gamma}\}$$

Let

$$I_1 = E|q_{n+1,s}(\mathbf{x}) - \hat{q}_{n+1,s}(\mathbf{x})|^{\gamma}$$

and

$$I_2 = |\frac{q_{n+1,s}}{p_{n+1}}|^{\gamma}.$$

Then

$$I_1 = E|\sum_{r=1}^{k} d_{rs}(p'_{n+1,r}(\mathbf{x}) - \hat{p}_{n+1,r}(\mathbf{x}))|^{\gamma} \quad (\text{where } d_{rs} = [A_{n+1}^{-1}\Sigma_{n+1}]_{rs})$$

$$\leq \sum_{r=1}^{k} |d_{rs}|^{\gamma} E|p'_{n+1,r}(\mathbf{x}) - \hat{p}'_{n+1,r}(\mathbf{x})|^{\gamma} \quad (\text{by } C_\delta - \text{ineq.})$$

$$\leq \sum_{r=1}^{k} |d_{rs}|^{\gamma}(E|p'_{n+1,r}(\mathbf{x}) - \hat{p}'_{n+1,r}|^{2})^{\gamma/2}$$

$$= \sum_{r=1}^{k} |d_{rs}|^{\gamma} O(n^{(\frac{-1}{1+\rho})\frac{\gamma}{2}}(\log n)^{(k+1)\frac{\gamma}{2}})$$

$$= O(n^{(-\frac{1}{1+\rho})\frac{\gamma}{2}}(\log n)^{(k+1)\frac{\gamma}{2}})$$

and

$$I_2 = |\frac{q_{n+1,s}}{p_{n+1}}|^{\gamma} = \frac{1}{|p_{n+1}|^{\gamma}}[|\sum_{r=1}^{k} d_{rs}p'_{n+1,r}(\mathbf{X}_{n+1})|^{\gamma}]$$

$$\leq \sum_{r=1}^{k} |d_{rs}|^{\gamma} |\frac{p'_{n+1,r}(\mathbf{X}_{n+1})}{p_{n+1}(\mathbf{X}_{n+1})}|^{\gamma}$$

$$\leq \sum_{r=1}^{k} |d_{rs}|^{\gamma} \|\Sigma_{n+1}^{-1}\|^{\gamma}[\|A_{n+1}\|T + \|\mathbf{X}_{n+1}\| + \|\mathbf{b}_{n+1}\|]^{\gamma} \quad (\text{where } T = \max\{|T_1|, |T_2|\})$$

$$\leq \sum_{r=1}^{k} |d_{rs}|^{\gamma} \|\Sigma_{n+1}^{-1}\|^{\gamma}(\|A_{n+1}\|^{\gamma}T^{\gamma} + \|\mathbf{X}_{n+1}\|^{\gamma} + \|\mathbf{b}_{n+1}\|^{\gamma})$$

Thus

$$R_n(\mathbf{t}_n, G) - R_{n+1}(G)$$
$$\leq \sum_{s=1}^{k} |T_2 - T_1|^{2-\gamma}2^{\gamma}E_{\mathbf{X}_{n+1}}[|p_{n+1}(\mathbf{X}_{n+1})|^{-\gamma}\{O(n^{-\frac{1}{1+\rho}\frac{\gamma}{2}}(\log n)^{(k+1)\frac{\gamma}{2}})\}]$$
$$+ \sum_{s=1}^{k} |T_2 - T_1|^{2-\gamma}2^{\gamma}E_{\mathbf{X}_{n+1}}[|p_{n+1}(\mathbf{X}_{n+1})|^{-\gamma}\{\sum_{r=1}^{k}(B_1 + B_2\|\mathbf{X}_{n+1}\|^{\gamma})\}O(n^{-\frac{1}{1+\rho}\frac{\gamma}{2}}(\log n)^{\frac{k\gamma}{2}})] = F(\text{set})$$

where

$$B_1 = \sum_{r=1}^{k} |d_{rs}|^{\gamma}\|\Sigma_{n+1}^{-1}\|^{\gamma}(\|A_{n+1}\|^{\gamma}T^{\gamma} + \|\mathbf{b}_{n+1}\|^{\gamma}) + |T_2 - T_1|^{\gamma} < \infty$$

and

$$B_2 = \sum_{r=1}^{k} |d_{rs}|^{\gamma}\|\Sigma_{n+1}^{-1}\|^{\gamma} < \infty.$$

$$F = \sum_{s=1}^{k} |T_2 - T_1|^{2-\gamma} 2^{\gamma} E_{\mathbf{X}_{n+1}}(|p_{n+1}(\mathbf{X}_{n+1})|^{-\gamma}) O(n^{-\frac{\gamma}{2(1+\rho)}} (\log n)^{\frac{\gamma(k+1)}{2}})$$
$$+ \sum_{s=1}^{k} |T_2 - T_1|^{2-\gamma} 2^{\gamma} k B_2 E_{\mathbf{X}_{n+1}}(|p_{n+1}(\mathbf{X}_{n+1})|^{-\gamma} \|\mathbf{X}_{n+1}\|^{\gamma}) O(n^{-\frac{\gamma}{2(1+\rho)}} (\log n)^{\frac{k\gamma}{2}})$$
$$+ \sum_{s=1}^{k} |T_2 - T_1|^{2-\gamma} 2^{\gamma} k B_1 E_{\mathbf{X}_{n+1}}(|p_{n+1}(\mathbf{X}_{n+1})|^{-\gamma}) O(n^{-\frac{\gamma}{2(1+\rho)}} (\log n)^{\frac{k\gamma}{2}})$$

$$= O(n^{-\frac{\gamma}{2(1+\rho)}} (\log n)^{\frac{(k+1)\gamma}{2}}).$$

This follows from the fact that $E(p_{n+1}^{\gamma}(\mathbf{X}_{n+1}))$ and $E(\|\mathbf{X}_{n+1}\|^{\gamma} p_{n+1}^{-\gamma}(\mathbf{X}_{n+1}))$ are uniformly bounded in n whenever $\|A_n\|$, $\|\mathbf{b}_n\|$, and $\|\Sigma_n^{-1}\|$ are uniformly bounded which in turn is due to the support of G being included in $[T_1, T_2]_k$ (compactness).

Proof of Lemma 2 :

Let $\mathbf{Z} = \mathbf{X} - \underline{\theta}_{n+1}$. Then

$$\|\mathbf{X} - \underline{\theta}_j\| \le \|\mathbf{Z}\| + 2m\alpha$$

where $\underline{\theta}_j \in [-\alpha, \alpha]_m$: subset of $R^m, j = 1, 2, ...$ and $\mathbf{X} \sim N(\underline{\theta}, I)$.

$$\begin{aligned} P_j(\mathbf{X}) &= (2\pi)^{-\frac{m}{2}} \exp - (\|\mathbf{X} - \underline{\theta}_j\|^2/2) \\ &\ge (2\pi)^{-\frac{m}{2}} \exp\{-(\|\mathbf{Z}\| + 2m\alpha)^2/2\} \end{aligned} \qquad (*)$$

Let M be the minimum value of $\|\mathbf{Z}\|$ such that the rhs of (*) $\le \beta$. Since

$$\begin{aligned} P[\|\mathbf{Z}\| > 2t] &\le \frac{e^{-bt}\int_0^{\infty} e^{-\frac{(1-b)u}{2}} u^{\frac{(m-2)}{2}} du}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})} \\ &= e^{-bt}(1-b)^{-\frac{m}{2}} \end{aligned}$$

(by the definition of the Gamma function) for all $t > 0$ and $b \in (0, 1)$, from (*) yields

$$\begin{aligned}\beta^{-a}P_{n+1}[p(\mathbf{y}) < \beta] &\le \beta^{-a}P_{n+1}[\|\mathbf{Z}\| > M] \\ &\le (2\pi)^{\frac{am}{2}} \exp\{a(M + 2m\alpha)^2/2\}e^{-bt}(1-b)^{\frac{-m}{2}} \\ &= (2\pi)^{\frac{am}{2}}(1-b)^{\frac{-m}{2}} \exp\{-bM^2/2 + a(M+2m\alpha)^2/2\} \\ &< \infty \quad \text{for } M < \infty \text{ and } 0 < a < b < 1.\end{aligned}$$

Thus we have proven

$$P_{n+1}[p(\mathbf{y}) < \beta] < c^* \beta^a$$

Proof of Lemma 3 :

First we prove that $P([p(\mathbf{y}) < \delta])$ is bounded :

$$\begin{aligned}P[p(\mathbf{y}) < \delta] &= P(p(\mathbf{y}) < \delta) \\ &= P(((2\pi)^k|\Sigma|)^{-1/2}\int e^{-\frac{1}{2}(\mathbf{y}-\underline{\theta})^T\Sigma^{-1}(\mathbf{y}-\underline{\theta})}dG(\underline{\theta}) < \delta) \quad \text{(by Lemma 2)} \\ &\le P(((2\pi)^k|\Sigma|)^{-1/2}e^{-\frac{1}{2}\int(\mathbf{y}-\underline{\theta})^T\Sigma^{-1}(\mathbf{y}-\underline{\theta})dG(\underline{\theta})} < \delta) \quad \text{(by Jensen inequality)} \\ &= P(\int(\mathbf{y}-\underline{\theta})^T\Sigma^{-1}(\text{-----})dG(\underline{\theta}) > -2\log((2\pi)^k|\Sigma|)^{1/2}\delta) \\ &\le \frac{E\int(\mathbf{y}-\underline{\theta})^T\Sigma^{-1}(\text{-----})dG(\underline{\theta})}{-2\log((2\pi)^k|\Sigma|)^{1/2}\delta} \quad \text{(by Markov inequality)} \\ &\le \frac{E\lambda_{\max}\int(\mathbf{y}-\underline{\theta})^T(\text{----})dG(\underline{\theta})}{-2\log((2\pi)^k|\Sigma|)^{1/2}\delta}, \quad \lambda_{\max} \text{ denotes maximum root of } \Sigma^{-1} \\ &= \frac{\lambda_{\max}(E\mathbf{y}^T\mathbf{y} - 2E\mathbf{y}E(\underline{\theta}) + E\|\underline{\theta}\|^2)}{-2\log((2\pi)^k|\Sigma|)^{1/2}\delta} \\ &\le \frac{\lambda_{\max}\{E\|\mathbf{y}\|^2 + 2E\|\mathbf{y}\|(E\|\underline{\theta}\|^2)^{1/2} + E\|\underline{\theta}\|^2\}}{-2\log((2\pi)^k|\Sigma|)^{1/2}\delta} \\ &< M_1(\log(((2\pi)^k|\Sigma|)^{1/2}\delta)^{-2})^{-1} \quad \text{(by assumption } E(\|\underline{\theta}\|^{2+\gamma}) < \infty \text{ and } \gamma > 0)\end{aligned}$$

Next

$$\begin{aligned}&E\{\|p'_{n+1}(\mathbf{X}_{n+1})/p_{n+1}(\mathbf{X}_{n+1})\|^2[p_{n+1}(\mathbf{X}_{n+1}) \le \delta]\} \\ &\le (E\|p'_{n+1}(\mathbf{X}_{n+1})/p_{n+1}(\mathbf{X}_{n+1})\|^{2+\gamma})^{\frac{2}{2+\gamma}}(E[p_{n+1}(\mathbf{X}_{n+1}) \le \delta]^{(2+\gamma)/\gamma})^{\frac{\gamma}{2+\gamma}} \quad \text{(by Holder Inequality)} \qquad [G]\end{aligned}$$

Since

$$\|p'_{n+1}(\mathbf{X}_{n+1})/p_{n+1}(\mathbf{X}_{n+1})\| = \|\Sigma^{-1}(\mathbf{X}_{n+1} - A_{n+1}\underline{\theta} - \mathbf{b}_{n+1})\|$$
$$\leq \|\Sigma_{n+1}^{-1}\|(\|\mathbf{X}_{n+1}\| + \|A_{n+1}\|\|\underline{\theta}\| + \|\mathbf{b}_{n+1}\|),$$

$$E\|p'_{n+1}(\mathbf{X}_{n+1})/p_{n+1}(\mathbf{X}_{n+1})\|^{2+\gamma} \leq \|\Sigma_{n+1}^{-1}\|^{2+\gamma}E\,(\|\mathbf{X}_{n+1}\| + \|A_{n+1}\|\|\theta\| + \|\mathbf{b}_{n+1}\|)^{2+\gamma}$$
$$\leq 2^{1+\gamma}\|\Sigma_{n+1}^{-1}\|^{2+\gamma}(E\|\mathbf{X}_{n+1}\|^{2+\gamma} + \|A_{n+1}\|^{2+\gamma}E\|\underline{\theta}\|^{2+\gamma} + \|\mathbf{b}_{n+1}\|^{2+\gamma})$$
$$< M_2 \quad \text{(by assumption}(A4),(A5),(A6))$$

Thus [G] becomes

$$< M_2 E\,[p_{n+1}(\mathbf{X}_{n+1}) \leq \delta]^{\frac{\gamma}{2+\gamma}}$$
$$< M_1^{\frac{\gamma}{2+\gamma}} M_2(-2\log(((2\pi)^k|\Sigma|)^{1/2}\delta))^{-\frac{\gamma}{2+\gamma}} \quad (E\,[p_{n+1}(\mathbf{X}_{n+1}) \leq \delta] = P[p_{n+1}(\mathbf{X}_{n+1}) \leq \delta])$$
$$= c\{-2\log(((2\pi)^k|\Sigma|)^{1/2}\delta)\}^{-\frac{\gamma}{2+\gamma}}$$
$$= c\{\log((2\pi)^k|\Sigma|\delta^2)^{-1}\}^{-\frac{\gamma}{2+\gamma}}.$$

Proof of Theorem 3 :

$$R_n(\mathbf{t}_n^*, G) - R_{n+1}(G)$$
$$= E\|\mathbf{t}_{n,G} - \mathbf{t}_n^*\|^2$$
$$= E\|A_{n+1}^{-1}\Sigma_{n+1}(\frac{p'_{n+1}(\mathbf{X}_{n+1})}{p_{n+1}(\mathbf{X}_{n+1})} - \frac{\hat{p}'_{n+1}(\mathbf{X}_{n+1})}{\max\{\hat{p}_{n+1}(\mathbf{X}_{n+1}), \delta_{n+1}\}})\|^2$$
$$\leq C^* \sum_{s=1}^{k} E(\frac{p'_{n+1,s}(\mathbf{X}_{n+1})}{p_{n+1}(\mathbf{X}_{n+1})} - \frac{\hat{p}'_{n+1,s}(\mathbf{X}_{n+1})}{\max\{\hat{p}_{n+1}(\mathbf{X}_{n+1}), \delta_{n+1}\}})^2 \quad (\text{where } C^* = a_*^{-2}k^2\lambda_{1,n+1}^2)$$
$$\leq C^* \sum_{s=1}^{k} 2\delta_{n+1}^{-2}\{E(p'_{n+1,s}(\mathbf{X}_{n+1}) - \hat{p}'_{n+1,s}(\mathbf{X}_{n+1}))^2\}$$
$$+ C^* \sum_{s=1}^{k} 2\delta_{n+1}^{-2}E\,[(\frac{p'^2_{n+1}(\mathbf{X}_{n+1})}{p^2_{n+1}(\mathbf{X}_{n+1})})\{(p_{n+1}(\mathbf{X}_{n+1}) - \hat{p}_{n+1}(\mathbf{X}_{n+1}))^2 + \delta_{n+1}^2[\hat{p}_{n+1}(\mathbf{X}_{n+1}) < \delta_{n+1}]\}]$$

(The last inequality follows from the following simple inequality :

$$[ab^{-1} - c\max\{d, \eta\}^{-1}]^2 \leq 2\eta^{-2}\{(a-c)^2 + a^2b^{-2}((b-d)^2 + \eta^2[d < \eta])\}$$

where $[d < \eta]$ denotes the indicator function and a, b (> 0), c, d and $\eta(>0)$ are reals .)

$$\leq C^* \sum_{s=1}^{k} 2\delta_{n+1}^{-2} E_{X_{n+1}}(E_{X_1,\ldots,X_n}(p'_{n+1,s}(X_{n+1}) - \hat{p}'_{n+1,s}(X_{n+1}))^2 \mid X_{n+1} = x)$$

$$+ 2C^*\delta_{n+1}^{-2} E_{X_{n+1}}\{\frac{p'^2_{n+1}(X_{n+1})}{p^2_{n+1}(X_{n+1})} E_{X_1,\ldots,X_n}[(p_{n+1}(X_{n+1}) - \hat{p}_{n+1}(X_{n+1}))^2 \mid X_{n+1} = x]\}$$

$$+ 2C^*[\hat{p}_{n+1}(X_{n+1}) < \delta_{n+1}]$$

$$\leq 2C^*\delta_{n+1}^{-2} \sum_{s=1}^{k} O(n^{-\frac{1}{1+\rho}}(\log n)^{k+1})$$

$$+ 2C^*\delta_{n+1}^{-2} \sum_{s=1}^{k} E_{X_{n+1}}\{\frac{p'^2_{n+1}(X_{n+1})}{p^2_{n+1}(X_{n+1})}\} O(n^{\frac{-1}{1+\rho}}(\log n)^{k})$$

$$+ 2C^* \sum_{s=1}^{k} E_{X_{n+1}}\{\frac{p'^2_{n+1}(X_{n+1})}{p^2_{n+1}(X_{n+1})}[\hat{p}_{n+1}(X_{n+1}) < \delta_{n+1}]\}$$

$$\leq 2C^*((2\pi)^k|\Sigma|)n^{\beta}kO(n^{-\frac{1}{1+\rho}}(\log n)^{k+1})$$

$$+ 2C^*1((2\pi)^k|\Sigma|)n^{\beta}kM_2O(n^{-\frac{1}{1+\rho}}(\log n)^{k})$$

$$+ 2C^*c(\log\{(2\pi)^k|\Sigma|\}^{-1}\delta_{n+1}^{-2})^{-\frac{\gamma}{2+\gamma}}$$

$$= M_3O(n^{\beta-\frac{1}{1+\rho}}(\log n)^{k+1})$$

$$+ M_4O(n^{\beta-\frac{1}{1+\rho}}(\log n)^{k})$$

$$+ M_5O(\log n^{\beta})^{-\frac{\gamma}{2+\gamma}}$$ (where M_3, M_4 and M_5 are suitable constants.)

$$= O((\log n)^{-\frac{\gamma}{2+\gamma}})$$

since by the assumption $\beta - \frac{1}{1+\rho} < 0$. This completes the proof of Theorem 3.

Chapter 4 Applications of Empirical Bayes Estimators to Time Series Regression Models

4.1 Introduction

In this chapter we attempt to apply the theory developed in the last chapter to estimate coefficients in econometric models such as the dynamic linear regression models (DLRM). Since most data obtained in econometric models are time series observations, we do the estimation in two ways: one by assuming panel data (see Section 4), and the other by assuming one realization with independent pieces. (see Section 2 and 3).

4.2 Empirical Bayes Estimation in the Multiple Linear Regression Model (MLRM) with Serially Correlated Disturbances

The classical linear regression model, written in vector notation, has the form

$$y_t = \mathbf{x}'_t\underline{\beta} + e_t, \quad t = 1, \dots, T.$$

A key assumption in the classical model is that the disturbances are uncorrelated. However, in dealing with time series, this may not always be reasonable. If the disturbance term is regarded as being made up of a number of omitted variables, then in a time series context it seems more than likely that it will exhibit serial correlation. Results to follow show that there is much to gain and little to lose by considering alternatives to the independent disturbance assumption of the classical linear regression model.

The classical linear regression model can be written in matrix form

$$\mathbf{y} = X\underline{\beta} + \mathbf{e}$$

where $\mathbf{y}$ is a $(T \times 1)$ vector of observations on a dependent variable, X is a $(T \times k)$ nonstochastic design matrix and $\mathbf{e}$ is a random vector with $E(\mathbf{e}) = \mathbf{0}$ and $E(\mathbf{ee}') = \Phi = \sigma^2\Psi$. Autocorrelation exists if the disturbance terms corresponding to different observations are correlated, that is, if Ψ is not diagonal. Although this model can occur with cross-sectional data that are based on some kind of natural ordering, or are not drawn from a random sample of cross-sectional units, it is generally associated with time series data. In this case one needs to estimate the correlation of e_t with itself over time, and since this involves T(T-1)/2 unknown parameters, some kind of restriction needs to be placed on Ψ . A common assumption that leads to a reduction in the number of parameters is that the e_t are observations on a particular stationary stochastic process.

4.2.1 EBE in MLRM with AR(1) Disturbance

The main attraction of handling serially correlated disturbances by the stationary AR (1) process is its ease of estimation.

4.2.1.1 The Model

The model with AR(1) disturbance can be written as

$$\mathbf{y} = X\underset{\sim}{\beta} + \mathbf{e}, \quad e_t = \rho e_{t-1} + u_t, \quad t = 1, 2, \dots, T.$$

where

$$\mathbf{y} = (y_1, \dots, y_T)', \quad \underset{\sim}{\beta} = (\beta_1, \dots, \beta_k)'$$

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ \cdots & & & \\ \cdots & & & \\ \cdots & & & \\ x_{T1} & x_{T2} & \cdots & x_{Tk} \end{bmatrix}, \quad \mathbf{e} = (e_1, \dots, e_T)'$$

with assumptions:

(i) $|\rho| < 1$ (for stationarity)

(ii) u_t are i i d with zero mean and constant variance σ_u^2 .

4.2.1.2 Estimation of the parameters

From the above assumptions, we know that

$$E(\mathbf{ee}') = \sigma_u^2\Omega = \sigma_u^2\frac{1}{1-\rho^2}\begin{bmatrix} 1 & \rho & \dots & \rho^{T-1} \\ \rho & 1 & \dots & \rho^{T-2} \\ \dots & & & \\ \rho^{T-1} & & \dots & \rho \quad 1 \end{bmatrix}$$

Since the covariance matrix of $\mathbf{e}$ is nonspherical (i.e., not a scalar multiple of the identity matrix), the ordinary least squares estimator, though unbiased, is inefficient relative to the generalized least squares estimator by Aitken's theorem. The generalized least squares estimator (GLSE) is

$$\tilde{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\mathbf{y}$$

where

$$\Omega^{-1} = \begin{bmatrix} 1 & -\rho & 0 & & \\ -\rho & 1+\rho^2 & -\rho & & \\ 0 & -\rho & 1+\rho^2 & & \mathbf{0} \\ & & \dots & & \\ \mathbf{0} & -\rho & 1+\rho^2 & -\rho & \\ & \dots & -\rho & 1 & \end{bmatrix}$$

If ρ is known, the GLSE can be easily computed and under normality of disturbance u_t has the small sample property of minimum variance, unbiased and the large sample property of asymptotic efficiency. More frequently, however, ρ is unknown and must be estimated consistently by an estimate $\hat{\rho}$. Replacement of ρ with $\hat{\rho}$ in the generalized least squares formula provides a feasible generalized least squares estimator which is asymptotically consistent to $\tilde{\beta}$ if certain regularity conditions are satisfied.

There are several alternative methods for finding estimators of $\underline{\beta}$. They are, for example, the maximum likelihood search and the Cochrane-Orcutt iterative procedure, for details see Harvey (1981a).

There are also several methods of finding a consistent estimator for ρ when it is unknown, such as the Cochrane-Orcutt method and the Durbin estimator (see Fomby et. al. 1984).

The large sample properties of the ML (maximum likelihood) estimator of $\underline{\beta}$, ρ and σ_u^2 in the presence of normal disturbances are unaffected by the initial conditions. The simplest approach is to take y_1 as fixed. Taking the expectation of the matrix of second derivatives of Log L (L denotes likelihood function) yields a block diagonal matrix. The MLE of $\underline{\beta}$ is therefore distributed independent of the estimators of ρ and σ_u^2. Thus, asymptotically, $(\tilde{\underline{\beta}}', \tilde{\rho}, \tilde{\sigma}_u^2)'$ has a multivariate normal distribution with mean $(\underline{\beta}', \rho, \sigma_u^2)'$ and asymptotic covariance matrix

$$Avar(\underline{\beta}, \rho, \sigma_u^2) = \begin{bmatrix} \sigma_u^2(X'\Omega^{-1}X)^{-1} & 0 & 0 \\ 0 & \dfrac{(1-\rho^2)}{T} & 0 \\ 0 & 0 & \dfrac{2\sigma_u^4}{T} \end{bmatrix}$$

Note that $(X'\Omega^{-1}X)^{-1}$ will not be identical, if X does not remain unchanged.

Stating that a particular estimator has a certain limiting distribution acts as a guide indicating the approximate results which can be expected in finite samples. That the conditions could be conceived as holding as $T \to \infty$, however, is neither necessary nor sufficient for the asymptotic approach to be reasonable in practice.

4.2.1.3 EBE of the Parameters

Consider a MLRM $\{y_t \mid \mathbf{X}, \underline{\theta}\}$ with AR(1) disturbances where $\underline{\theta} = (\underline{\beta}', \rho)'$ is a vector of parameters in our model. Suppose that there are n independent realizations of $\{y_t^{(j)} \mid \mathbf{X}^{(j)}, \underline{\theta}^{(j)}\}$ each with length T_j and $\underline{\theta} = \underline{\theta}^{(j)}$, $j = 1, 2, \ldots, n$. Here the $\underline{\theta}^{(j)}$ are randomly and independently drawn from a common distribution G($\underline{\theta}$). The problem is to estimate $\underline{\theta}^{(j)}$ using information from all n realizations. We can consider those realizations to be independent since the time gaps between successive periods is assumed large.

Let $y^{(j)} = (y_1^{(j)}, \dots , y_{T_j}^{(j)})'$. Denote by $\hat{\underline{\theta}}^{(j)} = (\hat{\underline{\beta}}^{(j)'}, \hat{\rho}^{(j)})'$ an asymptotically efficient estimator of $\underline{\theta}$ using only the information in $y^{(j)}$, for example, one of estimators given in subsection 4.2.1.2 . It is further assumed that G($\underline{\theta}$) is absolutely continuous. In the next paragraph an empirical Bayes estimator for $\underline{\theta}^{(j)} = (\underline{\beta}^{(j)'}, \rho^{(j)})'$ is obtained for large enough T_j and n .

Since $\hat{\underline{\theta}}^{(j)} = (\hat{\underline{\beta}}^{(j)'}, \hat{\rho}^{(j)})'$ is asymptotically sufficient , for large T_j the Bayes estimator $\bar{\underline{\theta}}^{(j)} = (\bar{\underline{\beta}}^{(j)'}, \bar{\rho}^{(j)})'$ given $y^{(j)}$ under a squared error loss can be expressed as

$$\bar{\underline{\theta}}^{(j)} = E(\underline{\theta}^{(j)} \mid \hat{\underline{\theta}}^{(j)}) \tag{4.1}$$

It is known that for large T_j the asymptotic probability density for $\hat{\underline{\theta}}^{(j)} = (\hat{\underline{\beta}}^{(j)'}, \hat{\rho}^{(j)})'$ given $\underline{\theta}^{(j)} = (\underline{\beta}^{(j)'}, \rho^{(j)})'$ is

$$f(\hat{\underline{\theta}}^{(j)} \mid \underline{\theta}^{(j)}) = \frac{1}{(2\pi)^{k/2} \mid \Sigma_{(j)} \mid^{1/2}} \exp\{ - \frac{1}{2}(\hat{\underline{\theta}}^{(j)} - \underline{\theta}^{(j)})'\Sigma_{(j)}^{-1}(\hat{\underline{\theta}}^{(j)} - \underline{\theta}^{(j)})\} \tag{4.2}$$

where

$$\Sigma_{(j)} = \begin{bmatrix} \sigma_u^2(X^{(j)'}\Omega_{(j)}^{-1}X^{(j)})^{-1} & 0 \\ 0 & \dfrac{(1 - \rho^{(j)2})}{T_j} \end{bmatrix}$$

Note that $\hat{\underline{\theta}}^{(j)} \mid \underline{\theta}^{(j)} \sim N(\underline{\theta}^{(j)}, \Sigma_{(j)}^{-1})$, where it should be noted that $\Sigma_{(j)}$ will not be identical for each j.

Following the Miyasawa's result (see chapter 3) we have

$$\bar{\underline{\theta}}^{(j)} = \hat{\underline{\theta}}^{(j)} + \hat{\Sigma}_{(j)}\frac{f_G'(\hat{\underline{\theta}}^{(j)})}{f_G(\hat{\underline{\theta}}^{(j)})}, \tag{4.3}$$

where $f_G(\hat{\underline{\theta}}^{(j)}) = \int f(\hat{\underline{\theta}}^{(j)} \mid \underline{\theta}^{(j)}) dG(\underline{\theta}^{(j)})$ and $f_G'(\hat{\underline{\theta}}^{(j)})$ is the derivative of the marginal density $f_G(\hat{\underline{\theta}}^{(j)})$ evaluated at $\hat{\underline{\theta}}^{(j)}$. An empirical Bayes estimator can be obtained if we replace $f_G(\hat{\underline{\theta}})$ and $f_G'(\hat{\underline{\theta}})$ in (4.3) by their corresponding consistent estimators (see chapter 3). Note that for sufficiently large T_j, $\Sigma_{(j)}$ can be consistently estimated by $\hat{\Sigma}_{(j)}$ which is obtained by replacing $\rho^{(j)}$ and σ_u^2 with $\hat{\rho}^{(j)}$ and $\hat{\sigma}_u^2$ respectively.

4.2.1.4 EBE for MLRM with AR(p) Disturbances

The empirical Bayes estimator for the model with AR(p) disturbances when $p > 1$ can be obtained in a straightforward way as an extension of method of the last subsection under the corresponding assumptions.

4.2.2. EBE of MLRM with MA(1) Disturbances

The problem of estimating parameters in regression models with moving-average (MA) or mixing autoregressive moving-average (ARMA) disturbances is quite different from the problem of estimating parameters by pure autoregressive disturbances. Estimating parameters in regression models with MA(1) disturbances embodies difficulties which are not encountered in the pure AR case.

4.2.2.1 The model

The model with MA(1) disturbances can be written as

$$\mathbf{y} = X\underline{\beta} + \mathbf{e}, \quad e_t = u_t + \alpha u_{t-1}, \quad t = 1, 2, \ldots, T.$$

where $\mathbf{y}, X, \underline{\beta}$, $\mathbf{e}$ are the same as described in subsection 4.2.1.2 , with assumptions:

(i) $|\alpha| < 1$ (the invertibility condition)

(ii) u_t are i i d with zero mean and constant variance σ_u^2

4.2.2.2 Estimation of the Parameters

From the above assumptions, we have

$$E(\mathbf{ee}') = \sigma_u^2\Psi = \sigma_u^2 \begin{bmatrix} (1+\alpha^2) & \alpha & & & & \\ \alpha & (1+\alpha^2) & \alpha & & \mathbf{0} & \\ & & \cdots & & & \\ & \mathbf{0} & & \alpha & (1+\alpha^2) & \alpha \\ & & & & \alpha & (1+\alpha^2) \end{bmatrix}$$

since $\sigma_e^2 = \sigma_u^2(1 + \alpha^2)$, $\rho_1 = \dfrac{\alpha}{1+\alpha^2}$, $\rho_s = 0$ for $s \geq 2$. For known α the generalized least squares estimator

$$\bar{\beta}(\alpha) = (X'\Psi^{-1}X)^{-1}X'\Psi^{-1}\mathbf{y}$$

is the best linear unbiased estimator for the MA(1) model. With normality, $\bar{\beta}$ becomes minimum variance unbiased (MVU) and hence efficient.

For unknown α , which is the usual case , we estimate α by the sample autocorrelation coefficient

$$r_1 = \frac{\sum_{t=2}^{T} \hat{e}_t \times \hat{e}_{t-1}}{\sum_{t=1}^{T} \hat{e}_t^2} \qquad (\hat{e}_t\text{:OLS residual})$$

using the relationship

$$r_1 = \frac{\hat{\alpha}}{1+\hat{\alpha}^2} \quad (i.e. \quad \hat{\alpha} = \frac{1-\sqrt{1-4r_1^2}}{2r_1}).$$

Since r_1 is consistent, $\hat{\alpha}$ is consistent by Slusky's theorem. Note that $|r_1| \leq 0.5$ since $|\alpha| < 1$. The feasible generalized least squares estimator

$$\tilde{\underline{\beta}}(\hat{\alpha}) = (X'\tilde{\Psi}^{-1}X)^{-1}X'\tilde{\Psi}^{-1}\mathbf{y}$$

is asymptotically efficient in the presence of normality and nonstochastic X. Note that $\tilde{\Psi}$ is Ψ with α replaced by $\hat{\alpha}$.

There are also several alternative methods for finding estimators of $\underline{\beta}$, see Harvey (1981a).

Under certain conditions (see Pierce, 1971) all the estimators asymptotically will have a normal distribution, will be consistent, and will have a covariance matrix that can be estimated using the inverse of the matrix of second derivatives of the log likelihood function.

4.2.2.3 EBE of the Parameters

Consider a MLRM $\{y_t \mid X, \underline{\theta}\}$ with MA(1) disturbances where $\underline{\theta} = (\underline{\beta}', \alpha)'$ is a vector of parameters in the model. Suppose that there are n independent realizations of $\{y_t \mid X, \underline{\theta}\}$ each with length T_j and $\underline{\theta} = \underline{\theta}^{(j)}, j = 1, \dots, n$. Here the $\underline{\theta}^{(j)}$ are randomly and independently drawn from a common distribution G ($\underline{\theta}$). The problem is to estimate $\underline{\theta}^{(j)}$ using information from all n realizations. We can consider those realizations to be independent since the time gaps between successive periods is taken to be large.

Let $\mathbf{y}^{(j)} = (y_1^{(j)}, \dots, y_{T_{(j)}}^{(j)})'$. Denote by $\hat{\underline{\theta}}^{(j)} = (\hat{\underline{\beta}}^{(j)\prime}, \hat{\alpha}^{(j)})'$ an asymptotically efficient estimator of $\underline{\theta}^{(j)}$ using only the information in $\underline{y}^{(j)}$, for example, one of estimators given in subsection 4.2.2.2 . It is further assumed that G($\underline{\theta}$) is absolutely continuous. In the next paragraph an empirical Bayes estimator for $\underline{\theta}^{(j)} = (\underline{\beta}^{(j)\prime}, \alpha^{(j)})'$ is obtained for large $T_{(j)}$ and n.

Since $\hat{\underline{\theta}}^{(j)}$ is asymptotically sufficient , for large T_j the Bayes estimator $\underline{\tilde{\theta}}^{(j)}$ given $\mathbf{y}^{(j)}$ under a squared error loss can be expressed as

$$\underline{\theta}^{(j)} = E(\underline{\theta}^{(j)} \mid \hat{\underline{\theta}}^{(j)}). \tag{4.4}$$

It is known that, for large T_j the asymptotic probability density for $\hat{\underline{\theta}}^{(j)}$ given $\underline{\theta}^{(j)}$ is

$$f(\hat{\underline{\theta}} \mid \underline{\theta}^{(j)}) = \frac{1}{(2\pi)^{k/2} |\Sigma_{(j)}|^{1/2}} \exp\{ -\frac{1}{2}(\hat{\underline{\theta}}^{(j)} - \underline{\theta}^{(j)})'\Sigma_{(j)}^{-1}(\hat{\underline{\theta}}^{(j)} - \underline{\theta}^{(j)})\} \tag{4.5}$$

where

$$\Sigma_{(j)} = \begin{bmatrix} \sigma_u^2(X^{(j)'}\Omega_{(j)}^{-1}X^{(j)})^{-1} & 0 \\ 0 & \dfrac{1}{T_j(1 + \alpha^{(j)^2})} \end{bmatrix}.$$

Note that $\hat{\underline{\theta}}^{(j)} \mid \underline{\theta}^{(j)} \sim N(\underline{\theta}^{(j)}, \Sigma_{(j)})$, where $\Sigma_{(j)}$ will not be identical for each j.

Following the Miyasawa's result (see chapter 3) we have

$$\underline{\theta}^{(j)} = \hat{\underline{\theta}}^{(j)} + \hat{\Sigma}_{(j)} \frac{f_G'(\hat{\underline{\theta}}^{(j)})}{f_G(\hat{\underline{\theta}}^{(j)})}, \tag{4.6}$$

where $\hat{\Sigma}_{(j)}$ is the estimator of $\Sigma_{(j)}$ using $\hat{\alpha}^{(j)}$ and $\hat{\sigma}^2$, $f_G(\hat{\underline{\theta}}^{(j)}) = \int f(\hat{\underline{\theta}}^{(j)} \mid \underline{\theta}^{(j)})dG(\underline{\theta}^{(j)})$ and $f_G'(\hat{\underline{\theta}}^{(j)})$ is the derivative of the marginal density $f_G(\hat{\underline{\theta}}^{(j)})$ evaluated at $\hat{\underline{\theta}}^{(j)}$. An empirical Bayes estimator can be obtained if we replace $f_G(\hat{\underline{\theta}}^{(j)})$ and $f_G'(\hat{\underline{\theta}}^{(j)})$ in (4.6) by their corresponding consistent estimators. (see chapter 3) Note that $\hat{\Sigma}_{(j)}$ is the estimator of $\Sigma_{(j)}$ with $\alpha^{(j)}$ and σ_u^2 replaced by $\hat{\alpha}^{(j)}$ and $\hat{\sigma}_u^2$ respectively.

4.2.2.4 Generalization

The results of last subsection can obviously be extended to the models with ARMA(1,1) disturbances.

4.3 EBE in DLRM with Lagged Dependent Variables

Most econometric models that are applied to time series data in economics are constructed in such a way that time lags play an important role in the relationships between variables. For example, consumption in one year varies not only according to the level of disposable income during that year, but also according to the level of consumption already reached in preceding years. Models that fit this general description are often called dynamic and several different types of specification are in regular use. Many of these come within the framework of statistical models that are well known in mathematical statistics, although the detailed specification of the econometric equations often raises new and intriguing problems for statistical inference. In this section, we will be looking at single-equation models with autoregressive disturbances and their empirical Bayes estimators.

4.3.1 EBE of DLRM with Lagged Dependent Variables and Non-autocorrelated Disturbances

4.3.1.1 The Model

The general form of the lagged dependent variable model is given by

$$y_t = \sum_{l=1}^{G} \alpha_l \, y_{t-l} + \mathbf{z}'_t \underline{\gamma} + u_t, \quad t = 1, \ldots, T. \tag{4.7}$$

where the y_t are observations on the dependent variable y, the α_l are scalar parameters, $\mathbf{z}_t$ is a $(K - G) \times 1$ vector of the tth observations on the (K-G) non-stochastic explanatory variables, $\underline{\gamma}$ is the $(K - G) \times 1$ vector of corresponding parameters, and u_t are the random disturbances. (Note that sometimes (4.7) are called linear stochastic difference equations).

Assumptions:

(i) The u_t are independent and normally distributed with zero mean and variance σ_u^2. (*i.e.* $u \sim NID(0, \sigma^2)$) .

(ii) All G roots of the polynomial equation in λ

$$\lambda^G - \alpha_1\lambda^{G-1} - \dots - \alpha_{G-1}\lambda - \alpha_G = 0$$

are less than one in absolute value. (This is the requirement of dynamic stability)

(iii) The following assumption guarantees the convergence of moment matrices of the non-stochastic variables. Let $\eta = 0, 1, \dots$. The matrix

$$Q_\eta = \lim_{T\to\infty} \frac{1}{T-\eta} \sum_{t=1}^{T-\eta} \mathbf{z}_t \mathbf{z}'_{t+\eta}$$

exists for all η and is nonsingular for $\eta = 0$.

(4.7) can be written in familiar matrix form as

$$\mathbf{y} = X\underline{\beta} + \mathbf{u}$$

where

$$\mathbf{y}' = (y_1, \dots, y_T), \quad \underline{\beta}' = (\alpha_1, \dots, \alpha_G, \gamma_1, \dots, \gamma_{(k-G)})$$

$\mathbf{u}' = (u_1, \dots, u_T)$ and

$$X = \begin{bmatrix} y_0 & y_{-1} & \cdots & y_{-G+1} & z_{11} & z_{12} & \cdots & z_{1(k-G)} \\ y_1 & y_0 & \cdots & y_{-G+2} & z_{21} & z_{22} & \cdots & z_{2(k-G)} \\ \cdots & & & & & & & \\ y_{T-1} & y_{T-2} & \cdots & y_{-G+T} & z_{T1} & z_{T2} & \cdots & z_{T(k-G)} \end{bmatrix}$$

Note that the total number of observations on y_t is T + G, not T alone. In what follows the pre-sample values $y_0, y_{-1}, \dots, y_{-G+1}$ will be treated as fixed with all probability statements conditional on these values.

4.3.1.2 Estimation of the Parameters

Under the aforementioned assumptions we can obtain the following properties:

(1) The matrix $Q = \operatorname{plim}(X'X/T)$ is finite and nonsingular and $X'\mathbf{u}/\sqrt{T}$ is asymptotically distributed as $N(0, \sigma^2 Q)$. (here $Q = \operatorname{plim}(X'X/T)$ is defined as the random variables $X'X/T$ converges in probability to Q.)

(2) The ordinary least squares estimators $\hat{\beta} = (X'X)^{-1}X'\mathbf{y}$ and $\hat{\sigma}^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}/T - k$ are consistent estimators of β and σ^2 respectively, where $\hat{\mathbf{u}} = \mathbf{y} - X\hat{\beta}$ is the residual vector. In addition, $\sqrt{T}(\hat{\beta} - \beta)$ has a limiting distribution which is $N(0, \sigma^2 Q^{-1})$. For proof see Schonfelt (1971) .

(3) Under the normality of the $\mathbf{u}$ (disturbances) the ordinary least squares estimator of β is the maximum likelihood estimator conditional on $y_0, \dots, y_{-G+1}$. Thus the ordinary least squares estimator of β is asymptotically efficient.

Note that given a dynamically stable lagged dependent variable model the method of treating the initial conditions makes no difference asymptotically. Suppose, rather than having T + G observations, we have only T observations, the values $y_0, \dots, y_{-G+1}$ not being available. We obtain the same asymptotic results of consistency, normality and/or asymptotic efficiency whether we drop the first G observations and run ordinary least squares on the remaining T − G observations or simply assign arbitrary values (possible zero) to $y_0, \dots, y_{-G+1}$.

Moreover, in the lagged dependent variable model, unlike in the classical normal linear model, the distribution of y_t is not directly deducible. The reason is that the regressors $\mathbf{x}_t$ are no longer nonstochastic. It is therefore necessary to apply Jacobian (determinant) techniques in analyzing the maximum likelihood properties of $\hat{\beta}$.

4.3.1.3 BE and EBE

Consider a dynamically stable lagged dependent variable regression model $\{ y_t \mid X, \underline{\beta}\}$ where $\underline{\beta} = (\alpha_1, \dots, \alpha_G, \gamma_1, \dots, \gamma_{(K-G)})'$ is a vector of parameters in our model. Suppose that there are n independent realizations of $\{ y_t \mid X, \underline{\beta} \}$ each with length T_j and $\underline{\beta} = \underline{\beta}^{(j)}$, $j = 1, 2, \dots, n$. Here the $\underline{\beta}$ are randomly and independently drawn from a common distribution G($\underline{\beta}$). The problem is to estimate $\underline{\beta}^{(j)}$ using information from all n realizations. We can consider those realizations to be independent since the time gaps between successive periods can be assumed large.

Let $\mathbf{y}^{(j)} = (y_1^{(j)}, \dots, y_{T_{(j)}}^{(j)})'$ be the observations of the jth realization. Denote by $\hat{\underline{\beta}}^{(j)}$ an asymptotically efficient estimator of $\underline{\beta}^{(j)}$ using only the information in $\mathbf{y}^{(j)}$, for example, one of estimators given in subsection 4.3.1.2 . It is further assumed that G($\underline{\beta}$) is absolutely continuous. In the next paragraph an empirical Bayes estimator for $\underline{\beta}^{(j)}$ is obtained for large T_j and n.

Since $\hat{\underline{\beta}}^{(j)}$ is asymptotically sufficient for large T_j the Bayes estimator $\bar{\underline{\beta}}^{(j)}$ given $\mathbf{y}^{(j)}$ under a squared error loss can be expressed as

$$\bar{\underline{\beta}}^{(j)} = E(\underline{\beta}^{(j)} \mid \hat{\underline{\beta}}^{(j)}). \tag{4.8}$$

It is known that for large T_j the asymptotic probability density for $\hat{\underline{\beta}}^{(j)} = (\hat{\alpha}_1^{(j)}, \dots, \hat{\alpha}_G^{(j)}, \hat{\gamma}(j)_1, \dots, \hat{\gamma}_{(k-G)}^{(j)})'$ given $\underline{\beta}^{(j)}$ is

$$f(\hat{\underline{\beta}}^{(j)} \mid \underline{\beta}^{(j)}) = \frac{1}{(2\pi\sigma^2)^{k/2} \mid Q_{(j)}^{-1} \mid^{1/2}} \exp\{ -\frac{1}{2}(\hat{\underline{\beta}}^{(j)} - \underline{\beta}^{(j)})' Q_{(j)} (\hat{\underline{\beta}}^{(j)} - \underline{\beta}^{(j)})\} \tag{4.9}$$

where

$$Q_{(j)} = \lim_{T_j \to \infty} \frac{X'_{(j)} X_{(j)}}{T_j}$$

and

$$
X_{(j)} = \begin{bmatrix} y_0^{(j)} & y_{-1}^{(j)} & \cdots & y_{-G+1}^{(j)} & z_{11} & z_{12} & \cdots & z_{1(k-G)} \\ & & & \cdot & \cdot & \cdot & & \\ & & & \cdot & \cdot & \cdot & & \\ & & & \cdot & \cdot & \cdot & & \\ y_{T_j-1}^{(j)} & y_{T_j-2}^{(j)} & \cdots & y_{-G+T_j}^{(j)} & z_{T_j1} & z_{T_j2} & \cdots & z_{T_j(k-G)} \end{bmatrix}
$$

Note that $\hat{\beta}^{(j)} \mid \beta^{(j)} \sim N(\beta^{(j)}, \sigma^2 Q_{(j)}^{-1})$, where $Q_{(j)}^{-1}$ will not be identical for each j.

Following the Miyasawa's result (see chapter 3) we have

$$
\bar{\beta}^{(j)} = \hat{\beta}^{(j)} + \hat{\sigma}_{(j)}^2 Q_{(j)}^{-1} \frac{f_G'(\hat{\beta}^{(j)})}{f_G(\hat{\beta}^{(j)})}, \tag{4.10}
$$

where $f_G(\hat{\beta}^{(j)}) = \int f(\hat{\beta}^{(j)} \mid \beta^{(j)}) dG(\beta^{(j)})$ and $f_G'(\hat{\beta}^{(j)})$ is the derivative of the marginal density $f_G(\hat{\beta}^{(j)})$ evaluated at $\hat{\beta}^{(j)}$. An empirical Bayes estimator can be obtained if we replace $f_G(\hat{\beta}^{(j)})$ and $f_G'(\hat{\beta}^{(j)})$ in (4.10) by their corresponding consistent estimators (see chapter 3).

4.3.2 EBE of DLRM with Lagged Dependent Variables and Autocorrelated Disturbances

4.3.2.1 The Model with AR(1) disturbances

The main issues which arise in estimating the lagged dependent variable model with AR(1) disturbances can be discussed with respect to the simple model

$$
\begin{aligned}
y_t &= \alpha y_{t-1} + \beta x_t + e_t, \quad t = 2, \dots, T. \\
e_t &= \rho e_{t-1} + u_t
\end{aligned}
$$

where

(i) $|\alpha| < 1, |\rho| < 1$ (for stationary).

(ii) $u_t \sim NID(0, \sigma^2)$.

(iii) $|x_t| < B$, all t, and $\lim_{t \to \infty} \frac{1}{T} \sum_{t=1}^{T-\tau} x_t x_{t+\tau} = q_\tau$ exists and q_0 is positive.

The best-known of such models are the partial adjustment and the adaptive expectations models in econometrics.

4.3.2.2 Estimation of the Parameters

Although ordinary least squares has very desirable properties when e_t is white noise (i i d disturbance, see subsection 4.2.1.2), it becomes less attractive in the presence of serial correlation. Since $E(e_t e_{t-1}) \neq 0$, e_t and y_{t-1} will be correlated even in large samples. As a consequence the ordinary least squares estimator is no longer consistent.

There are several methods for obtaining the ML estimator for $\underline{\theta} = (\alpha, \beta, \rho)'$ which is always asymptotically efficient, for example, iterative Gauss-Newton (non-linear optimazation) method and Hatanaka's two step method. There are also a number of ways to obtain starting values for an iterative ML procedure for this model, for example, using the method of instrumental variables for estimating α and β, and then computing an estimate of ρ from the residuals in the usual way.

Under the conditions given in subsection 4.3.2.1 the ML estimator for $\underline{\theta} = (\alpha, \beta, \rho)'$ is asymptotically normally distributed with a mean of $\underline{\theta}$ and a covariance matrix equal to the inverse of the asymptotic information matrix IA ($\underline{\theta}$) divided by T. That is

$$\hat{\underline{\theta}} = (\hat{\alpha}, \hat{\beta}, \hat{\rho})' \sim N(\underline{\theta}, \ \mathrm{Avar}(\hat{\underline{\theta}}))$$

where

$$\mathrm{Avar}(\hat{\theta}) = \frac{1}{T-1}\begin{bmatrix} \mathbf{V} & \begin{matrix}\frac{1}{1-\alpha\rho} \\ 0\end{matrix} \\ \frac{1}{1-\alpha\rho} \quad 0 & \frac{1}{1-\rho^2} \end{bmatrix}^{-1}_{3\times 3}$$

with

$$V \equiv \underset{T\to\infty}{\mathrm{plim}} \frac{1}{T\hat{\sigma}^2}(X'\hat{\Omega}^{-1}X), \quad \sigma^2\Omega = \mathrm{var}(\mathbf{e}).$$

Here Ω is the same as given in subsection 4.2.1.2 and

$$X = \begin{bmatrix} y_1 & x_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ y_{T-1} & x_T \end{bmatrix}$$

4.3.2.3 BE and EBE

Consider a dynamically stable lagged dependent variable regression model with AR(1) disturbances $\{ y_t \mid X, \underline{\theta} \}$ where $\underline{\theta} = (\alpha, \beta, \rho)'$ is a vector of parameters. Suppose that there are n independent realizations of $\{y_t^{(j)} \mid X^{(j)}, \underline{\theta}^{(j)}\}$ each with length T_j and $\underline{\theta} = \underline{\theta}^{(j)}, j = 1, 2, \ldots, n$. Here the $\underline{\theta}^{(j)}$ are randomly and independently drawn from a common distribution G($\underline{\theta}$). The problem is to estimate $\underline{\theta}^{(j)}$ using information from all n realizations. We can consider those realizations to be independent since the time gaps between successive periods can be considered large.

Let $\mathbf{y}^{(j)} = (y_1^{(j)}, \ldots, y_{T_j}^{(j)})'$ be the jth realization of observations. Denote by $\hat{\underline{\theta}}^{(j)}$ an asymptotically efficient estimator of $\underline{\theta}^{(j)}$ using only the information in $\mathbf{y}^{(j)}$, for example, one of estimators given in subsection 4.3.2.2 . It is further assumed that G($\underline{\theta}$) is absolutely continuous. In the next paragraph an empirical Bayes estimator for $\underline{\theta}^{(j)} = (\alpha^{(j)}, \beta^{(j)}, \rho^{(j)})'$ is obtained for large T_j and n .

Since $\hat{\underline{\theta}}^{(j)}$ is asymptotically sufficient , for large T_j, the Bayes estimator $\underline{\overline{\theta}}^{(j)} = (\overline{\alpha}^{(j)}, \overline{\beta}^{(j)}, \overline{\rho}^{(j)})'$ given $y^{(j)}$ under a squared error loss can be expressed as

$$\underline{\overline{\theta}}^{(j)} = E(\underline{\theta}^{(j)} \mid \hat{\underline{\theta}}^{(j)}). \tag{4.11}$$

It is known that for large T_j the asymptotic probability density for $\hat{\underline{\theta}}^{(j)} = (\hat{\alpha}^{(j)}, \hat{\beta}^{(j)}, \hat{\rho}^{(j)})'$ given $\underline{\theta}^{(j)} = (\alpha^{(j)}, \beta^{(j)}, \rho^{(j)})'$ is

$$f(\hat{\underline{\theta}}^{(j)} \mid \underline{\theta}^{(j)}) = \frac{1}{(2\pi)^{k/2} |\Sigma_{(j)}|^{1/2}} \exp\{ -\frac{1}{2}(\hat{\underline{\theta}}^{(j)} - \underline{\theta}^{(j)})'\Sigma_{(j)}^{-1}(\hat{\underline{\theta}}^{(j)} - \underline{\theta}^{(j)})\} \tag{4.12}$$

where

$$\Sigma_{(j)} = \frac{1}{T-1} \begin{bmatrix} V_{(j)} & \begin{matrix} \dfrac{1}{1-\alpha^{(j)}\rho^{(j)}} \\ 0 \end{matrix} \\ \dfrac{1}{1-\alpha^{(j)}\rho^{(j)}} \quad 0 & \dfrac{1}{1-\rho^{(j)^2}} \end{bmatrix}^{-1}_{3\times 3}$$

and

$$V_{(j)} = \operatorname{plim} \frac{1}{T_j\hat{\sigma}^2}(X'_{(j)}\hat{\Omega}_{(j)}^{-1}X_{(j)}).$$

Note that $\hat{\underline{\theta}}^{(j)} \mid \underline{\theta}^{(j)} \sim N(\theta^{(j)}, \Sigma_{(j)})$, where $\Sigma_{(j)}$ will not be identical for each j.

Following the Miyasawa's result (see chapter 3) we have

$$\underline{\overline{\theta}}^{(j)} = \hat{\underline{\theta}}^{(j)} + \hat{\Sigma}^{(j)}\frac{f_G'(\hat{\underline{\theta}}^{(j)})}{f_G(\hat{\underline{\theta}}^{(j)})}, \tag{4.13}$$

where $f_G(\hat{\underline{\theta}}^{(j)}) = \int f(\hat{\underline{\theta}}^{(j)} \mid \underline{\theta}^{(j)}) dG(\underline{\theta}^{(j)})$ and $f_G'(\hat{\underline{\theta}}^{(j)})$ is the derivative of the marginal density $f_G(\hat{\underline{\theta}}^{(j)})$ evaluated at $\hat{\underline{\theta}}^{(j)}$. An empirical Bayes estimator can be obtained if we replace $f_G(\hat{\underline{\theta}})$ and $f_G'(\hat{\underline{\theta}})$ in (4.13) by their corresponding consistent estimators (see chapter 3). Note that $\hat{\Sigma}_{(j)}$ is obtained from $\Sigma_{(j)}$ by replacing the corresponding sample estimates.

The empirical Bayes estimator for the more general model of a dynamically stable lagged dependent variable with AR(1) disturbances can be obtained in a straightforward way as an extension of the method of the last subsection under the corresponding assumptions.

4.3.2.4 The model with ARMA disturbances and its EBE

In more general formulations of dynamically stable lagged dependent variable models, the disturbances are generated by an ARMA (p,q) process. However, the main principles involved in estimating such models can be illustrated by taking the simple model

$$y_t = \alpha y_{t-1} + \beta x_t + u_t + \theta u_{t-1} \quad t = 1, 2, \dots, T$$

allowing the disturbance term to follow an MA(1) process, when u_t is distributed as $NID(0, \sigma^2)$. This model which was an alternative form of the Koyck or geometric distributed lag model falls within the class of ARMAX models.

The maximum likelihood estimator for (α, β, θ) can be obtained if the Gauss-Newton iterative method is employed, where the initial consistent estimators of α and β can be obtained by the method of instrumental variables, and the residuals used to construct a consistent estimator of θ using

$$\hat{\theta} = [1 - (1 - 4r_1^2)^{1/2}]/2r_1$$

where $r_1 = \frac{\sum_{t=1}^{T} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=0}^{T} \hat{u}_t^2}$, and $\hat{u}_t$ are residuals.

Now we can find the EBE for this model in a manner similar to that given in subsection 4.3.2.3 .

4.4 EBE in TSRM with Panel Data (Temporal Cross-section Data)

Linear regression models with coefficients regarded as random variables have very wide applicability in the analysis of economic, sociological, biological and industrial panel data. Traditionally, interest has been centered mainly on estimating the distributional parameters (sometimes called 'hyper' or 'secondary' parameters) of the random coefficients. (see Swamy 1971 and 1974). In this section we exploit the EB approach to discuss the random coefficient TSRM model as it is applied to panel data in a time series context. In this regard, only two papers have been found (see Li and Hui 1983 a&b). They just mentioned the use of the EB approach to estimate random coefficients in the ARMA model. Our main interest is to use all 'past' information to estimate the current individual parameters, in an optimal way.

In such a problem, we typically have a panel of short time series, e.g., growth curves of plants or animals, or annual observations on some economic indicator from various regions of a country. It seems reasonable to suppose that the time series parameters across the individual units may be random samples from some population.

4.4.1 EBE in TSRM with non-autocorrelected disturbances

4.4.1.1 The Model and Estimation of the Parameters

Consider the following dynamic linear regression model with lagged dependent variables used to analyze a time series of cross-sections

$$y_{it} = \beta_{i0} + \sum_{k=0}^{k-1} \delta_{ik} z_{kit} + \sum_{g=1}^{G} \gamma_{ig} y_{i(t-g)} + u_{it} \quad i = 1, \dots, n;\ t = 1, \dots, T_i.$$

or in matrix form

$$\mathbf{y}_i = X_i \underline{\beta}_i + \mathbf{u}_i \quad i = 1, \dots, n \tag{4.14}$$

where

$$\mathbf{y}_i = (y_{i1}, \dots, y_{iT_i})',$$

$$X_i = \begin{bmatrix} 1 & z_{0i1} & z_{1i1} & \cdots & z_{(k-1)i1} & y_{i0} & \cdots & y_{i(1-G)} \\ 1 & z_{0i2} & z_{1i2} & \cdots & z_{(k-1)i2} & y_{i1} & \cdots & y_{i(2-G)} \\ \cdot & \cdot & \cdot & & & & & \\ \cdot & \cdot & \cdot & & & & & \\ \cdot & \cdot & \cdot & & & & & \\ 1 & z_{0iT_i} & z_{1iT_i} & \cdots & z_{(k-1)iT_i} & y_{i(T_i-1)} & \cdots & y_{i(T_i-G)} \end{bmatrix},$$

$$\underline{\beta}_i = (\beta_{i0}, \delta_{i0}, \dots, \delta_{i(k-1)}, \gamma_{i1}, \dots, \gamma_{iG})',$$

and

$$\mathbf{u}_i = (u_{i1}, \dots, u_{iT_i})'.$$

Suppose we have independent observations on the y's and z's for n individual units taken over T_i periods. The individual units here may be firms, consumers, communities or regions. The subscript i refers to an observation on an individual unit, while the subscript t refers to an observation for a year.

We assume

(1) $u_{it} \sim N(0, \sigma^2)$, $i = 1, 2, \dots n$; $t = 1, \dots T_i$ where σ^2 is a known constant variance.

(2) $\underline{\beta}_i \equiv (\beta_{i0}, \delta_{i0}, \dots, \delta_{i(k-1)}, \gamma_{i1}, \dots, \gamma_{iG})'$ is a realization in the ith individual of the random vector $\underline{\beta}$ with unknown and unspecified prior distribution G ($\underline{\beta}$).

(3) The model is dynamically stable for each i, i= 1,.....,n.

Let $\hat{\underline{\beta}}_i$ denote the least square or maximum likelihood estimator using only the information in y_i and X_i, (see EBE in MLRM). In addition, we know that for each i, i= 1,....,n.

(i) $\hat{\underline{\beta}}_i$ is consistent

(ii) $\sqrt{T_i}(\hat{\underline{\beta}}_i - \underline{\beta}_i) \sim N(0, \sigma^2 Q_i^{-1})$ (asymptotically) where $Q_i = \text{plim}\dfrac{X'_i X_i}{T_i}$.

4.4.1.2 Bayes Estimators for the Coefficients

Let us consider estimating $\underline{\beta}_i$ with a squared error loss. Since $\hat{\underline{\beta}}_i$ is an asymptotically sufficient statistic for $\underline{\beta}_i$ in the ith individual unit the Bayes estimator for $\underline{\beta}_i$, under a squared error loss, can be expressed as

$$\bar{\underline{\beta}}_i = E(\underline{\beta}_i \mid y_i) = E(\underline{\beta}_i \mid \hat{\underline{\beta}}_i). \tag{4.15}$$

The Bayes risk incurred in using $\bar{\underline{\beta}}_i$ is given by

$$E(\bar{\underline{\beta}}_i - \underline{\beta}_i)'(\bar{\underline{\beta}}_i - \underline{\beta}_i) = E \text{ tr var } (\underline{\beta}_i \mid \hat{\underline{\beta}}_i). \tag{4.16}$$

From (ii) in last subsection, we have

$$f(\hat{\beta}_i \mid \beta_i) = \frac{1}{(2\pi\sigma^2)^{p/2} \mid Q_i \mid^{1/2}} \exp\{ -\frac{1}{2\sigma^2}(\hat{\beta}_i - \beta_i)' Q_i (\hat{\beta}_i - \beta_i)\} \tag{4.17}$$

where p = K + G.

Now

$$\frac{\partial \log f(\hat{\beta}_i \mid \beta_i)}{\partial \hat{\beta}_i} = -\sigma^{-2} Q_i(\hat{\beta}_i - \beta_i) \tag{4.18}$$

or

$$\frac{\partial f(\hat{\beta}_i \mid \beta_i)}{\partial \hat{\beta}_i} = -\sigma^{-2} Q_i(\hat{\beta}_i - \beta_i) f(\hat{\beta}_i \mid \beta_i)$$

or

$$\beta_i f(\hat{\beta}_i \mid \beta_i) = \hat{\beta}_i f(\hat{\beta}_i \mid \beta_i) + (\sigma^{-2} Q_i)^{-1} \frac{\partial f(\hat{\beta}_i \mid \beta_i)}{\partial \hat{\beta}_i}. \tag{4.19}$$

Thus

$$\begin{aligned} E(\beta_i \mid \hat{\beta}_i) &= f_G(\hat{\beta}_i)^{-1} \int \beta_i f(\hat{\beta} \mid \beta_i) dG(\beta_i) \\ &= \hat{\beta}_i + (\sigma^{-2} Q_i)^{-1} f_G(\hat{\beta}_i)^{-1} \int \frac{\partial f(\hat{\beta}_i \mid \beta_i)}{\partial \hat{\beta}_i} dG(\beta_i) \\ &= \hat{\beta}_i + (\sigma^{-2} Q_i)^{-1} f_G(\hat{\beta}_i)^{-1} \frac{\partial f_G(\hat{\beta}_i)}{\partial \hat{\beta}_i}. \end{aligned}$$

The last equality follows by interchanging integration and differentiation (see Theorem 2.9 of Lehmann (1959)). That is, the Bayes estimator exhibits as

$$\bar{\beta}_i = E(\beta_i \mid \hat{\beta}_i) = \hat{\beta}_i + \sigma^2 Q_i^{-1} \frac{f_G'(\hat{\beta}_i)}{f_G(\hat{\beta}_i)} \tag{4.20}$$

where we have let

$$f_G'(\hat{\beta}_i) = \frac{\partial f_G(\hat{\beta}_i)}{\partial \hat{\beta}_i} = \left(\frac{\partial f_G(\hat{\beta}_i)}{\partial \hat{\beta}_{i1}}, \dots, \frac{\partial f_G(\hat{\beta}_i)}{\partial \hat{\beta}_{iT}}\right)$$

be the vector derivative of the marginal density $f_G(\hat{\beta}_i)$.

4.4.1.3 EBE for β_n

Since the prior distribution $G(\beta)$ is not known in practice, the marginal density $f_G(\hat{\beta}_n)$ appearing in (4.20) cannot be used directly. By estimating $f_G(\hat{\beta}_n)$ and $f_G'(\hat{\beta}_n)$ consistently, however, we obtain an empirical Bayes estimator for β_n. If we let $f_n(\hat{\beta}_n)$ and $f_n'(\hat{\beta}_n)$ represent consistent estimators for $f_G(\hat{\beta}_n)$ and $f_G'(\hat{\beta}_n)$ respectively, then the empirical Bayes estimator for β_{n+1} is given by

$$\tilde{\beta}_{n+1} = \hat{\beta}_{n+1} + \sigma^2 Q_{n+1}^{-1} \frac{f_{n+1}'(\hat{\beta}_{n+1})}{f_{n+1}(\hat{\beta}_{n+1})} \tag{4.21}$$

which provides a consistent estimator for the Bayes estimator given by (4.20). It can be shown (see chapter 3) that the risk obtained in using (4.21) converges to the Bayes risk associated with (4.20) as n, the number of individual units, tends toward infinity. By estimating $f_G(\hat{\beta}_n)$ and $f'_G(\hat{\beta}_n)$ consistently we by-pass exact knowledge of $G(\beta\ )$ thus forming the empirical Bayes estimator for β_n .

Consider now the situation in which $(\hat{\beta}_1,\beta_1), \dots, (\hat{\beta}_n,\beta_n), \dots$ is a sequence of independent pairs of random p (= K + G) vectors, where $\hat{\beta}_i$ is the maximum likelihood estimator of β_i in ith individual unit. Since the β_i are known, the estimates $\hat{\beta}$ are distributed according to the marginal density $f_G(\hat{\beta})$ which is equal to $\int f(\hat{\beta} \mid \beta) dG(\beta)$. We can then use the sequence of estimates $\{ \hat{\beta}_i ; i = 1, \dots, n\}$ to estimate $f_G(\hat{\beta})$. Following (3.11) in chapter 3 we obtained the EBE of β . Note that $\hat{\beta}_n \mid \beta_n \sim N(\beta_n, \sigma^2 Q_n^{-1})$. Thus $\{\hat{\beta}_n\}$ are not identically distributed (since Q_j varies with j), unlike the situation considered by Martz and Krutchkoff (1969). However, their results still hold if the sequence $\{ \hat{\beta}_n \}$ are independent (see chapter 3).

4.4.2 EBE in TSRM with autocorrelated disturbances

This case is similar to the DLRM with autocorrelected disturbances given in section 4.3.2 , and thus will not be repeated here.

Chapter 5 Empirical Linear Bayes Estimator

5.1 Introduction

In contrast with Bayes estimation which assumes that parameters are random variables distributed with some known distribution function G, the linear Bayes estimation uses only the first two moments of distribution known. The linear Bayes estimator of the parameters turns out to be an elementary function of these two moments only. In the empirical linear Bayes (ELB) approach, we attempt to find estimators of those two moments using past independent experiments. In 1971, Griffin and Krutchkoff first used such kind of EBE, called the linear optimal estimates to solve the problem of estimating the binomial parameter, since there is no EBE for that situation.

As we shall see, when the prior G is normal, the Bayes estimator and the linear Bayes estimator coincide. We anticipate that LBE will be a 'good' linear approximation to Bayes Estimator when G is not normal. In engineering and econometrics, LBE has been used rather frequently since G, although it is unknown, is commonly specified as a wide sense distribution function.

5.2 Linear Bayes Estimation

Let $\underline{\theta}$ be q-dim random vector and Y be k-dim random vector . The Bayes estimator is $E(\underline{\theta} \mid \mathbf{Y})$ outlined in Chapter 3 is indeed the best with respect to the mean-square loss function. However, $E(\underline{\theta} \mid \mathbf{Y})$ is a nonliear function of Y (for the general case), and it is often difficult to obtain this exact relationship. Since very often $f_{\underline{\theta}\mathbf{Y}}(\underline{\theta}, \mathbf{Y})$ is not available, the $E(\underline{\theta} \mid \mathbf{Y})$ may not be achievable either.

Consequently we shall do the next best thing and introduce a constraint that $\hat{\underline{\theta}}(\mathbf{Y})$ has a linear form in Y . That is,

$$\hat{\underline{\theta}}(\mathbf{Y}) = A\mathbf{Y} + \mathbf{b} \tag{5.1}$$

where A is an $q \times k$ matrix and **b** is a q-dim vector. With the constraint (5.1), we have a risk function which looks like

$$E\,\|\underline{\theta} - A\mathbf{Y} - \mathbf{b}\|^2 = E\,(\underline{\theta} - A\mathbf{Y} - \mathbf{b})'(\underline{\theta} - A\mathbf{Y} - \mathbf{b}) \tag{5.2}$$

Now we can choose A and **b** (unknown constant quantities) such that (5.2) is minimized. Let us denote the optimal values of A and **b** as A_0 and $\mathbf{b}_0$ respectively , then $E(\underline{\theta} \mid \mathbf{Y})$ shall be given by

$$\bar{\underline{\theta}}(\mathbf{Y}) = A_0\mathbf{Y} + \mathbf{b}_0 \tag{5.3}$$

To minimize the risk function given by (5.2), we shall calculate A_0 and $\mathbf{b}_0$ in the usual manner by setting

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{b}} E\|\underline{\theta} - A\mathbf{Y} - \mathbf{b}\|^2 &= \frac{\partial}{\partial \mathbf{b}} E(\underline{\theta} - A\mathbf{Y} - \mathbf{b})'(\text{----}) = 0 \\
\frac{\partial}{\partial A} E\|\underline{\theta} - A\mathbf{Y} - \mathbf{b}\|^2 &= \frac{\partial}{\partial A} E(\underline{\theta} - A\mathbf{Y} - \mathbf{b})'(\text{----}) = 0.
\end{aligned}$$

Without any loss of generality, assume temporarily that $\underline{\theta}$ and Y have zero mean. From the first equation, we find:

$$\mathbf{b}_0 = 0$$

and, from the second,

$$A_0 = E(\underline{\theta}\mathbf{Y})[E(\mathbf{YY}')]^{-1} = \mathrm{cov}(\underline{\theta}, \mathbf{Y})(\mathrm{var}\mathbf{Y})^{-1} \tag{5.4}$$

because $\underline{\theta}$ and $\mathbf{Y}$ were assumed to have zero mean. Hence

$$\underline{\overline{\theta}}(\mathbf{Y}) = \mathrm{cov}(\underline{\theta}, \mathbf{Y})(\mathrm{var}\mathbf{Y})^{-1}\mathbf{Y}. \tag{5.5}$$

Now if $\underline{\theta}$ and $\mathbf{Y}$ do have nonzero mean, the random vector $\underline{\theta} - \mathbf{m}_{\underline{\theta}}$ and $\mathbf{Y} - \mathbf{m}_{\mathbf{Y}}$ has zero means. Applying (5.5) yields

$$\overline{\underline{\theta}(\mathbf{Y}) - \mathbf{m}_{\underline{\theta}}} = \mathrm{cov}(\underline{\theta}, \mathbf{Y})(\mathrm{var}\mathbf{Y})^{-1}(\mathbf{Y} - \mathbf{m}_{\mathbf{Y}}) \tag{5.6}$$

or, equivalently,

$$\underline{\overline{\theta}} = \mathbf{m}_{\underline{\theta}} + \mathrm{cov}(\underline{\theta}, \mathbf{Y})(\mathrm{var}\mathbf{Y})^{-1}(\mathbf{Y} - \mathbf{m}_{\mathbf{Y}}).$$

This is the multivariate result for LBE (see chapter 2).

5.3 Orthogonality Principle

In this section we shall assume, without loss of generality, that all parameters have zero mean, unless specified otherwise. For example, if the mean of $\underline{\theta}$ is non-zero, then we shall introduce a new random variable $\underline{\theta}^* = \underline{\theta} - \mathbf{m}_{\underline{\theta}}$ which will have zero mean (as in the previous section).

The concept of orthogonality is extremely important in the theory of linear Bayes estimation. (or linear mean-square estimation). We shall show that the orthogonality principle will serve as a necessary and sufficient condition that the linear Bayes estimate $\underline{\bar{\theta}}$ be optimal. The orthogonality principle states that if the measurement **Y** is orthogonal to the error $\bar{\mathbf{e}} = \underline{\theta} - \underline{\bar{\theta}}$, i.e.

$$E[(\underline{\theta} - \underline{\bar{\theta}})\mathbf{Y}'] = E(\bar{\mathbf{e}}\mathbf{Y}') = 0 \tag{5.7}$$

then the estimate $\underline{\bar{\theta}}$ is the best linear mean-square estimate.

Definition An estimate $\underline{\bar{\theta}}$ is *optimal* if it is the best linear mean- square estimate.

Let H be a vector space generated by the set of all random variables X such that

$$E(X) = 0 \quad \text{and} \quad E|X|^2 < \infty$$

and the random variables X_1 and X_2 are equivalent if $E|X_1 - X_2|^2 = 0$. Note that $|X|$ denotes vector norm of X.

We can assume without any loss of generality that the random variables are real. Thus the inner product

$$(\underline{\theta}, \mathbf{Y}) = E(\underline{\theta}\mathbf{Y}) \tag{5.8}$$

and

$$\|\underline{\theta}\|_{q.m.} = (\underline{\theta}, \underline{\theta})^{1/2} = (E|\underline{\theta}|^2)^{1/2} \tag{5.9}$$

Let M denote the subspace of H generated by $X_1, \ldots, X_n$ assumed to be linearly independent. Then H can be decomposed into the direct sum of M and M^+:

$$H = M + M^+.$$

That is, every vector $\underline{\theta}$ belonging to H is given by:

$$\underline{\theta} = \underline{\eta}_1 + \underline{\eta}_2$$

where $\underline{\eta}_1 \in M$ and $\underline{\eta}_2 \in M^+$. Note that we say M and M^+ are orthogonal to each other.

Recall that the projection of $\underline{\theta}$ denoted by P on M is given by:

$$P\underline{\theta} = \underline{\eta}_1$$

and the projection of $\underline{\theta}$ denoted by Q on M^+ is given by:

$$Q\underline{\theta} = \underline{\eta}_2$$

where

$$P + Q = I$$

and I is the identity operator. Note that $(\underline{\eta}_1, \underline{\eta}_2) = E(\underline{\eta}_1 \underline{\eta}_2') = 0$.

Theorem 1 Let $\underline{\theta}$ be a variable in H, and let Z be a vector in M. Then

$$\|\underline{\theta} - \mathbf{Z}\|^2_{q.m.} = E[(\underline{\theta} - \mathbf{Z})'(\underline{\theta} - \mathbf{Z})]$$

attains its minimum if and only if

$$\mathbf{Z} = P\underline{\theta}.$$

Proof For any $\underline{\theta} \in H$, we have

$$\underline{\theta} = \underline{\eta}_1 + \underline{\eta}_2$$

where $\underline{\eta}_1 \in M, \underline{\eta}_2 \in M^+$, and $\underline{\eta}_1 = P\underline{\theta}, \underline{\eta}_2 = (I - P)\underline{\theta}$. Now

$$\begin{aligned} \|\underline{\theta} - \mathbf{Z}\|^2_{q.m.} &= E(\underline{\theta} - \mathbf{Z})'(\underline{\theta} - \mathbf{Z}) \\ &= E\{[(\underline{\theta} - \underline{\eta}_1) + (\underline{\eta} - \mathbf{Z})]'[(\underline{\theta} - \underline{\eta}_1) + (\underline{\eta}_1 - \mathbf{Z})]\}. \end{aligned} \tag{5.11}$$

In the above equation $\underline{\theta} - \underline{\eta}_1$ is orthogonal to M, i.e. $\underline{\theta} - \underline{\eta}_1 \in M^+$, while $\underline{\eta}_1, \mathbf{Z}, \underline{\eta}_1 - Z$ are all members of M. Utilizing these facts in (5.11) yields:

$$\|\underline{\theta} - \mathbf{Z}\|^2_{q.m.} = \|\underline{\theta} - \underline{\eta}_1\|^2_{q.m.} + \|\eta_1 - \mathbf{Z}\|^2_{q.m.} \tag{5.12}$$

It is obvious that

$$\|\underline{\theta} - \mathbf{Z}\|^2_{q.m.} \geq \|\underline{\theta} - \underline{\eta}_1\|^2_{q.m.} \tag{5.13}$$

since $\|\underline{\eta}_1 - \mathbf{Z}\|^2_{q.m.} \geq 0$. Thus, the inequality in (5.13) becomes an equality if and only if

$$\mathbf{Z} = \underline{\eta}_1 = P\underline{\theta}.$$

This completes the proof.

Assume that we have received m measurements that are linearly independent, say, the random vectors $\mathbf{Y}_1, \dots, \mathbf{Y}_m$. Let M be the vector space spanned by the set of all linear combinations of $\mathbf{Y}_1, \dots, \mathbf{Y}_m$. According to the theorem, $\|\underline{\theta} - \mathbf{Z}\|^2_{q.m.}$ is minimized if and only if

$$\mathbf{Z} = P\underline{\theta} \in M.$$

Thus, Z can be written as the linear combination of $\mathbf{Y}_1, \dots, \mathbf{Y}_m$, which means that Z is optimal.

Remark 1 Let $\mathbf{Y}_1, \dots, \mathbf{Y}_m$ be the measurement vectors (observations), let M denote the vector space generated by these measurement vectors. Then vector $\underline{\theta}^*$ is an optimal estimate of $\underline{\theta}$ if and only if $\underline{\theta}^*$ is the projection of $\underline{\theta}$ onto M.

Remark 2 The vector $\underline{\theta}^*$ is an optimal estimate of $\underline{\theta}$ if and only if the error $\bar{\mathbf{e}} = \underline{\theta} - \underline{\theta}^*$ is orthogonal to the observation vectors $\mathbf{Y}_1, \dots, \mathbf{Y}_m$,i.e.,

$$E[(\underline{\theta} - \underline{\theta}^*)\mathbf{Y}'_i] = E[\bar{\mathbf{e}}\mathbf{Y}'_i] = 0 \quad \text{for} \quad i = 1, \dots, m.$$

Note that Remark 2 follows from Remark 1, because if $\underline{\bar{\theta}}$ is the projection of $\underline{\theta}$ onto M, then $\underline{\theta} - \underline{\bar{\theta}} \in M^+$.

5.4 Empirical Linear Bayes Estimation (ELBE)

Using (5.6), let

$$\underline{\delta}^* = \underline{\theta}^*(\mathrm{Y}) = \mathbf{m}_{\underline{\theta}} + \mathrm{cov}(\underline{\theta}, \mathrm{Y})(\mathrm{var}\mathrm{Y})^{-1}(\mathrm{Y} - \mathbf{m}_{\mathrm{Y}})$$

Specifically, assume that

$$\mathrm{Y} \sim WS(A\underline{\theta}, \sigma^2\Sigma) \quad i.e. \quad \mathrm{Y} = A\underline{\theta} + \underline{\varepsilon}, \underline{\varepsilon} \sim WS(0, \sigma^2\Sigma)$$

$$\underline{\theta} \sim WS(\underline{\mu}, T) \quad i.e. \quad \underline{\theta} = \underline{\mu} + \underline{v}, \underline{v} \sim WS(0, T)$$

where A, Σ are known, and $\underline{\varepsilon}$ and v are uncorrelated.

Now

$$\mathrm{cov}(\underline{\theta}, \mathrm{Y}) = \mathrm{cov}(\underline{\mu} + \underline{v}, A\underline{\mu} + A\underline{v} + \underline{\varepsilon}) = TA'$$

and

$$\mathrm{var}\mathrm{Y} = \mathrm{var}(A\underline{\mu} + A\underline{v} + \underline{\varepsilon}) = ATA' + \sigma^2\Sigma.$$

Thus

$$\underline{\delta}^* = \underline{\theta}^*(\mathrm{Y}) = \underline{\mu} + TA'(\sigma^2\Sigma + ATA')^{-1}(\mathrm{Y} - A\underline{\mu}).$$

By matrix manipulation, it holds that

$$\begin{aligned}\underline{\delta}^* &= (\sigma^2 T^{-1} + A'\Sigma^{-1}A)^{-1}(A'\Sigma^{-1}\mathbf{Y} + \sigma^2 T^{-1}\underline{\mu}) \\ &= (\sigma^2 T^{-1} + A'\Sigma^{-1}A)^{-1}(A'\Sigma^{-1}A\hat{\underline{\theta}} + \sigma^2 T^{-1}\underline{\mu})\end{aligned} \tag{5.14}$$

where $\hat{\underline{\theta}}$ is the generalized least square estimator of the model $\mathbf{Y} = A\underline{\theta} + \underline{\varepsilon}$.

Theorem 2 $\underline{\delta}^*$ is the linear Bayes estimator for $\underline{\theta}$, and

$$R(\pi, \underline{\delta}^*) = R(\pi, \hat{\underline{\theta}}) - \mathrm{tr}\{Q(T + Q)^{-1}Q\}$$
$$R(\pi, \hat{\underline{\theta}}) = \mathrm{tr}\{Q\}$$

where $Q = (\sigma^{-2}A'\Sigma^{-1}A)^{-1}$ and π denotes any distribution with the specified first two moments. Thus $R(\pi, \underline{\delta}) < R(\pi, \hat{\underline{\theta}})$.

Proof: (i) From (5.14), we know that

$$\underline{\delta}^* = (\sigma^2 T^{-1} + A'\Sigma^{-1}A)^{-1}(A'\Sigma^{-1}A\hat{\underline{\theta}} + \sigma^2 T^{-1}\underline{\mu})$$

is the linear Bayes estimator for $\underline{\theta}$, which follows from (5.7).

(ii) By definition, we have

$$\begin{aligned}R(\pi, \hat{\underline{\theta}}) &= \|\hat{\underline{\theta}} - \underline{\theta}\|^2_{q.m.} \\ &= \|(A'(\sigma^2\Sigma)^{-1}A)^{-1}A'(\sigma^2\Sigma)^{-1}Y - \underline{\theta}\|^2_{q.m.} \\ &= E\ \mathrm{tr}\ [(A'(\sigma^2\Sigma)^{-1}A)^{-1}A'(\sigma^2\Sigma)^{-1}\underline{\varepsilon\varepsilon}'(\sigma^2\Sigma)^{-1}A(A'(\sigma^2\Sigma)^{-1}A)^{-1}] \\ &= \mathrm{tr}\ (A'(\sigma^2\Sigma)^{-1}A)^{-1} = \mathrm{tr}\ Q\end{aligned}$$

(iii) Let $\Delta = (\sigma^2 T^{-1} + A'\Sigma^{-1}A)$, then

$$
\begin{aligned}
R(\pi, \underline{\delta}^*) &= \|\underline{\delta}^* - \underline{\theta}\|^2_{q.m.} \\
&= \|\Delta^{-1}(A'\Sigma^{-1}A\hat{\underline{\theta}} + \sigma^2 T^{-1}\underline{\mu}) - (\underline{\mu} + \underline{v})\|^2_{q.m.} \\
&= \|\Delta^{-1}[A'\Sigma^{-1}A\hat{\underline{\theta}} + \sigma^2 T^{-1}\underline{\mu} - (\Delta\underline{\mu}) - (\Delta\underline{v})]\|^2_{q.m.} \\
&= \|\Delta^{-1}[A'\Sigma^{-1}\varepsilon - \sigma^2 T^{-1}\underline{v}]\|^2_{q.m.} \\
&= \text{ tr } E\Delta^{-1}(A'\Sigma^{-1}\underline{\varepsilon\varepsilon}'\Sigma^{-1}A + \sigma^2\underline{vv}'T^{-1}\sigma^2)(\Delta^{-1})' \\
&= \text{ tr } \sigma^2\Delta^{-1}(A'\Sigma^{-1}A + \sigma^2T^{-1})(\Delta^{-1})' \\
&= \text{ tr } \sigma^2(\Delta^{-1}) \\
&= \text{ tr } T(Q + T)^{-1}Q \\
&= \text{ tr } Q - \text{ tr } Q(Q + T)^{-1}Q.
\end{aligned}
$$

Thus $R(\pi, \hat{\underline{\theta}}) > R(\pi, \underline{\delta}^*)$ since tr Q $(Q + T)^{-1}Q > 0$.

If the prior parameters $\underline{\mu}$, T, σ^2 and Σ are unknown, then the LBE $\underline{\delta}^*$ is not applicable. However, if there are earlier observations and the component problems occur independently, we can attempt to estimate those prior parameters from the data. Let us assume that we have, additionally, independent observations from N experiments:

$$
\begin{aligned}
\underline{\theta}_1 &: Y_1 = (Y_{11}, Y_{12}, \dots, Y_{1m_1}) \\
\underline{\theta}_2 &: Y_2 = (Y_{21}, Y_{22}, \dots, Y_{2m_2}) \\
&\cdot \\
&\cdot \\
&\cdot \\
\underline{\theta}_N &: Y_N = (Y_{N1}, Y_{N2}, \dots, Y_{Nm_N})
\end{aligned}
$$

with

$$Y_i = A_i\underline{\theta}_i + \underline{\varepsilon}_i \quad A_i\text{: known } m_i \times q \text{ matrix.}$$

$$\underline{\varepsilon}_i \sim WS(0, \sigma^2\Sigma_i) \quad \Sigma_i\text{: known } m_i \times m_i \text{ matrix.}$$

$$\underline{\theta} = \underline{\mu} + \underline{\nu}_i, \quad i = 0, 1, \dots, N.$$

$$\underline{\nu}_i \sim WS(0, T)$$

Furthermore, $\{\underline{\nu}_i, \underline{\varepsilon}_i, i = 0, 1, \dots\}$ is a set of independent random vectors .

Now we attempt to use these data to estimate the prior parameters $\underline{\mu}$, T and σ^2 in order to estimate $\underline{\delta}^* = \underline{\theta}^*(\mathrm{Y})$ for the current stage ($i = 0$) .

From the relation $EY_i = A_i\underline{\theta}_i$, we know that Y_i is an unbiased estimator of $A_i\underline{\theta}_i$. Thus we propose (i) $\overline{\underline{\mu}}_N$ as the estimator for $\underline{\mu}$

$$\overline{\underline{\mu}}_N = \frac{1}{N}\sum_{i=1}^{N}\hat{\underline{\theta}}_i, \quad \text{where } \hat{\underline{\theta}}_i = (A'_i\Sigma_i^{-1}A_i)^{-1}A'_i\Sigma_i^{-1}Y_i \quad i = 1, \dots, N.$$

(ii) $\overline{\sigma}_N^2$ as the estimator for σ^2

$$\overline{\sigma}_N^2 = \frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_i^2$$

where

$$\hat{\sigma}_i^2 = \frac{1}{m_j - q}(Y_i - A_i\hat{\underline{\theta}}_i)'\Sigma_i^{-1}(Y_i - A_i\hat{\underline{\theta}}_i)$$

(iii) $\hat{T}_N$ as the estimator for T

$$\hat{T}_N = \frac{1}{N-1}\sum_{i=1}^{N}(\hat{\underline{\theta}}_i - \overline{\underline{\mu}}_N)(\hat{\underline{\theta}}_i - \overline{\underline{\mu}}_N)' - \frac{1}{N}\overline{\sigma}_N^{2}\sum_{i=1}^{N}(A'_i\Sigma_i^{-1}A_i)^{-1}$$

Theorem 3 (1) As proposed above, $\overline{\underline{\mu}}_N, \hat{T}_N$ and $\overline{\sigma}_N^2$ are unbiased.

(2) If

(i)

$$\lim_{N\to\infty} \frac{1}{N^2}\sum_{i=0}^{N}(A'_i\Sigma_i^{-1}A_i)^{-1} = 0$$

(ii)

$$\lim_{N\to\infty} \frac{1}{N^2}\sum_{i=1}^{N} E\{(\hat{\theta}_i^r - \mu^r)^2(\hat{\theta}_i^s - \mu^s)^2\} = 0 \quad \text{for} \quad r, s = 1, \dots, k.$$

(iii)

$$\lim_{N\to\infty} \frac{1}{N^2}\sum_{i=1}^{N}\frac{1}{(m_i - q)^2}E\{(\underline{\varepsilon}'_i V_i \underline{\varepsilon}_i)^2\} = 0$$

where $\hat{\theta}_i^r$ and μ^r denote the rth component of $\hat{\underline{\theta}}_i$ and $\underline{\mu}$ respectively, and $V_i = \Sigma_i^{-1} - \Sigma_i^{-1}A_i(A'_i\Sigma_i^{-1}A_i)^{-1}A'_i\Sigma_i^{-1}$, then $\overline{\underline{\mu}}_N, \hat{T}_N$ and $\overline{\sigma}_N^2$ are consistent for $\underline{\mu}$, T and σ^2 respectively.

Proof: (1)

$$\begin{aligned} E(\overline{\underline{\mu}}_N) &= \frac{1}{N}\sum_{i=1}^{N} E(\hat{\underline{\theta}}_i) = \frac{1}{N}\sum_{i=1}^{N}(A'_i\Sigma_i^{-1}A_i)^{-1}A'_i\Sigma_i^{-1}EY_i \\ &= \frac{1}{N}\sum_{i=1}^{N}(A'_i\Sigma_i^{-1}A_i)^{-1}A'_i\Sigma_i^{-1}A_i\mu = \underline{\mu} \end{aligned}$$

$$\begin{aligned} E(\overline{\sigma}_N^2) &= \frac{1}{N}\sum_{i=1}^{N} E\hat{\sigma}_i^2 \\ &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{m_i - q}E(Y_i - A_i\hat{\underline{\theta}}_i)'\Sigma_i^{-1}(Y_i - A_i\hat{\underline{\theta}}_i) \\ &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{m_i - q}E[\underline{\varepsilon}'_i(I - A_i(A'_i\Sigma_i^{-1}A_i)^{-1}A'_i\Sigma_i^{-1})'\Sigma_i^{-1}(\text{------})\underline{\varepsilon}_i] \\ &= \frac{1}{N}\sum_{i=1}^{n}\frac{1}{m_i - q}\ \mathrm{tr}\ (\Sigma_i^{-1} - \Sigma_i^{-1}A_i(A'_i\Sigma_i^{-1}A_i)^{-1}A'_i\Sigma_i^{-1})E\underline{\varepsilon}\underline{\varepsilon}' \\ &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{m_i - q}\sigma^2\ \mathrm{tr}\ (I - \Sigma_i^{-1}A_i(A'_i\Sigma_i^{-1}A_i)^{-1}A'_i) \\ &= \frac{1}{N}\sigma^2\sum_{i=1}^{N}\frac{1}{m_i - q}(m_i - q) = \sigma^2. \end{aligned}$$

$$
\begin{aligned}
E\hat{T}_N &= \frac{1}{N-1}\sum_{i=1}^{N} E(\hat{\underline{\theta}} - \overline{\underline{\mu}}_N)(\text{-----})' - \frac{1}{N}E(\overline{\sigma}_N^2)\sum_{i=1}^{N}(A'_i\Sigma_i^{-1}A_i)^{-1} \\
&= \frac{1}{N-1}\sum_{i=1}^{N} E(\hat{\underline{\theta}}_i - \underline{\mu} + \underline{\mu} - \overline{\underline{\mu}}_i)(\text{-----})' - \frac{1}{N}\sigma^2\sum_{i=1}^{N}(A'_i\Sigma_i^{-1}A_i)^{-1} \\
&= \frac{1}{N-1}\sum_{i=1}^{N} E(\hat{\underline{\theta}}_i - \underline{\mu})(\text{----})' + \frac{1}{N-1}\sum_{i=1}^{N} E(\underline{\mu} - \overline{\underline{\mu}}_N)(\text{----})' - \frac{1}{N}\sigma^2\sum_{i=1}^{N}(A'_i\Sigma_i^{-1}A_i)^{-1} \\
&= \frac{1}{N-1}\sum_{i=1}^{N}((A'_i\Sigma_i^{-1}A_i)^{-1}\sigma^2 + T) - \frac{1}{N-1}\sum_{i=1}^{N}\frac{1}{N}[(A'_i\Sigma_i^{-1}A_i)^{-1}\sigma^2 + T] \\
&\qquad - \frac{1}{N}\sigma^2\sum_{i=1}^{N}(A'_i\Sigma_i^{-1}A_i)^{-1} \\
&= \frac{1}{N}[\sum_{i=1}^{N}\sigma^2(A'_i\Sigma_i^{-1}A_i)^{-1} + NT] - \frac{1}{N}\sigma^2\sum_{i=1}^{N}(A'_i\Sigma_i^{-1}A_i)^{-1} \\
&= T.
\end{aligned}
$$

(2) Since the covariance matrix of $\overline{\underline{\mu}}_N$ is

$$
\operatorname{cov}\ \overline{\underline{\mu}}_N = \frac{1}{N^2}\sum_{i=1}^{N}\operatorname{cov}\hat{\underline{\theta}}_i = \frac{1}{N^2}\sum_{i=1}^{N}[(A'_i\Sigma_i^{-1}A_i)^{-1}\sigma^2 + T],
$$

the consistency of $\overline{\underline{\mu}}_N$ follows by assumption (i) and the WLLN.

Similarly, by assumption (iii) and the WLLN, the consistency of $\overline{\sigma}_N^2$ follows.

Write $\hat{T}_N$ in the following form with four terms:

$$
\begin{aligned}
\hat{T}_N = \frac{1}{N-1}\{S_N - (\overline{\underline{\mu}}_N - \underline{\mu})(\overline{\underline{\mu}}_N - \underline{\mu})'\} &+ \frac{\sigma^2}{(N-1)N^2}\sum_{j=1}^{N}(A'_j\Sigma_j^{-1}A_j)^{-1} \\
&+ \frac{1}{N(N-1)}(\sigma^2 - \overline{\sigma}_N^2)\sum_{j=1}^{N}(A'_j\Sigma_j^{-1}A_j)^{-1}
\end{aligned}
\tag{5.15}
$$

where

$$
S_N = \frac{1}{N}\sum_{j=1}^{N}\{(\hat{\underline{\theta}}_i - \underline{\mu})(\hat{\underline{\theta}}_i - \underline{\mu})' - \sigma^2(A'_i\Sigma_i^{-1}A_i)^{-1}\}.
$$

Due to assumption (i) the third term on the r.h.s. of (5.15) tends to zero. The second term on the r.h.s. of (5.15) tends in probability to zero because of the consistency of $\underline{\bar{\mu}}_N$. Using $S_N^{r,s}$ for the element of the matrix S_N , it follows that

$$\mathrm{var} S_N^{r,s} \leq \frac{1}{N^2} \sum_{i=1}^{N} E(\hat{\theta}_i^r - \mu^r)^2 (\hat{\theta}_i^s - \mu^s)^2. \tag{5.16}$$

Since $ES_N = T$, the first term on the r.h.s. of (5.15) S_N tends to T in probability by (5.16) and assumption (ii). Finally, by the consistency of $\bar{\sigma}_N^2$ the last term on the r.h.s. of (5.15) tends to zero in probability.

Improvement of The Estimator $\hat{T}_N$

Since T is the covariance matrix of $\underline{\theta}$ then $T \in M_q$ (the set of all symmetric positive semidefinite $q \times q$ matrices). But by definition $\hat{T}_N$ is not necessary in M_q . Here we attempt to improve the estimator $\hat{T}_N$ by projection onto the convex set M_q .

First we introduce the projection $\bar{T}_N = P_{M_q} \hat{T}_N$ by

$$\underset{T \in M_q}{\mathrm{lub}} \; \|\bar{T}_N - T\| = \|\hat{T}_N - \bar{T}_N\|$$

where $\|A\|^2 = \mathrm{tr}AA'$. Because of the convexity of M_q , the projection is unique and we have

$$\|T - \bar{T}_N\| \leq \|T - \hat{T}_N\| \quad \text{for all } T \in M_q$$

such that the consistency of $\hat{T}_N$ implies the consistency of $\bar{T}_N$.

Next, the problem is how to find $\bar{T}_N$ from $\hat{T}_N$. Because of the symmetry of $\hat{T}_N$ we can represent $\hat{T}_N$ as

$$\hat{T}_N = \sum_{i=1}^{k} \lambda_i^N U_i^N (U_i^N)'$$

with the eigenvalues $\lambda_i^N; i = 1, \dots, q$, of $\hat{T}_N$ and with corresponing eigenvectors $U_i^N; i = 1, \dots, q$. It follows that $\bar{T}_N = P_{M_q}(\hat{T}_N)$ has the representation

$$\bar{T}_N = \sum_{i, \lambda_i^N > 0} \lambda_i^N U_i^N (U_i^n)'.$$

.

Now we consider the linear Bayes estimator $\underline{\delta}^*$ in (5.14) and replace the unknown prior parameters by their corresponding consistent estimators. Thus we have

$$\underline{\delta}_N = \underline{\theta}_N(\mathrm{Y}) = \underline{\bar{\mu}}_N + \bar{T}_N A(\bar{\sigma}_N^2 \Sigma + A\bar{T}_N A')^{-1}(\mathrm{Y} - A\underline{\bar{\mu}}_N) \tag{5.17}$$

as an empirical linear Bayes estimator for $\underline{\theta}$. It is easy to show that $\underline{\delta}_N$ is consistent to LBE $\underline{\delta}^*$ (by Slusky theorem). Now we investigate the asymptotic properties of the risk function

$$R_N(\pi, \underline{\delta}_N) = E\, l\, (\underline{\theta}, \underline{\delta}_N(Y; Y_1, \dots, Y_N))$$

of the corresponding sequence of decision functions

$$\underline{\delta}_N \in D^{\overline{m}} = \{\underline{\delta}:(R^m \times R^{\overline{m}}, B^m \times B^{\overline{m}}) \rightarrow (R^q, B^q)\},$$

where $\overline{m} = \sum_{i=1}^{N} m_i.$

Definition 2 A sequence $\{\underline{\delta}_N\}$ of decision functions $\underline{\delta}_N \in D^{\overline{m}}$ is called *asymptotically linearly optimal* for any specified prior distribution moments π if

$$\lim_{N \to \infty} R_N(\pi, \underline{\delta}_N) = \operatorname*{lub}_{\underline{\delta} \in D_l} R(\pi, \underline{\delta}) \quad \text{for all} \quad \pi.$$

where $D_l = \{\underline{\delta}(\mathrm{y}) = \mathbf{a} + L\mathrm{y}, \mathbf{a} \in R^q, L$ is a $q \times n$ matrix, $\mathrm{y} \in R^n\}$

Theorem 4 Under the assumptions of Theorem 3, $\{\underline{\delta}_N\}$ defined by (5.17) is asymptotically linearly optimal.

Proof: From Theorem 1, to prove this theorem it is sufficient to show that

$$\lim_{N\to\infty}\{R_N(\pi, \underline{\delta}_N) - R(\pi, \underline{\delta})\} = 0 \quad \text{for all} \quad \pi.$$

Now

$$\begin{aligned} &R_N(\pi, \underline{\delta}_N) - R(\pi, \underline{\delta}) \\ &= E\{\|\hat{\underline{\theta}} - \underline{\theta} - \underline{\eta}_N\|^2 - \|\hat{\underline{\theta}} - \underline{\theta} - \underline{\eta}\|^2\} \\ &= E\{\|\underline{\eta}_N\|^2 - \|\underline{\eta}\|^2\} - 2E(\hat{\underline{\theta}} - \underline{\theta}, \underline{\eta}_N - \underline{\eta}) \end{aligned}$$

where

$$\begin{aligned} \underline{\eta}_N &= \bar{\sigma}_N^2(A'\Sigma^{-1}A)^{-1}[(\bar{\sigma}_N^2 A'\Sigma^{-1}A)^{-1} + \tilde{T}_N]^{-1}(\hat{\underline{\theta}} - \bar{\underline{\mu}}_N) \\ \underline{\eta} &= \sigma^2(A'\Sigma^{-1}A)^{-1}[(\sigma^2 A'\Sigma^{-1}A)^{-1} + T]^{-1}(\hat{\underline{\theta}} - \underline{\mu}) \end{aligned}$$

and (A, B) = tr A B′.

Since $\tilde{T}_N \to T, \bar{\underline{\mu}} \to \underline{\mu}$, and $\bar{\sigma}_N^2 \to \sigma^2$ (all in probability) implies $\underline{\eta}_N \to \eta$ (in probability), we have

$$U_N = \|\underline{\eta}_N\|^2 - \|\underline{\eta}\|^2 \to 0 \quad \text{(in probability)}$$

and

$$F_N = (\hat{\underline{\theta}} - \underline{\theta}, \underline{\eta}_N - \underline{\eta}) \to 0 \quad \text{(in probability)}$$

We now want to show that $E\,U_N \to 0$ and $E\,F_N \to 0$

(i) Let

$$\begin{aligned} \tilde{Q}_N &= \bar{\sigma}_N^2(A'\Sigma^{-1}A)^{-1} \\ \tilde{W}_N &= [(\bar{\sigma}_N^2 A'\Sigma^{-1}A)^{-1} + \tilde{T}_N]^{-1} \\ Q &= \sigma^2(A'\Sigma^{-1}A)^{-1} \\ W &= [(\sigma^2 A'\Sigma^{-1}A)^{-1} + T]^{-1}. \end{aligned}$$

Then

$$
\begin{aligned}
E\,U_N &= E \text{ tr } \{\tilde{Q}_N \tilde{W}_N(\hat{\underline{\theta}} - \bar{\underline{\mu}}_N)(\hat{\underline{\theta}} - \bar{\underline{\mu}}_N)' \tilde{W}_N \tilde{Q}_N - QW(\hat{\underline{\theta}} - \underline{\mu})(\hat{\underline{\theta}} - \underline{\mu})'WQ\} \\
&= A_N - \text{ tr } QWQ \quad \text{(set)}
\end{aligned}
$$

where

$$
\begin{aligned}
A_N &= \text{ tr } E\{(\hat{\underline{\theta}} - \bar{\underline{\mu}}_N)(\hat{\underline{\theta}} - \bar{\underline{\mu}}_N)' \tilde{W}_N \tilde{Q}_N^2 \tilde{W}_N\} \\
&= \text{ tr } \{W^{-1} E\, \tilde{W}_N \tilde{Q}_N^2 \tilde{W}_N\} + \text{ tr } E\{(\underline{\mu} - \bar{\underline{\mu}}_N) \tilde{W}_N \tilde{Q}_N^2 \tilde{W}_N (\underline{\mu} - \bar{\underline{\mu}}_N)\} \\
&= B_N + C_N.
\end{aligned}
$$

Since

$$
\begin{aligned}
0 &\le \text{ tr } (W^{-1/2} \tilde{W}_N \tilde{Q}_N^2 \tilde{W}_N W^{-1/2}) \\
&\le \text{ tr } \{W^{-1/2} \tilde{W}_N (\tilde{T}_N + \tilde{Q}_N)^2 \tilde{W}_N W^{-1/2}\} \qquad (5.18) \\
&= \text{ tr } W^{-1} \quad (\textit{bounded})
\end{aligned}
$$

it follows that $\lim_{N\to\infty} B_N = \text{ tr } QWQ$, by using the dominated Lebesgue convergence theorem, since $\tilde{W}_N \to W$ (in probability) and $\tilde{Q}_N \to Q$ (in probability).

Since $\tilde{Q}_N > 0$ (due to $\bar{\sigma}_N^2 > 0$ and $A'\Sigma^{-1}A > 0$ by assumption), we have

$$
\begin{aligned}
0 \le C_N &\le \text{ tr } E\{(\mu - \bar{\mu}_N)(\mu - \bar{\mu}_N)'\} \quad \text{(using (5.18))} \\
&= \text{ tr } \left(\frac{1}{N^2}\sum_{i=1}^{N} Q_i + \frac{1}{N} T\right)
\end{aligned}
$$

Thus by assumption (i) in Theorem 3

$$
\lim_{N\to\infty} C_N = 0.
$$

From these we have

$$\lim_{N\to\infty} E\,U_N = \lim(A_N - \quad \text{tr } QWQ)$$
$$= \lim_{N\to\infty} \quad (B_N + C_N - \quad \text{tr } QWQ)$$
$$= 0.$$

(ii)

$$E\,F_N = E \text{ tr } \{(\hat{\underline{\theta}} - \underline{\theta})[(\hat{\underline{\theta}} - \bar{\underline{\mu}}_N)'\bar{W}_N\bar{Q}_N - (\hat{\underline{\theta}} - \underline{\mu})'WQ]\}$$
$$= \text{ tr } [E\{(\hat{\underline{\theta}} - \underline{\theta})\hat{\underline{\theta}}'\}\{E(\bar{W}_N\bar{Q}_N) - WQ\}]$$
$$= \text{ tr } Q\{E(\bar{W}_N\bar{Q}_N) - WQ\}$$
$$< \text{ tr} Q\{E(\bar{Q}_N^{-1}\bar{Q}_N) - WQ\} \quad (\text{since } \bar{W}_N < \bar{Q}_N^{-1})$$
$$< \text{ tr } \{Q - QWQ\} \quad (Q - QWQ: \text{ positive definite})$$

Thus by the dominated Lebesgue convergence theorem and since $\bar{W}_N \to W$(in probability) and $\bar{Q}_N \to Q$(in probability), we have

$$\lim_{N\to\infty} E\,F_N = 0.$$

Since $E\,U_N \to 0$ and $E\,F_N \to 0$ the proof of Theorem 4 is completed.

Chapter 6 Applications of Empirical Linear Bayes Estimator to Time Series Regression Models

6.1 The Empirical Linear Bayes Approach to MLRM

Consider the following multiple linear regression model (MLRM):

$$\mathbf{y} = X\underline{\beta} + \underline{\varepsilon}$$

where $\underline{\beta}$ and $\underline{\varepsilon}$ are independent vectors with

$$\underline{\varepsilon} \sim WS(0, \Sigma)$$
$$\underline{\beta} \sim WS(\underline{\mu}, T), \ \ \underline{\beta} = \underline{\mu} + \underline{v}$$
$$\underline{v} \sim WS(0, T)$$

Further X is a $n \times k$ matrix with rank k.

By Aitken's theorem the generalized least squares estimator (GLSE) $\hat{\underline{\beta}}$ of $\underline{\beta}$ is given by

$$\hat{\underline{\beta}} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\mathbf{y}$$

and from last chapter the linear Bayes estimator of $\underline{\beta}$ is given by

$$\tilde{\underline{\beta}} = \hat{\underline{\beta}} - (X'\Sigma^{-1}X)^{-1}(T + (X'\Sigma^{-1}X)^{-1})^{-1}(\hat{\underline{\beta}} - \underline{\mu}).$$

Let $A \equiv X$, and $\underline{\theta} \equiv \underline{\beta}$. Then the theory of the previous chapter is the MLRM case and thus applicable to the MLRM.

6.2 The Empirical Linear Bayes Approach to Autoregressive Linear Models with Lagged Dependent Variables

In this section the empirical linear Bayes parameter estimators of autoregressive linear models are constructed and discussed. Some properties such as consistency and asymptotic optimality are investigated. Note that the autoregressive linear models are also known as lagged dependent variable models in econometrics, and dynamic linear stochastic models in system control theory.

6.2.1 The Model and its Estimation

The general form of the lagged dependent variable model is given by

$$y_t = \sum_{l=1}^{G} \alpha_l \, y_{t-l} + \mathbf{z}'_t\underline{\gamma} + \varepsilon_t, \quad t = 1, \dots , T \tag{6.1}$$

where the y_t are (scalar) observations on the dependent variable y, the α_l are scalar parameters, $\mathbf{z}_t$ is a $(K - G) \times 1$ vector of the tth observations on the (K-G) non-stochastic explanatory variables, $\underline{\gamma}$ is the $(K - G) \times 1$ vector of corresponding parameters, and ε_t are the random disturbances. (Note that sometimes (6.1) are called a linear stochastic difference equations).

The following assumptions are made:

(i) The ε_t are i i d as $WS(0, \sigma^2)$ with finite fourth moments. Note that the notation $\varepsilon_t \sim WS(0, \sigma^2)$ indicates the wide sense model in which ε_t has any distribution with zero mean and variance σ^2 .

(ii) All G roots of the polynomial equation in λ

$$\lambda^G - \alpha_1\lambda^{G-1} - \cdots - \alpha_{G-1}\lambda - \alpha_G = 0$$

are less than one in absolute value. (This is the requirement of dynamic stability)

(iii) The following assumption guarantees the convergence of moment matrices of the non-stochastic variables. Let $\eta = 0, 1, \ldots$. The matrix

$$Q_\eta = \lim_{T\to\infty} \frac{1}{T-\eta} \sum_{t=1}^{T-\eta} \mathbf{z}_t \mathbf{z}'_{t+\eta}$$

exists for any integer η and is nonsingular for $\eta = 0$.

(6.1) can be written in familiar matrix form

$$\mathbf{y} = X\underline{\beta} + \underline{\varepsilon} \tag{6.2}$$

where

$$\mathbf{y}' = (y_1, \ldots, y_T), \quad \underline{\beta}' = (\alpha_1, \ldots, \alpha_G, \gamma_1, \ldots, \gamma_{(k-G)})$$
$$\underline{\varepsilon}' = (\varepsilon_1, \ldots, \varepsilon_T)$$

and

$$X = \begin{bmatrix} y_0 \; y_{-1} & \cdots & y_{-G+1} \; z_{11} \; z_{12} & \cdots & z_{1(k-G)} \\ y_1 \; y_0 & \cdots & y_{-G+2} \; z_{21} \; z_{22} & \cdots & z_{2(k-G)} \\ \cdots & & & & \\ y_{T-1} \; y_{T-2} & \cdots & y_{-G+T} \; z_{T1} \; z_{T2} & \cdots & z_{T(k-G)} \end{bmatrix}$$

Note that the total number of observations on y_t is T+G, not T alone. In what follows the pre-sample values $y_0, y_{-1}, \ldots, y_{-G+1}$ will be treated as fixed with all probability statements conditional on these values.

Under the aforementioned assumptions we can obtain the following properties:

(1) The matrix $Q = \text{plim}(X'X/T)$ is finite and nonsingular and $X'\underline{\varepsilon}/\sqrt{T}$ is asymptotically distributed as $N(0, \sigma^2 Q)$. (here $Q = \text{plim}\,(X'X/T)$ is defined as the random variables $X'X/T$ converges in probability to Q.)

(2) The ordinary least squares estimators $\hat{\underline{\beta}} = (X'X)^{-1}X'\text{y}$ and $\hat{\sigma}^2 = \hat{\underline{\varepsilon}}'\hat{\underline{\varepsilon}}/T - k$ are consistent estimators of $\underline{\beta}$ and σ^2 respectively, where $\hat{\underline{\varepsilon}} = \text{y} - X\hat{\underline{\beta}}$ is the residual vector. In addition, $\sqrt{T}(\hat{\underline{\beta}} - \underline{\beta})$ has a limiting distribution which is $N(0, \sigma^2 Q^{-1})$. For proof see Schonfelt (1971)).

Note that $\hat{\underline{\beta}}$ is biased since $E[(X'X)^{-1}X'\underline{\varepsilon}] \neq 0$. Though $E(\varepsilon_t \mid X_t) = E(\varepsilon_t) = 0$, $E(\underline{\varepsilon} \mid X) \neq 0$. This is because X_t contains y_{t-1} and is certainly not independent of ε_{t-1}.

6.2.2 Bayes Estimation with a Normal Prior Distribution

Under the model (1) suppose that

$$\begin{bmatrix} \underline{\beta} \\ \underline{\varepsilon} \end{bmatrix} \sim N\left[\begin{bmatrix} \underline{\mu} \\ 0 \end{bmatrix}, \begin{bmatrix} V & 0 \\ 0 & \Sigma \end{bmatrix}\right] \tag{6.3}$$

with $V > 0$, and $\Sigma > 0$ (positive definite). Then the Bayes estimator for $\underline{\beta}$ under a squared error loss is given by

$$\begin{aligned} \underline{\beta}^* &= E(\underline{\beta} \mid \hat{\underline{\beta}}) \\ &= (V^{-1} + K)^{-1}(V^{-1}\underline{\mu} + K\hat{\underline{\beta}}) \end{aligned} \tag{6.4}$$

where $K = (X'\Sigma^{-1}X)$.

We prove (6.4) by employing the theorem from Hartigan, which is concise and simple.

Proof: Since

$$f(\mathrm{y}, \underset{\sim}{\beta} \mid X) = f(\mathrm{y} \mid \underset{\sim}{\beta}, X) f(\underset{\sim}{\beta} \mid X)$$
$$= f(\underset{\sim}{\beta} \mid \mathrm{y}, X) f(\mathrm{y} \mid X)$$

(this is just Bayes′ theorem), taking logarithm on both sides yields

$$\log f(\mathrm{y} \mid \underset{\sim}{\beta}, X) + \log f(\underset{\sim}{\beta} \mid X) = \log f(\underset{\sim}{\beta} \mid \mathrm{y}, X) + \log f(\mathrm{y} \mid X).$$

Now $f(\mathrm{y} \mid \underset{\sim}{\beta}, X), f(\underset{\sim}{\beta} \mid X), f(\underset{\sim}{\beta} \mid \mathrm{y}, X)$ and $f(\mathrm{y} \mid X)$ are all normal density functions (under the assumptions). Hence we have

$$\begin{aligned} & K_1 - \frac{1}{2}(\mathrm{y} - X\underset{\sim}{\beta})' V^{-1}(\mathrm{y} \mid \underset{\sim}{\beta}, X)(\mathrm{y} - X\underset{\sim}{\beta}) \\ & \quad + K_2 - \frac{1}{2}(\underset{\sim}{\beta} - \underset{\sim}{\mu})' V^{-1}(\underset{\sim}{\beta} \mid X)(\beta - \underset{\sim}{\mu}) \\ = & K_3 - \frac{1}{2}(\underset{\sim}{\beta} - \underset{\sim}{\beta}^*)' V^{-1}(\beta \mid \mathrm{y}, X)(\underset{\sim}{\beta} - \underset{\sim}{\beta}^*) \\ & \quad + K_4 - \frac{1}{2}(\mathrm{y} - X\underset{\sim}{\mu})' V^{-1}(\underset{\sim}{y} \mid X)(\mathrm{y} - X\underset{\sim}{\mu}) \end{aligned}$$

where K_1, K_2, K_3, K_4 are suitable constants with respect to $\underset{\sim}{\beta}$. Now comparing the coefficients for $\underset{\sim}{\beta}'$ and $\underset{\sim}{\beta}'\underset{\sim}{\beta}$ respectively, we obtain

$$V^{-1}(\underset{\sim}{\beta} \mid \mathrm{y}, X)\underset{\sim}{\beta}^* = X' V^{-1}(\mathrm{y} \mid \underset{\sim}{\beta}, X)\mathrm{y} + V^{-1}(\underset{\sim}{\beta} \mid X)\underset{\sim}{\mu},$$

and

$$X' V^{-1}(\mathrm{y} \mid \underset{\sim}{\beta}, X) X + V^{-1}(\underset{\sim}{\beta} \mid X) = V^{-1}(\underset{\sim}{\beta} \mid \mathrm{y}, X).$$

That is,

$$\begin{aligned} \underset{\sim}{\beta}^* & = V(\underset{\sim}{\beta} \mid \mathrm{y}, X)\, [X' V^{-1}(\mathrm{y} \mid \underset{\sim}{\beta}, X)\, \mathrm{y} + V^{-1}(\underset{\sim}{\beta} \mid X)\, \underset{\sim}{\mu}] \\ & = (X' V^{-1}(\mathrm{y} \mid \underset{\sim}{\beta}, X)\, X + V^{-1}(\beta \mid X))^{-1} [V^{-1}(\underset{\sim}{\beta} \mid X)\, \underset{\sim}{\mu} + X' V^{-1}(\mathrm{y} \mid \underset{\sim}{\beta}, X)\, X\hat{\underset{\sim}{\beta}}] \\ & = (X'\Sigma^{-1} X + V^{-1})^{-1} [V^{-1}\underset{\sim}{\mu} + X'\Sigma^{-1} X\hat{\underset{\sim}{\beta}}] \end{aligned}$$

where in our terminology (by assumption)

$$V^{-1}(\mathrm{y} \mid \underline{\beta}, X) = V^{-1}(\underline{\varepsilon}) = \Sigma^{-1}$$
$$V^{-1}(\underline{\beta} \mid X) = V^{-1}(\underline{\beta}) = V^{-1}$$

6.2.3 Linear Bayes Estimator with a Wide Sense Prior Distribution

In this subsection we assume that under the model (6.1)

$$\begin{bmatrix} \underline{\beta} \\ \underline{\varepsilon} \end{bmatrix} \sim WS\left[\begin{bmatrix} \underline{\mu} \\ 0 \end{bmatrix}, \begin{bmatrix} V & 0 \\ 0 & \sigma^2\Sigma \end{bmatrix}\right] \tag{6.5}$$

where $V > 0$, and $Sig > 0$. Then the linear Bayes estimator for $\underline{\beta}$ is given by

$$\bar{\underline{\beta}} = (V^{-1} + K)^{-1}(V^{-1}\underline{\mu} + K\hat{\underline{\beta}}) \tag{6.6}$$

where $K = \sigma^{-2}Q$, with $Q = (X'\Sigma^{-1}X)$, and $\hat{\underline{\beta}}$ is GLSE for fixed $\underline{\beta}$. (see section 6.1). Note that we call $\bar{\underline{\beta}}$ *the linear Bayes estimator* since only the first two moments (including product moments) of observations and parameters are assumed known. Let π be any distribution function with mean $\underline{\mu}$ and variance V.

Proof of (6.6):

Let $R(\pi, \underline{\delta})$ be the risk incurred by the linear decision rule $\underline{\delta} = \mathbf{a} + L\mathrm{y}$ where $\mathbf{a}$ (a constant vector) and L (a matrix) are to be determined. Then

$$\begin{aligned} R(\pi, \underline{\delta}) &= E(\underline{\beta} - \mathbf{a} - L\mathrm{y})'(\underline{\beta} - \mathbf{a} - L\mathrm{y}) \\ &= E(\underline{\mu} + \underline{v} - \mathbf{a} - L(X\underline{\mu} + X\underline{v} + \underline{\varepsilon}))'(\text{----}) \\ &= \mathrm{tr}E[(\underline{\mu} - LX\underline{\mu} - \mathbf{a}) + (I - LX)\underline{v} - L\underline{\varepsilon}][(\text{----})'] \\ &= \ \mathrm{tr}\ \ E_X(E_{\underline{v}, \underline{\varepsilon} \mid X}[(\text{----})(\text{----})']). \end{aligned}$$

To minimize $R(\pi, \underline{\delta})$ we minimize

$$
\begin{aligned}
&E_{\underline{v},\underline{\varepsilon}|X}(\text{----})(\text{----})' \\
&= E_{\underline{v},\underline{\varepsilon}|X}[(\underline{\mu} - LX\underline{\mu} - \mathbf{a}) + (I - LX)\underline{v} - L\underline{\varepsilon}][\text{----}]' \\
&= (\underline{\mu} - LX\underline{\mu} - \mathbf{a})(\text{----})' + (\underline{\mu} - LX\underline{\mu} - \mathbf{a})E((I - LX)\underline{v} - L\underline{\varepsilon})' \\
&+ E((I - LX)\underline{v} - L\underline{\varepsilon})(\underline{\mu} - LX\underline{\mu} - \mathbf{a})' \\
&+ E(I - LX)\underline{v}\underline{v}'(I - LX)' + E(L\underline{\varepsilon}\underline{\varepsilon}'L') \\
&= (\underline{\mu} - LX\underline{\mu} - \mathbf{a})(\text{----})' + (I - LX)V(I - LX)' + L\sigma^2\Sigma L'.
\end{aligned}
$$

By Chipman's method we have

$$\hat{\mathbf{a}} = (I - \hat{L}X)\underline{\mu}$$

and then minimizing

$$(I - LX)V(I - LX)' + L\Sigma L'$$

yields (by taking differentials with respect to L)

$$2LXVX' - 2VX' + 2L\sigma^2\Sigma = 0.$$

Thus

$$\hat{L} = VX'(XVX' + \sigma^2\Sigma)^{-1}.$$

We obtain the LBE:

$$
\begin{aligned}
&\hat{\mathbf{a}} + \hat{L}\mathbf{y} \\
&= \underline{\mu} + VX'(XVX' + \sigma^2\Sigma)^{-1}(\mathbf{y} - X\underline{\mu}) \\
&= \underline{\mu} + VX'[\sigma^{-2}\Sigma^{-1} - \sigma^{-2}\Sigma^{-1}X(V^{-1} + X'\sigma^{-2}\Sigma^{-1}X)^{-1}X'\sigma^{-2}\Sigma^{-1}](\mathbf{y} - X\underline{\mu}) \\
&= \underline{\mu} + V[I - X'\sigma^{-2}\Sigma^{-1}X(V^{-1} + X'\sigma^{-2}\Sigma^{-1}X)^{-1}]X'\sigma^{-2}\Sigma^{-1}(\mathbf{y} - X\underline{\mu}) \\
&= \underline{\mu} + (V^{-1} + X'\sigma^{-2}\Sigma^{-1}X)^{-1}X'\sigma^{-2}\Sigma^{-1}X(\hat{\underline{\beta}} - \underline{\mu}) \\
&= (V^{-1} + X'\sigma^{-2}\Sigma^{-1}X)^{-1}(V^{-1}\underline{\mu} + X'\sigma^{-2}\Sigma^{-1}X\hat{\underline{\beta}}).
\end{aligned}
$$

Notice that (6.6) coincides with (6.4).

Theorem 1 Assume that

$$\begin{bmatrix}\beta \\ \varepsilon\end{bmatrix} \sim WS\left[\begin{bmatrix}\mu \\ 0\end{bmatrix}, \begin{bmatrix}V & 0 \\ 0 & \sigma^2\Sigma\end{bmatrix}\right], \quad \begin{matrix} V > 0 \\ \Sigma > 0. \end{matrix}$$

and

(i) $\underset{T\to\infty}{\text{plim}}\ \lambda_{\min}(X'_T X_T) = \infty$

(ii) $\lambda_{\max}(\Sigma) \le k_1 < \infty$ all $T \ge 1$ (where $\lambda_{\min}(A)$ means minimum eigenvalue of matrix A. Similarly for the $\lambda_{\max}(A)$). Then

$$\lim_{T\to\infty} E\,\|\tilde{\beta} - \beta\|^2 = 0 \tag{a}$$

and

$$\underset{T\to\infty}{\text{plim}}\ \hat{\beta} = \beta \tag{b}$$

where $\tilde{\beta}$ is the linear Bayes estimator, $\hat{\beta}$ is GLSE for β and $\|Z\|^2 = \operatorname{tr}(ZZ')$.

Proof :

a) Let $\mathbf{y}_T = (y_1, \dots, y_T)'$. Frst consider

$$\begin{aligned} E\{\|\tilde{\beta} - \beta\|^2 \mid \mathbf{y}_T\} &= \operatorname{tr}\ [\Sigma_{\beta \mid \mathbf{y}_T}] \quad \text{where } \Sigma_{\beta \mid \mathbf{y}_T} = (V^{-1} + K)^{-1} \\ &\le \operatorname{tr}\ [(X'_T\sigma^{-2}\Sigma^{-1}X_T)^{-1}] \\ &\le \sigma^2(\lambda_{\min}(\Sigma^{-1}))^{-1}\operatorname{tr}(X_{T'}X_T)^{-1} \\ &\le \sigma^2 \operatorname{rank} X_T\ \lambda_{\max}(\Sigma)\ (\lambda_{\min}(X'_T X_T))^{-1} \end{aligned}$$

which by assumption (i) and (ii) yields

$$\operatorname*{plim}_{T\to\infty} E\{\|\hat{\underline{\beta}} - \underline{\beta}\| \mid \mathbf{y}_T\} = 0.$$

Since for all $T \geq 1$, $\operatorname{tr} V < \infty$, and

$$E\{\operatorname{tr} \Sigma_{\underline{\beta} \mid \mathbf{y}_T}\} < \infty$$

Thus by the Lebesgue's dominated convergence theorem

$$\begin{aligned}
&\lim_{T\to\infty} E\{\|\tilde{\underline{\beta}} - \underline{\beta}\|^2\} \\
&= \lim_{T\to\infty} E\{E[\|\tilde{\underline{\beta}} - \underline{\beta}\|^2 \mid \mathbf{y}_T]\} \\
&= 0
\end{aligned}$$

b) From $\tilde{\underline{\beta}} = (V^{-1} + K)^{-1}(V^{-1}\underline{\mu} + K\hat{\underline{\beta}})$ one can derive

$$\hat{\underline{\beta}} = \tilde{\underline{\beta}} + K^{-1}V^{-1}(\tilde{\underline{\beta}} - \underline{\mu}).$$

Now

$$\begin{aligned}
&\operatorname*{plim}_{T\to\infty} \|K^{-1}V^{-1}(\tilde{\underline{\beta}} - \underline{\mu})\| \\
&\leq \operatorname*{plim}_{T\to\infty} \{\|K^{-1}\|\,\|V^{-1}\|\,[\|\tilde{\underline{\beta}} - \underline{\beta}\| + \|\underline{\beta} - \underline{\mu}\|]\}' \\
&= 0
\end{aligned}$$

since $\operatorname*{plim}_{T\to\infty} \|\tilde{\underline{\beta}} - \underline{\beta}\| = 0$ by part a) and

$$\begin{aligned}
\operatorname*{plim}_{T\to\infty} \|K^{-1}\| &= \operatorname*{plim}_{T\to\infty} \operatorname{tr} (K^{-1}K^{-1'})^{\frac{1}{2}} \\
&= \operatorname*{plim}_{T\to\infty} [\sigma^2 \operatorname{rank} X \lambda_{\max} \Sigma (\lambda_{\min}(X'_T X_T))^{-1}] \\
&= 0.
\end{aligned}$$

Thus

$$\operatorname{plim} \hat{\underline{\beta}} = \operatorname{plim} \tilde{\underline{\beta}} = \underline{\beta},$$

since $\|\underline{\beta} - \underline{\mu}\|$ is independent of T. Notice that Theorem 1 holds for $\underline{\beta}^{*}$ (Bayes estimator with normal prior distribution).

6.2.4 Empirical Linear Bayes Estimators

If the prior parameters $\underline{\mu}$, V and σ^2 are unknown, we are unable to employ the linear Bayes estimate. However, if there are earlier observations, that is, the dynamic linear regression problem occurs repeatedly, we can attempt to estimate these prior parameters from those data. Thus let us assume that we have observations from N independent periods of a dynamic linear regression model:

$$y^{(j)} = X^{(j)}\underline{\beta}^{(j)} + \underline{\varepsilon}^{(j)} \quad j = 1, 2, \dots, N \tag{6.7}$$

with

$$\begin{bmatrix} \underline{\beta}^{(j)} \\ \underline{\varepsilon}^{(j)} \end{bmatrix} \sim WS\left[\begin{bmatrix} \underline{\mu} \\ 0 \end{bmatrix}, \begin{bmatrix} V & 0 \\ 0 & \sigma^2\Sigma_j \end{bmatrix}\right], \quad \Sigma_j > 0 \quad (known)$$

and $X^{(j)} \in M_{n_j \times (N+H)}$ (set of matrices), rank $X^{(j)} = N + H$ and $\{\underline{\beta}, \underline{\varepsilon}, \underline{\beta}^{(j)}, \underline{\varepsilon}^{(j)}, j = 1, \dots, N\}$ a set of independent random vectors.

Note that in the above model (6.7) $X^{(j)}$ cannot remain fixed for each j. Thus the model differs from the classical one. Moreover, in each period the sample sizes are not necessary the same. However, the empirical Bayes approach is still applicable (see chapter 3).

We now confine ourself to the estimation of the current $\underline{\beta}^{(0)}$, using the observations from past history, to estimate the unknown parameters. For them we introduce the following estimators for $\underline{\mu}$, V and σ^2 which depend on $\omega = \{N, n_1, \dots, n_N\}$.

$$\hat{\underline{\mu}}_{\omega} = \frac{1}{N}\sum_{j=1}^{N} \hat{\underline{\beta}}^{(j)} \tag{6.8}$$

where

$$\hat{\underline{\beta}}^{(j)} = \sigma^{-2} K_j X^{(j)} \Sigma_j^{-1},$$

$$K_j = (\sigma^{-2} X^{(j)T} \Sigma_j^{-1} X^{(j)})^{-1},$$

$$\hat{V}_\omega = \frac{1}{N} \sum_{j=0}^{N} (\hat{\underline{\beta}}^{(j)} - \hat{\mu}_\omega)(----)' \tag{6.9}$$

and

$$\hat{\sigma}_\omega^2 = \frac{1}{N} \sum_{i=1}^{N} \hat{\sigma}_i^2, \tag{6.10}$$

where

$$\hat{\sigma}_i^2 = \frac{1}{n_i} (\mathrm{y}^{(i)} - X^{(i)} \hat{\underline{\beta}}^{(i)})' \Sigma_i^{-1} (\mathrm{y}^{(i)} - X^{(i)} \hat{\underline{\beta}}^{(i)}) .$$

In order to achieve asymptotic results we need establish the convergence properties for $\hat{\underline{\mu}}_\omega$, $\hat{V}_\omega$ and $\hat{\sigma}_\omega^2$ with respect to ω . Since $\beta^{(j)}$ is not unbiased in the jth period, we are forced to demand that not only the number (N) of periods increase but also that the sample size n_j in period j increases. Thus we state a 'large-sense' convergence in probability as follows:

Definition 1

A sequence $\{ I_\omega = I(\mathrm{y}^{(1)}, \dots, \mathrm{y}^{(N)}) \}$ of random vectors is called *a large-sense convergence in probability* to a given I, denoted by $\operatorname*{plim}_\omega I_\omega = I$, if for all positive number η_1, η_2 there are positive integers $N^*(\eta_1,\eta_2)$, $n_j^*(N, \eta_1,\eta_2)$, $j = 1, 2, \dots, N$, such that

$$P\{\| I_\omega - I \| > \eta_1\} < \eta_2$$

for all $N > N^*$, $n_j \geq n_j^*, j = 1, \dots, N$.

Theorem 2 (1) If $\underset{n_j \to \infty}{\text{plim}}\ \hat{\beta}^{(j)} = \beta^{(j)}$ $(j = 1, 2, \dots)$, then

$$\underset{\omega}{\text{plim}}\ \hat{\mu}_\omega = \mu \quad \text{and} \quad \underset{\omega}{\text{plim}}\ \hat{V}_\omega = V$$

(2) If $\lim_{n_j \to \infty} E\|\hat{\beta}^{(j)} - \beta^{(j)}\|^2 = 0, j = 1, 2, \dots$, then

$$\lim_{\omega} \quad E\|\hat{\mu}_\omega - \mu\|^2 = 0$$

(3) If

$$\underset{n_j \to \infty}{\text{plim}} \frac{1}{\sqrt{n_j}} K_j^{1/2}(\hat{\beta}^{(j)} - \beta^{(j)}) = 0$$

and

$$\text{var}\ (\varepsilon^{(j)\prime}\Sigma_{(j)}^{-1}\varepsilon^{(j)}) \to o\,(n_j^2)$$

then

$$\underset{\omega \to \infty}{\text{plim}}\ \hat{\sigma}^2_\omega = \sigma^2$$

Proof: (1) First we prove that $\underset{\omega}{\text{plim}}\ \hat{\mu}_\omega = \mu$.

Let η_1, η_2 be any two positive numbers. By the weak law of large number there exists a number $N^*(\eta_1,\eta_2)$ such that

$$P\{\|N^{-1}\sum_{j=1}^{N}\beta^{(j)} - \mu\| > \frac{\eta_1}{2}\} < \frac{\eta_2}{2} \quad \text{for} \quad N \geq N^*.$$

Now from the assumption $\underset{n_j \to \infty}{\text{plim}}\ \hat{\beta}^{(j)} = \beta^{(j)}$, $j = 1, \dots, N$, we know that for each $N \geq N^*$ there exist $n_j^*(N, \eta_1,\eta_2)$ such that

$$P\{\|\hat{\beta}^{(j)} - \beta^{(j)} > \frac{\eta_1}{2}\} < \frac{\eta_2}{2N} \quad \text{for} \ n_j \geq n_j^*,\ j = 1, 2, \dots, N.$$

Since

$$\|\hat{\underline{\mu}}_\omega - \underline{\mu}\| \le N^{-1}\sum_{j=1}^{N}\|\hat{\underline{\beta}}^{(j)} - \underline{\beta}^{(j)}\| + \|N^{-1}\sum_{j=1}^{N}\underline{\beta}^{(j)} - \underline{\mu}\|,$$

The proof is completed by combining the above two facts:

$$\begin{aligned}
&P\{\|\hat{\underline{\mu}}_\omega - \underline{\mu}\| > \eta_1\}\\
&\le P\{N^{-1}\sum_{j=1}^{N}\|\hat{\underline{\beta}}^{(j)} - \underline{\beta}^{(j)}\| + \|N^{-1}\sum_{j=1}^{N}\underline{\beta}^{(j)} - \underline{\mu}\| > \eta_1\}\\
&\le P\{N^{-1}\sum_{j=1}^{N}\|\hat{\underline{\beta}}^{(j)} - \underline{\beta}^{(j)}\| > \frac{\eta_1}{2} \text{ or } \|n^{-1}\sum_{j=1}^{N}(\beta^{(j)} - \mu)\| > \frac{\eta_1}{2}\}\\
&< N\frac{\eta_2}{2N} + \frac{\eta_2}{2} = \eta_2 \quad \text{for all } N > N^*, n_j \ge n_j^*.
\end{aligned}$$

Next we prove that $\operatorname*{plim}_{\omega} \hat{V}_\omega = V$. It is easy to see that

$$\hat{V}_\omega = A_\omega + B_\omega + C_\omega + D_\omega$$

where

$$\begin{aligned}
A_\omega &= N^{-1}\sum_{j=1}^{N}(\hat{\underline{\beta}}^{(j)} - \underline{\beta}^{(j)})(\text{----})'\\
B_\omega &= N^{-1}\sum_{j=1}^{N}(\underline{\beta}^{(j)} - \underline{\mu})(\text{----})'\\
C_\omega &= (\hat{\underline{\mu}}_\omega - \underline{\mu})(\text{----})'
\end{aligned}$$

and

$$D_\omega = \text{ the sum of all mixed (cross) terms}$$

Analogously to the proof of the first part we obtain

$$\operatorname*{plim}_{\omega} A_\omega = \mathbf{0}$$

and the WLLN (weak law of large number) gives

$$\operatorname*{plim}_{\omega} B_\omega = V.$$

Moreover the first part of the theorem implies

$$\operatorname*{plim}_{\omega} C_\omega = \mathbf{0}.$$

Finally, by the Cauchy - Schwarz inequality all the cross terms tend in probability to zero (in large sense):

$$\operatorname*{plim}_{\omega} D_\omega = \mathbf{0}.$$

Thus we have proven that $\operatorname*{plim}_{\omega} \hat{V}_\omega = V.$

(2) Applying the C_δ - inequality we have

$$\begin{aligned}
E\,\|\hat{\underline{\mu}}_\omega - \underline{\mu}\|^2 &\le \frac{2}{N^2} E\,\|\sum_{j=1}^{N} (\hat{\underline{\beta}}^{(j)} - \underline{\beta}^{(j)})\|^2 + \frac{2}{N^2} E\|\sum_{j=1}^{N} (\underline{\beta}^{(j)} - \underline{\mu})\|^2 \\
&= \frac{2}{N^2}\sum_{j=1}^{N} E\,\|\hat{\underline{\beta}}^{(j)} - \underline{\beta}^{(j)}\|^2 + \frac{2}{N^2}\sum_{j\ne i} E(\hat{\underline{\beta}}^{(j)} - \underline{\beta}^{(j)})'E\,(\hat{\underline{\beta}}^{(i)} - \underline{\beta}^{(i)}) + \frac{2}{N^2}\sum_{j=1}^{N} E\,\|\underline{\beta}^{(j)} - \underline{\mu}\|^2 + \mathbf{0} \\
&= \frac{2}{N^2}\sum_{j=1}^{N} E\,\|\hat{\underline{\beta}}^{(j)} - \underline{\beta}^{(j)}\|^2 + \frac{2}{N^2}\sum_{j\ne i} (E\,\hat{\underline{\beta}}^{(j)} - \underline{\mu})'(E\,\hat{\underline{\beta}}^{(i)} - \underline{\mu}) + \frac{2}{N^2}\sum_{j=1}^{N} E\,\|\underline{\beta}^{(j)} - \underline{\mu}\|^2
\end{aligned}$$

which from the assumptions yields

$$\lim_{n_j \to \infty} E\,\hat{\underline{\beta}}^{(j)} = \underline{\mu}, j = 1, \ldots, N,$$

and

$$\frac{2}{N^2}\sum_{j=1}^{N} E\|\underline{\beta}^{(j)} - \underline{\mu}\|^2 = \frac{2}{N}\ \mathrm{tr}T \to 0 \quad \text{as} \quad N \to \infty$$

Thus

$$\lim_{\omega} E\,\|\hat{\underline{\mu}}_\omega - \underline{\mu}\|^2 = 0$$

which completes the proof.

(3) Since

$$\begin{aligned}
\hat{\sigma}^2_{(j)} &= \frac{1}{n_j}(\mathbf{y}^{(j)} - X_j\hat{\underline{\beta}}^{(j)})^T\Sigma_j^{-1}(\mathbf{y}^{(j)} - X_j\hat{\underline{\beta}}^{(j)}) \\
&= \frac{1}{n_j}(\mathbf{y}^{(j)} - X_j\underline{\beta}^{(j)} + X_j(\underline{\beta}^{(j)} - \hat{\underline{\beta}}^{(j)}))^T\Sigma_j^{-1}(\text{----}) \\
&= \frac{1}{n_j}[\underline{\varepsilon}^{(j)^T}\Sigma_j^{-1}\underline{\varepsilon}^{(j)} + (\underline{\beta}^{(j)} - \hat{\underline{\beta}}^{(j)})^T X_j^T\Sigma_j^{-1}X_j(\underline{\beta}^{(j)} - \hat{\underline{\beta}}^{(j)})] \\
&+ \frac{1}{n_j}[\underline{\varepsilon}^{(j)^T}\Sigma_j^{-1}X_j(\underline{\beta}^{(j)} - \hat{\underline{\beta}}^{(j)}) + (\underline{\beta}^{(j)} - \hat{\underline{\beta}}^{(j)})^T X_j^T\Sigma_j^{-1}\underline{\varepsilon}^{(j)})] \\
&= \frac{1}{n_j}[\underline{\varepsilon}^{(j)^T}\Sigma_j^{-1}\underline{\varepsilon}^{(j)} - (\hat{\underline{\beta}}^{(j)} - \underline{\beta}^{(j)})^T X_j^T\Sigma_j^{-1}X_j(\hat{\underline{\beta}}^{(j)} - \underline{\beta}^{(j)})]
\end{aligned}$$

then

$$\begin{aligned}
&\frac{1}{n_j}\underline{\varepsilon}^{(j)^T}\Sigma_j^{-1}\underline{\varepsilon}^{(j)} = \frac{1}{n_j}\sum_{i=1}^{n_j}\tilde{\underline{\varepsilon}}^{(j)^T}\tilde{\underline{\varepsilon}}^{(j)} \\
&\to \sigma^2 \quad \text{(in probability)} \quad \text{as} \quad n_j \to \infty .
\end{aligned}$$

Since the components of $\tilde{\underline{\varepsilon}}^{(j)} = \Sigma_j^{-1/2}\underline{\varepsilon}^{(j)}$ are independent with zero mean, variance σ^2 and fourth monent finite (by assumption), j = 1, ..., N, the above result follows by the WLLN . Moreover by another assumption :

$$\frac{1}{\sqrt{n_j}}(X_j^T\Sigma_j^{-1}X_j)^{1/2}(\hat{\underline{\beta}}^{(j)} - \underline{\beta}^{(j)}) \to \mathbf{0} \quad \text{(in probability)} \quad \text{as} \quad n_j \to \infty$$

which completes the proof of (3).

Clearly, a necessary condition for $\hat{V}_\omega > \mathbf{0}$ is $N \geq G + H$. Thus the positive definiteness of $\hat{V}_\omega$ with probability one follows from the regularity of $X^{(j)}$, j = 1, 2,.... (that is: $X^{(j)T}X^{(j)} > \mathbf{0}$ or rank $X^{(j)}$ = G + H) (see appendix (4) in chapter 1)

Using the estimators $\hat{\underline{\mu}}_\omega$, $\hat{V}_\omega$ and $\hat{\sigma}^2_\omega$, we obtain an empirical linear Bayes estimator for $\underline{\beta}^{(0)}$ by

$$\underline{\beta}_{\omega}^{(0)} = (\hat{V}_{\omega}^{-1} + \hat{K}_{\omega}^{(0)})^{-1}(\hat{V}_{\omega}^{-1}\hat{\underline{\mu}}_{\omega} + \hat{K}_{\omega}^{(0)}\hat{\underline{\beta}}^{(0)})$$

where

$$\hat{K}_{\omega}^{(0)} = \hat{\sigma}_{\omega}^{-2}X^{(0)^T}\Sigma_{(0)}^{-1}X^{(0)} = \hat{\sigma}_{\omega}^{-2}Q_{(0)}.$$

Improvement of Estimators $\hat{\underline{\mu}}_{\omega}$ and $\hat{V}_{\omega}$

Let Γ be subset of invertible matrices with order G+H such that it is a convex set with the properties

$$0 < \tau_1(= \inf_{V \in \Gamma} \lambda_{\min}(V)), \tau_2(= \operatorname{lub}_{V \in \Gamma} \lambda_{\max}(V)) < \infty$$

where $\lambda_{\min}(V)$ and $\lambda_{\max}(V)$ denote the smallest and the largest eigenvalue of V. Furthermore, let B be the subset of R^{G+H} such that it is a convex and bounded set with

$$\operatorname{lub}_{\underline{\mu} \in B} \|\underline{\mu}\|^2 = \alpha < \infty$$

If we assume further that the prior parameters $\underline{\mu}$ and V in our model belong to B and Γ respectively then from the convexity of both sets the estimators $\hat{\underline{\mu}}_{\omega}$ and $\hat{V}_{\omega}$ can be uniformly improved upon by use of the corresponding projections $\tilde{\underline{\mu}}_{\omega}$ and $\tilde{V}_{\omega}$ onto the closure of B and Γ respectively.

Under these assumptions we choose

$$\bar{\beta}_{\omega}^{(0)} = (\tilde{V}_{\omega}^{-1} + \hat{K}_{\omega}^{(0)})^{-1}(\tilde{V}_{\omega}^{-1}\tilde{\underline{\mu}}_{\omega} + \hat{k}_{\omega}^{(0)}\hat{\underline{\beta}}^{(0)})$$

as the corrected empirical linear Bayes estimator for $\underline{\beta}^{(0)}$.

If $\underline{\mu}$ and V belong to the convex bounded sets B and Γ respectively, then the consistency of $\tilde{\underline{\mu}}_{\omega}$ and $\tilde{V}_{\omega}$ follows from the consistency of $\hat{\underline{\mu}}_{\omega}$ and $\hat{V}_{\omega}$.

Next we want to investigate the asymptotic properties of the ELBE. Before doing this let us first establish the asymptotic equivalence between the empirical linear Bayes estimator and the linear Bayes estimator:

Theorem 3 (i) If $\operatorname*{plim}_{n_0 \to \infty} \|K_0^{-1}\| = 0$, then

$$\operatorname*{plim}_{n_0 \to \infty} (\underline{\beta}_\omega^{(0)} - \underline{\beta}^{(0)}) = 0.$$

(ii) If estimators $\hat{\underline{\mu}}_\omega, \hat{V}_\omega$ and $\hat{\sigma}_\omega^2$ of prior parameters are consistent, then

$$\operatorname*{plim}_{\omega} (\underline{\beta}_\omega^{(0)} - \tilde{\underline{\beta}}^{(0)}) = 0.$$

Proof (i) Since

$$\underline{\beta}_\omega^{(0)} - \tilde{\underline{\beta}}^{(0)} = Q_\omega \underline{\delta}_\omega + (V^{-1} + K_0)^{-1} V^{-1} (\hat{\underline{\mu}}_\omega - \underline{\mu}), \quad \text{(see appendix (2))}$$

where

$$\begin{aligned} Q_\omega &= (\hat{V}_\omega^{-1} + \hat{K}_\omega^{(0)})^{-1} \hat{V}_\omega^{-1} - (V^{-1} + K_0)^{-1} V^{-1} \\ &= (V^{-1} + K_0)^{-1} (\hat{\sigma}_\omega^2 \sigma^{-2} \hat{V}_\omega^{-1} - V^{-1}) \hat{V}_\omega (\hat{V}_\omega + K_0^{-1})^{-1} \quad \text{(see appendix (1))} \end{aligned}$$

and

$$\underline{\delta}_\omega = \hat{\underline{\mu}}_\omega - \hat{\underline{\beta}}_\omega^{(0)}$$

we obtain

$$\begin{aligned} &\|\underline{\beta}_\omega^{(0)} - \tilde{\underline{\beta}}^{(0)}\| \\ &\leq \|Q_\omega\| \|\delta\| + \|(V^{-1} + K_0)^{-1}\| \|V^{-1}\| \|\hat{\underline{\mu}}_\omega - \underline{\mu}\| \\ &\leq (\|\hat{K}_\omega^{(0)-1}\| \|\hat{V}_\omega^{-1}\| + \|K_0^{-1}\| \|V^{-1}\|) \|\underline{\delta}_\omega\| + \|K_0^{-1}\| \|V^{-1}\| \|\hat{\underline{\mu}}_\omega - \underline{\mu}\| \\ &\leq \|K_0^{-1}\| [(\hat{\sigma}_\omega^2 \sigma^{-2} \|\hat{V}_\omega^{-1}\| + \|V^{-1}\|) \|\underline{\delta}_\omega\| + \|V^{-1}\| \|\hat{\underline{\mu}}_\omega - \underline{\mu}\|]. \end{aligned}$$

Thus

$$\|\underline{\beta}_\omega^{(0)} - \bar{\underline{\beta}}^{(0)}\| < \frac{\varepsilon}{2}, \quad \text{(with probability 1)} \quad \text{as } n_0 \text{ is large enough.} \tag{6.11}$$

since

$$\operatorname*{plim}_{n_0 \to \infty} \underline{\delta}_\omega = \hat{\underline{\mu}}_\omega - \underline{\beta}^{(0)}. \text{ (by Theorem 1 (ii))}$$

Now from the relation

$$\|\underline{\beta}_\omega^{(0)} - \underline{\beta}^{(0)}\| \le \|\underline{\beta}_\omega^{(0)} - \bar{\underline{\beta}}^{(0)}\| + \|\bar{\underline{\beta}}^{(0)} - \underline{\beta}^{(0)}\|$$

we have

$$\operatorname*{plim}_{n_0 \to \infty} (\underline{\beta}_\omega^{(0)} - \underline{\beta}^{(0)}) = \mathbf{0} \text{ (by Theorem 1 (i) and (6.11))}$$

(ii)

$$\begin{aligned}
&\underline{\beta}_\omega^{(0)} - \bar{\underline{\beta}}^{(0)} \\
&= Q_\omega \underline{\delta}_\omega + (V^{-1} + K_0)^{-1} V^{-1}(\hat{\underline{\mu}}_\omega - \underline{\mu}) \\
&= (V^{-1} + K_0)^{-1}(\hat{\sigma}_\omega^2 \sigma^{-2} \hat{V}_\omega^{-1} - V^{-1})\hat{V}_\omega(\hat{V}_\omega + K_0^{-1})^{-1}\underline{\delta}_\omega + (V^{-1} + K_0)^{-1}V^{-1}(\hat{\underline{\mu}}_\omega - \underline{\mu})
\end{aligned}$$

Thus under the assumption that $\hat{\underline{\mu}}_\omega, \hat{V}_\omega$ and $\hat{\sigma}_\omega$ are consistent and

$$\operatorname*{plim}_{\omega} \underline{\delta}_\omega = \underline{\mu} - \underline{\beta}^{(0)}$$

we obtain

$$\operatorname*{pilm}_{\omega} (\underline{\beta}_\omega^{(0)} - \bar{\underline{\beta}}^{(0)}) = \mathbf{0}.$$

In general it is very complicated to obtain results about the risk of these empirical procedures for finite samples. Sometimes a Monte Carlo study is used to investigate approximately how good an empirical linear Bayes estimation procedure works for small sample sizes. Here we restrict our consideration to the case of large N, and n_j $(j = 1, \ldots, N)$.

One criteria, which is only a minimal requirement for an empirical linear Bayes estimator, is the property of asymptotic optimality. We define the risk of estimator $\bar{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}^{(0)}$ as follows:

$$
\begin{aligned}
R(\bar{\boldsymbol{\beta}}) &= E\,\|\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}\| = \operatorname{tr} E\,[(\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(0)})(\text{----})'] \\
R(\bar{\boldsymbol{\beta}} \mid \mathbf{y}_{\omega}) &= E\,\{\|\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(0)}\|^2 \mid \mathbf{y}_{\omega}\}
\end{aligned}
$$

where

$$
\mathbf{y}'_{\omega} = (\mathbf{y}^{(1)'}, \ldots, \mathbf{y}^{(N)'})
$$

Notice that $\bar{\boldsymbol{\beta}}$ depends on $\mathbf{y}_{\omega}$ and $\mathbf{y}^{(0)}$.

Definition 2 A sequence $\{\boldsymbol{\beta}_{\omega}^{(0)}\}$ of ELBEs is said to be

(i) *conditionally asymptotically linearly optimal* if

$$
\operatorname*{plim}_{\omega} R(\boldsymbol{\beta}_{\omega}^{(0)} \mid \mathbf{y}_{\omega}) = R(\tilde{\boldsymbol{\beta}}^{(0)})
$$

(ii) *asymptotically linearly optimal* if

$$
\operatorname*{plim}_{\omega} R(\boldsymbol{\beta}_{\omega}^{(0)}) = R(\tilde{\boldsymbol{\beta}}^{(0)})
$$

Theorem 4 If the estimators $\hat{\boldsymbol{\mu}}_{\omega}$, $\hat{V}_{\omega}$ and $\hat{\sigma}_{\omega}$ are consistent and $\lim_{n_0 \to \infty} E\,\|\hat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}^{(0)}\|^2 = 0$, then

(a) the ELBE $\boldsymbol{\beta}_{\omega}^{(0)}$ is conditionally asymptotically optimal

(b) the corrected ELBE $\bar{\boldsymbol{\beta}}_{\omega}^{(0)}$ is asymptotically optimal provided $\hat{\boldsymbol{\mu}}_{\omega} \in \mathrm{B}$ and $\hat{V}_{\omega} \in \Gamma$.

Proof: (a) Since

$$
\begin{aligned}
R(\underline{\beta}_{\omega}^{(0)} \mid \mathbf{y}_{\omega}) &= E\{\|\underline{\beta}_{\omega}^{(0)} - \underline{\beta}^{(0)}\|^2 \mid \mathbf{y}_{\omega}\} \\
&= E\{\|\underline{\beta}_{\omega}^{(0)} - \tilde{\underline{\beta}}^{(0)} + \tilde{\underline{\beta}}^{(0)} - \beta^{(0)}\|^2 \mid \mathbf{y}_{\omega}) \\
&\le E\|\tilde{\underline{\beta}}^{(0)} - \underline{\beta}^{(0)}\|^2 + E(\|\underline{\beta}_{\omega}^{(0)} - \tilde{\underline{\beta}}^{(0)}\|^2 \mid \mathbf{y}_{\omega}) \\
&\quad + 2[E\{\|\underline{\beta}_{\omega}^{(0)} - \tilde{\underline{\beta}}^{(0)}\|^2 \mid \mathbf{y}_{\omega}\} E\{\|\tilde{\underline{\beta}}^{(0)} - \underline{\beta}^{(0)}\|^2\}]^{1/2},
\end{aligned}
$$

it will suffice to show that

$$
E\{\|\underline{\beta}_{\omega}^{(0)} - \tilde{\underline{\beta}}(0)\| \mid \mathbf{y}_{\omega}\} \to 0 \ \ \text{(in probability)} \ \ \text{as} \ \ \omega \to \infty.
$$

From the relation (see appendix)

$$
\underline{\beta}_{\omega}^{(0)} - \tilde{\underline{\beta}}^{(0)} = Q_{\omega}\underline{\delta}_{\omega} + (V^{-1} + K_0)^{-1} V^{-1}(\hat{\underline{\mu}}_{\omega} - \underline{\mu})
$$

we obtain

$$
\|\underline{\beta}_{\omega}^{(0)} - \tilde{\underline{\beta}}^{(0)}\|^2 \le 2\|Q_{\omega}\underline{\delta}_{\omega}\|^2 + 2\|(V^{-1} + K_0)^{-1} V^{-1}(\hat{\underline{\mu}}_{\omega} - \underline{\mu})\|^2.
$$

Now we show that each term on the right hand side tends to zero in suitable convergence.

(1)

$$
\begin{aligned}
&E\{\|(V^{-1} + K_0)^{-1} V^{-1}(\hat{\underline{\mu}}_{\omega} - \underline{\mu})\|^2 \mid \mathbf{y}_{\omega}\} \\
&= \|(V^{-1} + K_0)^{-1} V^{-1}(\hat{\underline{\mu}}_{\omega} - \mu)\|^2 E[1 \mid \mathbf{y}_{\omega}] \\
&\le \|V\|^2 \|V^{-1}\|^2 \|\hat{\underline{\mu}}_{\omega} - \underline{\mu}\|^2 \\
&\text{(by the Schwarz inequality and } \|(V^{-1} + K_0)^{-1}\| \le \|V\|).
\end{aligned}
$$

Thus

$$
E\{\|(V^{-1} + K_0)^{-1} V^{-1}(\hat{\underline{\mu}}_{\omega} - \underline{\mu})\|^2 \mid \mathbf{y}_{\omega}\} \to 0 \ \ \text{in probability} \ \ \text{as} \ \ \hat{\underline{\mu}}_{\omega} \to \underline{\mu} \ \ \text{in probability.}
$$

(2)

$$
\begin{aligned}
\|Q_\omega \underline{\delta}_\omega\|^2 &\le \|Q_\omega\|^2 \|\hat{\underline{\mu}}_\omega - \hat{\underline{\beta}}^{(0)}\|^2 \\
&= \|Q_\omega\|^2 \|\hat{\underline{\mu}} - \underline{\mu} + \underline{\mu} - \hat{\underline{\beta}}^{(0)}\|^2 \\
&\le \|Q_\omega\|^2 [2\|\hat{\underline{\mu}} - \underline{\mu}\|^2 + 2\|\underline{\mu} - \hat{\underline{\beta}}^{(0)}\|^2] \text{ (using } C_\delta - \text{ineq.)} \\
&\le \|Q_\omega\|^2 [2\|\hat{\underline{\mu}}_\omega - \underline{\mu}\|^2 + 4\|\underline{\mu} - \underline{\beta}^{(0)}\|^2 + 4\|\underline{\beta}^{(0)} - \hat{\underline{\beta}}^{(0)}\|^2]
\end{aligned}
$$

where

$$
\begin{aligned}
\|Q_\omega\|^2 &\le \|(V^{-1} + K_0)^{-1}\|^2 \|a\hat{V}_\omega^{-1} - V^{-1}\|^2 \|\hat{V}_\omega\|^2 \|(\hat{V}_\omega + K_0^{-1})^{-1}\|^2 \\
&\le \|V\|^2 \|\hat{V}_\omega\|^2 \|\hat{V}_\omega^{-1}\|^2 \|\hat{V}_\omega^{-1} - V^{-1}\|^2.
\end{aligned}
\qquad \text{(where } a = \hat{\sigma}_\omega^2 \sigma^{-2}\text{)}
$$

Let

$$
C(\mathbf{y}_\omega) = \|V\|^2 \|\hat{V}_\omega\|^2 \|\hat{V}_\omega^{-1}\|^2 \|\hat{V}_\omega^{-1} - V^{-1}\|^2,
$$

then $C(\mathbf{y}_\omega)$ does not depend on K and converges to zero in probability under the assumptions. (Notice that $\operatorname*{plim}_\omega V_\omega = V$ implies $\operatorname*{plim}_\omega V_\omega^{-1} = V^{-1}$.). Thus we have

$$
\|Q_\omega \underline{\delta}_\omega\|^2 \le C(\mathbf{y}_\omega)[2\|\hat{\underline{\mu}}_\omega - \underline{\mu}\|^2 + 4\|\underline{\mu} - \underline{\beta}^{(0)}\|^2 + 4\|\underline{\beta}^{(0)} - \hat{\underline{\beta}}^{(0)}\|^2]
$$

which implies

$$
E\{\|Q_\omega \underline{\delta}_\omega\|^2 \mid \mathbf{y}_\omega\} \le C(\mathbf{y}_\omega)[2\|\hat{\underline{\mu}} - \underline{\mu}\|^2 + 4\|\underline{\mu} - \underline{\beta}^{(0)}\|^2 + 4\|\underline{\beta}^{(0)} - \hat{\underline{\beta}}^{(0)}\|^2].
$$

Thus we have (by the Lebesgue dominated convergence Theorem)

$$
\operatorname*{plim}_\omega E(\|Q_\omega \underline{\delta}_\omega\|^2 \mid \mathbf{y}_\omega) = 0
$$

as $C(\mathbf{y}_\omega) \to 0$ in probability and that

$$
2\|\hat{\underline{\mu}}_\omega - \underline{\mu}\|^2 + 4\|\underline{\mu} - \underline{\beta}^{(0)}\|^2 + 4\|\underline{\beta}^{(0)} - \hat{\underline{\beta}}^{(0)}\|^2
$$

is bounded in probability. This completes the proof of (a).

(b) Analogous to the proof of (a), it suffices to show that

$$\lim_{\omega} E\|\bar{\underline{\beta}}_{\omega}^{(0)} - \tilde{\underline{\beta}}^{(0)}\|^2 = 0.$$

Similarly to the above result we have

$$\|(V^{-1} + K_0)^{-1} V^{-1}(\tilde{\underline{\mu}}_{\omega} - \underline{\mu})\|^2 \leq \|V\|^2 \|V^{-1}\|^2 \|\tilde{\underline{\mu}}_{\omega} - \underline{\mu}\|^2$$

Letting $\tilde{C}(\mathbf{y}_{\omega}) = \|V\|^2 \|V^{-1}\|^2 \|\tilde{\underline{\mu}}_{\omega} - \underline{\mu}\|^2$ and using

$$\|V\|^2 \leq c\tau_2^2, \|V^{-1}\| \leq c\tau_1^{-2}, \|\tilde{\underline{\mu}}_{\omega}\| \leq \alpha, \|\underline{\mu}\| \leq \alpha$$

we obtain

$$\tilde{C}(\mathbf{y}_{\omega}) \leq 4c^2[\frac{\tau_2}{\tau_1}]^2 \alpha < \infty$$

Hence, again by the Lebesgue dominated convergence theorem we conclude

$$E\|(V^{-1} + K_0)^{-1} V^{-1}(\tilde{\underline{\mu}}_{\omega} - \underline{\mu})\|^2 \rightarrow 0 \quad \text{as} \quad \omega \rightarrow \infty$$

Now denoting $\tilde{Q}_{\omega}$ and $\tilde{\underline{\delta}}_{\omega}$ analogously to Q_{ω} and $\underline{\delta}_{\omega}$ respectively by replacing $\hat{V}_{\omega}$ and $\hat{\underline{\mu}}_{\omega}$ with $\tilde{V}_{\omega}$ and $\tilde{\underline{\mu}}_{\omega}$, we have

$$\begin{aligned}
\|\tilde{Q}_{\omega}\|^2 &\leq 2[\|(\tilde{V}_{\omega}^{-1} + K_0)^{-1} \tilde{V}_{\omega}^{-1}\|^2 + \|(V^{-1} + K_0)^{-1} V^{-1}\|^2] \\
&\leq 2[\|\tilde{V}_{\omega}\|^2 \|\tilde{V}_{\omega}^{-1}\|^2 + \|V\|^2 \|V^{-1}\|^2] \\
&\leq 4c^2[\frac{\tau_2}{\tau_1}]^2.
\end{aligned}$$

Thus from

$$
\begin{aligned}
\|\tilde{Q}_\omega \tilde{\underline{\delta}}_\omega\|^2 &\le \|\tilde{Q}_\omega\|^2 [\|\tilde{\underline{\mu}}_\omega - \hat{\underline{\beta}}^{(0)}] \\
&\le 2\|\tilde{Q}_\omega\|^2 [\|\tilde{\underline{\mu}}_\omega\|^2 + \|\hat{\underline{\beta}}^{(0)}\|^2] \\
&\le 2\|\tilde{Q}_\omega\|^2 [\|\tilde{\underline{\mu}}_\omega\|^2 + 2\|\underline{\beta}^{(0)}\|^2 + 2\|\hat{\underline{\beta}}^{(0)} - \underline{\beta}^{(0)}\|^2] \\
&\le 8c^2 [\frac{\tau_2}{\tau_1}]^2 [\alpha + 2\|\underline{\beta}^{(0)}\|^2 + 2\|\hat{\underline{\beta}}^{(0)} - \underline{\beta}^{(0)}\|^2] \\
&\le 8c^2 [\frac{\tau_2}{\tau_1}]^2 [3\alpha + 2\|\hat{\underline{\beta}}^{(0)} - \underline{\beta}^{(0)}\|^2]
\end{aligned}
$$

we have (by LDC)

$$E\|\tilde{Q}_\omega \tilde{\underline{\delta}}_\omega\|^2 \to 0 \quad \text{as} \quad \|\tilde{Q}_\omega \tilde{\underline{\delta}}_\omega\|^2 \to 0 \quad \text{(in probability)}$$

(due to assumption $E\|\hat{\underline{\beta}}^{(0)} - \underline{\beta}^{(0)}\| \to 0$).

6.3 ELBE in TSRM with AR (1) Disturbances

Consider the dynamic linear regression model with lagged dependent variables

$$y_t = \sum_{i=1}^{G} \alpha_i y_{t-i} + \sum_{j=1}^{L} \gamma_j z_{t-j} + u_t \tag{6.12}$$

and with AR (1) disturbances

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad t = 1, \dots, T.$$

where

$$\varepsilon_t \sim WS(0, \sigma^2), \quad \sigma^2\text{: known}$$

or , in matrix form :

$$\mathbf{y} = X\underline{\beta} + \mathbf{u}$$

with

$$U = \quad \text{var}\,\mathbf{u} = \sigma^2\Sigma = \frac{\sigma^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \dots & & \\ \dots & & & & \\ \rho^{T-1} & \rho^{T-2} & & \dots & 1 \end{bmatrix}.$$

Suppose that $\underline{\beta} = \underline{\mu} + \underline{v}$ where $\underline{v} \sim WS(0, V)$. Analogously to the previous procedures we can obtain the LBE $\tilde{\underline{\beta}}$ for $\underline{\beta}$ under a squared loss function

$$(\tilde{\underline{\beta}} - \underline{\beta})'(\tilde{\underline{\beta}} - \underline{\beta}). \tag{6.13}$$

Now let us procede as follows. Let

$$\tilde{\underline{\beta}} = A\mathbf{y} + \mathbf{b}. \tag{6.14}$$

This would be a linear Bayesian estimator if A and $\mathbf{b}$ minimize the risk

$$\begin{aligned} R &= E(\tilde{\underline{\beta}} - \underline{\beta})'(\text{----}) \\ &= \text{tr}\,E(\tilde{\underline{\beta}} - \underline{\beta})(\text{----})' \\ &= \text{tr}\,E_X E[(\tilde{\underline{\beta}} - \underline{\beta})(\text{----})' \mid X]. \end{aligned}$$

To minimize R is tantamount to minimizing

$$E[(\tilde{\underline{\beta}} - \underline{\beta})(\text{----})' \mid X] = \Delta(\text{ say}),$$

which by assumptions can be simplified as follows :

$$\Delta = (AX - I)V(AX - I)' + AUA' + (\mathbf{b} - (I - AX)\underline{\mu})(\text{----})'.$$

Note that X and $\underline{v}$ are independent , while $\mathbf{u}$ and $\underline{v}$ are also independent.

To minimize Δ , we proceed sequentially according to Chipman (1964) as follows :

i) Since **b** is contained in the last term only ,

$$\hat{\mathbf{b}} = (I - AX)\underline{\mu}. \tag{6.15}$$

Thus

$$\Delta = (AX - I)V(AX - I)' + AUA'. \tag{6.16}$$

ii) Taking the differential of Δ with respect to A and setting it equal to zero yields

$$\hat{A} = VX'(XVX' + U)^{-1}. \tag{6.17}$$

Thus we find the linear Bayes estimator given **y** :

$$\tilde{\underline{\beta}} = \underline{\mu} + VX'(XVX' + U)^{-1}(\mathbf{y} - X\underline{\mu}). \tag{6.18}$$

Now the problem is how to find the estimators for $\underline{\mu}$, V, and $\sigma^2\Sigma(= U)$. Σ can be estimated by the usual estimation method --- just replacing ρ by $\hat{\rho}$. The estimators of $\underline{\mu}$, V and σ^2 are obtained same as in the previous section, but need further be estimated by replacing ρ_j with $\hat{\rho}_j, j = 1, \dots, N$. When past N independent component problems are available, we obtain the ELBE for current $\underline{\beta}^{(0)}$ which can be given by

$$\tilde{\underline{\beta}}^{(0)}(\hat{\rho}_0) = \hat{\underline{\mu}}_\omega + \hat{V}_\omega X^{(0)\prime}(X^{(0)}\hat{V}_\omega X^{(0)\prime} + \hat{\sigma}^2_\omega\hat{\Sigma}_{(0)})^{-1}(\mathbf{y}^{(0)} - X^{(0)}\hat{\underline{\mu}}_\omega) \tag{6.19}$$

where $\hat{\Sigma}_{(0)}$ is $\Sigma_{(0)}$ with ρ_0 replaced by $\hat{\rho}_0$ and

$$\hat{\rho}_j = \frac{\sum_{t=2}^{T_j} \hat{u}_t^{(j)\prime}\hat{u}_{t-1}^{(j)}}{\sum_{t=2}^{T_j} \hat{u}_{t-1}^{(j)2}}.$$

Here $\hat{u}_t^{(j)}$ denotes residual and j = 0, 1, 2, ... N. Notice that we can show that $\tilde{\beta}^{(0)}(\hat{\rho})$ is consistent and asymptotically optimal to the LBE since $\hat{\mu}_\omega, \hat{V}_\omega$ and $\hat{U}_\omega (= \hat{\sigma}_\omega^2 \Sigma_{(0)})$ are all consistent for large samples .

Appendix :

1. Let $Q_\omega = (\hat{V}_\omega^{-1} + \hat{K}_\omega^{(0)})^{-1}\hat{V}_\omega^{-1} - (V^{-1} + K_0)^{-1}V^{-1}$, then

$$Q_\omega = (V^{-1} + K_0)^{-1}(a\hat{V}_\omega^{-1} - V^{-1})\hat{V}_\omega(\hat{V}_\omega + \hat{K}_\omega^{(0)-1})^{-1}.$$

where $a = \hat{\sigma}_\omega^2\sigma^{-2}$.

Proof:

$$\begin{aligned} Q_\omega &= (\hat{V}_\omega^{-1} + \hat{K}_\omega^{(0)})^{-1}\hat{V}_\omega^{-1} - (V^{-1} + K_0)^{-1}V^{-1} \\ &= (V^{-1} + K_0)^{-1}[(V^{-1} + K_0)(\hat{V}_\omega^{-1} + \hat{K}_\omega^{(0)})^{-1}\hat{V}_\omega^{-1} - V^{-1}] \\ &= (V^{-1} + K_0)^{-1}[(V^{-1} + K_0)\hat{K}_\omega^{(0)-1}(\hat{V}_\omega + \hat{K}_\omega^{(0)-1})^{-1} - V^{-1}] \\ &= (V^{-1} + K_0)^{-1}[(V^{-1} + K_0)\hat{K}_\omega^{(0)-1} - V^{-1}(\hat{V}_\omega + \hat{K}_\omega^{(0)-1})](\hat{V}_\omega + \hat{K}_\omega^{(0)-1})^{-1} \\ &= (V^{-1} + K_0)^{-1}[K_0\hat{K}_\omega^{(0)-1} - V^{-1}\hat{V}_\omega](\hat{V}_\omega + \hat{K}_\omega^{(0)-1})^{-1} \\ &= (V^{-1} + K_0)^{-1}(\hat{\sigma}_\omega^2\sigma^{-2}\hat{V}_\omega^{-1} - V^{-1})\hat{V}_\omega(\hat{V}_\omega + \hat{K}_\omega^{(0)-1})^{-1} \end{aligned}$$

2. Prove that $\beta_\omega^{(0)} - \tilde{\beta}^{(0)} = Q_\omega\delta_\omega + (V^{-1} + K_0)^{-1}V^{-1}(\hat{\mu}_\omega - \mu)$

Proof:

$$
\begin{aligned}
&\underline{\beta}^{(0)}_{\omega} - \bar{\underline{\beta}}^{(0)} \\
&= (\hat{V}_{\omega}^{-1} + \hat{K}_{\omega}^{(0)})^{-1}(\hat{V}_{\omega}^{-1}\hat{\underline{\mu}}_{\omega} + \hat{K}^{(0)}\hat{\underline{\beta}}^{(0)}) - (V^{-1} + K_0)^{-1}(V^{-1}\underline{\mu} + K_0\hat{\underline{\beta}}^{(0)}) \\
&= [(\hat{V}_{\omega}^{-1} + \hat{K}_{\omega}^{(0)})^{-1}\hat{K}_{\omega}^{(0)} - (V^{-1} + K_0)^{-1}K_0]\hat{\underline{\beta}}^{(0)} + (\hat{V}_{\omega}^{-1} + \hat{K}_{\omega}^{(0)})^{-1}\hat{V}_{\omega}^{-1}\hat{\underline{\mu}}_{\omega} - (V^{-1} + K_0)^{-1}V^{-1}\underline{\mu} \\
&= [(V^{-1} + K_0)^{-1}K_0 - (\hat{V}_{\omega}^{-1} + \hat{K}_{\omega}^{(0)})^{-1}\hat{K}_{\omega}^{(0)}](\hat{\underline{\mu}}_{\omega} - \hat{\underline{\beta}}^{(0)}) + (\hat{V}_{\omega}^{-1} + \hat{K}_{\omega}^{(0)})^{-1}\hat{V}_{\omega}^{-1}\hat{\underline{\mu}}_{\omega} \\
&\quad - (V^{-1} + K_0)^{-1}V^{-1}\underline{\mu} - ((V^{-1} + K_0)^{-1}K_0)^{-1}K_0\hat{\underline{\mu}}_{\omega} + (\hat{V}_{\omega}^{-1} + \hat{K}_{\omega}^{(0)})^{-1}\hat{K}_{\omega}^{(0)}\hat{\underline{\mu}}_{\omega} \\
&= [(V^{-1} + K_0)^{-1}(V^{-1} + K_0 - V^{-1}) - (\hat{V}_{\omega}^{-1} + \hat{K}_{\omega}^{(0)})^{-1}(\hat{V}_{\omega}^{-1} + \hat{K}_{\omega}^{(0)} - \hat{V}_{\omega}^{-1})](\hat{\underline{\mu}}_{\omega} - \hat{\underline{\beta}}^{(0)}) \\
&\qquad\qquad + \hat{\underline{\mu}}_{\omega} - (V^{-1} + K_0)^{-1}K_0\hat{\underline{\mu}}_{\omega} - (V^{-1} + K_0)^{-1}V^{-1}\underline{\mu} \\
&= [(\hat{V}_{\omega}^{-1} + \hat{K}_{\omega}^{(0)})^{-1}\hat{V}_{\omega}^{-1} - (V^{-1} + K_0)^{-1}V^{-1}](\hat{\underline{\mu}}_{\omega} - \hat{\underline{\beta}}^{(0)}) + (V^{-1} + K_0)^{-1}V^{-1}(\hat{\underline{\mu}}_{\omega} - \underline{\mu}) \\
&= Q_{\omega}\underline{\delta}_{\omega} + (V^{-1} + K_0)^{-1}V^{-1}(\hat{\underline{\mu}}_{\omega} - \underline{\mu})
\end{aligned}
$$

Chapter 7 Predictions Using The Empirical Bayes Estimator and The Empirical Linear Bayes Estimator

7.1 Introduction

Since economic time series data are often analyzed using autoregressive models, the methods employed to estimate the parameters of these models is of importance. Using the estimated model to predict is often the purpose of observing time series data. Of course, the predictions could be required for one period of time in the future or for many periods ahead. However, estimated models which exhibit certain properties for predicting one period ahead may not exhibit these same properties for multiperiod predictions. Therefore, when estimating the model, it is reasonable to consider the number of periods ahead for which the prediction is desired.

In the context of a classical regression model, a prediction is a estimate of a future value of the dependent variable conditional on the corresponding future values of the independent variables. Thus, having estimated the parameters of the model using the observations $y_1, \ldots, y_T$, the problem is to estimate y_{T+l} given $\mathbf{x}_{T+l}$ for $l = 1, 2, \ldots$.

If the model obeys the classical assumptions, the OLS (ordinary least squares) estimator, $\hat{\beta}$, is the BLUE of β . An obvious predictor is therefore

$$\hat{y}_{T+l|T} = \mathbf{x}'_{T+l}\hat{\beta} \quad l = 1, 2, \dots .$$

The estimator $\tilde{y}_{T+l|T}$ is sometimes known as the best linear unbiased predictor (BLUP) of y_{T+l}.

In time series analysis, the main reason for building an ARIMA time series model is to make predictions of future observations. For example, consider the ARMA(1,1) process,

$$y_t = \varphi y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1} \quad t = 1, \dots T$$

where the ε_t are i i d with zero mean and finite variance σ^2 , then the BLUP of y_{T+1} is

$$\tilde{y}_{T+1|T} = \hat{\varphi} y_T + \hat{\theta}\hat{\varepsilon}_T \tag{7.1}$$

where $\hat{\varphi}$ and $\hat{\theta}$ are ML estimator and $\hat{\varepsilon}_T$ is the residual at t = T. Further predictions are built up from recursive equation

$$\tilde{y}_{T+l|T} = \hat{\varphi}\tilde{y}_{T+l-1|T}, \quad l = 2, 3, \dots$$

with formula (7.1) providing the starting value.

In this chapter we will describe the predictions of various dynamic regression model having been presented in previous chapters. Moreover, their Bayes predictions and linear Bayes predictions with the corresponding predictions using EBE and ELBE are discussed.

7.2 Recursive Prediction

Consider the classical multiple regression model

$$y_t = \mathbf{x}'_t\underline{\beta} + u_t, \quad t = 1, \ldots, T \tag{7.2}$$

with the ARMA disturbance

$$u_t = \varphi_1 u_{t-1} + \cdots + \varphi_p u_{t-p} = \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q}$$

$$(i.e. u_t = \frac{\psi(L)}{\varphi(L)}\varepsilon_t) \quad \text{where} \quad \underline{\varepsilon} \sim WS(0, \sigma^2 I).$$

Given future values of the explanatory values, the dependent variable in a regression model (7.2) with disturbances may be predicted recursively. Multiplying both sides of (7.2) by AR(p) polynomial $\varphi(L)$ yields

$$\begin{aligned} y_t = \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + (x_t - \varphi_1 x_{t-1} - \cdots - \varphi_p x_{t-p})'\underline{\beta} \\ + \varepsilon_t + \theta_t\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q}. \end{aligned} \tag{7.3}$$

This equation forms the basis for the recursive calculation of the optimal predictors of future values of y_t.

[1] When $\underline{\beta}$, $\underline{\varphi}$ and $\underline{\theta}$ are known

In this case we obtain the predictor of y_{T+l} by setting future disturbances equal to their expected value of zero :

$$\begin{aligned} \tilde{y}_{T+l|T} &= \varphi_1 \tilde{y}_{T+l-1|T} + \cdots \varphi_p \tilde{y}_{T+l-p|T} \\ &+ (x_{T+l} - \varphi_1 x_{T+l-1} - \cdots - \varphi_p x_{T+l-p})'\underline{\beta}, \quad l = 1, 2, \ldots \\ &+ \tilde{\varepsilon}_{T+l|T} + \cdots + \theta_q \tilde{\varepsilon}_{T+l-q|T} \end{aligned} \tag{7.4}$$

where

$$\tilde{y}_{T+j|T} = \tilde{y}_{T+j|T} \text{ for } j > 0$$
$$= y_{T+j} \quad j \leq 0$$

$$\tilde{\varepsilon}_{T+j|T} = 0 \quad \text{for} j > 0$$
$$= \varepsilon_{T+j} \quad j \leq 0$$

and

$$MSE\,(\tilde{y}_{T+l|T}) = E\,(\tilde{y}_{T+l|T} - y_{T+l})^2 = (1 + \psi_1^2 + \cdots + \psi_{l-1}^2)\sigma^2$$

where ψ_j is the coefficient of ε_{t-j} in the MA representation of the ARMA process.

Since any stationary ARMA process may be expressed as an infinite moving average,

$$y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + \sum_{j=0}^{\infty} \psi_j X_{t-j} \tag{7.5}$$

and any predictor linear in the observations must also be linear in the disturbance, i.e.

$$\hat{y}_{T+l|T} = \sum_{j=0}^{\infty} \psi_j X_{T+l-j} + \psi_l^* \varepsilon_T + \psi_{l+1}^* \varepsilon_{T-1} + \cdots \qquad l = 1, 2, \ldots \tag{7.6}$$

as futures of ε_t are unknown. Since

$$\begin{aligned} y_{T+l} - \hat{y}_{T+l|T} = \varepsilon_{T+l} + \psi_1 \varepsilon_{T+l-1} + \cdots + \psi_{l-1} \varepsilon_{T+1} \\ + (\psi_l - \psi_l^*)\varepsilon_T + (\psi_{l+1} - \psi_{l+1}^*)\varepsilon_{T-1} + \cdots \end{aligned} \tag{7.7}$$

has zero expectation, $\hat{y}_{T+l|T}$ is unconditionally unbiased (u - unbiased) and its MSE is given by

$$MSE\,(\hat{y}_{T+l|T}) = \sigma^2(1 + \psi_1^2 + ... + \psi_{l-1}^2) + \sigma^2 \sum_{j=0}^{\infty} (\psi_{l+j} - \psi_{l+j}^*)^2. \tag{7.8}$$

This is minimized by setting $\varphi_{l+j}^* = \psi_{l+j}$. Thus the MMSLE of y_{T+l} is therefore

$$\tilde{y}_{T+l|T} = \sum_{j=0}^{\infty} \psi_{l+j}\varepsilon_{T-j} + \sum_{j=0}^{\infty} \psi_j \mathbf{x}_{T+l-j} \tag{7.9}$$

and

$$MSE(\tilde{y}_{T+l|T}) = (1 + \psi_1^2 + \cdots + \psi_{l-1}^2)\sigma^2. \tag{7.10}$$

For the AR(1) model it follows that

$$\begin{aligned}
MSE\,(\tilde{y}_{T+l|T}) &= E\,(\tilde{y}_{T+l|T} - y_{T+l})^2 \\
&= E\begin{bmatrix} [\varphi^l y_T + (y_{T+l} + \varphi x_{T+l-1} + \cdots + \varphi^{l-1} x_{T+1})'\underline{\beta} \\ - \varphi^l y_T - (x_{T+l} + \varphi x_{T+l-1} + \cdots + \varphi^{l-1} x_{T+1})'\underline{\beta} \\ - \varepsilon_{T+l} - \varphi\varepsilon_{T+l-1} - \cdots - \varphi^{l-1}\varepsilon_{T+1}]^2 \end{bmatrix} \\
&= E\,(\varepsilon_{T+l} + \varphi\varepsilon_{T+l-1} + \cdots + \varphi^{l-1}\varepsilon_{T+1})^2 \\
&= \sigma^2 + \varphi^2\sigma^2 + \cdots + \varphi^{2(l-1)}\sigma^2 \\
&= \sigma^2 \frac{1 - \varphi^{2l}}{1 - \varphi^2}.
\end{aligned} \tag{7.11}$$

[2] When $\underline{\beta}$ is unknown, but $\underline{\varphi}$ and $\underline{\theta}$ are known

In this case, predictions are computed from the recursion

$$\begin{aligned}
\tilde{y}_{T+l|T} &= \varphi\tilde{y}_{T+l-1|T} + (x_{T+l} - \cdots - \varphi_p x_{T+l-p})'\hat{\underline{\beta}} \\
&\quad + \tilde{\varepsilon}_{T+l} + \cdots + \theta_q \tilde{\varepsilon}_{T+l-q|T}, \qquad l = 1, 2, \dots .
\end{aligned} \tag{7.12}$$

where $\hat{\underline{\beta}}$ is the generalized least squares estimator of $\underline{\beta}$ and $\tilde{y}_{T|T} = y_T$. Thus when $\underline{\beta}$ is unknown, the predictions in the regression model are subject to an additional source of variation as compared with the pure time series case. This causes from the estimation of $\underline{\beta}$. The effect may be illustrated for AR(1) disturbances. Here, predictions are computed from the recursion

$$\tilde{y}_{T+l|T} = \varphi\tilde{y}_{T+l-1|T} + (\mathbf{x}_{T+l} - \varphi\mathbf{x}_{T+l-1})'\hat{\underline{\beta}}, \quad l = 1, 2, \dots . \tag{7.13}$$

Repeated substitution for predicted values of y_t gives

$$
\begin{aligned}
\tilde{y}_{T+l|T} &= (\mathbf{x}_{T+l} - \varphi^l \mathbf{x}_T)'\hat{\underline{\beta}} + \varphi^l y_T \\
&= \mathbf{x}'_{T+l}\hat{\beta} + \varphi^l \tilde{u}_T
\end{aligned}
\tag{7.14}
$$

where $\tilde{u}_T = y_T - \mathbf{x}'_T\hat{\underline{\beta}}$. The last expression shows how the predictor of $\tilde{y}_{T+l}$ is made up of two components , one the BLUE of $E\,(y_{T+l})$, and the other the BLUP of $\tilde{u}_{T+l}$.

To show the computation of the MSE of the prediction error, we still use AR(1) disturbance for simplicity. The true value of y_{T+l} is given by

$$
\begin{aligned}
y_{T+l} &= \mathbf{x}'_{T+l}\underline{\beta} + u_{T+l} \\
&= \mathbf{x}'_{T+l}\underline{\beta} + \sum_{j=0}^{l-1}\varphi^j\varepsilon_{T+l-j} + \varphi^l(y_T - \mathbf{x}'_T\underline{\beta}) \\
&= (\mathbf{x}_{T+l} - \varphi^l\mathbf{x}_T)'\underline{\beta} + \sum_{j=0}^{l-1}\varphi^j\varepsilon_{T+l-j} + \varphi^l y_T
\end{aligned}
\tag{7.15}
$$

Subtracting (7.15) from (7.13) gives

$$
y_{T+l} - \tilde{y}_{T+l|T} = \sum_{j=0}^{l-1}\varphi^j\varepsilon_{T+l-j} + (\mathbf{x}_{T+l} - \varphi^l\mathbf{x}_T)'(\underline{\beta} - \hat{\underline{\beta}})
\tag{7.16}
$$

The two components in (7.16) are independent, and so the MSE of the prediction error consists of a contribution from the estimation error, $\underline{\beta} - \hat{\underline{\beta}}$, plus a contribution from future values of the disturbance term,i.e.

$$
\begin{aligned}
MSE\,(\tilde{y}_{T+l|T}) &= E\,(y_{T+l} - \tilde{y}_{T+l|T})^2 \\
&= \sigma^2\frac{1-\varphi^{2l}}{1-\varphi^2} + \sigma^2(\mathbf{x}_{T+l} - \varphi^l\mathbf{x}_T)'(X'V^{-1}X)^{-1}(\mathbf{x}_{T+l} - \varphi^l\mathbf{x}_T)
\end{aligned}
\tag{7.17}
$$

where

$$V = \begin{bmatrix} 1 & \varphi & \dots & \varphi^{T-1} \\ \varphi & 1 & \dots & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \varphi^{T-1} & \dots & \varphi & 1 \end{bmatrix}.$$

[3] When $\underline{\beta}$, $\underline{\varphi}$ and $\underline{\theta}$ are all unknown

In this case, predictions are computed from the recursion

$$\begin{aligned} \tilde{y}^{*}_{T+l|T} &= \hat{\varphi}_1 \tilde{y}^{*}_{T+l-1|T} + \dots + \hat{\varphi}_p \tilde{y}^{*}_{T+l-p|T} \\ &+ (\mathbf{x}_{T+l} - \hat{\varphi}_1 \mathbf{x}_{T+l-1} - \dots - \hat{\varphi}_p \mathbf{x}_{T+l-p})'\hat{\underline{\beta}} \\ &+ \tilde{\varepsilon}_{T+l|T} + \hat{\theta}\tilde{\varepsilon}_{T+l-1|T} + \dots + \hat{\theta}_q \tilde{\varepsilon}_{T+l-q|T}, \quad l = 1, 2, \dots . \end{aligned} \tag{7.18}$$

where $\hat{\varphi}_j$ and $\hat{\theta}_j$ are consistent estimators of φ and θ respectively, while $\hat{\beta}$ is GLSE as in [2], and $\tilde{y}^{*}_{T|T} = y_T$.

In practice, predictions are almost invariably made with $\underline{\beta}$, $\underline{\varphi}$ and $\underline{\theta}$ replaced by their estimates. This, of course, creates an additional source of variability, which should ideally be incorporated in the expression for the prediction MSE. In order to show how φ causes the problem, let us consider the AR(1) process.

When φ is known, the MMSLE for l periods ahead is given by

$$\tilde{y}_{T+l|T} = \varphi^l y_T. \tag{7.19}$$

When φ is unknown it will be replaced by its ML estimator, $\tilde{\varphi}$ or by an estimator which is asymptotically equivalent. The actual predictor is therefore

$$\tilde{y}^*_{T+l|T} = \tilde{\varphi}^l y_T. \tag{7.20}$$

In more general cases, $\tilde{y}^*_{T+l|T}$ can be computed by a difference equation having exactly the same form as (7.18).

The prediction error for (7.19) may be decomposed into two parts by writing

$$y_{T+l} - \tilde{y}^*_{T+l|T} = (y_{T+l} - \tilde{y}_{T+l|T}) + (\tilde{y}_{T+l|T} - \tilde{y}^*_{T+l|T}) \tag{7.21}$$

The first term on the right-hand side of (7.21) is the prediction error when φ is known, while the second term represents the error arising from the estimation of φ . This decomposition is appropriate for all regression models with ARMA disturbances. In the present case it specializes in view of (7.20) to

$$y_{T+l} - \tilde{y}^*_{T+l|T} = (y_{T+l} - \tilde{y}_{T+l|T}) + (\dot{\varphi}^l - \tilde{\varphi}^l)y_T. \tag{7.22}$$

Now consider the one-step ahead predictor. The MSE may be written as

$$MSE\,(\tilde{y}^*_{T+1|T}) \doteq MSE\,(\tilde{y}_{T+1|T}) + y_T^2 E\,(\tilde{\varphi} - \varphi)^2. \tag{7.23}$$

In formulating the contribution of the estimation error to (7.23), y_T is taken as fixed, whereas $\tilde{\varphi}$ is a random variable. This may appear to be contradictory, as y_T is actually used to construct $\tilde{\varphi}$. However, (7.23) provides a sensible definition of MSE in this context, since any prediction is always made conditional on sample observations being known. Replacing $E\,(\hat{\varphi} - \varphi)^2$ by its asymptotic variance gives an approximation to the mean square error, i.e.

$$MSE\,(\tilde{y}^*_{T+l|T}) \cong \sigma^2 + y_T^2(1 - \varphi^2)/T. \tag{7.24}$$

The AR(1) model is often estimated by OLS. Applying the usual regression formula for estimating the MSE of $\tilde{y}^*_{T+1|T}$ gives

$$mse(\tilde{y}^*_{T+l|T}) = s^2(1 + y_T^2 / \sum_{t=2}^{T} y_T^2). \tag{7.25}$$

Note that

$$\hat{\varphi} = \varphi + \frac{\Sigma \varepsilon_t y_{t-1}}{\Sigma y_{t-1}^2}$$

implies that

$$E(\hat{\varphi} - \varphi)^2 = \frac{\Sigma y_{t-1}^2}{(\Sigma y_{t-1}^2)^2} \sigma^2.$$

This is closely related to (7.24) since

$$\sum_{t=2}^{T} y_{t-1}^2 \cong T\sigma^2/(1 - \varphi^2) \quad (= E(y_t^2))$$

in large samples.

When $l \geq 1$,the last term in (7.23) is $y_T^2 E(\tilde{\varphi}^l - \varphi^l)^2$. Writing

$$\varphi^l - \tilde{\varphi}^l = \varphi^l - \varphi^l[1 - \frac{(\varphi - \tilde{\varphi})}{\varphi}]^l$$

and expanding the term in square brackets yields

$$\varphi^l - \tilde{\varphi}^l \cong l\varphi^{l-1}(\varphi - \tilde{\varphi})$$

when higher order terms are ignored. Therefore,

$$E(\tilde{\varphi}^l - \varphi^l)^2 \cong l^2\varphi^{2(l-1)}E(\tilde{\varphi} - \varphi)^2.$$

Together with the result in (7.17) this gives

$$
\begin{aligned}
MSE(\tilde{y}^{*}_{T+l|T}) \cong{}& \sigma^2 \frac{1-\varphi^{2l}}{1-\varphi^2} \\
&+ \sigma^2(\mathbf{x}_{T+l} - \varphi^l \mathbf{x}_T)'(X'V^{-1}X)^{-1}(\mathbf{x}_{T+l} - \varphi^l \mathbf{x}_T) \\
&+ y_T^2 \frac{(1-\varphi^2)l^2\varphi^{2(l-1)}}{T}.
\end{aligned} \tag{7.26}
$$

Expression (7.26) is an approximation to the MSE of the multistep predictor for a particular sample. In order to get some idea of the MSE of such a predictor on the average, y_T^2 is replaced by its expected value $\sigma^2/(1-\varphi^2)$. The resulting expression is known as the asymptotic mean square error (AMSE). Thus

$$
\begin{aligned}
AMSE(\tilde{y}^{*}_{T+l|T}) \cong{}& \sigma^2 \frac{1-\varphi^{2l}}{1-\varphi^2} \\
&+ \sigma^2(\mathbf{x}_{T+l} - \varphi l \mathbf{x}_T)'(X'V^{-1}X)^{-1}(\mathbf{x}_{T+l} - \varphi^l \mathbf{x}_T) \\
&+ \frac{\sigma^2}{T} l^2 \varphi^{2(l-1)}.
\end{aligned} \tag{7.27}
$$

For the special case of $l = 1$,

$$
AMSE(\tilde{y}^{*}_{T+l|T}) = \sigma^2(1 + (\mathbf{x}_{T+1} - \varphi \mathbf{x}_T)'(X'V^{-1}X)^{-1}(\mathbf{x}_{T+1} - \varphi \mathbf{x}_T) + T^{-1}). \tag{7.28}
$$

In both (7.27) and (7.28), the contribution arising from the error in estimating φ is a term of $O(T^{-1})$. For more details see Baillie (1979).

Optimal predictions of future values of y_t in a transfer function model

$$
y_t = \frac{B(L)}{A(L)}\mathbf{x}_{t-v} + \frac{\theta(L)}{\varphi(L)}\varepsilon_t \tag{7.29}
$$

may be computed recursively. The required expression is obtained by multiplying (7.29) by $A(L)\varphi(L)$ to yield

$$
\varphi(L)A(L)y_t = \varphi(L)B(L)\mathbf{x}_{t-v} + A(L)\theta(L)\varepsilon_t \tag{7.30}
$$

and rearranging so that only y_t appears on the left-hand side.

7.3 The Bayes Predictor and The Empirical Bayes Predictor

Consider the simple dynamic linear regression model

$$y_t = \alpha y_{t-1} + \beta x_t + u_t \tag{7.31}$$

where $u_t \sim NID(0, \sigma^2)$ and $|\alpha| < 1$. The initial value y_0 is treated as fixed. The parameters are the coefficient $\pi' = (\alpha, \beta)$, and varience σ^2 of the disturbances. The problem is that, having observed $\mathbf{y}' = (y_1, \ldots, y_T)$, one wishes to predict y_{T+l}. For the predictor $\hat{y}_{T+l} = \hat{y}_{T+l}(\mathbf{y})$ of y_{T+l}, assume the risk function to be

$$\begin{aligned} R(\pi, \sigma^2) &= E(\hat{y}_{T+l} - y_{T+l})^2 \\ &= \int(\hat{y}_{T+l} - y_{T+l})^2 p(\mathbf{y}, y_{T+l} \mid \pi, \sigma^2) d\mathbf{y} dy_{T+l}. \end{aligned} \tag{7.32}$$

The Bayesian approach is to choose the prediction function $\hat{y}_{T+l}(\mathbf{y})$ such that the expected risk is minimum, given a prior density $p(\pi, \sigma^2)$ of the parameters.

The derivation will be simplified considerably if we rewrite the risk (7.32) as an expectation taken over the future disturbances $u_{T+1}, \ldots, u_{T+l}$, rather than the future y_{T+l} . By the model (7.31), and on repeated substitutions of u's for y's, we have

$$\begin{aligned} y_{T+l} = \alpha^l y_T &+ (\alpha^{l-1}\beta x_{T+1} + \alpha^{l-2}\beta x_{T+2} + \cdots + \beta x_{T+l}) \\ &+ (\alpha^{l-1} u_{T+1} + \cdots + u_{T+l}) \end{aligned} \tag{7.33}$$

Then, noting the fact that the future u's are uncorrelated with the observed $\mathbf{y}$, or with the parameters, we can write the risk (7.32) as

$$
\begin{aligned}
& E\,(\hat{y}_{T+l} - y_{T+l})^2 \\
& = E[\hat{y}_{T+l} - \alpha^l y_T - (\alpha^{l-1}\beta x_{T+1} + \cdots + \beta x_{T+l})]^2 + E(\alpha^{l-1}u_{T+1} + \cdots + u_{T+l})^2
\end{aligned} \tag{7.34}
$$

Since the prediction function $\hat{y}_{T+l}(\mathrm{y})$ affects the first term on the right hand side of (7.34), the Bayesian approach amounts to minimizing, with respect to $\hat{y}_{T+l}$, the expectation of this first term, taken over the prior density $p(\pi, \sigma^2)$:

$$
\int E\,[\hat{y}_{T+l} - \alpha^l y_T - (\alpha^{l-1}\beta x_{T+1} + \cdots + \beta x_{T+l})]^2 d\pi d\sigma^2
$$

$$
= \int\int [\hat{y}_{T+l} - \alpha^l y_T - (\alpha^{l-1}\beta x_{T+1} + \cdots + \beta x_{T+l})]^2 p(\mathrm{y} \mid \pi, \sigma^2) d\mathrm{y} \times p(\pi, \sigma^2) d\pi d\sigma^2. \tag{7.35}
$$

The above expression, treated as an integral over y, will be minimized by minimizing the integrand

$$
\int [\hat{y}_{T+l} - \alpha^l y_T - (\alpha^{l-1}\beta x_{T+1} + \cdots + \beta x_{T+l})]^2 p(\mathrm{y} \mid \pi, \sigma^2) p(\pi, \sigma^2) d\pi d\sigma^2. \tag{7.36}
$$

Setting the derivative of (7.36) with respect to $\hat{y}_{T+l}$ equal to zero, we obtain

$$
\begin{aligned}
\hat{y}_{T+l} &= \frac{1}{\int p(\mathrm{y} \mid \pi, \sigma^2) d\pi d\sigma^2} \int [\alpha^l y_T + (\alpha^{l-1}\beta x_{T+1} + \cdots + \beta x_{T+l})] p(\mathrm{y} \mid \pi, \sigma^2) p(\pi, \sigma^2) d\pi d\sigma^2 \\
&= [\textstyle\int \alpha^l p(\pi \mid \mathrm{y}) d\pi] y_T + [\textstyle\int \alpha^{l-1}\beta p(\pi \mid \mathrm{y}) d\pi] x_{T+1} + \cdots + [\textstyle\int \beta p(\pi \mid \mathrm{y}) d\pi] x_{T+l} \\
&= E(\alpha^l \mid \mathrm{y}) y_T + E(\alpha^{l-1}\beta \mid \mathrm{y}) x_{T+1} + \cdots + E(\beta \mid \mathrm{y}) x_{T+l}.
\end{aligned} \tag{7.37}
$$

The predictor (7.37) could have been obtained by appealing directly to the well-known result in Bayesian statistics that for a squared loss function, the mean of y_{T+l} given y_T , i.e. $\alpha^l y_T + (\alpha^{l-1}\beta x_{T+1} + \cdots + \beta x_{T+l})$ evaluated by the posterior density $p(\alpha, \beta \mid \mathrm{y}) = p(\pi \mid \mathrm{y})$ is optimal. Thus the optimal predictor $\hat{y}_{T+l}$ consists of two parts, the first being the posterior mean of α^l times the initial value y_T, and the second being the inner product of posterior mean vector of $(\alpha^{l-1}\beta, \ldots, \beta)$ and $(x_{T+1}, \ldots, x_{T+l})$.

If we replace Bayesian estimators in (7.37) by their corresponding empirical Bayes estimators described in chapter 3, we obtain the empirical Bayes predictor $\tilde{y}^*_{T+l}$ of y_{T+l} given y_T ,i.e.

$$\tilde{y}^*_{T+l} = (\alpha_n^l)^* y_T + x_{T+1}(\alpha_n^{l-1}\beta_n)^* + \cdots + x_{T+l}(\beta_n)^* \tag{7.38}$$

where $(\alpha_n^{l-i}\beta_n)^*$ denotes the EBE of $(\alpha_n^{l-i}\beta_n)$ with n past experiences, and $i = 0, 1, \ldots, l$. Note that in general $(\alpha_n^l)^* \neq (\alpha_n^*)^l$ for $l > 1$.

From the result derived above, we conclude that the prediction of y_{T+l} using the EB approach is obtained by just substituting each coefficient in (7.33) by its corresponding EBE and letting $u_{T+l} = 0$, for $l = 1, 2, \ldots$.

When the disturbances are autocorrelated, the above result can be generalized in a straightforward way. For simplicity, let us consider the model with AR(1) disturbance

$$\begin{aligned}(1 - \alpha L)y_t &= x_t\gamma + u_t \\ (1 - \rho L)u_t &= \varepsilon_t, \ \ \varepsilon \sim WS(0, \sigma^2)\end{aligned}$$

Following Baillie's method (1979, p.181), the partial moving average form of the model is given by

$$y_{T+l} = \sum_{j=0}^{l-1} \delta_j \varepsilon_{T+l-j} + \mathbf{q}(l)' X_{T+l}\gamma + \xi(l)'\mathbf{y}_T \tag{7.39}$$

where

$$\begin{aligned}
X'_{T+l} &= [x_{T+l}, \ldots, x_T], \ \mathbf{y}_T = [y_T, y_{T-1}] \\
\delta_j &= \frac{\alpha^{j+1} - \rho^{j+1}}{\alpha - \rho} \\
\mathbf{q}(l)' &= [q_0, \ldots, q_l], \ \text{where} \ \begin{aligned} q_j &= \alpha^j, j = 0, 1, \ldots, l - 1 \\ &= -\rho\frac{\alpha^j - \rho^j}{\alpha - \rho}, j = l \end{aligned} \\
\xi(l)' &= (\frac{\alpha^{l+1} - \rho^{l+1}}{\alpha - \rho}, -\alpha\rho\frac{\alpha^l - \rho^l}{\alpha - \rho}).
\end{aligned}$$

Treating (7.39) as (7.33), and following the exact same procedure leads to the Bayes prediction using EBE in the lagged dependent variable model with autocorrelated disturbances.

From (7.37), for l = 1, and by repeated substitution,

$$\begin{aligned}
\hat{y}_{T+1|T} &= E(\alpha \mid \mathrm{y})y_T + E(\beta \mid \mathrm{y})x_{T+1} + 0 \\
\hat{y}_{T+2|T} &= E(\alpha \mid \mathrm{y})\hat{y}_{T+1|T} + E(\beta \mid \mathrm{y})x_{T+2} + 0 \\
&\vdots \\
\hat{y}_{T+l|T} &= E(\alpha \mid \mathrm{y})\hat{y}_{T+l-1|T} + E(\beta \mid \mathrm{y})x_{T+l} + 0.
\end{aligned}$$

Thus

$$\hat{y}_{T+l|T} = [E(\alpha \mid \mathrm{y})]^l y_T + E(\beta \mid \mathrm{y})[x_{T+l} + E(\alpha \mid \mathrm{y})x_{T+l-1} + \cdots + E(\alpha \mid \mathrm{y})^{l-1}x_{T+1}] \quad (7.40)$$

is called *the recursive Bayes predictor* of y_{T+l}. Thus when each posterior mean is replaced by its EBE , we obtain the predictor of y_{T+l} using EBE . Moreover, it is easy to show that the resulting predictor is consistent to the linear Bayes predictor (LBP). Note that as the sample size T approaches infinity, the difference between the predictors (7.38) and (7.40) can be shown to be $O(T^{-1})$. This follows from the fact that the asymptotic variance is $O(T^{-1})$. In most time series models, the predictions are built up recursively.

7.4 The Best Linear Unbiased Predictor When the Disturbances of the Model are Serially Correlated

Consider the multiple linear regression model

$$y_t = \mathbf{x}'_t\underline{\beta} + u_t, \quad t = 1,\ldots,T \quad (7.41)$$

or

$$\mathbf{y} = X\underline{\beta} + \mathbf{u}$$

where

$$\mathbf{u} = (u_1, \dots, u_T)' \sim WS(0, V), \quad V \neq I, \quad V > \mathbf{0}.$$

$$i.e. \;\; Eu_su_t \neq 0 \quad \text{for any} \quad s,t = 1,\dots,T.$$

Let $\tilde{y}_{T+l|T}$ be the best linear unbiased predictor (BLUP) of y_{T+l}, then

$$\tilde{y}_{T+l|T} = \mathbf{c}'\mathbf{y}, \tag{7.42}$$

where $\mathbf{c}$ is $T \times 1$ constant vector. Because of the unbiasedness of $\tilde{y}_{T+l|T}$,

$$E\,(y_{T+l} - \tilde{y}_{T+l|T}) = 0$$

$$i.e. \qquad E\,(\mathbf{x}'_{T+l}\beta + u_{T+l} - \mathbf{c}'(X\underline{\beta} + \mathbf{u})) = 0,$$

which implies that

$$\mathbf{x}'_{T+l} - \mathbf{c}'X = 0. \tag{7.43}$$

Moreover, to minimize the MSE $E(y_{T+l} - \tilde{y}_{T+l|T})^2$ under (7.43) , we have to minimize

$$E\,(u_{T+l} - \mathbf{c}'\mathbf{u})^2 + 2\underline{\lambda}(\mathbf{x}'_{T+l} - \mathbf{c}'X)$$

where $\underline{\lambda}$ is the Lagrange's multiplier. After simplifying this yields

$$\sigma^2_{T+l} + \mathbf{c}'V\mathbf{c} - 2\mathbf{c}'\mathbf{w} + 2\underline{\lambda}(\mathbf{x}'_{T+l} - \mathbf{c}'X)$$

where $\mathbf{w} = E(u_{T+l}\mathbf{u})$ and $E\,u^2_{T+l} = \sigma^2_{T+l}$. Taking derivatives with respect to $\mathbf{c}$ and $\underline{\lambda}$ and setting them equal zero yields

$$2V\hat{\mathbf{c}} - 2\mathbf{w} + 2\hat{\underline{\lambda}}X = 0$$
$$\mathbf{x}'_{T+l} - \hat{\mathbf{c}}'X = 0$$

or

$$\begin{bmatrix} V & X \\ X' & 0 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{c}} \\ \hat{\underline{\lambda}} \end{bmatrix} = \begin{bmatrix} \mathbf{w} \\ \mathbf{x}'_{T+l} \end{bmatrix}.$$

Thus

$$\begin{bmatrix} \hat{\mathbf{c}} \\ \hat{\underline{\lambda}} \end{bmatrix} = \begin{bmatrix} V & X \\ X' & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{w} \\ \mathbf{x}'_{T+l} \end{bmatrix},$$

from which we have

$$\hat{\mathbf{c}} = V^{-1}[I - X(X'V^{-1}X)^{-1}X'V^{-1}]\mathbf{w} + V^{-1}X(X'V^{-1}X)^{-1}\mathbf{x}_{T+l}.$$

Thus the BLUP of y_{T+l} is given by

$$\tilde{y}_{T+l|T} = \hat{\mathbf{c}}'\mathbf{y} = \mathbf{x}'_{T+l}\hat{\underline{\beta}} + \mathbf{w}'V^{-1}(\mathbf{y} - X\hat{\underline{\beta}}). \tag{7.44}$$

Its MSE is

$$\begin{aligned}
MSE(\tilde{y}_{T+l|T}) &= E(y_{T+l} - \tilde{y}_{T+l|T})^2 = \sigma^2_{T+l} + \hat{\mathbf{c}}'V\hat{\mathbf{c}} - 2\hat{\mathbf{c}}'\mathbf{w} \\
&= \sigma^2_{T+l} + \mathbf{x}'_{T+l}(X'V^{-1}X)^{-1}\mathbf{x}_{T+l} + \mathbf{w}'(V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1})\mathbf{w} \\
&\qquad - 2[\mathbf{x}'_{T+l}(X'V^{-1}X)^{-1}X'V^{-1}\mathbf{w} + \mathbf{w}'\{I - V^{-1}X(X'V^{-1}X)^{-1}X'\}V^{-1}\mathbf{w}] \\
&= \sigma^2_{T+l} + \mathbf{x}'_{T+l}(X'V^{-1}X)^{-1}\mathbf{x}_{T+l} - 2\mathbf{x}'_{T+l}(X'V^{-1}X)^{-1}X'V^{-1}\mathbf{w} \\
&\qquad - \mathbf{w}'[I - V^{-1}X(X'V^{-1}X)^{-1}X']V^{-1}\mathbf{w} \\
&= \sigma^2_{T+l} + \mathbf{x}'_{T+l}(X'V^{-1}X)^{-1}\mathbf{x}_{T+l} - 2\mathbf{x}'_{T+l}(X'V^{-1}X)^{-1}X'V^{-1}\mathbf{w} \\
&\qquad - \mathbf{w}'V^{-1}\mathbf{w} + \mathbf{w}'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}\mathbf{w} \\
&= \sigma^2_{T+l} + (\mathbf{x}_{T+l} - X'V^{-1}\mathbf{w})'(X'V^{-1}X)^{-1}(\mathbf{x}_{T+l} - X'V^{-1}\mathbf{w}) - \mathbf{w}'V^{-1}\mathbf{w}.
\end{aligned}$$

7.5 The Empirical Linear Bayes Predictor (ELBP)

Consider the multiple linear regression model

$$\mathbf{y} = X\underline{\beta} + \mathbf{u}, \quad \mathbf{u} \sim WS(0, \sigma^2\Sigma) \tag{7.45}$$

$$\underline{\beta} \sim WS(\underline{\mu}, V) \quad \text{i.e. } \underline{\beta} = \underline{\mu} + \underline{\nu}, \underline{\nu} \sim WS(0, V)$$

where $\mathbf{u}$, and $\underline{\nu}$ are uncorrelated. Given $\mathbf{y} = (y_1, \ldots, y_T)'$ and X, let

$$\hat{y}_{T+l|T} = a + \mathbf{c}'\mathbf{y} \tag{7.46}$$

be a linear Bayes predictor (LBP) for y_{T+l}. Then

$$\begin{aligned}
\text{MSE}\,(\hat{y}_{T+l|T}) &= E\,(\hat{y}_{T+l|T} - y_{T+l})^2 \\
&= E\,[a + (\mathbf{c}'X - \mathbf{x}'_{T+l})\underline{\beta} + \mathbf{c}'\mathbf{u} - u_{T+l}]^2 \\
&= E\,[a + (\mathbf{c}'X - \mathbf{x}'_{T+l})\underline{\mu} + (\mathbf{c}'X - \mathbf{x}'_{T+l})\underline{\nu} + \mathbf{c}'\mathbf{u} - u_{T+l}]^2 \\
&= [a + (\mathbf{c}'X - \mathbf{x}'_{T+l})\underline{\mu}]^2 + (\mathbf{c}'X - \mathbf{x}'_{T+l})V(\mathbf{c}'X - \mathbf{x}'_{T+l})' \\
&\qquad + \sigma^2\mathbf{c}'\Sigma\mathbf{c} + \sigma^2_{T+l} - 2\mathbf{c}'\mathbf{w}
\end{aligned} \tag{7.47}$$

where $\mathbf{w} = \text{cov}(\mathbf{u}, u_{T+l})$, $\sigma^2_{T+l} = \text{Var}\, u_{T+l}$. Using the Chipman's method (1964) to minimize MSE ($\hat{y}_{T+l|T}$) yields

$$\hat{a} = (\mathbf{x}'_{T+l} - \mathbf{c}'X)\underline{\mu}. \tag{7.48}$$

And from

$$\frac{\partial \text{MSE}\,(\hat{y}_{T+l|T})}{\partial \mathbf{c}} = 2\sigma^2\Sigma\mathbf{c} - 2\mathbf{w} + 2XV(X'\mathbf{c} - \mathbf{x}_{T+l}) = 0,$$

we have

$$\hat{\mathbf{c}} = (\sigma^2\Sigma + XVX')^{-1}(\mathbf{w} + XV\mathbf{x}_{T+l}) \tag{7.49}$$

Thus

$$\begin{aligned} \tilde{y}_{T+l|T} &= \mathbf{x}'_{T+l}\underline{\mu} + (\mathbf{w}' + \mathbf{x}'_{T+l}VX')(\sigma^2\Sigma + XVX')^{-1}(\mathbf{y} - X\underline{\mu}) \\ &\overset{(*)}{=} \mathbf{x}'_{T+l}\tilde{\underline{\beta}} + \sigma^{-2}\mathbf{w}'\Sigma^{-1}(\mathbf{y} - X\tilde{\underline{\beta}}) \end{aligned} \tag{7.50}$$

where $\tilde{\underline{\beta}}$ is LBE of $\underline{\beta}$. Note that $y_{T+l|T}$ does not include σ^2 since $\mathbf{w}$ contains σ^2 too.

The equality (*) in (7.50) is gotten as follows:

$$\begin{aligned} \hat{y}_{T+l|T} &= \mathbf{x}'_{T+l}(\underline{\mu} + VX'(\sigma^2\Sigma + XVX')^{-1}(\mathbf{y} - X\underline{\mu})) \\ &\quad + \mathbf{w}'(\sigma^2\Sigma + XVX')^{-1}(\mathbf{y} - X\underline{\mu}) \\ &= \mathbf{x}'_{T+l}\tilde{\underline{\beta}} + \mathbf{w}'\sigma^{-2}\Sigma^{-1}(\mathbf{y} - X\tilde{\underline{\beta}}). \end{aligned}$$

where

$$(\sigma^2\Sigma + XVX')^{-1}(\mathbf{y} - X\underline{\mu}) = \sigma^{-2}\Sigma^{-1}(\mathbf{y} - X\tilde{\underline{\beta}})$$

The last equality follows from the following equalities. Now

$$\begin{aligned} \tilde{\underline{\beta}} &= \underline{\mu} + VX'(\sigma^2\Sigma + XVX')^{-1}(\mathbf{y} - X\underline{\mu}), \\ X\tilde{\underline{\beta}} &= X\underline{\mu} + XVX'(\sigma^2\Sigma + XVX')^{-1}(\mathbf{y} - X\underline{\mu}) \\ \mathbf{y} - X\tilde{\underline{\beta}} &= (\mathbf{y} - X\underline{\mu}) - XVX'(\sigma^2\Sigma + XVX')^{-1}(\mathbf{y} - X\underline{\mu}) \\ &= \sigma^2\Sigma(\sigma^2\Sigma + XVX')^{-1}(\mathbf{y} - X\underline{\mu}) \end{aligned}$$

which completes the result.

This shows that the LBP (7.50) for y_{T+l} coincides with the BLUP of y_{T+l} when the GLSE of $\underline{\beta}$ for the latter model is replaced by the linear Bayes estimator of $\underline{\beta}$.

If w and V are replaced by their corresponding consistent estimators (see chapter 5) and $\tilde{\beta}$ by its ELBE, we obtain the predictor of y_{T+l} using ELBE or ELBP of y_{T+l}. Moreover, it is easily shown that the resulting predictor is consistent to the LBP.

By combining the methods discussed in Section 6.3 and this section we can obtain the empirical linear Bayes predictor for y_{T+l} given $\mathbf{y}_T = (y_1, \dots, y_T)'$ in the autoregressive linear model with autocorrelated disturbances.

Chapter 8 Conclusions

I. Conclusions

In this study, we considered the EB approach to the squared error loss estimation problem in the multinormal case with varying mean and variance. Using the information contained in $(X_1,\ldots, X_n)$ from past component problems and X_0 from the present problem, we exhibited, for each n > 0, two classes of estimators $\hat{\underline{\theta}}$ and $\tilde{\underline{\theta}}$ for $\underline{\theta}$, one for the case when nothing is known about the support of the prior distribution and the other for the case when it is known that the prior distribution has a compact support. We have shown that for $\gamma > 0, \hat{\underline{\theta}}$ is a.o. $O((\log n)^{\frac{-\gamma}{2+\gamma}})$ uniformly over the class of all prior distributions satisfying certain moment conditions dependent on γ, whereas for $0 < \gamma' < 1, \tilde{\underline{\theta}}$ is shown to be a.o. $O(n^{\frac{-\gamma'}{2(1+\rho)}}(\log n)^{(k+1)\gamma'/2})$ uniformly over the class of all prior distributions with compact support. Thus, we have given procedures for constructing EB estimators asymptotically optimal in the multinormal case. It should be noted that it is no longer necessary that each component problem be identical distributed, that is, the mean and variance may vary with each component problem.

Different applications of the above results are exhibited. They not only extend Martz and Krutchkoff's results (1969), but also generalize Li and Hui's results (1983 a, b).

In contrast to the EB approach, we presented the ELB approach to the squared error loss estimation problem in wide sense distribution with mean and variance specified but unknown. Using the information contained in $(X_1, \ldots, X_N)$ from past component problems and X_0 from the present problem, we exhibited, for each N > 0, the ELB estimator $\underline{\delta}_N$ for $\underline{\theta}$. We showed that $\underline{\delta}_N$ is consistent to LB estimator $\underline{\delta}^*$ for $\underline{\theta}$ and is asymptotically optimal which satisfies certain moment conditions.

Various applications of the above results are exhibited. They not only extend Swamy's results (1971 , 1974) , but also generalize Bunke and Gladitz's results (1979) .

Predictions using EBE and ELBE are presented. Both of them are asymptotically optimal to the Bayes predictor and the linear Bayes predictor, respectively, since the corresponding EBE and ELBE are asymptotically optimal to the BE and LBE, respectively.

II. Comparisons:

1. EBE and ELBE: which is "better"?

(1) To approximate $\bar{\mathbf{t}}(X)$ (ELBE) may be simpler than to approximate $\mathbf{t}^*(X)$ (EBE), since $\bar{\mathbf{t}}(X)$ involves only the constants E (X), Var X , E ($\underline{\theta}$) and cov ($\underline{\theta}$, X) , which for some families f (x, $\underline{\theta}$) are easy to estimate from observations $\mathbf{x}_1, \ldots, \mathbf{x}_N$.

(2) For G belonging (or being close to belonging) to some special parametric family of distribution functions, $\mathbf{t}^*(X)$ is in fact a linear function of X , so that $\bar{\mathbf{t}}(X)$ and $\mathbf{t}^*(X)$ are identical.

(3) For $N \to \infty$, the EB approach, when feasible, would seem to be better than the ELB approach, but for moderate N the ELB approach with rapid convergence (as N tends infinity) may be better than the EB approach which converges slowly.

(4) For the EB and the ELB approaches, the assumptions on the vectors $\mathbf{x}_i$ can be generalized as only independent, not necessarily identically, distributed.

(5) The empirical Bayes approach to statistical decision theory is applicable when one is confronted with an independent sequence of Bayes decision problems having similar structure. The structural similarity includes the assumption that there is a prior distribution G on the parameter space, and that G is unknown. Such an empirical Bayes approach, in the words of Robbins , "offers certain advantages over any approach which ignores the fact that the parameter is itself a random variable as well as over any approach which assumes a personal or conventional distribution of the parameter not subject to change with experience"(Robbins, 1964).

(6) Monte Carlo simulations performed by Clemmer and Krutchkoff (1968), Griffin and Krutchkoff (1971), Maritz and Lwin (1975), and Martz and Krutchkoff(1969), among others, have shown, how for certain priors the EB procedures often perform at least as good and often better than the usual procedures whenever there is even only one past experience of the problem.

III. Suggestion

No simulation has been performed in this study. It is suggested that simulations be done in order to compare the results with those of the iterative MLE's.

Appendix A. References

1. Baillie, R. T. (1979) The asymptotic mean squared error of multistep prediction from the regression model with autoregressive errors. JASA 74 175-184

2. Bellach, J. (1974) Bayessche Schatzungen und Vorhersagen bei stochastischen linearen differenzengleichungssystemen. Math. Operationsforsch. Statist. 5, 599-623

3. Bellach, J. (1979) Consistency of the LSE for linear stochastic difference equations. Math. Operationsforsch. Statist. 10, 79-106

4. Billingsley, P. (1979) Probability and Measure Wiley & Sons New York

5. Bunke, H. and J. Gladitz (1974) Empirical Bayes decision rules for a sequence of linear models with different regressor matrices. Math. Operationsforsch. Statist. 5, 235-244

6. .... (1979) Empirical Bayes Approach to parameter identification in linear stochastic difference equations. Math. Operationsforsch. Statist. 10, 63-78

7. Cauculos, T. (1966) Estimation of a multivariate density. Ann. of Statist. Math. 18 179-189

8. Chipman, J. S. (1964) On least squares with insufficient observations, JASA 59 1078-1112

9. Chow, G. C. (1973) Multiperiod predictions from stochastic difference equations by Bayesian methods. Econometrica 41, 109-118

10. Clemmer, B. A. and Krutchkoff, R. G. (1968) The use of empirical Bayes in a linear regression model. Biometrika 55 525-34

11. Duncan D. B. and S. D. Horn (1972) Linear Dynamic Recursive Estimation from the view-point of regression Analysis. JASA 67 815-821

12. Fomby, T. B., R. C. Hill and S. R. Johnson (1984) Advanced Econometric Methods. Springer-Verlag (chap. 10 and chap. 11)

13. Fuller, W. A. (1976) Introduction to Statistical Time Series. John Wiley & Sons. New York. (chap. 9)

14. Goldberger, A. S. (1962) Best linear unbiased prediction in the generalized linear regression model. JASA 57 369-375

15. Good, I. J. (1965) The Estimation of Probabilities. MIT Press, Boston, Massachusetts.

16. Goodwin, G. C. and R. L. Payne (1977) Dynamic System Identification. Academic, New York (chap. 4 and 5)

17. Griffin, B. and R. G. Krutchkoff (1971) Optimal linear estimators: An empirical Bayes version with application to the binomial distribution. Biometrika, 58 195-203

18. Hartigan, J. (1969) Linear Bayes Methods J. Roy. Staist. Soc. Ser. B 31 446-54

19. Harvey, A. C. (1981 a) The Econometric Analysis of Time Series. Philip Allan. Oxford (chap. 6,7 and 8)

20. .... (1981 b) Time Series Models Philip Allan Oxford (chap 5 and 6)

21. Khoshgoftaar, T. M. (1982) Empirical Bayes methods in time series analysis, Ph. D. Thesis VPI, Blacksburg, Va.

22. Krutchkoff, R. G. (1972) Empirical Bayes estimation The Amer. Staistician 26 14-16

23. Lai, T. L. and C. Z. Wei (1982) Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems Ann. statist. 10 154-166

24. Launer, R.L. (1970) Empirical Bayes estimation in time analysis, Ph. D. thesis VPI , Blacksburg, Va.

25. Lehmann, E.L. (1959) Testing of Statistical Hypothesis Wiley, New York (chap. 2)

26. Li, W.K. and Y.V. Hui (1983a) Estimation of random coefficient autoregressive process: An empirical Bayes approach. J. of Time Series Analysis 4 89-94

27. .... (1983b) Estimation of random coefficient time series: An empirical Bayes approach. Amer. Statist. Asso. Proc. of Business and Economics. Statist. Sec. 371-374

28. Loeve, M. (1963) Probability Theory, New Jersey, Van Nostrand.

29. Maritz, J.S. and T. Lwin (1975) Construction of simple Bayes estimators. J.R. Statist. Soc B 75 421-425

30. Martz, H. and R.G. Krutchkoff (1969) Empirical Bayes estimators in a multiple linear regression model. Biometrika 56 367-374

31. Mittal, Y. (1984) Stochastic Process II (class notes), VPI, Blacksburg.

32. Miyasawa, K. (1960) An empirical Bayes estimator of mean of a normal population Bull. Inter. Statist. Inst. 38 181-7

33. Morris, C.N. (1983) Parametric empirical Inference: Theory and Applications. JASA 78 47-55 (with discussions)

34. O'Bryan, T.E. and V. Susarla (1975) An empirical Bayes estimation problem with nonidentical components involving normal distributions. Communications in Statist. 4 1033-1040

35. .... (1976) Rates in the empirical Bayes estimation problem with nonidentical components. Case of normal distributions. Ann. Int. statist. Math. 28 389-397

36. Pesaran, M.H. and L.J. Slater (1980) Dynamic Regression: Theory and Algorithms Ellis Horwood, London

37. Pierce, D. A. (1971) Least squares estimation in the regression model with autoregressive-moving average errors. Biometrika, 58 299-312

38. Rai, B. and A. Ullah (1981) Econometrics---a varying coefficients approach. St. Martins Press., New York

39. Rao, C.R. (1973) Linear Statistical Inference and its Applications Second edition Wiley, New York

40. Robbins, H. (1983) Some thoughts on empirical Bayes estimation Ann. of Staist. 11 713-723

41. ....(1964) The Empirical Bayes Approach to Statistical Decision Problems Ann. Math. Statis. 35 1-20

42. .... (1956) An empirical Bayes approach to statistics Proc. Third Berkeley Symp. Math. Statist. Probab. 1 157-163 Univ. of Calif. Press.

43. Schonfelt, P. (1971) A useful central limit theorem for m- dependent random variables. Metrika 17 116-128

44. Singh, R.S. (1977) Applications of estimators of a density and its derivatives to certain statistical problems. J. R. Statist. Soc. B 39 357-363

45. .... (1985) Empirical Bayes estimation in a multiple linear regression model Ann. Inst. Statist. Math. 37 Part A 71-86

46. Susarla, V. (1974) Rate of convergence in the sequence-compound squared-distance loss estimation problem for a family of m-variate normal distribution. Ann. Statist. 2 118-133

47. Swamy, P.A.V.B. (1971) Statistical Inference in Random Coefficient Regression Models. Springer-Verlag, New York.

48. .... (1974) Linear models with random coefficients. In Frontiers in Econometrics, chap. V by P. Zarembka (ed.)

49. .... and P.A. Tinsley (1980) Linear Prediction and Estimation methods for Regression models with stationary stochastic coefficients. J. of Econometrics 12 103-142

50. Theil, H. (1971) Principles of Econometrics North-Holland, Amsterdam (chapter 8)

51. Yamamoto, T. (1976) A note on the asymptotic mean square error of predicting more than one step ahead using the regression method. Applied Statist. 25 123-127

Vita

The vita has been removed from the scanned document