Journal of Economies 15 (1981) 341–356. North-Holland Publishing Company
etrics

A BAYESIAN APPROACH TO TIME-VARYING CROSS-SECTIONAL REGRESSION MODELS*

Lon-Mu LIU and Dominique M. HANSSENS

University of California, Los Angeles, CA 90024, USA

Received December 1979, final version received August 1980

This paper presents a Bayesian approach to regression models with time-varying parameters, or state vector models. Unlike most previous research in this field the model allows for multiple observations for each time period. Bayesian estimators and their properties are developed for the general case where the regression parameters follow an ARMA(s,q) process over time. This methodology is applied to the estimation of time-varying price elasticity for a consumer product, using biweekly sales data for eleven domestic markets. The parameter estimates and forecasting performance of the model are compared with various alternative approaches.

1. Introduction

Regression models with time-varying parameters have received a great deal of attention in econometrics. In this article, we are interested in the following time-varying regression model:

$$Y_{it} = X_{it}\beta_t + \varepsilon_{it}, \qquad t=1,2,\ldots,n, \quad i=1,2,\ldots,m_t, \tag{1}$$

$$(\beta_t - \beta) = \Phi(\beta_{t-1} - \beta) + a_t, \qquad t=1,2,\ldots,n, \tag{2}$$

where Y_{it} is the ith observation at the tth time period; X_{it} is a $1\times p$ vector of independent variables corresponding to Y_{it}; ε_{it} is the corresponding random error independently and identically distributed as $\mathrm{N}(0,\sigma_\varepsilon^2)$; β_t, also referred to as a state variable, is a $p\times 1$ vector of regression coefficients of Y_{it} on X_{it}, and a_t is also independently and identically distributed as $\mathrm{N}(0, A)$ and is independent of ε_{it}. We refer to the model in (1) as measurement equation, and (2) as process equation. The first-order autoregressive parameter Φ, also referred to as transition matrix in an engineering context, is assumed to have

*The authors wish to thank Professor G.C. Tiao and Dr. S. Graves for helpful comments and suggestions, and Lyda Boyer and Janice Gammel for editorial assistance on this article. This research was supported in part by National Science Foundation Grant MCS78-16827.

all characteristic roots less than one in absolute value. Note that when $\boldsymbol{\Phi}=\boldsymbol{0}$, the model set forth in (1) and (2) is reduced to a random coefficient regression model.

The model described in (1) and (2) is one type of state vector models that has been studied for many years in the engineering literature. The state vector model was first analyzed by Kalman (1960) and Kalman and Bucy (1961), who originated an extensive literature in control theory and applied physical sciences, in which the optimal estimation methods are often referred to as Kalman, Kalman–Bucy, or Wiener–Kalman filters. In engineering applications, it is usually assumed that only one data point is observed at each time period and $\sigma_\varepsilon^2=0$. Thus, estimation of secondary parameters $\boldsymbol{\Phi}$, $\boldsymbol{\beta}$, and $\boldsymbol{A}$ is impossible and must be assumed known [Sarris (1973), Mehra (1974)]. This is a severe restriction when transferring from engineering control theory to statistics and econometrics.

In econometrics, most of the literature on state vector models also focuses on the situation where only one data point is available at each time period (i.e., $m_t=1$). In this context, classical estimation procedures for various models have been proposed by Hildreth and Houck (1968), Rosenberg (1968, 1972), Duncan and Horn (1972), Cooley and Prescott (1973), Cooper (1973), Belsley (1973), Pagan (1978), Engle (1978), and Ledolter (1979). Bayesian estimation for an $m_t=1$ situation can be found in Sarris (1973). Few authors, however, have studied models with $m_t \geqq 1$: Rosenberg (1973) studies time-varying parameters in panel data where a complicated model involving parameters varying within each cross-section and over a period of time is proposed. Harvey (1977) studies the model set forth in (1) and (2) with $m_t \geqq 1$. His main interest is an efficient computation of the state variable $\boldsymbol{\beta}_t$ using Kalman filter techniques. The GLS and maximum likelihood estimates of the parameters are discussed in Harvey's paper.

In this article, we are interested in the state vector model shown in (1) and (2) which is similar to the model in Harvey (1977). Instead of using a classical approach, we study this problem from a Bayesian point of view. Although Sarris (1973) also studies the model in (1) and (2) from a Bayesian viewpoint for an $m_t=1$ situation, discrimination between different state vector models for single series is difficult and a particular structure is usually imposed on the model or some of the secondary parameters must be assumed known [see Sarris (1973) and Harvey (1977)]. The availability of time series of cross-sections and panel data provides much more scope for investigating the nature of state vector models.

In single time-series situations, analysis is usually complicated when data are missing. However, use of the state vector model can go past unobserved data points without difficulty.

This paper is organized as follows: first, in section 2 we discuss a special case of model (1) and (2) in which $p=1$. We illustrate the results using a price elasticity example. The results can be easily extended to a situation where β_t follows a univariate AR(s), MA(q), or ARMA(s,q) model. We then discuss the model described in (1) and (2).

Discussion throughout the paper centers on the following topics:

(1) Inferences about past and current values of $\boldsymbol{\beta}_t$, $t=1,2,\ldots,n$. (In econometrics, the current value $\boldsymbol{\beta}_n$ usually receives special interest.)
(2) Prediction of $\boldsymbol{\beta}_{n+1}$.
(3) Estimation of the ratio $\boldsymbol{A}$ and σ_ε^2 and some secondary parameters such as $\boldsymbol{\Phi}$ and $\boldsymbol{\beta}$.

2. State vector model with $p=1$

In this section, we study a special case of the model described in (1) and (2) where $p=1$. This model can be expressed as

$$Y_{it}=\beta_t X_{it}+\varepsilon_{it}, \qquad \varepsilon_{it}\sim \mathrm{N}(0,\sigma_\varepsilon^2), \tag{3}$$

$$(\beta_t-\beta)=\phi(\beta_{t-1}-\beta)+a_t, \qquad a_t\sim \mathrm{N}(0,\sigma_a^2), \tag{4}$$

for $t=1,2,\ldots,n$ and $i=1,2,\ldots,m_t$.

A special case of the above model with $X_{it}=1$ and $m_t=m$ is discussed in Tiao and Ali (1971) and Box and Tiao (1974). The process equation (4) can be an ARMA(s,q) process in general. For simplicity, we use the following notations in the discussion: $\boldsymbol{b}'=[\beta_1,\beta_2,\ldots,\beta_n]$, $\boldsymbol{Y}_t'=[Y_{1t},Y_{2t},\ldots,Y_{m_t t}]$, and $\boldsymbol{Y}'=[\boldsymbol{Y}_1',\boldsymbol{Y}_2',\ldots,\boldsymbol{Y}_n']$.

2.1. Estimation of β_t's for known ϕ, β, σ_a^2, and σ_ε^2

From (3), the joint distribution of $\boldsymbol{Y}$ given $\boldsymbol{b}$ and σ_ε^2 can be expressed as

$$p(\boldsymbol{Y}\mid \boldsymbol{b},\sigma_\varepsilon^2)\propto(\sigma_\varepsilon^2)^{-m./2}\exp\left\{-\frac{1}{2\sigma_\varepsilon^2}[S_\varepsilon^2+(\boldsymbol{b}-\hat{\boldsymbol{b}})'\boldsymbol{D}(\boldsymbol{b}-\hat{\boldsymbol{b}})]\right\}, \tag{5}$$

where $\boldsymbol{D}$ is the $n\times n$ diagonal matrix with S_{xt}^2 as the tth diagonal element in which

$$S_{xt}^2=\sum_{i=1}^{m_t}X_{it}^2, \qquad \hat{\beta}_t=\left(\sum_{i=1}^{m_t}X_{it}Y_{it}\right)\Big/ S_{xt}^2,$$

$$S_\varepsilon^2=\sum_{t=1}^{n}\left(\sum_{i=1}^{m_t}Y_{it}^2-\hat{\beta}_t^2 S_{xt}^2\right), \qquad m.=\sum_{t=1}^{n}m_t.$$

From (4), the joint distribution of $\boldsymbol{b}$ given ϕ, β and σ_a^2 can be written as

$$p(\boldsymbol{b}\,|\,\phi,\beta,\sigma_a^2)\propto(\sigma_a^2)^{-n/2}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma_a^2}\left[(\boldsymbol{b}-\beta\boldsymbol{1}_n)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{b}-\beta\boldsymbol{1}_n)\right]\right\}, \tag{6}$$

where $\boldsymbol{1}_n$ is a $n\times 1$ vector of 1's, and $\sigma_a^2\boldsymbol{\Sigma}$ is the $n\times n$ variance–covariance matrix of β_t, $t=1,2,\ldots,n$. For an AR(1) model, $\boldsymbol{\Sigma}^{-1}$ can be expressed as

$$\boldsymbol{\Sigma}^{-1}=\begin{bmatrix} 1 & -\phi & . & . & . & 0 & 0 \\ -\phi & 1+\phi^2 & . & . & . & 0 & 0 \\ \vdots & \vdots & \vdots & . & . & \vdots & \vdots \\ 0 & 0 & . & . & . & 1+\phi^2 & -\phi \\ 0 & 0 & . & . & . & -\phi & 1 \end{bmatrix}.$$

Note that $|\boldsymbol{\Sigma}|^{-\frac{1}{2}}=\sqrt{1-\phi^2}$.

By combining (5) and (6), we obtain the posterior distribution of $\boldsymbol{b}$ given ϕ, β, σ_a^2 and σ_ε^2 which can be simplified to

$$p(\boldsymbol{b}\,|\,\phi,\beta,\sigma_a^2,\sigma_\varepsilon^2,\boldsymbol{Y})\propto\exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\left[(\boldsymbol{b}-\tilde{\boldsymbol{b}})'(\boldsymbol{D}+\boldsymbol{H})(\boldsymbol{b}-\tilde{\boldsymbol{b}})\right]\right\}, \tag{7}$$

where

$$\tilde{\boldsymbol{b}}'=[\tilde{\beta}_1,\tilde{\beta}_2,\ldots,\tilde{\beta}_n], \qquad \tilde{\boldsymbol{b}}=(\boldsymbol{D}+\boldsymbol{H})^{-1}(\boldsymbol{D}\hat{\boldsymbol{b}}+\beta\boldsymbol{H}\boldsymbol{1}_n),$$

$$\boldsymbol{H}=(1/w)\boldsymbol{\Sigma}^{-1}, \qquad w=\sigma_a^2/\sigma_\varepsilon^2, \tag{8}$$

i.e., *a posteriori*, $\boldsymbol{b}$ is normally distributed with mean $\tilde{\boldsymbol{b}}$ and covariance matrix $\sigma_\varepsilon^2(\boldsymbol{D}+\boldsymbol{H})^{-1}$.

Assuming $\phi=0$ and $\sigma_a^2\to\infty$, the posterior modal estimate of β_t is $\hat{\beta}_t$ and the posterior variance $\sigma_\varepsilon^2/S_{xt}^2$, that is, equivalent to the usual OLS estimates. Using the matrix identity $(\boldsymbol{D}+\boldsymbol{H})^{-1}=\boldsymbol{D}^{-1}-\boldsymbol{D}^{-1}(\boldsymbol{D}^{-1}+\boldsymbol{H}^{-1})\boldsymbol{D}^{-1}$, it is observed that the state variable β_t has smaller posterior variance for the state vector model. This result can easily be extended to the $p>1$ situation.

Posterior distribution of β_{n+1} *given* $(\beta,\phi,\sigma_a^2,\sigma_\varepsilon^2)$. Suppose in the last observational period $t=n$ we want to predict or make an inference about the future value β_{n+1}. The posterior distribution of β_{n+1} can be obtained as follows: Using

$$\dot{\boldsymbol{b}}' = [\boldsymbol{b}', \beta_{n+1}], \qquad \dot{\hat{\boldsymbol{b}}}' = [\hat{\boldsymbol{b}}', 0], \qquad \dot{\boldsymbol{1}}' = [\boldsymbol{1}_n', 1],$$

$$\dot{\boldsymbol{D}} = \begin{bmatrix} \boldsymbol{D} & 0 \\ 0 & 0 \end{bmatrix}, \qquad \boldsymbol{H} = \frac{1}{w}\left[\begin{array}{c|cc} \Sigma^{-1} & & \boldsymbol{0} \\ & & 0 \\ \hline \boldsymbol{0} & 0 & 0 \end{array}\right] + \left[\begin{array}{cc|c} \boldsymbol{0}_{n-1} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \phi^2 & -\phi \\ \hline \boldsymbol{0} & -\phi & 1 \end{array}\right],$$

the posterior distribution of $(\beta_{n+1}, \boldsymbol{b})$ given $(\beta, \phi, \sigma_a^2, \sigma_\varepsilon^2)$ can be expressed as

$$p(\beta_{n+1}, \boldsymbol{b} \mid \phi, \beta, \sigma_a^2, \sigma_\varepsilon^2, Y) \propto \exp\left\{ -\frac{1}{2\sigma_\varepsilon^2} [(\dot{\boldsymbol{b}} - \dot{\tilde{\boldsymbol{b}}})'(\dot{\boldsymbol{D}} + \dot{\boldsymbol{H}})(\dot{\boldsymbol{b}} - \dot{\tilde{\boldsymbol{b}}})] \right\},$$

with

$$\dot{\tilde{\boldsymbol{b}}} = [\tilde{\boldsymbol{b}}', \tilde{\beta}_{n+1}]' = (\dot{\boldsymbol{D}} + \dot{\boldsymbol{H}})^{-1}(\dot{\boldsymbol{D}}\dot{\hat{\boldsymbol{b}}} + \beta \dot{\boldsymbol{H}}\dot{\boldsymbol{1}}). \tag{9}$$

After integrating out $\boldsymbol{b}$, we obtain

$$p(\beta_{n+1} | \beta, \phi, \sigma_a^2, \sigma_\varepsilon^2, \boldsymbol{Y}) \propto \exp\left\{ -\frac{1}{2\sigma_\varepsilon^2} [(\beta_{n+1} - \tilde{\beta}_{n+1})^2 / g] \right\},$$

where g is the $(n+1)$th diagonal element of $(\dot{\boldsymbol{D}} + \dot{\boldsymbol{H}})^{-1}$. In the AR(1) model, $g = 1 + \phi^2 C_{nn}$ where C_{nn} is the nth diagonal element of $(\dot{\boldsymbol{D}} + \dot{\boldsymbol{H}})^{-1}$. Note that from (9), we have

$$\tilde{\beta}_{n+1} = \phi \tilde{\beta}_n + (1 - \phi)\beta, \tag{10}$$

i.e., $\tilde{\beta}_{n+1}$ is a weighted average of $\tilde{\beta}_n$ and β.

In the above discussion we concentrate on the question of predicting β_{n+1}. It is clear that following a similar argument, the posterior distribution of β_{n+l}, $l = 2, 3, \ldots$, can readily be obtained. Details are omitted.

2.2. *Posterior distributions of the primary and secondary parameters when $(\beta, \phi, w, \sigma_\varepsilon^2)$ are unknown*

In practice σ_ε^2 and the secondary parameters β, ϕ and σ_a^2 are unknown. When $(\beta, \phi, \sigma_a^2, \sigma_\varepsilon^2)$ are unknown, we must incorporate them as variable parameters in the model and make inferences through their posterior distribution, but first we must construct a prior distribution for the parameters. From the results obtained above, we find that $(\beta, \phi, w, \sigma_\varepsilon^2)$ is a natural parameterization in this problem. We assume that $p(\beta) \propto c$. For the remaining three parameters $(\phi, w, \sigma_\varepsilon^2)$, by applying Jeffreys' rule to the marginal likelihood function $(\beta, \phi, w, \sigma_\varepsilon^2)$, we find that

$$P(\phi, w, \sigma_\varepsilon^2) \propto \sigma_\varepsilon^{-2} f(\phi, w),$$

where f does not involve σ_ε^2. An exact expression for f is exceedingly complicated; simple approximations are being investigated. For a moderate-sized sample, precise choice of $f(\phi, w)$ is not critical for inference about (ϕ, w) and is even less important for the β_t's [Liu and Tiao (1980)]. In the following discussion we simply assume that the prior for (ϕ, w) is locally uniform.

Under the above prior assumptions for $(\beta, \phi, w, \sigma_\varepsilon^2)$, we combine (5) with (6) and obtain the joint posterior distribution $(\boldsymbol{b}, \beta, \phi, w, \sigma_\varepsilon^2)$. After integrating out β and σ_ε^2, we obtain

$$\begin{aligned} P(\boldsymbol{b}, \phi, w \mid Y) \propto\; & w^{-n/2} |\Sigma|^{-\frac{1}{2}} |\boldsymbol{1}_n' \boldsymbol{H} \boldsymbol{1}_n|^{-\frac{1}{2}} \\ & \times \{S_\varepsilon^2 + (\hat{\boldsymbol{b}} - \bar{\beta}\boldsymbol{1}_n) \boldsymbol{D}(\boldsymbol{D}+\boldsymbol{H})^{-1} \boldsymbol{H} (\hat{\boldsymbol{b}} - \bar{\beta}\boldsymbol{1}_n) \\ & + (\boldsymbol{b} - \bar{\boldsymbol{b}})' \boldsymbol{C}^{-1} (\boldsymbol{b} - \bar{\boldsymbol{b}})\}^{-(m.+n-1)/2}, \end{aligned}$$

with

$$\bar{\boldsymbol{b}}' = [\bar{\beta}_1, \bar{\beta}_2, \ldots, \bar{\beta}_n],$$

$$\bar{\boldsymbol{b}} = (\boldsymbol{D}+\boldsymbol{H})^{-1} (\boldsymbol{D}\hat{\boldsymbol{b}} + \bar{\beta} \boldsymbol{H} \boldsymbol{1}_n),$$

$$\bar{\beta} = \boldsymbol{1}_n' \boldsymbol{D}(\boldsymbol{D}+\boldsymbol{H})^{-1} \boldsymbol{H} \hat{\boldsymbol{b}} / (\boldsymbol{1}_n' \boldsymbol{D}(\boldsymbol{D}+\boldsymbol{H})^{-1} \boldsymbol{H} \boldsymbol{1}_n),$$

$$\begin{aligned} \boldsymbol{C} &= \{\boldsymbol{D} + \boldsymbol{H} - (\boldsymbol{1}_n' \boldsymbol{H} \boldsymbol{1}_n)^{-1} \boldsymbol{H} \boldsymbol{1} \boldsymbol{1}' \boldsymbol{H}\}^{-1} \\ &= (\boldsymbol{D}+\boldsymbol{H})^{-1} + (\boldsymbol{D}+\boldsymbol{H})^{-1} \boldsymbol{H} \boldsymbol{1}_n \boldsymbol{1}_n' \boldsymbol{H} (\boldsymbol{D}+\boldsymbol{H})^{-1} / (\boldsymbol{1}_n' \boldsymbol{D}(\boldsymbol{D}+\boldsymbol{H})^{-1} \boldsymbol{H} \boldsymbol{1}_n). \end{aligned}$$

The above expression can be factorized into $P(\boldsymbol{b} \mid \phi, w, Y)$ and $P(\phi, w \mid Y)$ where

$$P(\boldsymbol{b} \mid \phi, w, Y) \propto \left\{1 + \frac{(\boldsymbol{b} - \bar{\boldsymbol{b}})' \boldsymbol{C}^{-1} (\boldsymbol{b} - \bar{\boldsymbol{b}})}{(m.-1)\bar{\sigma}_\varepsilon^2}\right\}^{-(m.+n-1)/2}, \tag{11}$$

$$\bar{\sigma}_\varepsilon^2 = (S_\varepsilon^2 + (\hat{\boldsymbol{b}} - \bar{\beta}\boldsymbol{1}_n) \boldsymbol{D}(\boldsymbol{D}+\boldsymbol{H})^{-1} \boldsymbol{H} (\hat{\boldsymbol{b}} - \bar{\beta}\boldsymbol{1}_n)) / (m.-1),$$

and

$$\begin{aligned} P(\phi, w \mid Y) \propto\; & w^{-n/2} |\Sigma|^{-\frac{1}{2}} |\boldsymbol{D}+\boldsymbol{H}|^{-\frac{1}{2}} |\boldsymbol{1}_n' \boldsymbol{D}(\boldsymbol{D}+\boldsymbol{H})^{-1} \boldsymbol{H} \boldsymbol{1}_n|^{-\frac{1}{2}} \\ & \times \{S_\varepsilon^2 + (\hat{\boldsymbol{b}} - \bar{\beta}\boldsymbol{1}_n)' \boldsymbol{D}(\boldsymbol{D}+\boldsymbol{H})^{-1} \boldsymbol{H} (\hat{\boldsymbol{b}} - \bar{\beta}\boldsymbol{1}_n)\}^{-(m.-1)/2}. \end{aligned} \tag{12}$$

The distribution in (11) is the multivariate t distribution, $t_n(\bar{b}, \bar{\sigma}_\varepsilon^2 C, m.-1)$. Thus given (ϕ, w), β_t has the univariate t distribution $t(\bar{\beta}_i, \bar{\sigma}_\varepsilon^2 C_{ii}, m.-1)$ with C_{ii} the ith diagonal element of $\boldsymbol{C}$.

Note that in deriving the above formula, $\boldsymbol{D}$ is not required to be a non-singular matrix, therefore the results above are still valid even if some of the $S_{xt}^2=0$. As a natural extension of the result in (11), it is readily seen that β_{n+1} given (ϕ, w) is distributed as $t(\bar{\beta}_{n+1}, \bar{\sigma}_\varepsilon^2 \bar{g}, m.-1)$, where $\bar{\beta}_{n+1}$ is the $(n+1)$th element of $\dot{\bar{b}}$,

$$\dot{\bar{\boldsymbol{b}}} = [\bar{\boldsymbol{b}}, \bar{\beta}_{n+1}]' = (\dot{\boldsymbol{D}} + \dot{\boldsymbol{H}})^{-1}(\dot{\boldsymbol{D}}\hat{\boldsymbol{b}} + \bar{\beta}\dot{\boldsymbol{H}}\dot{\boldsymbol{1}}),$$

and $\bar{g}$ is the $(n+1)$th diagonal element of $\dot{\boldsymbol{C}}$, where

$$\dot{\boldsymbol{C}} = (\dot{\boldsymbol{D}} + \dot{\boldsymbol{H}})^{-1} + (\dot{\boldsymbol{D}} + \dot{\boldsymbol{H}})^{-1}\dot{\boldsymbol{H}}\dot{\boldsymbol{1}}\dot{\boldsymbol{1}}'\dot{\boldsymbol{H}}(\dot{\boldsymbol{D}} + \dot{\boldsymbol{H}})^{-1}/(\dot{\boldsymbol{1}}'\dot{\boldsymbol{D}}(\dot{\boldsymbol{D}} + \dot{\boldsymbol{H}})^{-1}\dot{\boldsymbol{H}}\dot{\boldsymbol{1}}).$$

The unconditional posterior distribution of $\beta_t, t=1,2,\ldots,n+1,\ldots$ can be obtained by the following integration

$$P(\beta_t|Y) = \iint P(\beta_t|\phi, w, Y)P(\phi, w|Y)\,\mathrm{d}\phi\,\mathrm{d}w. \tag{13}$$

Numerical evaluation of the distribution is computationally expensive. A simple approximation is

$$P(\beta_t|Y) \doteq P(\beta_t|\overset{*}{\phi}, \overset{*}{w}, Y), \tag{14}$$

where $(\overset{*}{\phi}, \overset{*}{w})$ is the joint modal estimate of (ϕ, w) in $P(\phi, w|Y)$. The posterior mode of β_t in (14) is denoted by $\overset{*}{\beta}_t$.

Missing data. Assuming that β_t in (4) is observable, the secondary parameter ϕ can be estimated based on the observed β_t's [see e.g., Box and Jenkins (1970)]. In this situation, the likelihood function in (6) cannot be evaluated if one or more β_t's are missing in the series. However, in the state vector model, we can still use (12) for making an inference on ϕ and w even if some of the time periods contain no data. To show this more explicitly, we can replace $(\boldsymbol{D}+\boldsymbol{H})^{-1}\boldsymbol{H}$ in (12) by its identity $I-(\boldsymbol{D}+\boldsymbol{H})^{-1}\boldsymbol{D}$. When $m_t=0$, there is no information contained in the design matrix and therefore $S_{xt}^2=0$. Thus the posterior distribution of (12) will remain the same no matter what $\hat{\beta}_t$ is assumed to be. The same argument can be applied for the estimation of β_t.

2.3. *An example: Price elasticity for a consumer product*

We now apply the above estimation and forecasting procedures to a realworld case in marketing, focusing on the relationship between product sales and unit price. The product is an inexpensive branded gift item which is distributed in eleven markets in the United States.[1] Data on product sales (Y) and local prices (X) were available for 37 biweekly periods, for a total of 403 observations. The sales data are divided by the percentage retail availability in each market, so the analysis can focus on the sales–price relationship. One additional marketing instrument, brand advertising, is not included because there is virtually no difference in expenditures across the eleven markets.

Price elasticity is estimated using a constant-elasticity model for each time period. The data are expressed in deviations from the mean in each period. Primes are used to denote natural logarithms,

$$y_{it}=\beta_t x_{it}+\varepsilon_{it}, \qquad t=1,2,\ldots,37, \quad i=1,2,\ldots,m_t,$$

with

$$y_{it}=Y'_{it}-\bar{Y}'_i \quad \text{and} \quad x_{it}=X'_{it}-\bar{X}'_i,$$

where β_t is the price elasticity at time t, and ε_{it} is the disturbance term assumed normally distributed around zero with constant variance.

The first step in the analysis is the specification of a tentative model for price elasticity over time. The procedure follows a suggestion by Pagan (1978): compute OLS estimates of β_t for each t and examine the autocorrelation and partial autocorrelation functions of the time series $\hat{\beta}_t$ ($t=1,\ldots,37$). The identification procedure reveals that the series $\hat{\beta}_t$ can be represented parsimoniously by an AR(1) process. The $\hat{\beta}_t$'s are shown in fig. 1: virtually all 37 parameters are negative and significant, explaining up to 56% of the variance in sales, depending on the time period.

We decided to use the first thirty biweeks as the development sample and the remaining seven periods as the holdout sample to evaluate different approaches. Since one-step-ahead forecast errors are used, the model is estimated for $n=30,31,\ldots,37$. The results are summarized in table 1. This table and fig. 1 show that the Bayesian estimates of the state variable are stable, in spite of the fact that some of the OLS estimates of elasticity over time ($\hat{\beta}_t$) are very volatile. The pooled estimate of price elasticity is around -2.2 which is realistic for this type of product.

The evaluation of the state vector model consists of two parts: first, we test if the model has the ability to generate reliable estimates of β_n

[1]The name of the brand and the actual data cannot be revealed for confidentiality reasons. The number of markets $m_t=11$, except for a few cases of zero availability: $m_{16}=m_{18}=10$ and $m_{17}=9$. This situation illustrates the fact that m_t need not be constant in the model.

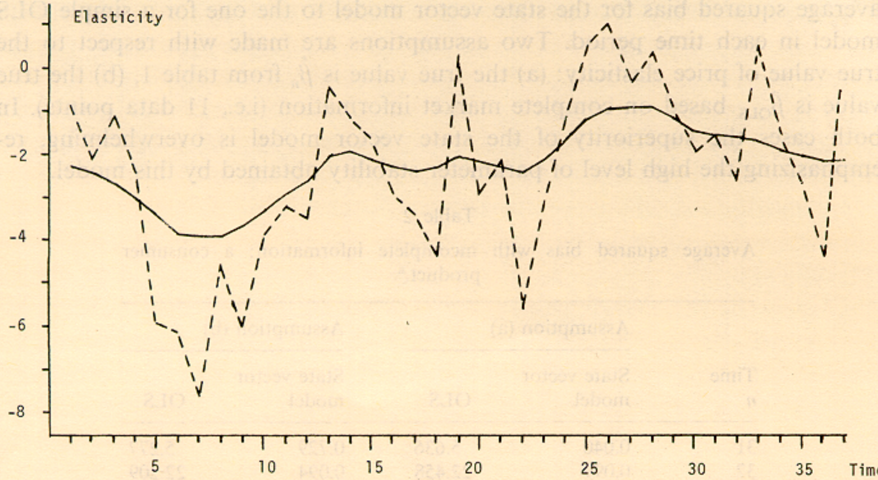

Fig. 1. Price elasticity over time: a consumer product; $\overset{*}{\beta}_t$: estimate under state vector model (———), and $\hat{\beta}_t$: OLS estimate (– – – –).

Table 1

Bayesian parameter estimates for a state vector model using various n: a consumer product.

Time n	$\overset{*}{\beta}$	$\overset{*}{\beta}_n$	$\overset{*}{\phi}$	$\overset{*}{w}$	$\overset{*}{\sigma}^2_\varepsilon$
30	−2.184	−1.295	0.862	2.511	0.252
31	−2.199	−1.469	0.855	2.510	0.248
32	−2.252	−1.800	0.839	2.453	0.251
33	−2.143	−1.408	0.857	2.511	0.246
34	−2.143	−1.517	0.858	2.431	0.244
35	−2.174	−1.746	0.875	1.793	0.240
36	−2.294	−2.360	0.826	2.469	0.238
37	−2.246	−2.123	0.829	2.465	0.233

($n=31, 32, \ldots, 37$) with incomplete market information, e.g., data on 5 out of 11 markets. Second, one-step-ahead forecast errors in the holdout sample are compared with those of three alternative models.

The first model evaluation procedure focuses on the quality of estimates of price elasticity when only partial cross-sectional data are available. The scenario is that at period n ($n=31, 32, \ldots, 37$), price and sales data from only five out of the eleven markets are available; estimates of price elasticity at time n are obtained using past observations plus five current observations. This scenario is simulated for all $\binom{11}{5}=462$ combinations of five markets.

The results of this exercise are summarized in table 2, which compares the average squared bias for the state vector model to the one for a simple OLS model in each time period. Two assumptions are made with respect to the true value of price elasticity: (a) the true value is $\overset{*}{\beta}_n$ from table 1, (b) the true value is $\hat{\beta}_{\text{OLS}}$ based on complete market information (i.e., 11 data points). In both cases the superiority of the state vector model is overwhelming, re-emphasizing the high level of parameter stability obtained by this model.

Table 2

Average squared bias with incomplete information: a consumer product.[a]

	Assumption (a)		Assumption (b)	
Time n	State vector model	OLS	State vector model	OLS
31	0.040	5.636	0.729	5.277
32	0.099	22.458	0.094	22.209
33	0.125	9.322	1.292	13.748
34	0.166	6.808	5.360	13.098
35	0.018	6.083	0.017	6.062
36	0.015	13.484	0.763	8.634
37	0.226	6.211	6.529	18.675

[a]Assumption (a): 'true' estimate of β_n is the $\overset{*}{\beta}_n$ in table 1. Assumption (b): 'true' estimate of β_n is the OLS estimate at time n based on all 11 observations.

The second test examines forecasting performance of the state vector model against three competing models:

(a) A state vector model with a random walk process equation (i.e., $\phi=1$); the best one-step ahead forecast for β_{n+1} at time n is the least squares estimate $\hat{\beta}_n$. This forecasting practice makes intuitive sense. The state vector model with this process equation was studied in Cooley and Prescott (1973).

(b) A single OLS model, which postulates that price elasticity is invariant over time and across markets,

$$y_{it}=\beta x_{it}+\varepsilon_{it} \quad \text{for all } i \text{ and } t,$$

where the deviations are taken with respect to the grand mean of Y'_{it} and X'_{it}.

(c) A model which postulates that each market has its own price elasticity, which is invariant over time,

$$y_{it}=\beta_i x_{it}+\varepsilon_{it},$$

where the deviations are now taken from the means of Y'_{it} and X'_{it} within each market. The OLS estimates of β_i are found to be very volatile.

Table 3

Mean square forecast errors for one-step ahead forecast: a consumer product.

Time n	State vector model	Random walk state vector model	Overall OLS	OLS by market
31	0.109	0.110	0.159	0.241
32	0.274	0.275	0.289	0.300
33	0.107	0.119	0.178	0.260
34	0.122	0.137	0.163	0.251
35	0.098	0.099	0.108	0.136
36	0.159	0.148	0.139	0.184
37	0.067	0.091	0.094	0.157

The results of the forecasting comparisons are summarized in table 3. In all but one case the state vector model outperforms the various OLS models in forecastability. The differences in mean squares forecast errors (MSFE) are usually rather substantial, e.g., in periods 31, 33, 34 and 37. The MSFE's for a random walk state vector [i.e., model (a)] are close to the state vector models with estimated ϕ. This is probably because the estimated ϕ, about 0.85, is close to 1. The improvement of MSFE under appropriate ϕ will be more significant if ϕ is not this close to 1.

2.4. *Extension of the results to other models*

When applying the state vector model, the AR(1) model for the process equation may not be adequate to represent the behavior of β_t. Thus we may need to consider a more general model, such as the ARMA(s,q). Analysis under this time series model hinges on the expression of the distribution of β_t's given $\boldsymbol{\phi}'=[\phi_1,\phi_2,\ldots,\phi_s]$, $\boldsymbol{\theta}'=[\theta_1,\theta_2,\ldots,\theta_q]$ and σ_a^2. The distribution of $\boldsymbol{b}'=[\beta_1,\beta_2,\ldots,\beta_n]$ given $(\boldsymbol{\phi},\boldsymbol{\theta},\beta,\sigma_a^2)$ in general can be expressed as in (6). The results in sections 2.1 and 2.2 can thus be extended to ARMA(s,q) situations. Ljung and Box (1978) show explicit expressions for Σ^{-1} and $|\Sigma|$ under AR(s), MA(q) and ARMA(s,q) that can be used in this study.

3. State vector models with $p>1$

In this section, we study the model set forth in (1) and (2) with $p>1$. Note that a multiple AR(1) process for the vector $\boldsymbol{\beta}_t$ implies a possible ARMA$(p, p-1)$ process for individual β_{jt}, $j=1,2,\ldots,p$. This is one reason why state vector models can be applied in many situations, despite their simplicity.

For mathematical convenience, we use the following notation:

$$b' = [\beta_1, \beta_2, \ldots, \beta_n],\ Y_t' = [Y_{1t}, Y_{2t}, \ldots, Y_{m_t t}], \text{ and } Y' = [Y_1', Y_2', \ldots, Y_n'].$$

3.1. *Estimation of β_t's for known Φ, β, A and σ_ε^2*

From (1), the joint distribution of Y given b and σ_ε^2 can be expressed as in (5), where D is the $np \times np$ diagonal matrix with $X_t'X_t$ as the tth diagonal block matrix, $\hat{\beta}_t = (X_t'X_t)^{-1}X_t'Y_t$, $S_\varepsilon^2 = \sum_{t=1}^n (Y_t'Y_t - \hat{\beta}_t'X_t'Y_t)$, and $m. = \sum_{t=1}^n m_t$.

The joint distribution of the β_t's given Φ, β, and A in (2) can be expressed as

$$P(b|\beta, \Phi, A) \propto |\Gamma|^{-\frac{1}{2}} |A|^{-(n-1)/2} \exp\{-\tfrac{1}{2}[(b - 1_n \otimes \beta)' \Sigma^{-1} (b - 1_n \otimes \beta)]\} \tag{15}$$

where

$$\Sigma^{-1} = \begin{pmatrix} \Phi'A^{-1}\Phi + \Gamma^{-1} & -\Phi'A^{-1} & 0 & \cdots & 0 & 0 \\ -A^{-1}\Phi & A^{-1} + \Phi'A^{-1}\Phi & -\Phi'A^{-1} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & & \vdots & \vdots \\ 0 & 0 & \cdots & & A^{-1} + \Phi'A^{-1}\Phi & -\Phi'A^{-1} \\ 0 & 0 & \cdots & & -A^{-1}\Phi & A^{-1} \end{pmatrix},$$

$$\mathscr{P}(\Gamma) = (I - \Phi \otimes \Phi)^{-1} \mathscr{P}(A),$$

and $\mathscr{P}$ is a pack operator that converts a matrix into a vector by stacking the column vectors in the matrix one after another.

Combining $P(Y|b, \sigma_\varepsilon^2)$ with (15), we obtain the posterior distribution of b given Φ, β, A and σ_ε^2, which can be simplified to

$$P(b|\beta, \Phi, W, \sigma_\varepsilon^2, Y) \propto \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}[(b - \tilde{b})'(D + H)(b - \tilde{b})]\right\},$$

with

$$\tilde{b} = (D + H)^{-1}(D\hat{b} + H(1_n \otimes \beta)), \quad H = \sigma_\varepsilon^2 \Sigma^{-1}, \quad W = A/\sigma_\varepsilon^2.$$

That is, *a posteriori*, b is normally distributed with mean $\tilde{b}$ and covariance matrix $\sigma_\varepsilon^2 (D + H)^{-1}$.

For one-step-ahead forecasting β_{n+1}, we can go through the same argument as that given in section 2.1 and obtain

$$\beta_{n+1} \big| \beta, \Phi, W, \sigma_\varepsilon^2, Y \sim N(\tilde{\beta}_n(1), \sigma_\varepsilon^2 G),$$

with

$$\tilde{\beta}_{n+1} = \Phi \tilde{\beta}_n + (I - \Phi)\beta,$$

and G as the $(n+1)$th diagonal block matrix of $(\dot{D}+\dot{H})^{-1}$, where $\dot{D}$ and $\dot{H}$ have a similar definition as in (9).

3.2. *Posterior distribution of the primary and secondary parameters when $(\beta, \Phi, W, \sigma_\varepsilon^2)$ are unknown*

Our aim here is to obtain the posterior distribution of β_t and the related secondary parameters. Following a similar argument as that given in section 2.2, we assume the prior distribution of $(\beta, \Phi, W, \sigma_\varepsilon^2)$ as follows:

$$P(\beta) \propto c, \qquad P(\Phi, W, \sigma_\varepsilon^2) \propto \sigma_\varepsilon^{-2}.$$

Under the prior distribution, we can combine $P(Y \big| b, \sigma_\varepsilon^2)$ with (15) and obtain the posterior distribution of $(b, \beta, \Phi, W, \sigma_\varepsilon^2)$. After integrating out β and σ_ε^2, we obtain

$$P(b, \Phi, W \big| Y) \propto \big|Q\big|^{-\frac{1}{2}} \big|W\big|^{-(n-1)/2} \big|(1_n' \otimes I_p) H (1_n \otimes I_p)\big|^{-\frac{1}{2}}$$
$$\times \{S_\varepsilon^2 + (\hat{b} - 1_n \otimes \bar{\beta})' D (D+H)^{-1} H (\hat{b} - 1_n \otimes \bar{\beta})$$
$$+ (b - \bar{b})' C^{-1} (b - \bar{b})\}^{-(m. + (n-1)p)/2},$$

with

$$Q = \sigma_\varepsilon^{-2} \Gamma,$$
$$\bar{b} = (D+H)^{-1} (D\hat{b} + H(1_n \otimes I_p)\bar{\beta}),$$
$$\bar{\beta} = [(1_n' \otimes I_p) D (D+H)^{-1} H (1_n \otimes I_p)]^{-1} (1_n' \otimes I_p) D (D+H) H \hat{b},$$
$$C = \{D + H - H(1_n \otimes I_p)[(1_n' \otimes I_p) H (1_n \otimes I_p)]^{-1} (1_n' \otimes I_p) H\}^{-1},$$
$$= (D+H)^{-1} + (D+H)^{-1} H (1_n \otimes I_p)$$
$$[(1_n' \otimes I_p) D (D+H)^{-1} H (1_n \otimes I_p)]^{-1} (1_n' \otimes I_p) H (D+H)^{-1}.$$

The expression above can be factored into $P(\boldsymbol{b}|\boldsymbol{\Phi}, W, Y)$ and $P(\boldsymbol{\Phi}, W|Y)$, where

$$P(\boldsymbol{b}|\boldsymbol{\Phi}, W, Y)\propto\left\{1+\frac{(\boldsymbol{b}-\bar{\boldsymbol{b}})'\boldsymbol{C}^{-1}(\boldsymbol{b}-\bar{\boldsymbol{b}})}{(m.-p)\bar{\sigma}_\varepsilon^2}\right\}^{-(m.+(n-1)p)/2}, \tag{16}$$

$$\bar{\sigma}_\varepsilon^2=[(S_\varepsilon^2+(\hat{\boldsymbol{b}}-\boldsymbol{1}_n\otimes\bar{\boldsymbol{\beta}})'\boldsymbol{D}(\boldsymbol{D}+\boldsymbol{H})^{-1}\boldsymbol{H}(\hat{\boldsymbol{b}}-\boldsymbol{1}_n\otimes\bar{\boldsymbol{\beta}})]/(m.-p),$$

and

$$P(\boldsymbol{\Phi}, \boldsymbol{W}|Y)\propto|\boldsymbol{Q}|^{-\frac{1}{2}}|\boldsymbol{W}|^{-\frac{1}{2}}|(\boldsymbol{1}_n'\otimes\boldsymbol{I}_p)\boldsymbol{D}(\boldsymbol{D}+\boldsymbol{H})^{-1}\boldsymbol{H}(\boldsymbol{1}_n\otimes\boldsymbol{I}_p)|^{\frac{1}{2}}$$
$$\times\{S_\varepsilon^2+(\hat{\boldsymbol{b}}-\boldsymbol{1}_n\otimes\bar{\boldsymbol{\beta}})'\boldsymbol{D}(\boldsymbol{D}+\boldsymbol{H})^{-1}\boldsymbol{H}(\hat{\boldsymbol{b}}-\boldsymbol{1}_n\otimes\bar{\boldsymbol{\beta}})\}^{-(m.-p)/2}.$$

The distribution in (16) is a $t_{np}(\bar{\boldsymbol{b}}, \bar{\sigma}_\varepsilon^2\boldsymbol{C}, m.-p)$ distribution. Thus, given $(\boldsymbol{\Phi}, \boldsymbol{W})$, $\boldsymbol{\beta}_t$ is distributed as $t(\bar{\boldsymbol{\beta}}_t, \bar{\sigma}_\varepsilon^2\boldsymbol{C}_{ii}, m.-p)$ with $\boldsymbol{C}_{ii}$ the ith diagonal block matrix of $\boldsymbol{C}$ if the $\boldsymbol{C}$ matrix is partitioned into n^2 $p\times p$ submatrices. For the unconditional posterior distribution of $\boldsymbol{\beta}_t$, we can integrate $P(\boldsymbol{\beta}_t|\boldsymbol{\Phi}, W, Y)$ over $P(\boldsymbol{\Phi}, \boldsymbol{W}|Y)$, or approximately

$$P(\boldsymbol{\beta}_t|Y)\doteq P(\boldsymbol{\beta}_t|\overset{*}{\boldsymbol{\Phi}}, \overset{*}{\boldsymbol{W}}, Y),$$

with $(\overset{*}{\boldsymbol{\Phi}}, \overset{*}{\boldsymbol{W}})$, the joint modal estimate of $(\boldsymbol{\Phi}, \boldsymbol{W})$ in $P(\boldsymbol{\Phi}, \boldsymbol{W}|Y)$.

Like the $p=1$ situation, the results in sections 3.1 and 3.2 can be extended to the situation where the process equation follows a multiple ARMA(s,q) model. The key point is the specification of the distribution for the process equation. The exact distribution of $\boldsymbol{b}$ for an MA(q) situation is discussed in Phadke and Kedem (1978), and Hillmer and Tiao (1979).

4. Discussion

Bayes' Theorem has been successfully applied in many statistical analyses. It is especially useful for analyzing random coefficient models [Liu (1978)]. A Bayesian approach to state vector models makes it easy to combine the probability density function of the first-stage (measurement equation) and second-stage (process equation) models. We can then obtain the posterior distribution of the primary and secondary parameters from the combined probability density function. The whole practice is rather straightforward. However, a sampling approach to the same problem, e.g., Harvey (1977), may involve many ad hoc procedures.

Harvey (1977) studies a state vector model similar to (1) and (2) where he assumes that $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')=\sigma_\varepsilon^2\boldsymbol{H}_t$ and $\boldsymbol{\varepsilon}_t'=[\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{m_t t}]'$. The Kalman filter technique is employed to obtain the GLS (generalized least squares) estimate of $\boldsymbol{\beta}$, the MMSE (minimum mean square error) estimate of $\boldsymbol{\beta}_t - \boldsymbol{\beta}$, and a set of prediction error vectors for given $\boldsymbol{\Phi}$ and $\boldsymbol{A}$. He then obtains the concentrated log-likelihood of $\boldsymbol{\Phi}$ and A using the results at the first stage of analysis. A modified Kalman filter technique is used to compute the log-likelihood. The whole computation procedure must be iterated in order to obtain stable estimates of the parameters. This approach of parameter estimation is equivalent to obtaining a joint modal estimate for the likelihood function of $(\boldsymbol{\beta}_t, t=1,\dots,n, \boldsymbol{\beta}, \boldsymbol{\Phi}, \sigma_\varepsilon^2, A)$, which is equivalent to the joint posterior distribution of all parameters under a constant prior. However, as noted by O'Hagan (1976), joint modal estimates may not be appropriate in many instances. In a Bayesian framework, we prefer to examine the entire posterior distribution rather than condensing it to one or a few moment estimates. When it is necessary to obtain point estimates we prefer to integrate out the nuisance parameters first and then consider, say, the mode of the marginal posterior of the parameters of interest.

In addition to the statistical advantages, the approach in this paper may also be more efficient computationally. Harvey's approach requires two Kalman filter recursions for each iteration that may be more complicated than the computations required in this paper. In addition, the Kalman filter recursion method is an approximation procedure that may be vulnerable when the root(s) of the transition matrix $\boldsymbol{\Phi}$ is close to 1. The key computation in the proposed approach is the inversion of the matrix $\boldsymbol{D}+\boldsymbol{H}$. Fortunately, $\boldsymbol{D}+\boldsymbol{H}$ is a diagonal band matrix (tridiagonal if $p=1$) that can be inverted very efficiently using, for example, an algorithm by Martin and Wilkinson (1965). Computer programs for inverting diagonal band matrices are available in LINPACK ([Dongarra et al. (1979)]. Phadke and Kedem (1978) and Ansley (1979) discuss inverting diagonal band matrices in time series context in more details. Also, the results in this paper can be readily extended to a general situation where the process equation is an ARMA process. It is not clear that Harvey's approach can be easily extended to such situations. At any rate, we feel it is important to obtain sound statistical results before computational efficiency is pursued.

In summary, state vector models provide a convenient framework for combining current observations with a previous forecast, and for forecasting future $\boldsymbol{\beta}_t$'s. Also, the state vector models help estimate current and past $\boldsymbol{\beta}_t$'s more accurately than other classical methods. This is especially true for the time periods where insufficient data are available.

References

Ansley, C.F., 1979, An algorithm for the exact likelihood of a mixed autoregressive-moving average process, Biometrika 66, 59–65.

Belsley, D.A., 1973, The applicability of the Kalman filter in the determination of systematic parameter variation, Annals of Economic and Social Measurement 2, 531–533.

Box, G.E.P. and G.M. Jenkins, 1970, Time series analysis: Forecasting and control (Holden Day, San Francisco, CA).

Box, G.E.P. and G.C. Tiao, 1974, Some comments on the problem of many means, Report of the Conference on Applications of Bayes Methods (Department of Statistics, Yale University, New Haven, CT).

Cooley, T.F. and E.C. Prescott, 1973, Varying parameter regression: A theory and some application, Annals of Economic and Social Measurement 2, 463–473.

Copper, J.P., 1973, Time-varying regression coefficient: A mixed estimation approach and operational limitations of the general Markov structure, Annals of Economic and Social Measurement 2, 525–530.

Dongarra, J.J., C.B. Moler, J.R. Bunch and G.W. Stewart, 1979, LINPACK user's guide (SIAM, Philadelphia, PA).

Duncan, D.B. and S.D. Horn, 1972, Linear dynamic recursive estimation from the viewpoint of regression analysis, Journal of the American Statistical Association 67, 815–821.

Engle, R. and W. Watson, 1978, A one factor multivariate time series model of metropolitan wage rates, Discussion Paper 78-15 (Department of Economics, University of California, San Diego, CA).

Harvey, A.C., 1977, The estimation of time varying parameters from panel data, Paper presented at the INSEE Symposium on the Econometrics of Panel Data, Paris.

Hildreth, C. and J.P. Houck, 1968, Some estimators for a linear model with random coefficients, Journal of the American Statistical Association 63, 584–595.

Hillmer, S.C. and G.C. Tiao, 1979, Likelihood function of stationary multiple autoregressive moving average models, Journal of the American Statistical Association 74, 652–660.

Kalman, R.E., 1960, A new approach to linear filtering and prediction problems, Transactions of ASME, Series D, Journal of Basic Engineering 82, 35–45.

Kalman, R.E. and R.S. Bucy, 1961, New results in linear filtering and prediction theory, Transactions of ASME, Series D, Journal of Basic Engineering 83, 95–108.

Ledolter, J., 1979, A recursive approach to parameter estimation in regression and time series models, Communications in Statistics A 12, 1227–1245.

Liu, L.M., 1978, A Bayesian approach to random coefficient models, Ph.D. Thesis (Department of Statistics, University of Wisconsin, Madison, WI).

Liu, L.M. and G.C. Tiao, 1980, Random coefficient first order autoregressive models, Journal of Econometrics 13, 305–325.

Ljung, M.J. and G.E.P. Box, 1978, The likelihood function of stationary autoregressive-moving average models, Biometrika 66, 265–270.

Martin, R.S. and J.H. Wilkinson, 1965, Symmetric decomposition of positive definite band matrices, Numerical Mathematics 7, 355–361.

Mehra, R.K., 1974, Topics in stochastic control theory: Identification in control and econometrics; similarities and differences, Annals of Economic and Social Measurement 3, 21–47.

O'Hagan, A., 1976, On posterior joint and marginal modes, Biometrika 63, 329–333.

Pagan, A., 1978, A unified approach to estimation and inference for stochastically varying coefficient regression models, Discussion Paper (Center for Operations Research and Econometrics, University of Louvain).

Phadke and Kedem, 1978, Computation of the exact likelihood function of multivariate moving average models, Biometrika 65, 511–519.

Rosenberg, B., 1968, Varying-parameter estimation, Ph.D. Thesis (Department of Economics, Harvard University, Cambridge, MA).

Rosenberg, B., 1972, The estimation of stationary stochastic regression parameters re-examined, Journal of the American Statistical Association 67, 650–654.

Rosenberg, B., 1973, The analysis of a cross-section of time series by stochastically convergent parameter regression, Annals of Economic and Social Measurement 2, 399–428.

Sarris, A.H., 1973, A Bayesian approach to estimation of time-varying regression coefficients, Annals of Economic and Social Measurement 2, 501–523.

Tiao, G.C. and M.M. Ali, 1971, Analysis of correlated random effects: Linear model with two random components, Biometrika 58, 37–51.