2 The Maximum Likelihood Estimation Method: Practical Issues

Part of the attraction of the ML theory is that it does offer estimator and test procedures that are technically almost universally applicable, provided one has a reasonably precise model. . . . There is a much better reason for using likelihood theory in that it provides a coherent framework for statistical inference in general.
—Cramer 1986, p. 8

The maximum likelihood estimation method is a way of obtaining estimators of a model when a specific distributional assumption is made about the vector of sample observations. Unlike least squares methods which use only the first two moments, maximum likelihood incorporates all the information in a model by working with the complete joint distribution of the observations.

Since the classical inferences for the models in chapters 3 through 6 depend heavily on maximum likelihood estimation, we provide a review of the method with a focus on related practical issues. For a more complete analysis of the maximum likelihood estimation method, readers are referred to Cramer 1986, Judge et al. 1982, Davidson and MacKinnon 1993, and Harvey 1990, from which this chapter draws.

2.1 Maximum Likelihood Estimation and the Covariance Matrix of $\hat{\theta}_{ML}$

A statistical model with a k-dimensional vector of parameters, θ, specifies a joint distribution for a vector of observations $\tilde{y}_T = [y_1 \quad y_2 \quad \ldots \quad y_T]'$:

$$\text{Joint Density Function:} \quad p(\tilde{y}_T|\theta), \tag{2.1}$$

which, in the discrete case, provides us with the probabilities of obtaining a particular set of values for $\tilde{y}_T$, given θ. The joint density therefore is a function of $\tilde{y}_T$ given θ.

In econometric practice, we have a realization of the $\tilde{y}_T$ vector, or the sample data, and we do not know the parameter vector θ of the underlying statistical model. In this case, the joint density in (2.1) is a function of θ given $\tilde{y}_T$, and it is called the likelihood function:

$$\text{Likelihood Function:} \quad L(\theta|\tilde{y}_T), \tag{2.2}$$

which is functionally equivalent to (2.1). Different values for θ result in different values for the likelihood function in (2.2). The likelihood function specifies the plausibility or likelihood of the data given the parameter vector θ. In the maximum likelihood estimation method, we are interested in choosing parameter estimates so as to maximize the probability of having generated the observed sample, by maximizing the log of the above likelihood function:

$$\hat{\theta}_{ML} = \text{Argmax} \ln L(\theta|\tilde{y}_T), \tag{2.3}$$

where ln L refers to the log likelihood function.

Maximizing the log likelihood function instead of the likelihood function itself enables us to estimate directly the asymptotic covariance matrix, $\text{Cov}(\hat{\theta}_{ML})$, of the maximum likelihood estimate, $\hat{\theta}_{ML}$. The expectation of the second derivatives of the log likelihood function provides us with the information matrix $I(\theta)$:

$$I(\theta) = -E\left[\frac{\partial^2 \ln L(\theta|\tilde{y}_T)}{\partial\theta\partial\theta'}\right], \tag{2.4}$$

which summarizes the amount of information in the sample. The inverse of this information matrix provides us with the lower bound for the covariance matrix of an unbiased estimator $\tilde{\theta}$, known as the Cramer-Rao inequality:

$$\text{Cov}(\tilde{\theta}) - I(\theta)^{-1} \quad \text{is positive semidefinite.} \tag{2.5}$$

In addition, it can be shown that the maximum likelihood estimator $\hat{\theta}_{ML}$ has the following asymptotic normal distribution:

$$\sqrt{T}(\hat{\theta}_{ML} - \theta) \longrightarrow N(0, (\bar{H})^{-1}), \tag{2.6}$$

where

$$-\frac{1}{T}\frac{\partial^2 lnL(\theta|\tilde{y}_T)}{\partial\theta\partial\theta'} \rightarrow \bar{H} = \lim \frac{1}{T}I(\theta).$$

For an easy proof of the Cramer-Rao inequality and the asymptotic normality of the maximum likelihood estimator, refer to Harvey 1990. Equation (2.6) suggests that the maximum likelihood estimator is consistent and asymptotically efficient in the sense that its covariance matrix reaches the Cramer-Rao lower bound. Equation (2.6) also provides us with an idea of how to estimate the covariance matrix of the maximum likelihood estimator, using the inverse of the negative of the second derivative of the log likelihood function (Hessian) evaluated at $\hat{\theta}_{ML}$:

$$\text{Cov}(\hat{\theta}_{ML}) = \left[-\frac{\partial^2 \ln L(\theta|\tilde{y}_T)}{\partial\theta\partial\theta'}\Big|_{\theta=\hat{\theta}_{ML}}\right]^{-1} \tag{2.7}$$

2.2 The Prediction Error Decomposition and the Likelihood Function

For maximum likelihood estimation, we need to derive the joint density function or the likelihood function (they are functionally equivalent) for the vector $\tilde{y}_T$, given a statistical model. For independent observations, its derivation is straightforward:

$$L(\theta \mid \tilde{y}_T) = \prod_{t=1}^{T} p(y_t \mid \theta), \tag{2.8}$$

where $p(y_t \mid \theta)$ is the marginal density of an individual observation. For dependent observations, products of the conditional densities allow us to achieve the same goal:

$$L(\theta \mid \tilde{y}_T) = \prod_{t=2}^{T} p(y_t \mid \tilde{y}_{t-1}, \theta) p(y_1 \mid \theta), \tag{2.9}$$

where $\tilde{y}_t = [\, y_1 \quad \ldots \quad y_t \,]$, $p(y_t \mid \tilde{y}_{t-1}, \theta)$, $t = 2, 3, \ldots, T$, is the conditional density, and $p(y_1 \mid \theta)$ is the marginal density of y_1. Notice that for the first observation, we have no information on which to condition.

However, the derivation of the conditional densities used in (2.9) for dependent observations may not always be straightforward. Consider, for example, the following unobserved-components model with normality assumptions:

$$y_t = x_t + e_t,\ e_t \sim \text{i.i.d.}N(0, \sigma_e^2) \quad t = 1, 2, \ldots, T, \tag{2.10}$$

$$x_t = \delta + \phi x_{t-1} + v_t,\ v_t \sim \text{i.i.d.}N(0, \sigma_v^2), \tag{2.11}$$

where e_t and v_t are independent and $\mid \phi \mid < 1$. The conditional density $p(y_t \mid \tilde{y}_{t-1}, \theta)$ is not directly obtained from the statistical model, where $\theta = [\delta \quad \sigma_e^2 \quad \sigma_v^2 \quad \phi]'$ in the present case. Taking advantage of the normality assumption, the vector of observations $\tilde{y}_T$ can be represented by the following multivariate normal distribution:

$$\tilde{y}_T \sim N(\mu, \Omega), \tag{2.12}$$

with the likelihood function:

$$L(\theta \mid \tilde{y}_T) = (2\pi)^{-\frac{T}{2}} \mid \Omega \mid^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\tilde{y}_T - \mu)'\Omega^{-1}(\tilde{y}_T - \mu)\}, \tag{2.13}$$

where all elements of μ and Ω are complicated functions of δ, σ_e, σ_v^2, and ϕ. Even when μ and Ω can be specified explicitly, maximizing the log of the likelihood function with respect to unknown parameters would be troublesome because of the inversion of the $T \times T$ matrix Ω. Harvey (1980) provides a solution to these difficulties, based on the prediction error decomposition obtained from applying the triangular factorization of the Ω matrix in (2.12).

Note that, for the $T \times T$ positive-definite, symmetric matrix Ω, there exists a unique triangular factorization of the following form:

$$\Omega = AfA', \tag{2.14}$$

where f is a diagonal matrix with positive elements and A is a lower triangular matrix with the following forms:

$$f = \begin{bmatrix} f_1 & 0 & 0 & \cdots & 0 \\ 0 & f_2 & 0 & \cdots & 0 \\ 0 & 0 & f_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & f_T \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_{21} & 1 & 0 & \cdots & 0 \\ a_{31} & a_{32} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{T1} & a_{T2} & a_{T3} & \cdots & 1 \end{bmatrix},$$

and where $f_t > 0$ for all t. Substituting (2.14) into (2.13), we have:

$$\begin{aligned} L(\theta \mid \tilde{y}_T) &= (2\pi)^{-\frac{T}{2}} \mid AfA' \mid^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\tilde{y}_T - \mu)'(AfA')^{-1}(\tilde{y}_T - \mu)\} \\ &= (2\pi)^{-\frac{T}{2}} \mid f \mid^{-\frac{1}{2}} \exp\{-\frac{1}{2}\eta' f^{-1}\eta\} \\ &= (2\pi)^{-\frac{T}{2}} \prod_{t=1}^{T} f_t^{-\frac{1}{2}} \exp\{-\frac{1}{2}\sum_{t=1}^{T}\eta_t' f_t^{-1}\eta_t\} \\ &= \prod_{t=1}^{T}\left[\frac{1}{\sqrt{2\pi f_t}} \exp\{-\frac{1}{2}\frac{\eta_t^2}{f_t}\}\right], \end{aligned} \tag{2.15}$$

where $\eta = A^{-1}(\tilde{y}_T - \mu)$ and η_t is the t-th element of the $T \times 1$ vector η. Because A is a lower trangular with ones in its diagonal elements, one can easily show that the t-th element of η can be rewritten as:

$$\eta_t = y_t - y_{t|t-1}, \tag{2.16}$$

where $y_{t|t-1}$ is the prediction of y_t conditional on $\tilde{y}_{t-1} = [\, y_1 \quad \ldots \quad y_{t-1}\,]'$, which is information up to $t-1$, since we have:

$$y_{t|t-1} = \sum_{i=1}^{t-1} a^*_{t,i} y_i, \quad t = 2, 3, \ldots, T, \tag{2.17}$$

where $a^*_{t,i}$ is the (t,i)-th element of A^{-1}. Notice that the argument in the bracket of the last line in (2.15) is a normal density function of y_t conditional on past information:

$$y_t \mid \tilde{y}_{t-1} \sim N(y_{t|t-1}, f_t), \tag{2.18}$$

where f_t is interepreted as the variance of the prediction error $\eta_t = y_t - y_{t|t-1}$.

To summarize, equations (2.15) and (2.18) suggest that when the observations are normally distributed, insofar as we have the prediction errors and their variances, the log likelihood value can be easily calculated. Thus, the success of maximum likelihood estimation for a complicated dynamic time series model with dependent observations may depend on the availability of the prediction errors and their variances. For the unobserved component model specified in equations (2.10) and (2.11), for example, the Kalman filter introduced in chapter 3 provides us with η_t and f_t used in the last line of equation (2.15). With normality assumptions, equation (2.15) is a general approach to deriving the likelihood function for dependent observations of a dynamic time series. For a multivariate case, it is easy to see that (2.15) is replaced by:

$$L(\theta \mid \tilde{y}_T) = \prod_{t=1}^{T} \left[\frac{1}{\sqrt{(2\pi)^n |f_t|}} \exp\{-\frac{1}{2}\eta_t' f_t^{-1} \eta_t\} \right], \tag{2.15$'$}$$

where n is the dimension of y_t, η_t is $n \times 1$, and f_t is $n \times n$.

2.3 Parameter Constraints and the Covariance Matrix of $\hat{\theta}_{ML}$

2.3.1 Constrained Optimization

The maximum likelihood estimator, $\hat{\theta}_{ML}$, can be obtained by setting the first derivative of the log likelihood function to 0:

$$\frac{\partial \ln L(\theta \mid \tilde{y}_T)}{\partial \theta} = 0. \tag{2.19}$$

In most cases, however, a closed-form solution for $\hat{\theta}_{ML}$ is not available. Thus, in general, we resort to a nonlinear numerical optimization procedure to maximize the log likelihood function. Given initial estimates (θ^{j-1}) of the parameters, new estimates (θ^{j}) are obtained using the information provided by the first derivatives (and sometimes, depending upon the algorithms employed, the second derivatives) of the log likelihood function evaluated at θ^{j-1}. New estimates are obtained such that the log likelihood value evaluated at the revised estimates is larger than that at the initial estimates. This process may be iterated until convergence is achieved to obtain the value of the parameters that maximize the log likelihood function. In some cases, the maximum may not be unique. For specific and easy expositions of various algorithms for numerical optimization, readers are referred to chapter 4 of Harvey 1990.

When numerical optimization is employed to maximize the log likelihood function with respect to θ, the computer searches over the parameter space that ranges between negative infinity and positive infinity. But some of the parameters may have to be constrained to lie in an interval. For example, if one of the elements in θ is a probability (p), then it must be constrained such that $0 < p < 1$. In general, such constraints may be imposed by the following transformations of a vector ψ that ranges between negative infinity and positive infinity:

$$\theta = g(\psi), \tag{2.20}$$

where $g(.)$ is a continuous function. Then the log likelihood function may be considered a function of ψ:

$$\ln L(\theta \mid \tilde{y}_T) = \ln L(g(\psi) \mid \tilde{y}_T) = \ln L(\psi \mid \tilde{y}_T), \tag{2.21}$$

and the unconstrained numerical optimization may be applied with respect to ψ.

For example, if θ_j, the j-th element of θ, represents a variance, then $\theta_j > 0$. Then we may use the transformations

$$\theta_j = \psi_j^2 \quad \text{or} \quad \theta_j = \exp(\psi_i).$$

If θ_j represents a probability term, then $0 < \theta_j < 1$. The transformation we may employ is

$$\theta_j = \frac{1}{1 + \exp(\psi_j^{-1})}.$$

If θ_j represents an autoregressive parameter in an AR(1) model, then we may want to constrain the parameter within the stationary region $-1 < \psi_j < +1$:

$$\theta_j = \frac{\psi_j}{1+|\psi_j|}.$$

If $\theta = [\phi_1 \quad \phi_2]'$, where ϕ_1 and ϕ_2 are the autoregressive coefficients of the model in an AR(2) model, we may want to constrain the values of ϕ_1 and ϕ_2 within the stationary region (roots of $(1 - \phi_1 L - \phi_2 L^2) = 0$ lie outside the unit circle). In this case, we may employ the transformation

$$z_1 = \frac{\psi_1}{1+|\psi_1|}, \quad z_2 = \frac{\psi_2}{1+|\psi_2|},$$

$$==> \phi_1 = z_1 + z_2, \quad \phi_2 = -1 * z_1 * z_2.$$

Notice, however, that the recommended procedure in fact imposes the further restriction that the roots of the AR(2) polynomial are real. Finally, consider the following generalized autoregressive conditional hetereskedasticity (GARCH)(1,1) model:

$$h_t = a_0 + a_1 e_{t-1}^2 + a_2 h_{t-1}.$$

We generally want $a_1 > 0$, $a_2 > 0$, and $0 < a_1 + a_2 < 1$. The following transformations achieve this goal:

$$a_1 = \frac{\exp(\psi_1)}{1 + \exp(\psi_1) + \exp(\psi_2)}, \quad a_2 = \frac{\exp(\psi_2)}{1 + \exp(\psi_1) + \exp(\psi_2)}.$$

2.3.2 Constrained Optimization and the Covariance Matrix of $\hat{\theta}_{ML}$

In section 2.3.1, we noted that applying *unconstrained* optimization to the log likelihood function in (2.18) with respect to ψ is equivalent to applying *constrained* optimization with respect to θ, the parameter of interest to us. Unconstrained optimization then results in $\hat{\psi}_{ML}$ and $\mathrm{Cov}(\hat{\psi}_{ML})$, the maximum likelihood (ML) estimate of ψ and its covariance matrix. But we actually want the parameter estimates and the covariance matrix for θ. As $\theta = g(\psi)$, the ML estimate for θ is easily obtained by

$$\hat{\theta}_{ML} = g(\hat{\psi}_{ML}). \qquad (2.22)$$

We can also obtain $\mathrm{Cov}(\hat{\theta}_{ML})$ based on $\mathrm{Cov}(\hat{\psi}_{ML})$ and $g(.)$ in the following way:

$$\mathrm{Cov}(\hat{\theta}_{ML}) = \left(\frac{\partial g(\hat{\psi}_{ML})}{\partial \psi}\right) \mathrm{Cov}(\hat{\psi}_{ML}) \left(\frac{\partial g(\hat{\psi}_{ML})}{\partial \psi}\right)'. \tag{2.23}$$

The following provides a proof of equation (2.23). Differentiating the log likelihood function $\ln L(\theta) = \ln L(g(\psi))$ with respect to ψ, we get

$$\frac{\partial \ln L(g(\psi))}{\partial \psi} = \frac{\partial \ln L(g(\psi))}{\partial \theta} \left(\frac{\partial g(\psi)}{\partial \psi}\right), \tag{2.24}$$

and then differentiating again,

$$\begin{aligned}&\left(\frac{\partial^2 \ln L(g(\psi))}{\partial \psi \partial \psi'}\right) \\ &= \left(\frac{\partial g(\psi)}{\partial \psi}\right)' \frac{\partial^2 \ln L(g(\psi))}{\partial \theta \partial \theta'} \left(\frac{\partial g(\psi)}{\partial \psi}\right) + \frac{\partial \ln L(g(\psi))}{\partial \psi} \left(\frac{\partial^2 g(\psi)}{\partial \psi \partial \psi'}\right).\end{aligned} \tag{2.25}$$

As

$$\frac{\partial \ln L(g(\psi_{ML}))}{\partial \psi} = 0,$$

(2.25) is written as

$$\begin{aligned}&\left(\frac{\partial^2 \ln L(g(\hat{\psi}_{ML}))}{\partial \psi \partial \psi'}\right) \\ &= \left(\frac{\partial g(\hat{\psi}_{ML})}{\partial \psi}\right)' \frac{\partial^2 \ln L(g(\hat{\psi}_{ML}))}{\partial \theta \partial \theta'} \left(\frac{\partial g(\hat{\psi}_{ML})}{\partial \psi}\right)\end{aligned} \tag{2.26}$$

Multiplying both sides of (2.26) by -1 and then taking the inverse of both sides, we have

$$\left(-\frac{\partial^2 \ln L(g(\hat{\psi}_{ML}))}{\partial\psi\partial\psi'}\right)^{-1}$$

$$= \left(\frac{\partial g(\hat{\psi}_{ML})}{\partial\psi}\right)^{-1}\left(-\frac{\partial^2 \ln L(g(\hat{\psi}_{ML}))}{\partial\theta\partial\theta'}\right)^{-1}\left(\frac{\partial g(\hat{\psi}_{ML})'}{\partial\psi}\right)^{-1}. \tag{2.27}$$

Arranging the terms in (2.27) and noting that

$$\text{Cov}(\hat{\theta}_{ML}) = (-\frac{\partial^2 \ln L(\hat{\theta}_{ML})}{\partial\theta\partial\theta'})^{-1}$$

and

$$\text{Cov}(\hat{\psi}_{ML}) = (-\frac{\partial^2 \ln L(\hat{\psi}_{ML})}{\partial\psi\partial\psi'})^{-1},$$

we get equation (2.23).

References

Cramer, J. S. 1986. *Econometric Applications of Maximum Likelihood Methods*. Cambridge: Cambridge University Press.

Davidson, Russell, and James G. MacKinnon. 1993. *Estimation and Inference in Econometrics*. Oxford, UK: Oxford University Press.

Harvey, Andrew C. 1981. *Time Series Models*. Oxford: Philip Allan and Humanities Press.

Harvey, Andrew C. 1990. *The Econometric Analysis of Time Series*. 2nd ed. Cambridge: MIT Press.

Judge, G. G., R. C. Hill, W. E. Griffiths, H. Lutkepohl, and T.-C. Lee. 1982. *Introduction to the Theory and Practice of Econometrics*. New York: John Wiley & Sons.

3 State-Space Models and the Kalman Filter

State-space models, which typically deal with dynamic time series models that involve unobserved variables, have a wide range of potential applications in econometrics, since economic theory often involves unobservable variables—for example, permanent income, expectations, the ex ante real rate of interest, and the reservation wage. Engle and Watson (1981) apply them to modeling the behavior of wage rates; Garbade and Wachtel (1978) and Antoncic (1986) apply them to modeling the behavior of ex ante real interest rates; Burmeister and Wall (1982) and Burmeister, Wall, and Hamilton (1986) apply it in estimating expected inflation; and Kim and Nelson (1989) apply it to modeling a time-varying monetary reaction function of the Federal Reserve. Stock and Watson's (1991) dynamic factor model of coincident economic indicators is a recent application of state-space models. For more surveys and applicability of state-space models, refer to Engle and Watson 1987; Harvey 1985, 1989, and 1990; and Hamilton 1994a and 1994b.

The basic tool used to deal with the standard state-space model is the Kalman filter, a recursive procedure for computing the estimator of the unobserved component or the state vector at time t, based on available information at time t. When the shocks to the model and the initial unobserved variables are normally distributed, the Kalman filter also enables the likelihood function to be calculated via the prediction error decomposition discussed in chapter 2.

This chapter reviews the state-space model and the Kalman filter with reference to applications, beginning with the time-varying-parameter model in section 3.1. We then discuss general state-space models in section 3.2, with a focus on the unobserved-components model. Sections 3.3 through 3.5 deal with specific applications of the state-space models and the Kalman filter to actual economic problems.

3.1 Time-Varying-Parameter Models and the Kalman Filter

Consider the following regression model, in which the regression coefficients are time varying with specific dynamics:

$$y_t = x_t\beta_t + e_t, \quad t = 1, 2, 3, \ldots, T \tag{3.1}$$

$$\beta_t = \tilde{\mu} + F\beta_{t-1} + v_t, \tag{3.2}$$

$$e_t \sim \text{i.i.d.}N(0, R), \tag{3.3}$$

$$v_t \sim \text{i.i.d.}N(0, Q), \tag{3.4}$$

where y_t is 1×1; x_t is a $1 \times k$ vector of exogenous or predetermined variables; and e_t and v_t are independent. We further assume that β_t is of dimension $k \times 1$; F is $k \times k$; and Q is $k \times k$. For example, with $\tilde{\mu} = 0$ and $F = I_k$, each regression coefficient in β_t follows a random walk. If F is a diagonal matrix, and the absolute values of its diagonal elements are less than 1, each regression coefficient follows a stationary AR(1) process. An extension of the model to a more general case is straightforward once one is acquainted with the general state-space model presented in section 3.2.

In sections 3.1.1 and 3.1.2, we consider two alternative ways of making inferences about β_t conditional on information available up to time t, assuming that all the hyperparameters of the model ($\tilde{\mu}$, F, R, and Q) are known. In section 3.1.1, we show that a sequence of generalized least squares (GLS) regressions enables us to achieve our goal. However, this method may be extremely inefficient in terms of its computational burden. This motivates us to consider making inferences about β_t by employing the Kalman filter in section 3.1.2. If some of these hyperparameters are not known, however, they have to be estimated first before making inferences on β_t. Section 3.1.3 discusses maximum likelihood estimation of the model's unknown hyperparameters.

3.1.1 GLS Estimation of β_t

For simplicity of the analysis, assume that $\tilde{\mu} = 0$. Then, from equation (3.2), we get

$$\begin{aligned}
\beta_t &= F\beta_{t-1} + v_t \\
&= F^2\beta_{t-2} + Fv_{t-1} + v_t \\
&\quad \cdots \\
&= F^{t-2}\beta_2 + F^{t-3}v_3 + F^{t-4}v_4 + \ldots + Fv_{t-1} + v_t \\
&= F^{t-1}\beta_1 + F^{t-2}v_2 + F^{t-3}v_3 + \ldots + Fv_{t-1} + v_t.
\end{aligned} \tag{3.5}$$

Using the above equation, we can solve $\beta_1, \beta_2, \ldots, \beta_{t-1}$ as functions of β_t and $v_t, v_{t-1}, v_{t-1}, \ldots$.

$$
\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{t-1} \\ \beta_t \end{bmatrix}
=
\begin{bmatrix} F^{-t+1}\beta_t - (F^{-1}v_2 + F^{-2}v_3 + \ldots + F^{-t+1}v_t) \\ F^{-t+2}\beta_t - (F^{-1}v_3 + F^{-2}v_4 + \ldots + F^{-t+2}v_t) \\ \vdots \\ F^{-1}\beta_t - F^{-1}v_t \\ \beta_t \end{bmatrix}
$$

$$
= \begin{bmatrix} F^{-t+1}\beta_t \\ F^{-t+2}\beta_t \\ \vdots \\ F^{-1}\beta_t \\ \beta_t \end{bmatrix}
- \begin{bmatrix} (F^{-1}v_2 + F^{-2}v_3 + \ldots + F^{-t+1}v_t) \\ (F^{-1}v_3 + F^{-2}v_4 + \ldots + F^{-t+2}v_t) \\ \vdots \\ F^{-1}v_t \\ 0 \end{bmatrix}. \tag{3.6}
$$

From (3.1) and (3.6), we get:

$$
\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{t-1} \\ y_t \end{bmatrix}
=
\begin{bmatrix} x_1F^{-t+1} \\ x_2F^{-t+2} \\ \vdots \\ x_{t-1}F^{-1} \\ x_t \end{bmatrix} \beta_t
- \begin{bmatrix} x_1F^{-1} & x_1F^{-2} & \ldots & x_1F^{-t+1} \\ 0 & x_2F^{-1} & \ldots & x_2F^{-t+2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & x_{t-1}F^{-1} \\ 0 & 0 & \ldots & 0 \end{bmatrix}
\begin{bmatrix} v_2 \\ \vdots \\ v_{t-1} \\ v_t \end{bmatrix}
+ \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_{t-1} \\ e_t \end{bmatrix}. \tag{3.7}
$$

Writing equation (3.7) in matrix notation, we have:

$$\tilde{y}_t = \tilde{X}_t^* \beta_t + \tilde{\epsilon}_t, \tag{3.7$'$}$$

where

$$E[\tilde{\epsilon}_t\tilde{\epsilon}_t'] = A_t(I_{t-1} \otimes Q)A_t' + RI_t = \Omega_t, \tag{3.8}$$

and where

$$
A_t = \begin{bmatrix} x_1F^{-1} & x_1F^{-2} & \ldots & x_1F^{-t+1} \\ 0 & x_2F^{-1} & \ldots & x_2F^{-t+2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{bmatrix}.
$$

One could apply GLS to the model (3.7)′ for $t = k+1, \ldots, T$, where k is the dimension of β_t. A major difficulty, however, is that each GLS application requires an inverse of a $t \times t$ matrix: simple

$$\beta_{t|t} = (\tilde{X}_t^{*\prime}\Omega_t^{-1}\tilde{X}_t^{*})^{-1}\tilde{X}_t^{*\prime}\Omega_t^{-1}\tilde{y}_t, \qquad t = k+1, k+2, \ldots, T, \tag{3.9}$$

where $\beta_{t|t}$ refers to an estimate of β_t conditional on information up to time t.

The Kalman filter approach discussed below can easily be implemented without having to invert such large matrices.

3.1.2 The Kalman Filter and Estimation of β_t

The Kalman filter is a recursive procedure for computing the optimal estimate of the unobserved-state vector β_t, $t = 1, 2, \ldots, T$, based on the appropriate information set, assuming that $\tilde{\mu}$, F, R, and Q are known. It provides a minimum mean squared error estimate of β_t given the appropriate information set. Depending upon the information set used, we have the *basic filter* and *smoothing*. The basic filter refers to an estimate of β_t based on information available up to time t, and smoothing to an estimate of β_t based on all the available information in the sample through time T.

Throughout this book, we use the following notation:

ψ	the information set.
$\beta_{t\mid t-1} = E[\beta_t\mid\psi_{t-1}]$	expectation (estimate) of β_t conditional on information up to $t-1$.
$P_{t\mid t-1} = E[(\beta_t - \beta_{t\mid t-1})(\beta_t - \beta_{t\mid t-1})']$	covariance matrix of β_t conditional on information up to $t-1$.
$\beta_{t\mid t} = E[\beta_t\mid\psi_t]$	expectation (estimate) of β_t conditional on information up to t.
$P_{t\mid t} = E[(\beta_t - \beta_{t\mid t})(\beta_t - \beta_{t\mid t})']$	covariance matrix of β_t conditional on information up to t.
$y_{t\mid t-1} = E[y_t\mid\psi_{t-1}] = x_t\beta_{t\mid t-1}$	forecast of y_t given information up to time $t-1$.
$\eta_{t\mid t-1} = y_t - y_{t\mid t-1}$	prediction error.
$f_{t\mid t-1} = E[\eta_{t\mid t-1}^2]$	conditional variance of the prediction error.

$\beta_{t|T} = E[\beta_t|\psi_T]$ expectation (estimate) of β_t conditional on information up to T (the whole sample).

$P_{t|T} = E[(\beta_t - \beta_{t|T})(\beta_t - \beta_{t|T})']$ covariance matrix of β_t conditional on information up to T (the whole sample).

Assuming that x_t is available at the beginning of time t and a new observation of y_t is made at the end of time t, the Kalman filter (basic filter) consists of the following two steps:

1. *Prediction:* At the beginning of time t, we may want to form an optimal predictor of y_t, based on all the available information up to time $t-1$: $y_{t|t-1}$. To do this, we need to calculate $\beta_{t|t-1}$.

2. *Updating:* Once y_t is realized at the end of time t, the prediction error can be calculated: $\eta_{t|t-1} = y_t - y_{t|t-1}$. This prediction error contains new information about β_t beyond that contained in $\beta_{t|t-1}$. Thus, after observing y_t, a more accurate inference can be made of β_t. $\beta_{t|t}$, an inference of β_t based on information up to time t, may be of the following form: $\beta_{t|t} = \beta_{t|t-1} + K_t\eta_{t|t-1}$, where K_t is the weight assigned to new information about β_t contained in the prediction error.

To be more specific, the basic filter is described by the following six equations:

Prediction

$$\beta_{t|t-1} = \tilde{\mu} + F\beta_{t-1|t-1}, \tag{3.10}$$

$$P_{t|t-1} = FP_{t-1|t-1}F' + Q, \tag{3.11}$$

$$\eta_{t|t-1} = y_t - y_{t|t-1} = y_t - x_t\beta_{t|t-1}, \tag{3.12}$$

$$f_{t|t-1} = x_tP_{t|t-1}x_t' + R, \tag{3.13}$$

Updating

$$\beta_{t|t} = \beta_{t|t-1} + K_t\eta_{t|t-1}, \tag{3.14}$$

$$P_{t|t} = P_{t|t-1} - K_tx_tP_{t|t-1}, \tag{3.15}$$

where $K_t = P_{t|t-1}x_t'f_{t|t-1}^{-1}$ is the Kalman gain, which determines the weight assigned to new information about β_t contained in the prediction error. Given the hyperparameters of the model, $\beta_{t|t}$ in equation (3.14) is the same as that in equation (3.9).

Derivation of the four equations in (3.10)–(3.13) is straightforward. To derive the smoothing equations in (3.14)–(3.15), we consider the following arguments. Let Z_1 and Z_2, conditional on ψ_{t-1}, be normally distributed as follows:

$$\begin{matrix} Z_1 \\ Z_2 \end{matrix} \mid \psi_{t-1} \sim MVN\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right). \tag{3.16}$$

Then the distribution of Z_1 given Z_2 and ψ_{t-1} is given by

$$Z_1|Z_2, \psi_{t-1} \sim N(\mu_{1|2}, \Sigma_{11|2}), \tag{3.17}$$

where

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(Z_2 - \mu_2), \tag{3.18}$$

$$\Sigma_{11|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \tag{3.19}$$

Thus if we let $Z_1 = \beta_t$ and $Z_2 = y_t - y_{t|t-1} = \eta_{t|t-1}$, then we have $\mu_1 = \beta_{t|t-1}$, $\Sigma_{11} = P_{t|t-1}$, $\Sigma_{22} = f_{t|t-1}$, and $\Sigma_{12} = P_{t|t-1}x_t'$. Therefore, equations (3.14) and (3.15) result. For more details on the derivation of the Kalman filter, refer to Hamilton 1993 and 1994a,b.

In equation (3.10), an inference on β_t given information up to time $t-1$ is a function of an inference on β_{t-1} given information up to time $t-1$, due to equation (3.2). Thus uncertainty underlying $\beta_{t|t-1}$ is a function of uncertainty underlying $\beta_{t-1|t-1}$ and Q, the covariance of the shocks to β_t. This is shown in equation (3.11). The prediction error in the time-varying-parameter model consists of two parts: the prediction error due to error in making an inference about β_t (i.e., $\beta_t - \beta_{t|t-1}$) and the prediction error due to e_t, a random shock to y_t in (3.1). Thus in equation (3.13), the conditional variance of the prediction error is a function of the uncertainty associated with $\beta_{t|t-1}$ and of R, the variance of e_t.

The updating equation in (3.14) suggests that $\beta_{t|t}$ is formed as a kind of weighted average of $\beta_{t|t-1}$ and new information contained in the prediction error $\eta_{t|t-1}$, the weight assigned to new information being the Kalman gain, K_t. Examining the Kalman gain more carefully, we notice that it is an inverse function of R, the variance of e_t; and given x_t, it is a positive function of the uncertainty underlying $\beta_{t|t-1}$. For simplicity, assume that β_t and x_t are 1×1. Then the Kalman gain can be rewritten as

Model: $y_t = \beta_t x_t + e_t,\ \ e_t \sim \text{i.i.d. } N(0, \sigma_e^2)$
$\beta_t = \mu + F\beta_{t-1} + v_t,\ \ v_t \sim \text{i.i.d. } N(0, \sigma_v^2)$

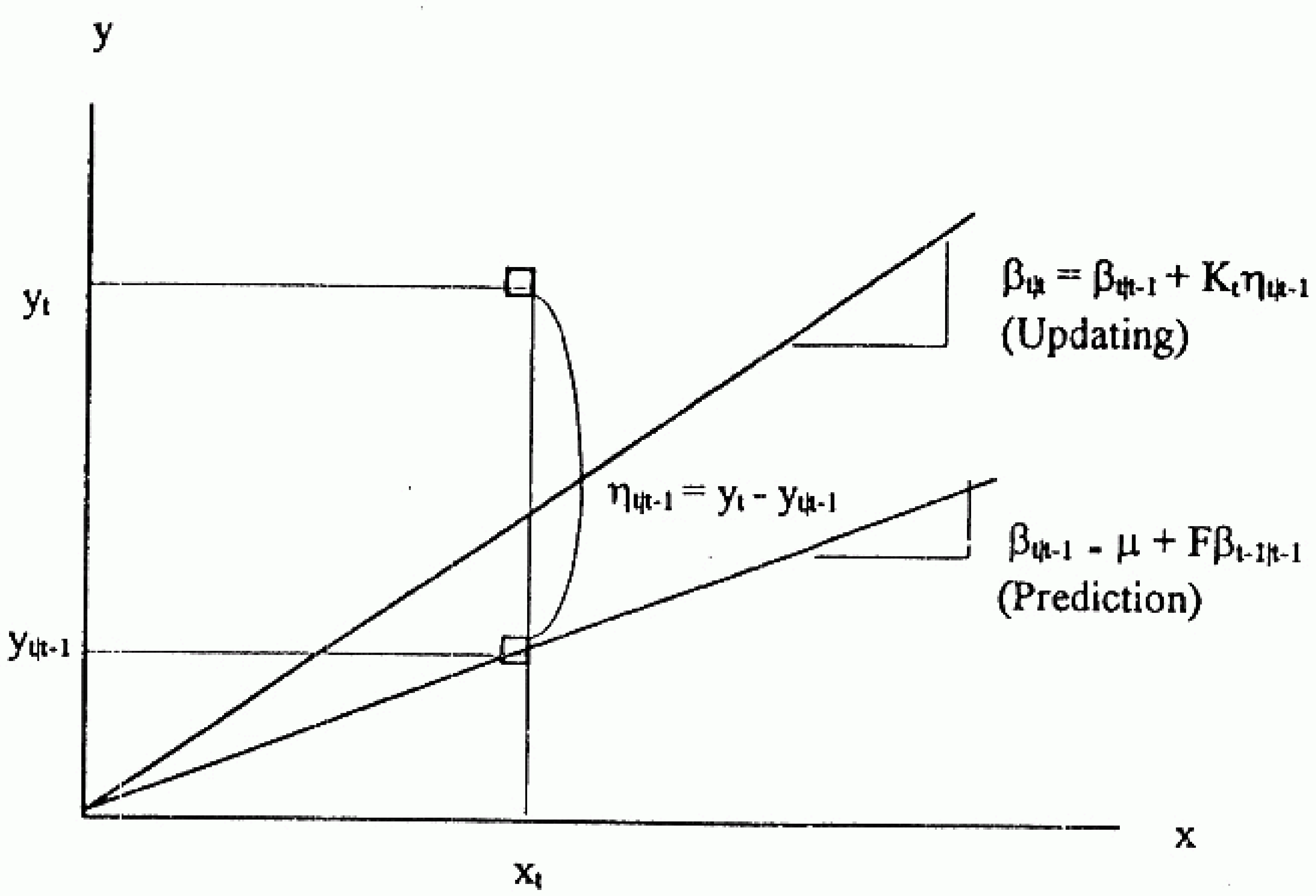

Figure 3.1
The Kalman filter: TVP model

$$K_t = \frac{1}{x_t}\frac{P_{t|t-1}x_t^2}{P_{t|t-1}x_t^2 + R}, \tag{3.20}$$

where $P_{t|t-1}x_t^2$ is the portion of the prediction error variance due to uncertainty in $\beta_{t|t-1}$ and R is the portion of the prediction error variance due to random shock e_t. We can easily see that

$$\left|\frac{\partial K_t}{\partial(P_{t|t-1}x_t^2)}\right| > 0,$$

suggesting that as uncertainty associated with $\beta_{t|t-1}$ increases, relatively more weight is given to new information in the prediction error, $\eta_{t|t-1}$. This is quite intuitive, since an increase in uncertainty in $\beta_{t|t-1}$ may be interpreted as a deterioration of the information content of $\beta_{t|t-1}$, relative to that of $\eta_{t|t-1}$.

The above prediction and updating steps of the Kalman filter may be understood more intuitively with the help of figure 3.1. The horizontal axis in the

$$\beta_{0|0},\ P_{0|0},\ l(\theta)=0$$

$$\Downarrow$$

$$\Downarrow$$

$$\beta_{t|t-1} = \tilde{\mu} + F\beta_{t-1|t-1}$$

$$P_{t|t-1} = FP_{t-1|t-1}F' + Q$$

$$\Downarrow$$

$$\eta_{t|t-1} = y_t - y_{t|t-1} = y_t - H_t\beta_{t|t-1} - Az_t$$

$$f_{t|t-1} = H_t P_{t|t-1} H_t' + R$$

$$\Downarrow$$

$$l(\theta) = l(\theta) - \frac{1}{2}\ln((2\pi)^n \,|\, f_{t|t-1} \,|) - \frac{1}{2}\eta_{t|t-1}' f_{t|t-1}^{-1} \eta_{t|t-1}$$

$$\Downarrow$$

$$\beta_{t|t} = \beta_{t|t-1} + P_{t|t-1} H_t' f_{t|t-1}^{-1} \eta_{t|t-1}$$

$$P_{t|t} = P_{t|t-1} - P_{t|t-1} H_t' f_{t|t-1}^{-1} H_t P_{t|t-1}$$

$$\Downarrow \qquad \text{t - 1 = t}$$

$$\Downarrow$$

$$l(\theta) = -\frac{1}{2}\sum \ln((2\pi)^n \,|\, f_{t|t-1} \,|) - \frac{1}{2}\sum \eta_{t|t-1}' f_{t|t-1}^{-1} \eta_{t|t-1}$$

Figure 3.2
Flowchart for the Kalman filter

figure represents the x variable available at the beginning of each period, and the vertical axis represents the y variable realized at the end of each period.

Given the initial values, $\beta_{0|0}$ and $P_{0|0}$, the six equations in the basic filter can be iterated for $t = 1, 2, \ldots, T$. (Refer to figure 3.2) This provides us with a minimum mean squared error estimate of β_t, $t = 1, 2, \ldots, T$, given information up to time $t-1$ or t. For stationary β_t in equation (3.2), the unconditional mean and covariance matrix of β_t may be employed as the initial values, $\beta_{0|0}$ and $P_{0|0}$. The unconditional mean of stationary β_t is derived as

$$E[\beta_t] = \tilde{\mu} + FE[\beta_{t-1}] + E[v_t], \tag{3.21}$$

$$\beta_{0|0} = \tilde{\mu} + F\beta_{0|0} \quad \text{(At Steady State)}, \tag{3.21$'$}$$

$$\beta_{0|0} = (I_k - F)^{-1}\tilde{\mu}. \tag{3.22}$$

The unconditional covariance matrix of stationary β_t is derived as

$$\mathrm{Cov}(\beta_t) = F\,\mathrm{Cov}(\beta_{t-1})F' + \mathrm{Cov}(v_t), \tag{3.23}$$

$$P_{0|0} = FP_{0|0}F' + Q, \quad \text{(At Steady-State)}, \tag{3.23$'$}$$

$$\mathrm{vec}(P_{0|0}) = \mathrm{vec}(FP_{0|0}F') + \mathrm{vec}(Q), \tag{3.24}$$

$$\mathrm{vec}(P_{0|0}) = (F \otimes F)\mathrm{vec}(P_{0|0}) + \mathrm{vec}(Q), \tag{3.25}$$

$$\mathrm{vec}(P_{0|0}) = (I - F \otimes F)^{-1}\mathrm{vec}(Q), \tag{3.26}$$

as $\mathrm{vec}(ABC) = (C' \otimes A)\mathrm{vec}(B)$.

For nonstationary β_t in equation (3.2), the unconditional mean and the covariance matrix of β_t do not exist. In this case, $\beta_{0|0}$ may be set at any arbitrary $k \times 1$ vector (wild guessing). But in order to assign very large uncertainty to this wild guess, we must assign very large values to the diagonal elements of $P_{0|0}$. For example, if $P_{t-1|t-1}$ is a very large positive definite matrix, most of the weight in the updating equation (3.14) is assigned to new information contained in the forecast error for y_t, and the information content in $\beta_{t|t-1}$ is treated as negligible. Alternatively, if we assume that $\beta_{0|0}$ is a vector of unknown constants, we can treat the elements of $\beta_{0|0}$ as additional parameters to be estimated. In this case, $P_{0|0}$ should be set equal to a $k \times k$ matrix of 0s when estimating the hyperparameters of the model via MLE, because $\beta_{0|0}$ is not a random variable. Once these parameters are estimated along with other hyperparameters of the model, we can run the Kalman filter again by setting $\beta_{0|0} = \hat{\beta}_{0|0,\mathrm{MLE}}$ and $P_{0|0} = \mathrm{Cov}(\hat{\beta}_{0|0,\mathrm{MLE}})$ for inferences on β_t, $t = 1, 2, \ldots, T$.

Smoothing ($\beta_{t|T}$) provides us with a more accurate inference on β_t, since it uses more information than the basic filter. The following two equations can be iterated backwards for $t = T - 1, T - 2, \ldots 1$, to get the smoothed estimates:

Smoothing

$$\beta_{t|T} = \beta_{t|t} + P_{t|t}F'P_{t+1|t}^{-1}(\beta_{t+1|T} - F\beta_{t|t} - \tilde{\mu}), \quad (3.2$$

$$P_{t|T} = P_{t|t} + P_{t|t}F'P_{t+1|t}^{-1}(P_{t+1|T} - P_{t+1|t})P_{t+1|t}^{-1\,\prime}FP_{t|t}', \quad (3.2$$

where $\beta_{T|T}$ and $P_{T|T}$, the initial values for the smoothing, are obtained fro the last iteration of the basic filter. Equations (3.27) and (3.28) can be deriv in the same way as the updating equations (3.14) and (3.15).

3.1.3 Maximum Likelihood Estimation of the Model Based on the Prediction Error Decomposition

The discussion in the previous section assumes that the model's paramete are known. However, some of these parameters are usually unknown. In th case, we need to estimate the parameters first; then the estimate of β_t, t $1, 2, \ldots, T$, is conditional on these estimated parameters. In chapter 2, w discussed the evaluation of the likelihood function based on the prediction err decomposition. That is, when the observations are normally distributed, insof as we have the prediction error and its variance, the log likelihood value can b calculated easily. For given parameters of the model, the Kalman filter provid us with prediction error ($\eta_{t|t-1}$) and its variance ($f_{t|t-1}$) in equations (3.1 and (3.13) as by-products. In addition, if β_0 and $\{e_t, v_t\}_{t=1}^{T}$ are Gaussian, th distribution of y_t conditional on ψ_{t-1} is also Gaussian:

$$y_t|\psi_{t-1} \sim N(y_{t|t-1}, f_{t|t-1}), \quad (3.29$$

and the sample log likelihood function is represented by

$$\ln L = -\frac{1}{2}\sum_{t=1}^{T} \ln(2\pi f_{t|t-1}) - \frac{1}{2}\sum_{t=1}^{T} \eta_{t|t-1}' f_{t|t-1}^{-1} \eta_{t|t-1}, \quad (3.30$$

which can be maximized with respect to unknown parameters of the model.

For nonstationary β_t in (3.2), the log likelihood function is evaluated fron observation $\tau + 1$ ($\tau \gg 1$):

$$\ln L = -\frac{1}{2}\sum_{t=\tau+1}^{T} \ln(2\pi f_{t|t-1}) - \frac{1}{2}\sum_{t=\tau+1}^{T} \eta_{t|t-1}' f_{t|t-1}^{-1} \eta_{t|t-1}, \quad (3.31$$

where τ is large enough. Notice that we start the Kalman filter with an arbitrary initial value $\beta_{0|0}$ and $P_{0|0}$, with large diagonal elements for nonstationary β_t. Iterating the filter starting from $t = 1$, then evaluating the log likelihood function from $t = \tau + 1$ minimizes the effect of the arbitrary initial value $\beta_{0|0}$ on the log likelihood value.

3.2 State-Space Models and the Kalman Filter

3.2.1 State-Space Models

State-space models, which were originally developed by control engineers (Kalman 1960), are useful tools for expressing *dynamic systems that involve unobserved state variables*. The reader is referred to Harvey 1989 and Hamilton 1994a and 1994b for other expositions. A state-space model consists of two equations: a transition equation (sometimes called a state equation) and a measurement equation.

Measurement Equation: An equation that describes the relation between observed variables (data) and unobserved state variables.

Transition Equation: An equation that describes the dynamics of the state variables. The transition equation has the form of a *first-order difference equation* in the state vector.

Consider the following representative state-space model:

Measurement Equation

$$y_t = H_t\beta_t + Az_t + e_t, \tag{3.32}$$

Transition Equation

$$\beta_t = \tilde{\mu} + F\beta_{t-1} + v_t, \tag{3.33}$$

$$e_t \sim \text{i.i.d.}N(0, R), \tag{3.34}$$

$$v_t \sim \text{i.i.d.}N(0, Q), \tag{3.35}$$

$$E(e_t v_s') = 0, \tag{3.36}$$

where y_t is an $n \times 1$ vector of variables observed at time t; β_t is a $k \times 1$ vector of unobserved state variables; H_t is an $n \times k$ matrix that links the observed y_t vector and the unobserved β_t; z_t is an $r \times 1$ vector of exogenous or predetermined observed variables; $\tilde{\mu}$ is $k \times 1$; v_t is $k \times 1$. Elements of the F matrix may be either data on exogenous variables or constant parameters. The positive-definiteness of R and Q in (3.34) and (3.35) is not always guaranteed. A usual practice is to write equations (3.33) and (3.35) alternatively as:

$$\beta_t = \tilde{\mu} + F\beta_{t-1} + Gv_t^*, \tag{3.33'}$$

$$v_t^* \sim \text{i.i.d.}N(0, Q^*), \tag{3.35'}$$

where G is $k \times g$; and v_t^* is $g \times 1$ $(g \leq k)$. In this representation of the transition equation, the positive definiteness of Q^* is guaranteed, and the relationship between Q in (3.35) and Q^* in (3.35)′ is given by $Q = GQ^*G'$.

In econometrics, the state-space representation of a general dynamic linear model includes autogressive integrated moving-average processes and classical regression models as special cases. Examples include

An AR(2) Model

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + w_t, \quad w_t \sim \text{i.i.d.}N(0, \sigma^2), \tag{3.37}$$

Measurement Equation

$$y_t = \delta^* + [\,1 \quad 0\,]\begin{bmatrix} \beta_{0t} \\ \beta_{0,t-1} \end{bmatrix}, \tag{3.38}$$

Transition Equation

$$\begin{bmatrix} \beta_{0t} \\ \beta_{0,t-1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \beta_{0,t-1} \\ \beta_{0,t-2} \end{bmatrix} + \begin{bmatrix} w_t \\ 0 \end{bmatrix}, \tag{3.39}$$

where

$$\delta^* = \frac{\delta}{1 - \phi_1 - \phi_2}.$$

An ARMA(2,1) model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + w_t + \theta w_{t-1}, \tag{3.40}$$

Measurement Equation

$$y_t = [\,1 \quad \theta\,]\begin{bmatrix} \beta_{1,t} \\ \beta_{2,t} \end{bmatrix}, \tag{3.41}$$

Transition Equation

$$\begin{bmatrix} \beta_{1,t} \\ \beta_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \beta_{1,t-1} \\ \beta_{2,t-1} \end{bmatrix} + \begin{bmatrix} w_t \\ 0 \end{bmatrix}. \tag{3.42}$$

An Unobserved-Components Model A typical application of the unobserved components model would be to decompose the log of real GDP into two independent components: a stochastic trend component and a cyclical component:

$$y_t = y_{1t} + y_{2t}, \tag{3.43}$$

$$y_{1t} = \delta + y_{1,t-1} + e_{1t}, \tag{3.44}$$

$$y_{2t} = \phi_1 y_{2,t-1} + \phi_2 y_{2,t-2} + e_{2t}, \tag{3.45}$$

$$e_{it} \sim \text{i.i.d.} N(0, \sigma_i^2), \quad i = 1, 2, \qquad E[e_{1t} e_{2s}] = 0 \quad \text{for all } t \text{ and } s, \tag{3.46}$$

where the roots of $(1 - \phi_1 L - \phi_2 L^2) = 0$ lie outside the unit circle.

Usually, there exists more than one way of writing a dynamic system in state-space models. In the current example, depending on which one of y_{1t} and y_{2t} (or both) may be treated as the state variable(s), we have at least three different representations. (Alternative representations of a time series imply the same observable moments of the time series.)

Representation 1 Suppose we want to treat y_{2t} as an unobserved state variable. In this case, we may want to transform the model so that y_{1t} does not show up in the model. This is done by taking a first difference of y_t in (3.43). Then we could design the measurement and transition equations based on the transformed model:

$$\Delta y_t = \Delta y_{1t} + \Delta y_{2t} \quad ==> \quad \Delta y_t = \delta + \Delta y_{2t} + e_{1t}, \tag{3.47}$$

Measurement Equation

$$\Delta y_t = \delta + [\,1 \quad -1\,] \begin{bmatrix} y_{2t} \\ y_{2,t-1} \end{bmatrix} + e_{1t}, \tag{3.48}$$

Transition Equation

$$\begin{bmatrix} y_{2t} \\ y_{2,t-1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y_{2,t-1} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} e_{2t} \\ 0 \end{bmatrix}. \tag{3.49}$$

REPRESENTATION 2 Suppose we want to treat y_{1t} as an unobserved state variable. In this case, we should transform the original model to eliminate the y_{2t} term. This is done by multiplying both sides of (3.43) by $(1 - \phi_1 L - \phi_2 L^2)$:

$$(1 - \phi_1 L - \phi_2 L^2) y_t = (1 - \phi_1 L - \phi_2 L^2) y_{1t} + (1 - \phi_1 L - \phi_2 L^2) y_{2t}, \tag{3.50}$$

$$\Longrightarrow\ y_t = (\phi_1 y_{t-1} + \phi_2 y_{t-2}) + (y_{1t} - \phi_1 y_{1,t-1} - \phi_2 y_{1,t-2}) + e_{2t}, \tag{3.51}$$

Measurement Equation

$$y_t = \begin{bmatrix} 1 & -\phi_1 & -\phi_2 \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{1,t-1} \\ y_{1,t-2} \end{bmatrix} + \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_{2t}, \tag{3.52}$$

Transition Equation

$$\begin{bmatrix} y_{1t} \\ y_{1,t-1} \\ y_{1,t-2} \end{bmatrix} = \begin{bmatrix} \delta \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{1,t-2} \\ y_{1,t-3} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ 0 \\ 0 \end{bmatrix}. \tag{3.53}$$

REPRESENTATION 3 Suppose we want to treat both y_{1t} and y_{2t} as unobserved state variables. In this case, we put both components in the state vector:

Measurement Equation

$$y_t = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{2,t-1} \end{bmatrix}, \tag{3.54}$$

Transition Equation

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{2,t-1} \end{bmatrix} = \begin{bmatrix} \delta \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & \phi_1 & \phi_2 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \\ 0 \end{bmatrix}. \tag{3.55}$$

NOTE When decomposing a nonstationary time series into a stochastic trend and a stationary component, one typically encounters the identification issue, which is carefully addressed in Nelson and Plosser 1982. In the preceding unobserved-component model, we assumed that the shocks to the two components, e_{1t} and e_{2t} are independent. How critical is this independence assumption? Let us address this issue in the simplest framework (a structural model) in which y_t is a sum of a random walk component and a white noise component:

$$y_t = y_{1t} + y_{2t}, \tag{3.56}$$

$$y_{1t} = y_{1,t-1} + e_{1t}, \tag{3.57}$$

$$y_{2t} = e_{2t}, \tag{3.58}$$

$$E(e_{1t}^2) = \sigma_1^2, \quad E(e_{2t}^2) = \sigma_2^2, \quad E(e_{1t}e_{2t}) = \sigma_{12}. \tag{3.59}$$

Suppose that the three parameters of the model, σ_1^2, σ_2^2, and σ_{12}, are not known. If the decomposed series y_{1t} and y_{2t} were observed, the parameters of the model could be easily estimated. However, only the sum of the two components, y_t, is observed. Thus, estimation of the parameters should be based on the observed series. We can easily show that y_t is an autoregressive integrated moving average (ARIMA)(0,1,1) process (a reduced-form model):

$$\Delta y_t = e_{1t} + e_{2t} - e_{2,t-1}, \tag{3.60}$$

$$==> \ \Delta y_t = \epsilon_t + \theta\epsilon_{t-1}, \quad E(\epsilon_t^2) = \sigma_\epsilon^2, \tag{3.61}$$

where θ and σ_ϵ^2 are functions of σ_1^2, σ_2^2, and σ_{12}. An identification problem arises because there exist more parameters in the structural model given by equations (3.56)–(3.59) than in the reduced-form model given by equation (3.61). We can estimate the parameters of the reduced-form model and then use these to indirectly estimate the parameters of the structural model. However, unless there is an identifying assumption in the structural model, leaving the same number of parameters in both models to be estimated, this is not possible. If the structural parameters are not identified, the decomposition is impossible.

The Time-Varying-Parameter Model The time-varying-parameter model that we discussed in section 3.1 is also a special case of the general state-space model in which H_t in the measurement equation (3.32) is replaced by a matrix of exogenous or predetermined variables. A specific example may be given by

$$y_t = \beta_{1t}x_{1t} + \beta_{2t}x_{2t} + \cdots + \beta_{kt}x_{kt} + e_t, \quad e_t \sim \text{i.i.d.}N(0, \sigma^2), \tag{3.62}$$

$$(\beta_{it} - \delta_i) = \phi_i(\beta_{i,t-1} - \delta_i) + v_{it}, \quad v_{it} \sim \text{i.i.d.}N(0, \sigma_i^2), \quad i = 1, 2, \ldots, k, \tag{3.63}$$

$$E(e_t v_{is}) = 0, \quad \text{for all } t \text{ and } s, i = 1, 2, \ldots, k, \tag{3.64}$$

where x_{it}, $i = 1, 2, \ldots, k$, are predetermined or exogenous variables.

Measurement Equation

$$y_t = [\, x_{1t} \quad x_{2t} \quad \ldots \quad x_{kt} \,] \begin{bmatrix} \beta_{1t} \\ \beta_{2t} \\ \vdots \\ \beta_{kt} \end{bmatrix} + e_t, \tag{3.65}$$

$$(y_t = x_t \beta_t + e_t), \tag{3.65$'$}$$

Transition Equation

$$\begin{bmatrix} \beta_{1t} \\ \beta_{2t} \\ \vdots \\ \beta_{kt} \end{bmatrix} = \begin{bmatrix} \delta_1^* \\ \delta_2^* \\ \vdots \\ \delta_k^* \end{bmatrix} + \begin{bmatrix} \phi_1 & 0 & \ldots & 0 \\ 0 & \phi_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \phi_k \end{bmatrix} \begin{bmatrix} \beta_{1,t-1} \\ \beta_{2,t-1} \\ \vdots \\ \beta_{k,t-1} \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \\ \vdots \\ v_{kt} \end{bmatrix}, \tag{3.66}$$

$$(\beta_t = \tilde{\mu} + F\beta_{t-1} + v_t), \tag{3.66$'$}$$

where $\delta_i^* = \delta_i(1 - \phi_i)$, $i = 1, 2, \ldots, k$. Comparing the measurement equations in (3.65)$'$ and (3.32), we notice that the time-varying-parameter model is a special case of the state-space model in which $H_t = x_t = [\, x_{1t} \quad x_{2t} \quad \ldots \quad x_{kt} \,]$. Here, x_t should be uncorrelated with e_t. If the observed z_t vector is included in the time-varying-parameter model, it may be correlated with x_t.

An Exercise Let us consider the following regression model, in which two of the regression coefficients are time varying (random walks) and the remaining coefficient is constant:

$$y_t = \beta_{1t} + \beta_{2t} x_{2t} + \beta_3 x_{3t} + e_t, \tag{3.67}$$

$$\beta_{it} = \beta_{i,t-1} + v_{it}, \qquad i = 1, 2. \tag{3.68}$$

The state-space representation of the model is given by

Measurement Equation

$$y_t = [\, 1 \quad x_{2t} \,] \begin{bmatrix} \beta_{1t} \\ \beta_{2t} \end{bmatrix} + \beta_3 x_{3t} + e_t, \tag{3.69}$$

$$(y_t = H_t \beta_t + \beta_3 x_{3t} + e_t), \tag{3.69$'$}$$

Transition Equation

$$\begin{bmatrix} \beta_{1t} \\ \beta_{2t} \end{bmatrix} = \begin{bmatrix} \beta_{1,t-1} \\ \beta_{2,t-1} \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix}, \tag{3.70}$$

$$(\beta_t = \beta_{t-1} + v_t). \tag{3.70$'$}$$

A Dynamic Factor Model Suppose we have two stationary variables, y_{1t}, and y_{2t} with a common component c_t:

$$y_{1t} = \gamma_1 c_t + z_{1t}, \tag{3.71}$$

$$y_{2t} = \gamma_2 c_t + z_{2t}, \tag{3.72}$$

$$c_t = \phi_1 c_{t-1} + v_t, \qquad v_t \sim \text{i.i.d.}N(0, 1), \tag{3.73}$$

$$z_{it} = \alpha_i z_{i,t-1} + e_{it}, \qquad e_{it} \sim \text{i.i.d.}N(0, \sigma_i^2), \tag{3.74}$$

where e_{1t}, e_{2t}, and v_t are independent of one another. Stock and Watson (1991) employ a general version of the above dynamic factor model to extract an experimental coincident index from four coincident economic variables. A state-space representation of the above model is given by

Measurement Equation

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \gamma_1 & 1 & 0 \\ \gamma_2 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_t \\ z_{1t} \\ z_{2t} \end{bmatrix}, \tag{3.75}$$

$$(y_t = H_t \beta_t), \tag{3.75$'$}$$

Transition Equation

$$\begin{bmatrix} c_t \\ z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} \phi_1 & 0 & 0 \\ 0 & \alpha_1 & 0 \\ 0 & 0 & \alpha_2 \end{bmatrix} \begin{bmatrix} c_{t-1} \\ z_{1,t-1} \\ z_{2,t-1} \end{bmatrix} + \begin{bmatrix} v_t \\ e_{1t} \\ e_{2t} \end{bmatrix}, \tag{3.76}$$

$$(\beta_t = F\beta_{t-1} + v_t). \tag{3.76$'$}$$

A Common Stochastic Trend Model Suppose that we have two integrated series, y_{1t} and y_{2t}. If the two series are cointegrated, then there exists a common stochastic trend. A typical example would be the spot and forward exchange rates. Hai, Mark, and Wu (1996) consider the following common stochastic trend model for the spot and forward exchange rates:

$$y_{1t} = z_t + x_{1t}, \tag{3.77}$$

$$y_{2t} = z_t + x_{2t}, \tag{3.78}$$

$$z_t = z_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{i.i.d.} N(0, \sigma_\epsilon^2), \tag{3.79}$$

$$x_{1t} = \mu_1 + \phi_{11}x_{1,t-1} + \phi_{12}x_{2,t-1} + e_{1t} + \theta_{11}e_{1,t-1} + \theta_{12}e_{2,t-1}, \tag{3.80}$$

$$x_{2t} = \mu_2 + \phi_{21}x_{1,t-1} + \phi_{22}x_{2,t-1} + e_{2t} + \theta_{21}e_{1,t-1} + \theta_{22}e_{2,t-1}, \tag{3.81}$$

$$\begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \sim \text{i.i.d.} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \right), \tag{3.82}$$

where z_t is a common stochastic trend and the stationary components x_{1t} and x_{2t} are assumed to be generated by a vector ARMA(1,1) process. A state-space representation of the above model is given by

Measurement Equation

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} z_t \\ x_{1t} \\ x_{2t} \\ e_{1t} \\ e_{2t} \end{bmatrix}, \tag{3.83}$$

Transition Equation

$$\begin{bmatrix} z_t \\ x_{1t} \\ x_{2t} \\ e_{1t} \\ e_{2t} \end{bmatrix} = \begin{bmatrix} 0 \\ \mu_1 \\ \mu_2 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \phi_{11} & \phi_{12} & \theta_{11} & \theta_{12} \\ 0 & \phi_{21} & \phi_{22} & \theta_{21} & \theta_{22} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} z_{t-1} \\ x_{1,t-1} \\ x_{2,t-1} \\ e_{1,t-1} \\ e_{2,t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_t \\ e_{1t} \\ e_{2t} \end{bmatrix}. \tag{3.84}$$

3.2.2 The Kalman Filter and MLE

Once a dynamic time series model is written in state-space form, the Kalman filter is readily available for inferences on the unobserved state vector β_t, conditional on the parameters of the model and the appropriate information set. The basic filtering and the smoothing algorithm are the same as those in section 3.1 with slight modification. For example, we have the H_t matrix in place of x_t of a time-varying-parameter model. The derivation and the intuition associated with the algorithms are the same. Treatment of the initial values, $\beta_{0|0}$ and $P_{0|0}$, is also exactly the same as in section 3.1.

Basic Filtering

Prediction

$$\beta_{t|t-1} = \tilde{\mu} + F\beta_{t-1|t-1}, \tag{3.85}$$

$$P_{t|t-1} = FP_{t-1|t-1}F' + Q, \tag{3.86}$$

$$\eta_{t|t-1} = y_t - y_{t|t-1} = y_t - H_t\beta_{t|t-1} - Az_t, \tag{3.87}$$

$$f_{t|t-1} = H_t P_{t|t-1} H_t' + R, \tag{3.88}$$

Updating

$$\beta_{t|t} = \beta_{t|t-1} + K_t\eta_{t|t-1}, \tag{3.89}$$

$$P_{t|t} = P_{t|t-1} - K_t H_t P_{t|t-1}, \tag{3.90}$$

where $K_t = P_{t|t-1}H_t' f_{t|t-1}^{-1}$ is the Kalman gain.

Smoothing ($t = T-1, T-2, \ldots, 1$)

$$\beta_{t|T} = \beta_{t|t} + P_{t|t}F'P_{t+1|t}^{-1}(\beta_{t+1|T} - F\beta_{t|t} - \tilde{\mu}), \tag{3.91}$$

$$P_{t|T} = P_{t|t} + P_{t|t}F'P_{t+1|t}^{-1}(P_{t+1|T} - P_{t+1|t})P_{t+1|t}^{-1\,\prime}FP_{t|t}', \tag{3.92}$$

where $\beta_{T|T}$ and $P_{T|T}$, the initial values for the smoothing, are obtained from the last iteration of the basic filter. The procedure for the maximum likelihood estimation of the model's parameters is exactly the same as in section 3.1.

3.3 Application 1: A Decomposition of Real GDP and the Unemployment Rate into Stochastic Trend and Transitory Components

Nelson and Plosser (1982) suggest that the nonstationarity in economic activity should be removed by first-differencing rather than linear detrending, making the trend component of real GDP or GNP a random walk with drift rather than a deterministic function of time. Their analysis further suggests that most of the innovation variance in annual real GNP should be allocated to the nonstationary trend component, with little variance left over the stationary cyclical component.

Noting that annual averages blur the pattern of economic activity apparent in quarterly or monthly data, Clark (1987) applies a version of the unobserved components model discussed in section 3.2 to quarterly real GNP and the monthly index of industrial production in order to evaluate the relative importance of the stochastic trend and the stationary cyclical components of economic activity. In this section, we apply Clark's unobserved components model to quarterly real GDP for the period 1947:II–1995:III and measure the stochastic trend and the cyclical components. By taking advantage of a standard version of Okun's law, the univariate model is then extended into a bivariate unobserved components model of real GDP and unemployment rate, as in Clark 1989, and the unemployment rate is decomposed into trend and cyclical components as well.

To distinguish between time trend and stochastic trend models of real output, Clark (1987) considers the following unobserved components model:

$$y_t = n_t + x_t, \tag{3.93}$$

$$n_t = g_{t-1} + n_{t-1} + v_t, \quad v_t \sim \text{i.i.d.} N(0, \sigma_v^2), \tag{3.94}$$

$$g_t = g_{t-1} + w_t, \quad w_t \sim \text{i.i.d.} N(0, \sigma_w^2), \tag{3.95}$$

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + e_t, \quad e_t \sim \text{i.i.d.} N(0, \sigma_e^2), \tag{3.96}$$

where y_t is the log of real GDP, n_t is a stochastic trend component, and x_t is a stationary cyclical component; v_t, w_t, and e_t are independent white noise processes. In the presence of the decline in the U.S. productivity growth in the 1970s and reduction of labor force growth in the 1980s, we follow Clark (1987) in modeling the drift term (g_t) in the stochastic trend component as a random walk. As in section 3.2.1, we have at least three alternative state-space representations of the above model. We employ the following representation:

$$y_t = [\,1 \quad 1 \quad 0 \quad 0\,] \begin{bmatrix} n_t \\ x_t \\ x_{t-1} \\ g_t \end{bmatrix}, \tag{3.97}$$

$$\begin{bmatrix} n_t \\ x_t \\ x_{t-1} \\ g_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & \phi_1 & \phi_2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} n_{t-1} \\ x_{t-1} \\ x_{t-2} \\ g_{t-1} \end{bmatrix} + \begin{bmatrix} v_t \\ e_t \\ 0 \\ w_t \end{bmatrix}. \tag{3.98}$$

The unemployment rate is also assumed to have both a random walk component and a stationary component. By imposing a version of Okun's law, the unemployment rate may be specified as

$$U_t = L_t + C_t, \tag{3.99}$$

$$L_t = L_{t-1} + v_{lt}, \quad v_{lt} \sim \text{i.i.d.}N(0, \sigma_{vl}^2), \tag{3.100}$$

$$C_t = \alpha_0 x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + e_{ct}, \quad e_{ct} \sim \text{i.i.d.}N(0, \sigma_{ec}^2), \tag{3.101}$$

where L_t is a trend component and C_t is a stationary component that is assumed to be a function of current and past transitory components of real output. Combined with the output equations (3.93)–(3.96), a state-space representation of the bivariate model is given by

$$\begin{bmatrix} y_t \\ U_t \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & \alpha_0 & \alpha_1 & \alpha_2 & 0 & 1 \end{bmatrix} \begin{bmatrix} n_t \\ x_t \\ x_{t-1} \\ x_{t-2} \\ g_t \\ L_t \end{bmatrix} + \begin{bmatrix} 0 \\ e_{ct} \end{bmatrix}, \tag{3.102}$$

$$\begin{bmatrix} n_t \\ x_t \\ x_{t-1} \\ x_{t-2} \\ g_t \\ L_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & \phi_1 & \phi_2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} n_{t-1} \\ x_{t-1} \\ x_{t-2} \\ x_{t-3} \\ g_{t-1} \\ L_{t-1} \end{bmatrix} + \begin{bmatrix} v_t \\ e_t \\ 0 \\ 0 \\ w_t \\ v_{lt} \end{bmatrix}. \tag{3.103}$$

Table 3.1 reports estimation results. As in Clark 1987, a significant portion of the quarter-to-quarter innovations in real GDP are cyclical from univariate and bivariate models. Figures 3.3 through 3.6 plot the log of real GDP along with its trend and cyclical components implied by the models. Figures 3.7 and 3.8 plot the unemployment rate and its trend and cyclical components implied by the bivariate model.

Table 3.1
Estimates of the unobserved components model of real GDP (1952:I–1995:III)

Parameters	Univariate model		Bivariate model	
σ_v	0.0056	(0.0013)	0.0049	(0.0006)
σ_e	0.0061	(0.0013)	0.0067	(0.0006)
σ_w	0.0002	(0.0002)	0.0003	(0.0002)
ϕ_1	1.5346	(0.1501)	1.4386	(0.0791)
ϕ_2	−0.5888	(0.1155)	−0.5174	(0.0569)
α_0	—		−0.3368	(0.0497)
α_1	—		−0.1635	(0.0310)
α_2	—		−0.0720	(0.0054)
σ_{vl}	—		0.0015	(0.0003)
σ_{ec}	—		0.0003	(0.0003)
Log likelihood	578.52		1566.99	

Note: Standard errors are in parentheses.

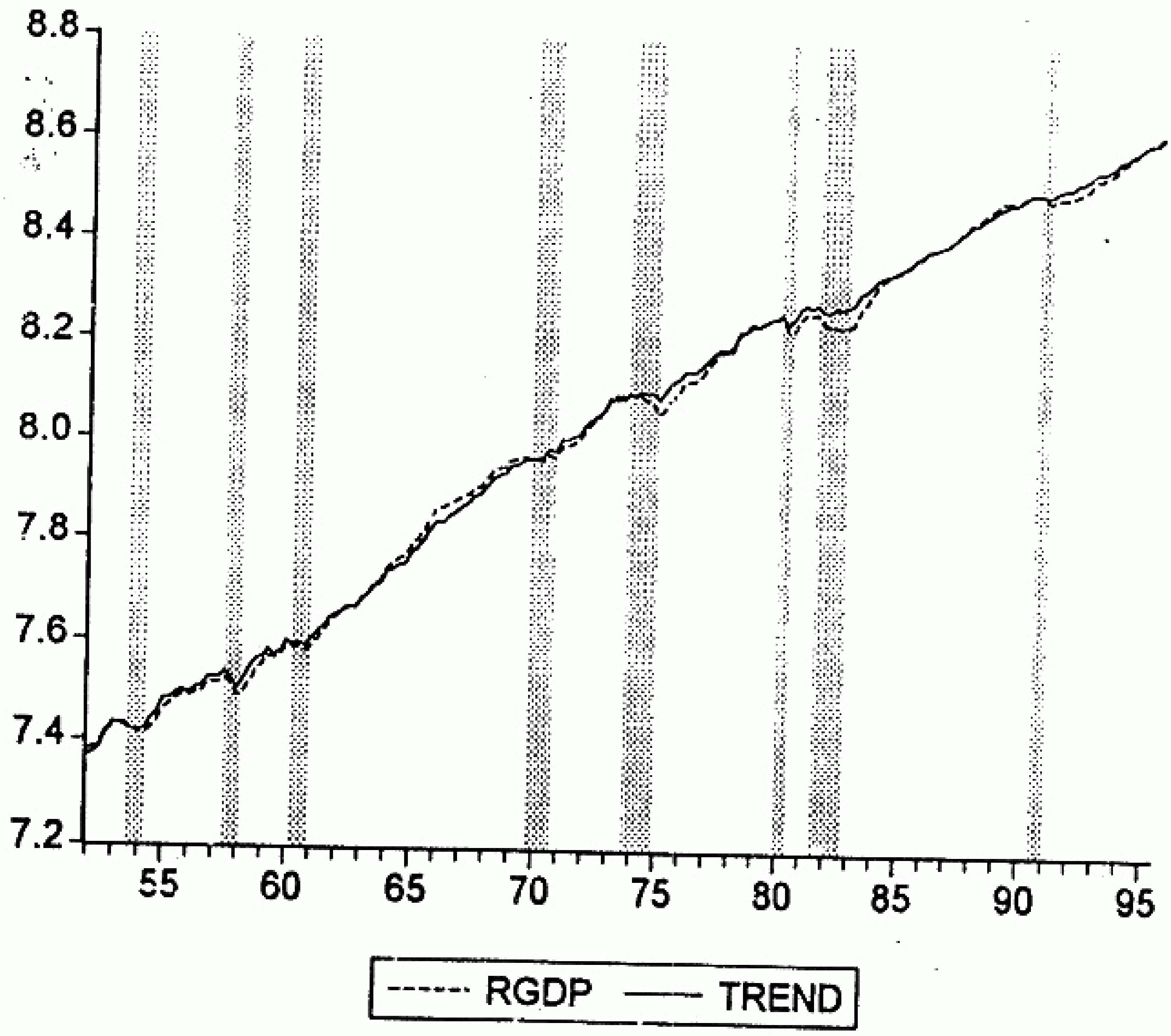

Figure 3.3
Real GDP and its trend component: univariate UC model

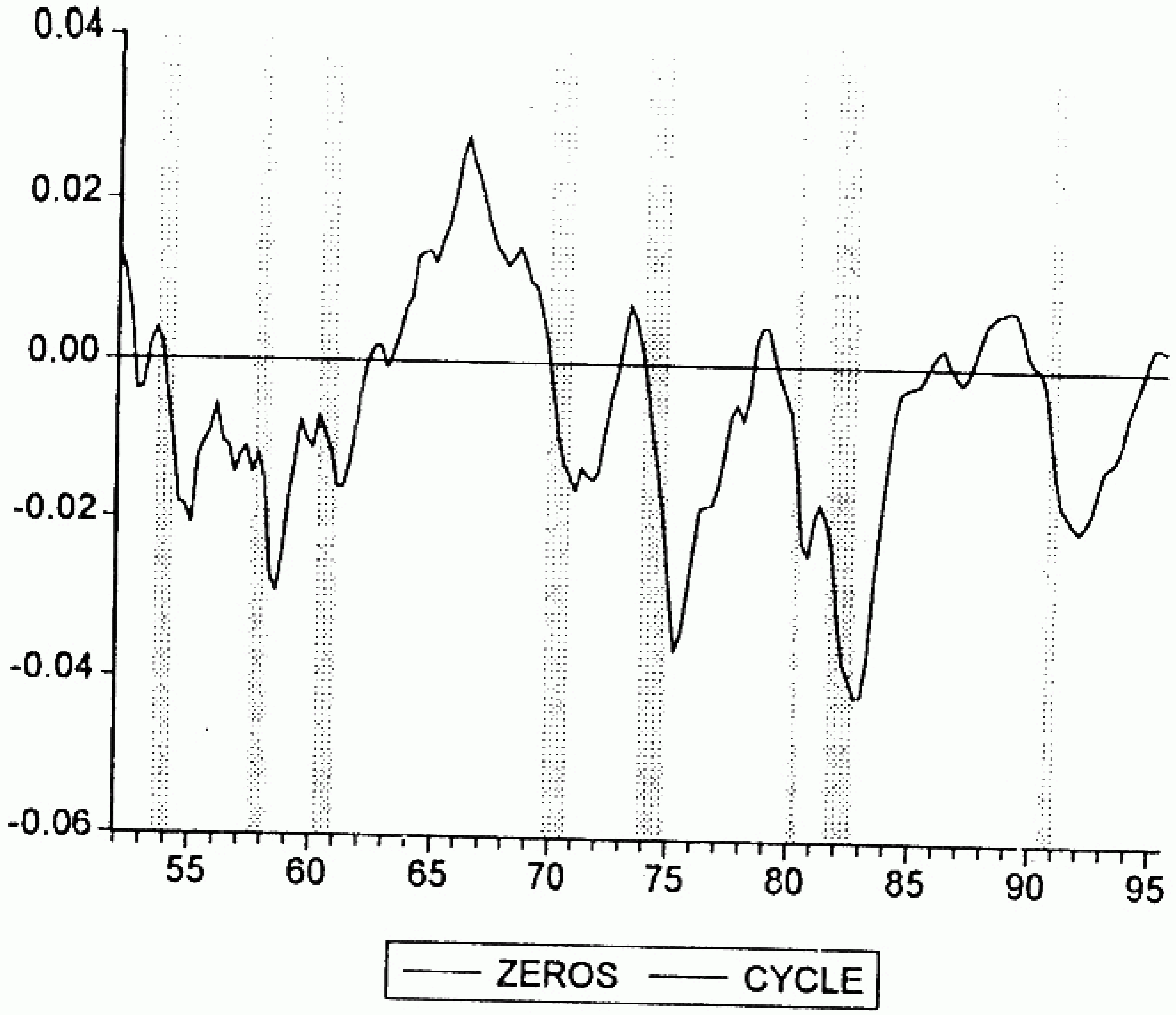

Figure 3.4
Cyclical component of real GDP: univariate UC model

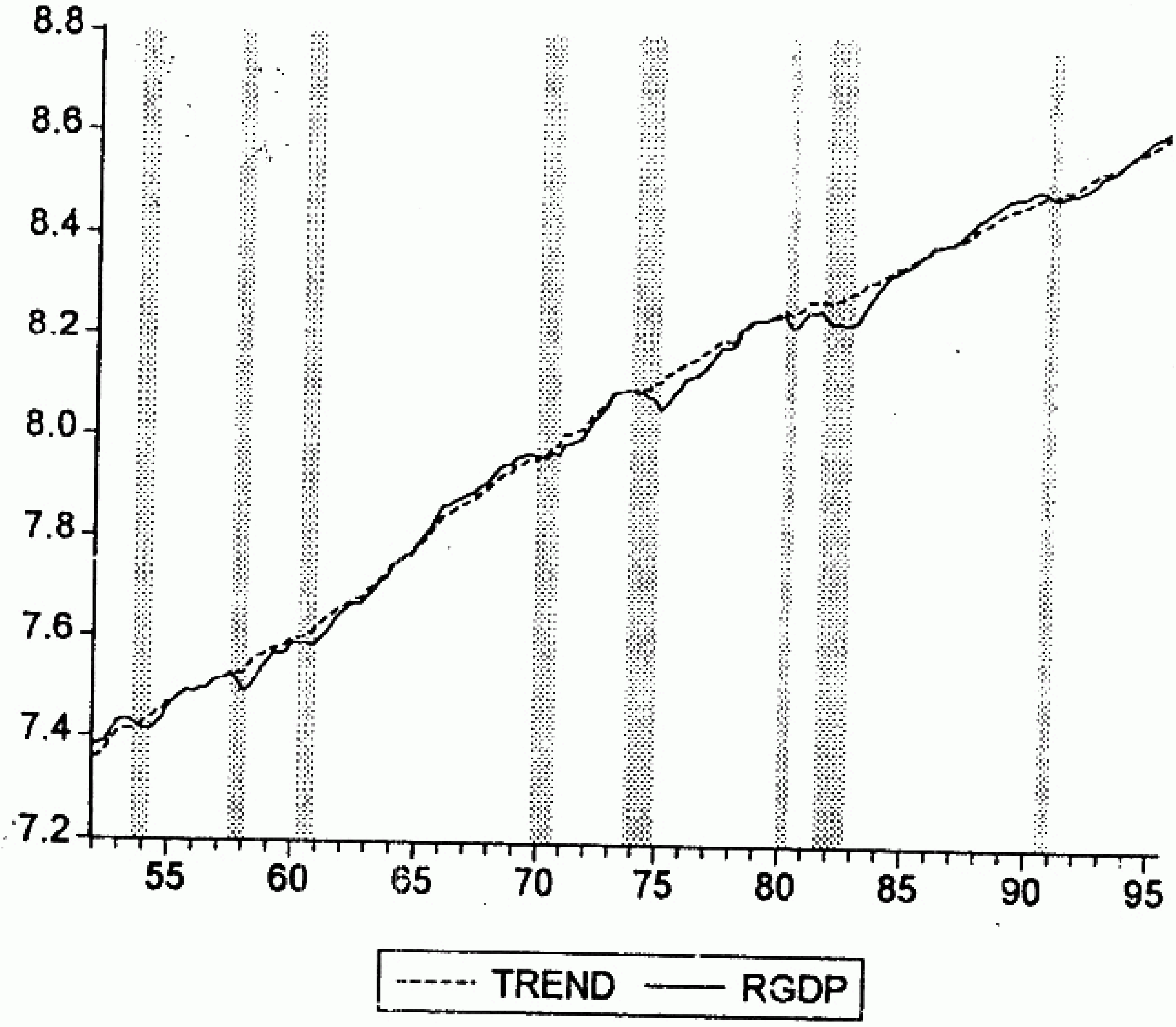

Figure 3.5
Real GDP and its trend component: bivariate UC model

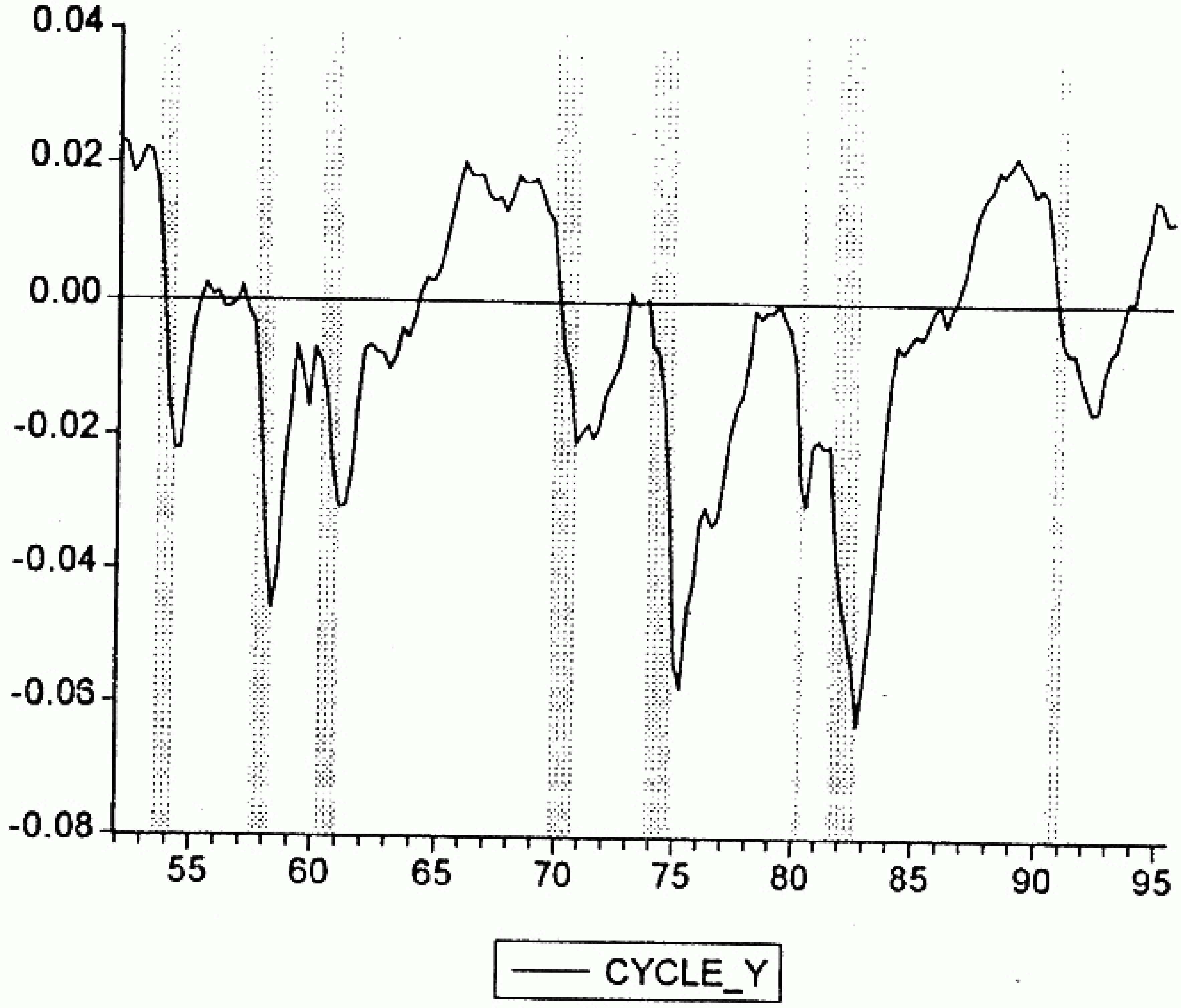

Figure 3.6
Cyclical component of real GDP: bivariate UC model

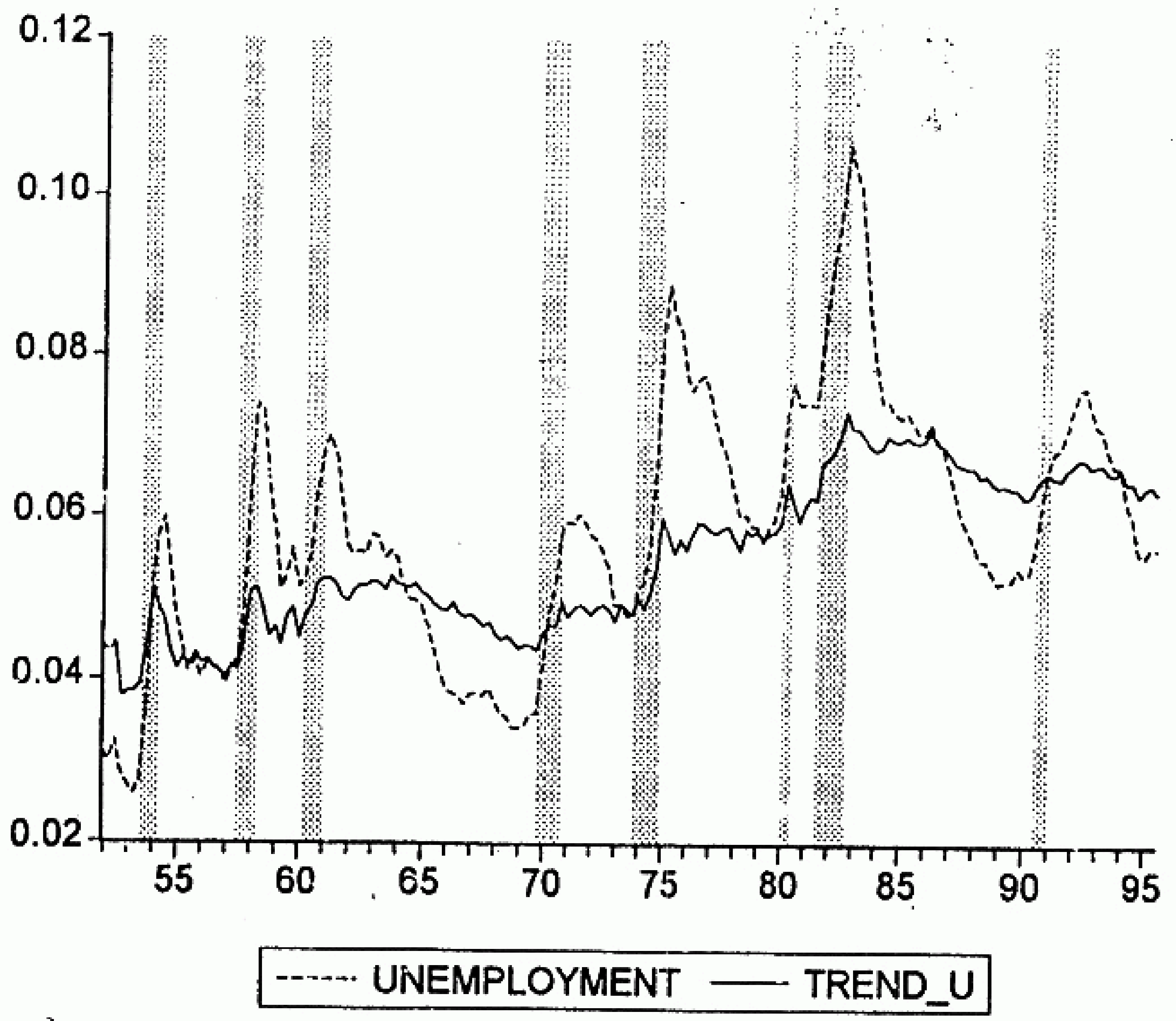

Figure 3.7
Unemployment rate and its trend component: bivariate UC model

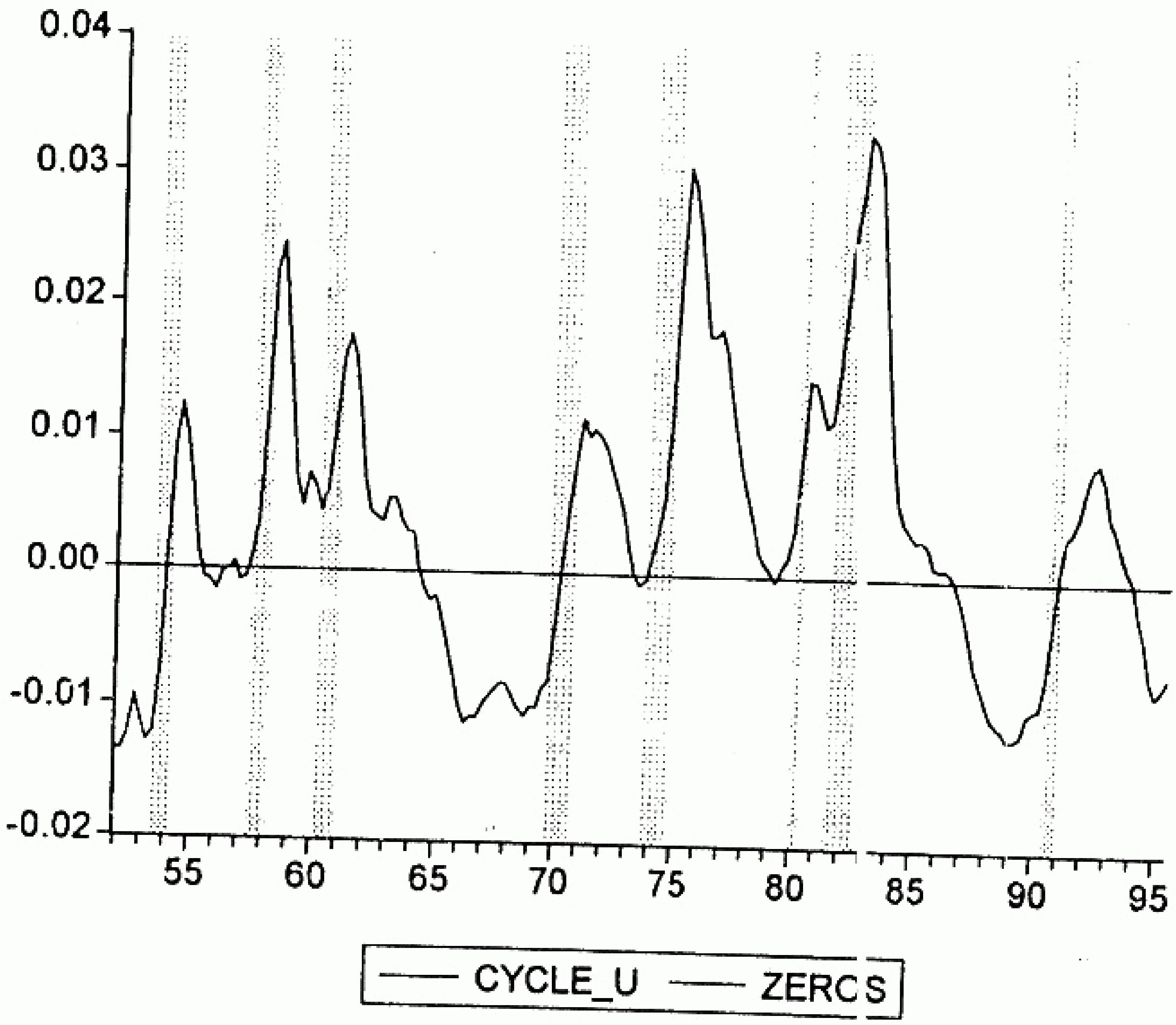

Figure 3.8
Cyclical component of unemployment rate: bivariate UC mode

Note, however, that the sum of the estimated AR coefficients for the cyclical component is close to 1. We suggest that caution should be exercised in drawing inferences about the nature of the decomposition in the presence of a highly persistent stationary component. First, Nelson (1988) argues that the decomposition may be spurious, just as detrending by linear regression is known to generate spurious trends and cycles in nonstationary time series. His Monte Carlo experiment confirms that when data are generated by a random walk, the state-space model tends to indicate (incorrectly) that the series consists of cyclical variations around a smooth trend. Second, Perron (1990) suggests that the standard unit root tests are biased toward nonrejection of the null of a unit root when the data-generating process is stationary with a switching mean. Suppose that the mean of the stationary AR component of real GDP is subject to regime shifts. Even when the sum of the AR coefficients is significantly lower than 1, ignoring regime shifts would result in a highly persistent stationary component. Recently, Kim and Nelson (1997) considered this possibility by explicitly considering asymmetry in the cyclical component of real GDP.

3.4 Application 2: An Application of the Time-Varying-Parameter Model to Modeling Changing Conditional Variance

Autoregressive conditional heteroskedasticity (ARCH), introduced by Engle (1982), explicitly characterizes the changing conditional variance of the regression disturbances. This class of models allows for the conditional variance to depend on the squares of previous innovations. In this section, we discuss an alternative source of changing conditional variance. Whereas Tsay (1987) and Bera and Lee (1993) show that parameter heterogeniety in the random coefficient autoregressive models is a source of changing conditional variance such as ARCH, we focus on Kim and Nelson's (1989) application of the time-varying-parameter model introduced in section 3.1 for modeling changing conditional variance or uncertainty in the U.S. monetary growth.

McNees (1986) states, "A policy reaction function is likely to be a fragile creature. Over time . . . the importance attached to conflicting objectives [of the policy] may change, [policy makers'] views on the structure of the economy may change" (p. 7). Thus, based on stability test results on the regression coefficients, Kim and Nelson (1989) consider the following time-varying-parameter model for the U.S. monetary growth function:

$$\Delta M_t = \beta_{0t} + \beta_{1t}\Delta i_{t-1} + \beta_{2t} INF_{t-1} + \beta_{3t} SURP_{t-1} + \beta_{4t}\Delta M_{t-1} + e_t, \tag{3.104}$$

$$\beta_{it} = \beta_{it-1} + v_{it}, \tag{3.105}$$

$$e_t \sim \text{i.i.d.} N(0, \sigma_e^2), \tag{3.106}$$

$$v_{it} \sim \text{i.i.d.} N(0, \sigma_{vi}^2), \quad i = 0, 1, \ldots, 4. \tag{3.107}$$

In matrix notation, we have

$$\Delta M_t = x_{t-1}\beta_t + e_t, \tag{3.108}$$

$$\beta_t = F\beta_{t-1} + v_t, \text{ with } F = I_5, \tag{3.109}$$

$$e_t \sim \text{i.i.d.} N(0, \sigma_e^2), \tag{3.110}$$

$$v_t \sim \text{i.i.d.} N(0, Q), \tag{3.111}$$

where ΔM, Δi, INF, and $SURP$ stand for U.S. figures for the quarterly M1 growth rate, changes in the interest rate as measured by the three-month

Table 3.2
Parameter estimates of the time-varying-parameter model of U.S. monetary growth (quarterly data, 1964:I–1985:IV)

σ_e	0.3712	(0.0633)
σ_{v0}	0.1112	(0.0627)
σ_{v1}	0.0171	(0.0341)
σ_{v2}	0.2720	(0.0607)
σ_{v3}	0.0378	(0.1634)
σ_{v4}	0.0224	(0.0375)
Log Likelihood	−97.0924	

Note: Standard errors are in parentheses.

T-bill rate, the inflation rate as measured by the CPI, and the detrended full employment budget surplus, respectively.

The Kalman filter is applied to make inferences on the changing regression coefficients. One nice thing about the Kalman filter is that it gives us insight into how a rational economic agent would revise his estimates of the coefficients in a Bayesian fashion when new information is available in a world of uncertainty, *especially* under a changing policy regime.

In an ARCH model, changing uncertainty about the future is focused on the changing conditional variance in the disturbance terms of the regression equation. As Harrison and Stevens (1976) state, however, a person's uncertainty about the future arises not simply because of future random terms but also because of uncertainty about current parameter values and the model's ability to link the present to the future. In the time-varying-parameter model above, uncertainty about current regression coefficients β_t results in changing conditional variance of monetary growth. This is well captured in an equation for the variance of the conditional forecast error in the Kalman filter:

$$f_{t|t-1} = x_{t-1} P_{t|t-1} x_{t-1}' + \sigma_e^2, \tag{3.13$'$}$$

where $P_{t|t-1}$ represents the degree of uncertainty associated with an inference on β_t conditional on information up to time $t-1$.

Table 3.2 reports parameter estimates of the model along with their standard errors for the sample period 1964:I–1985:IV (quarterly data). Figures 3.9 through 3.13 depict the Kalman filter inferences on the regression coefficients conditional on information up to time $t-1$. These show estimates of how the Federal Reserve has been changing its reaction to various macroeconomic variables in the presence of changing importance attached to potentially conflicting

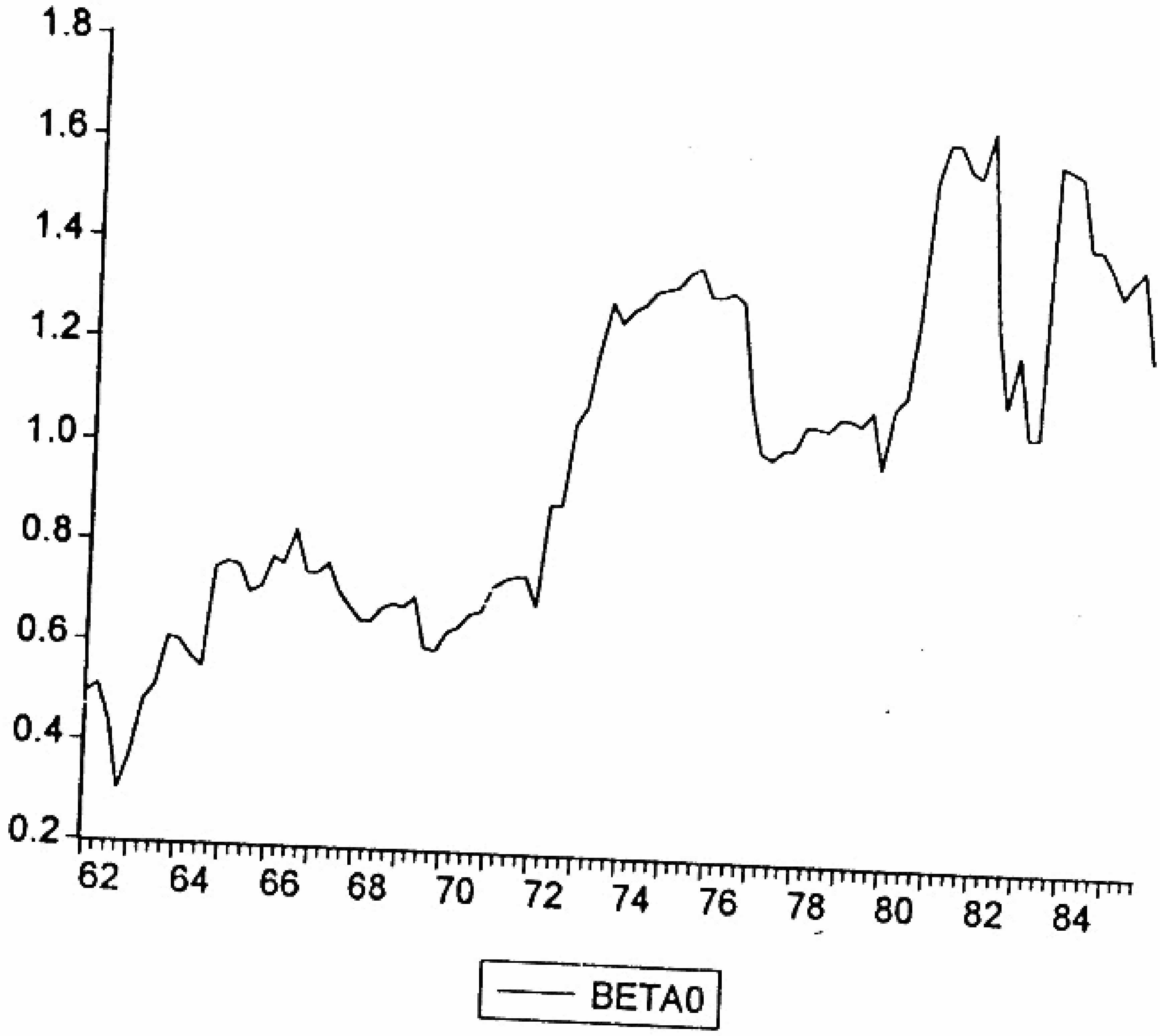

Figure 3.9
Time-varying regression coefficient: β_0

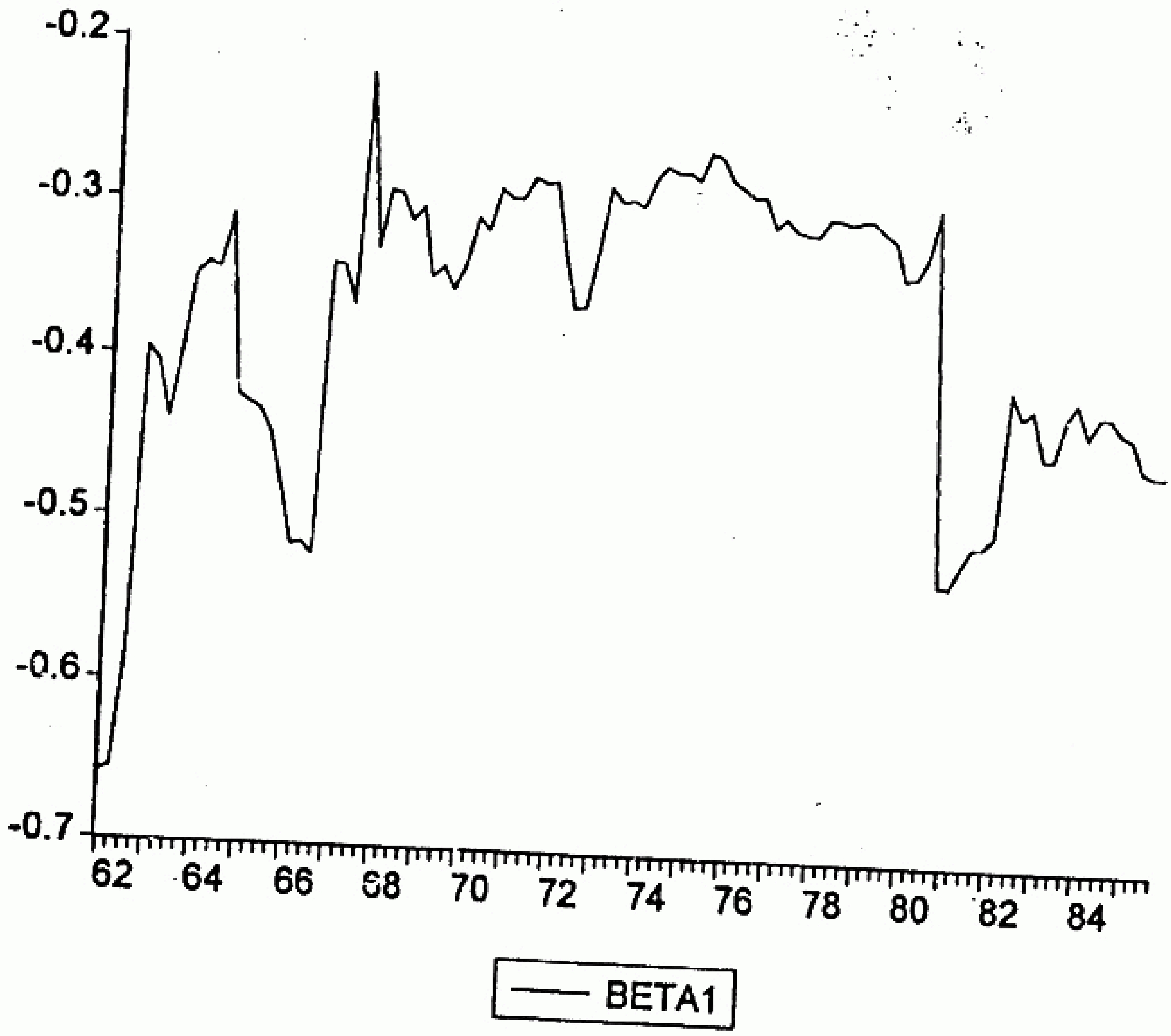

Figure 3.10
Time-varying regression coefficient: β_1

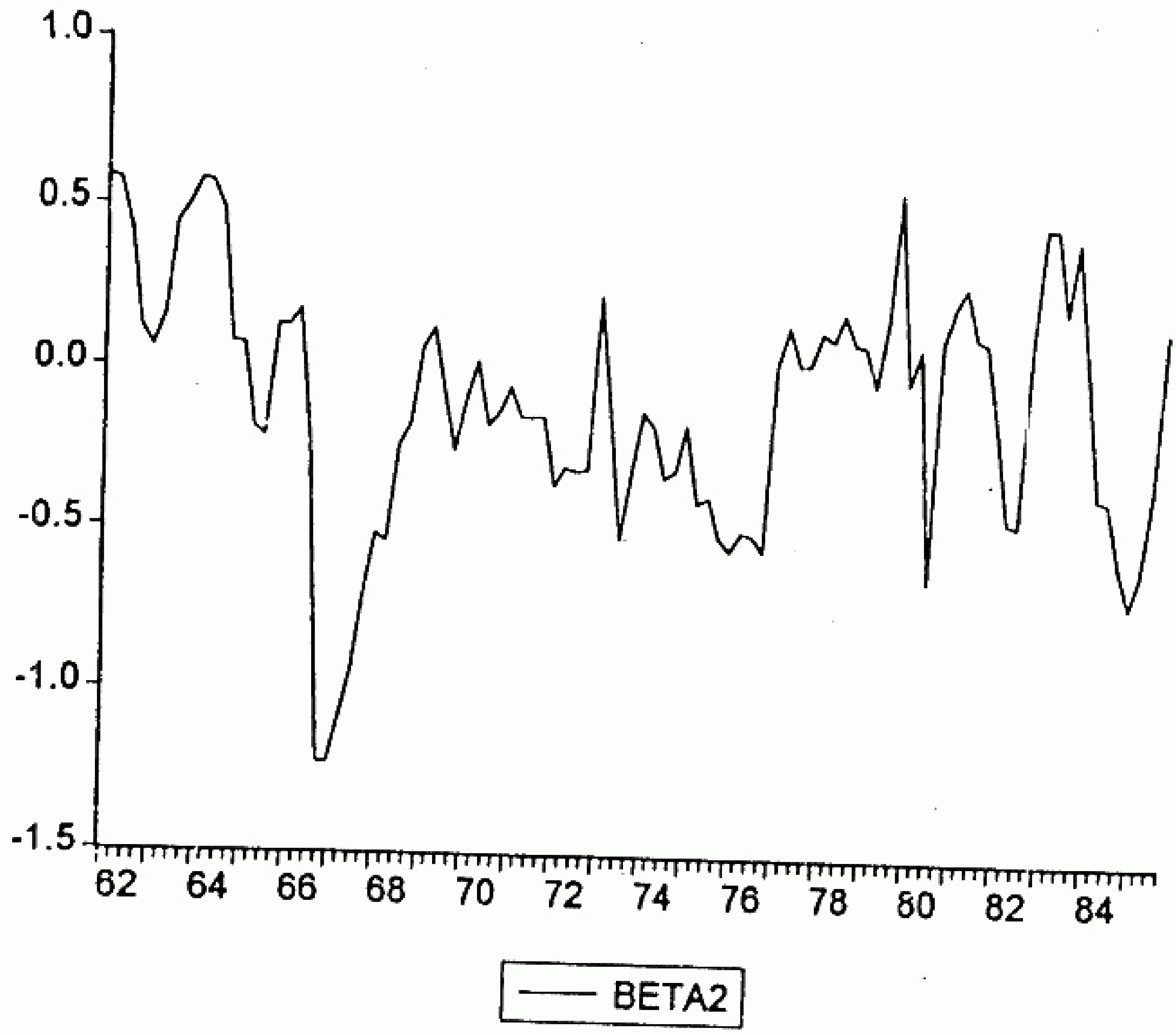

Figure 3.11
Time-varying regression coefficient: β_2

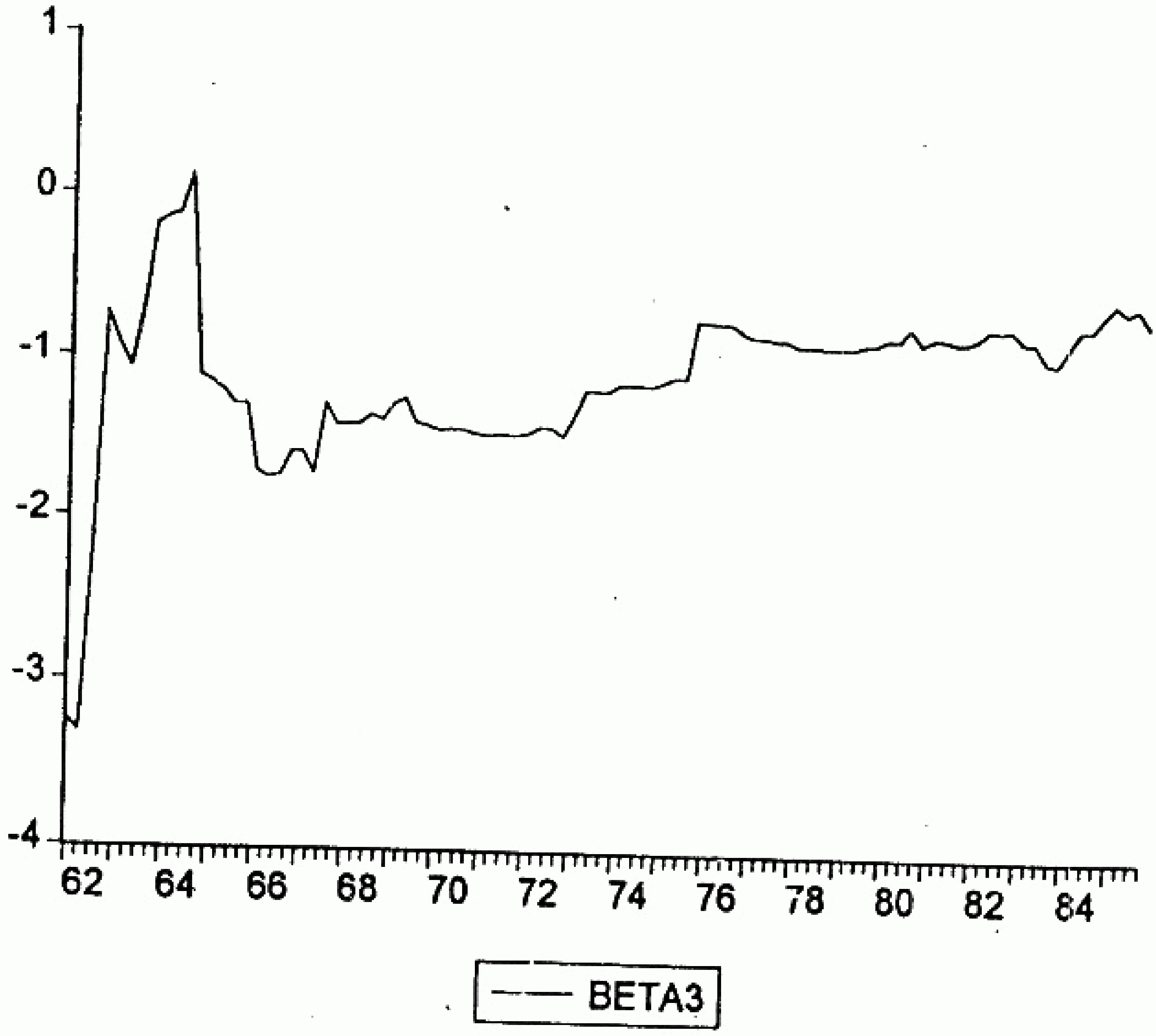

Figure 3.12
Time-varying regression coefficient: β_3

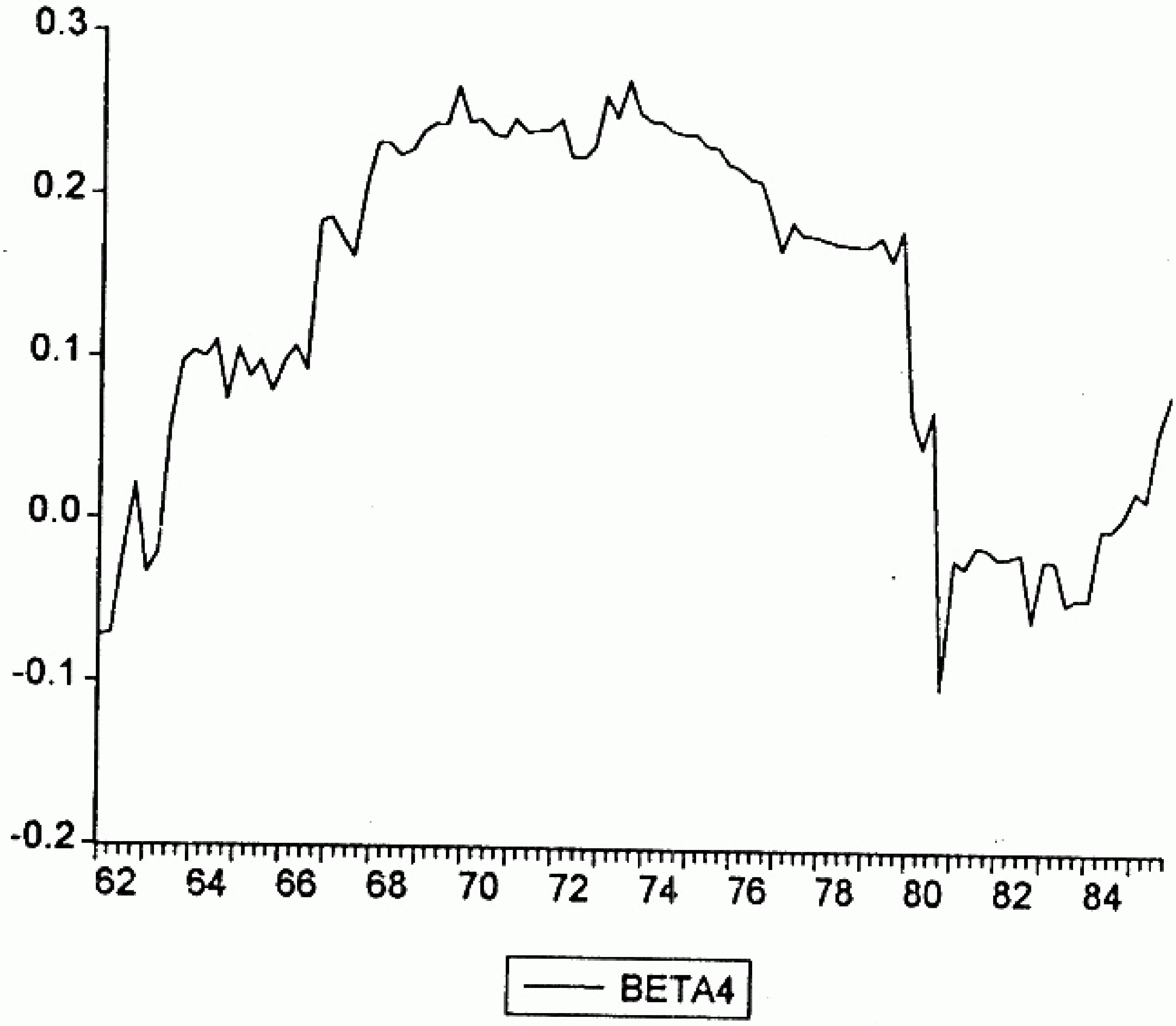

Figure 3.13
Time-varying regression coefficient: β_4

objectives of policy over time. The standardized forecast errors that result and their squares reveal no significant serial correlation, suggesting no evidence of model misspecification. Finally, figure 3.14 depicts changing conditional variance or uncertainty underlying monetary growth, as analytically given in (3.13)′.

Kim and Nelson (1989) further investigate the link between this measure of monetary growth uncertainty and economic activity. They document that uncertainty associated with monetary growth, as implied by the time-varying-parameter model, has a negative effect on the permanent component of real output.

3.5 Application 3: Stock and Watson's Dynamic Factor Model of the Coincident Economic Indicators

Ever since Burns and Mitchell (1946), there has been consensus about a stylized fact of the business cycle: that is, the comovement of many macroeco-

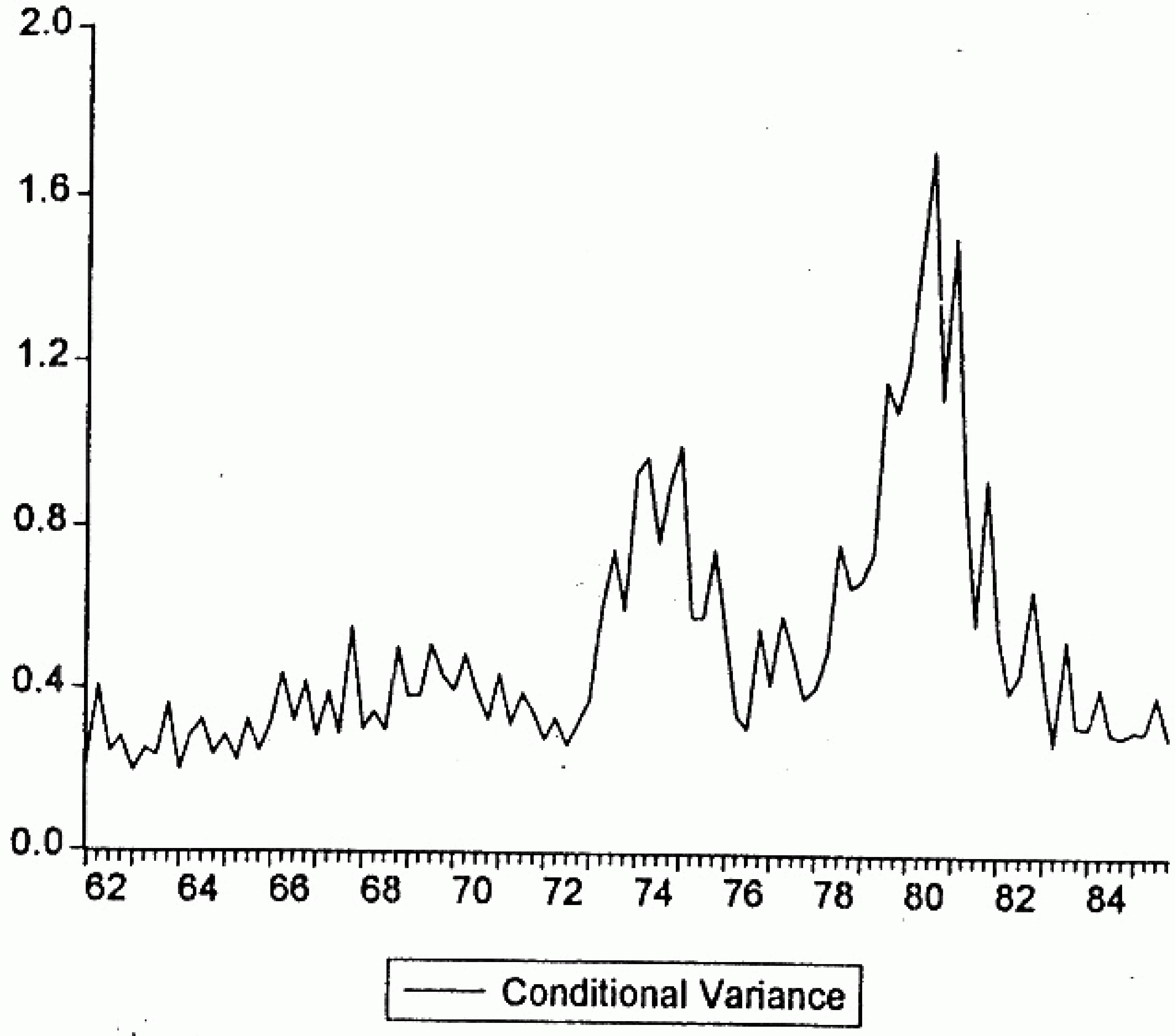

Figure 3.14
Conditional forecast error variance of monetary growth from a TVP model

nomic variables through the cycle. The index of coincident economic indicators issued by the Department of Commerce (DOC), developed for the purpose of summarizing the state of macroeconomic activity, is an example that exploits this stylized fact.

Recently, Stock and Watson (1991) developed a probability model of the coincident economic indicators, based on the notion that the comovements in many macroeconomic variables have a common element that can be captured by a single underlying, unobserved variable. As an appliction of the state-space model and the Kalman filter, this section focuses on Stock and Watson's (1991) dynamic factor model of the coincident economic indicators.

3.5.1 Model Specification and Identification Issue

Let Y_{1t}, Y_{2t}, Y_{3t}, and Y_{4t} be the logs of four coincident variables used to construct the coincident index: industrial production, personal income less transfer payments, total manufacturing and trade sales, and employees on nonagricultural payrolls. Unit root tests for each of these U.S. series suggest that one cannot reject the null hypothesis of a unit root. In addition, these four series do not seem to be cointegrated. Thus, Stock and Watson (1991) consider the following dynamic factor model in first differences of the four variables:

$$\Delta Y_{it} = D_i + \gamma_i \Delta C_t + e_{it}, \quad i = 1, 2, 3, 4, \tag{3.112}$$

$$(\Delta C_t - \delta) = \phi_1(\Delta C_{t-1} - \delta) + \phi_2(\Delta C_{t-2} - \delta) + w_t, \quad w_t \sim \text{i.i.d.} N(0, \sigma_w^2), \tag{3.113}$$

$$e_{it} = \psi_{i1} e_{i,t-1} + \psi_{i2} e_{i,t-2} + \epsilon_{it}, \quad \epsilon_{it} \sim \text{i.i.d.} N(0, \sigma_i^2), \quad i = 1, 2, 3, 4, \tag{3.114}$$

where ΔC_t is the common component and σ_w^2 is set to 1 in order to normalize the common component; roots of $(1 - \phi_1 L - \phi_2 L^2) = 0$ lie outside the unit circle; roots of $(1 - \psi_{i1} L - \psi_{i2} L^2) = 0$, $i = 1, 2, 3, 4$, lie outside the unit circle; and all the shocks are assumed to be independent.

In the above specification, the common component ΔC_t enters each equation in (3.112) with a different weight, γ_i, $i = 1, 2, 3, 4$. For each of the four series, $D_i + e_{it}$, $i = 1, 2, 3, 4$ represents the individual or the idiosyncratic component.

Note that the first population moment for the i-th indicator, ΔY_{it}, consists of two parameters:

$$E(\Delta Y_{it}) = D_i + \gamma_i \delta. \tag{3.115}$$

Given the corresponding sample first moment, $\Delta \bar{Y}_i$, however, the parameters D_i and δ are not separately identified. As a way to avoid the identification problem in the maximum likelihood estimation of the model, Stock and Watson suggest writing the model in deviation from means, thus concentrating the $D_i + \gamma_i \delta$ terms, $i = 1, 2, 3, 4$, out of the likelihood function:

Model in Deviation from Means

$$\Delta y_{it} = \gamma_i \Delta c_t + e_{it}, \quad i = 1, 2, 3, 4, \tag{3.116}$$

$$\Delta c_t = \phi_1 \Delta c_{t-1} + \phi_2 \Delta c_{t-2} + w_t, \quad w_t \sim \text{i.i.d.} N(0, 1), \tag{3.117}$$

$$e_{it} = \psi_{i1} e_{i,t-1} + \psi_{i2} e_{i,t-2} + \epsilon_{it}, \quad \epsilon_{it} \sim \text{i.i.d.} N(0, \sigma_i^2), \quad i = 1, 2, 3, 4, \tag{3.118}$$

where $\Delta y_{it} = \Delta Y_{it} - \Delta \bar{Y}_i$ and $\Delta c_t = \Delta C_t - \delta$.

Once the above model in deviation from means is written in state-space form, the Kalman filter is readily available for maximum likelihood estimation of the model based on the prediction error decomposition, as well as for inferences on Δc_t. A state-space representation of the model is given by

Measurement Equation

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \\ \Delta y_{3t} \\ \Delta y_{4t} \end{bmatrix} = \begin{bmatrix} \gamma_1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \gamma_2 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \gamma_3 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \gamma_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \Delta c_t \\ \Delta c_{t-1} \\ e_{1t} \\ e_{1,t-1} \\ e_{2t} \\ e_{2,t-1} \\ e_{3t} \\ e_{3,t-1} \\ e_{4t} \\ e_{4,t-1} \end{bmatrix}, \tag{3.119}$$

$(\Delta y_t = H\beta_t)$,

Transition Equation

$$\begin{bmatrix} \Delta c_t \\ \Delta c_{t-1} \\ e_{1t} \\ e_{1,t-1} \\ e_{2t} \\ e_{2,t-1} \\ e_{3t} \\ e_{3,t-1} \\ e_{4t} \\ e_{4,t-1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \psi_{11} & \psi_{12} & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & \psi_{41} & \psi_{42} \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \Delta c_{t-1} \\ \Delta c_{t-2} \\ e_{1,t-1} \\ e_{1,t-2} \\ e_{2,t-1} \\ e_{2,t-2} \\ e_{3,t-1} \\ e_{3,t-2} \\ e_{4,t-1} \\ e_{4,t-2} \end{bmatrix} + \begin{bmatrix} w_t \\ 0 \\ \epsilon_{1t} \\ 0 \\ \epsilon_{2t} \\ 0 \\ \epsilon_{3t} \\ 0 \\ \epsilon_{4t} \\ 0 \end{bmatrix}, \tag{3.120}$$

$(\beta_t = F\beta_{t-1} + v_t)$

Thus, one can estimate the parameters of the model using the MLE method based on the prediction error decomposition, and given the estimates of the parameters, one can run the Kalman filter to get $\beta_{t|t}$. Then, the (1, 1) element of $\beta_{t|t}$ is $\Delta c_{t|t}$.

3.5.2 Decomposing $\bar{Y}_i$ into D_i and δ, $i = 1, 2, 3, 4$

Given $\Delta c_{t|t}$, $t = 1, 2, \ldots, T$, we need an estimate of δ to construct a new coincident index, $C_{t|t}$, $t = 1, 2, \ldots, T$, since we have

$$C_{t|t} = C_{t|t-1} + \Delta c_{t|t} + \delta. \tag{3.121}$$

Suppose all the parameters of the model are known and the Kalman filter is applied to a model given by (3.112)–(3.114) to obtain $C_{t|t}$, $t = 1, 2, \ldots, T$. The relationship between $C_{t|t}$ and $\Delta Y_t = [\ \Delta Y_{1t} \quad \Delta Y_{2t} \quad \Delta Y_{3t} \quad \Delta Y_{4t}\ ]'$ can be given by

$$\Delta C_{t|t} = W(L)\Delta Y_t, \tag{3.122}$$

which states that $\Delta C_{t|t}$ is a function of current and past values of ΔY_1, ΔY_2, ΔY_3, and ΔY_4. Taking expectations on both sides of the above equation, we have:

$$E[\Delta C_{t|t}] = E[W(L)\Delta Y_t], \tag{3.123}$$

$$==> \ \delta = W(1)E(\Delta Y), \tag{3.124}$$

$$==> \ \hat{\delta} = W(1)\Delta\bar{Y}, \tag{3.125}$$

from which we can see that, once $W(L)$ is identified, δ is easily estimated given $\Delta\bar{Y}$. But note that the relation between $\Delta c_{t|t}$ and Δy_t is also given by

$$\Delta c_{t|t} = W(L)\Delta y_t, \tag{3.126}$$

which suggests that $W(L)$, and thus, $W(1)$, may be identified from the model in deviation from means. The following explains Stock and Watson's (1991) approach to identifying $W(1)$.

From the Kalman filter recursion applied to the state-space model written in deviation from means, we have

$$\beta_{t|t} = \beta_{t|t-1} + K_t\eta_{t|t-1}, \tag{3.127}$$

$$\beta_{t|t} = \beta_{t|t-1} + K_t(\Delta y_t - H\beta_{t|t-1}), \tag{3.128}$$

$$\beta_{t|t} = F\beta_{t-1|t-1} + K_t\Delta y_t - K_t H\beta_{t|t-1}, \tag{3.129}$$

$$\beta_{t|t} = F\beta_{t-1|t-1} + K_t\Delta y_t - K_t HF\beta_{t-1|t-1}, \tag{3.130}$$

$$\beta_{t|t} = (I - K_t H)F\beta_{t-1|t-1} + K_t\Delta y_t. \tag{3.131}$$

Harvey (1989) shows that for a stationary transition equation, the Kalman gain, K_t, approaches a steady-state Kalman gain, K, as $t \to \infty$. Given parameter estimates of the model, apply the Kalman filter to the model in deviation from means. The Kalman gain at the last iteration, K_T, is the steady-state Kalman Gain. If one prints K_t for $t = 1, 2, \ldots, T$, one will notice that K_t converges to a steady-state value reasonably fast. Thus at a steady state, we have $\beta_{t|t} = \beta_{t-1|t-1}$ and $K_t = K$, and from equation (3.131), we get

$$\beta_{t|t} = (I - (I - KH)FL)^{-1}K\Delta y_t, \tag{3.132}$$

where L is the lag operator. Because the (1,1) element of $\beta_{t|t}$ is $\Delta c_{t|t}$, $W(L)$ is given by the (1,1) element of $(I - (I - KH)FL)^{-1}K$. Thus, $W(1)$ in (3.125) is given by the (1,1) element of $(I - (I - KH)F)^{-1}K$.

3.5.3 Empirical Results

In this section, we apply Stock and Watson's (1991) model to the four coincident variables for the updated sample period 1960.1–1995.1. In addition, Δy_{4t} in equation (3.116) is replaced by

$$\Delta y_{4t} = \gamma_{40}\Delta c_t + \gamma_{41}\Delta c_{t-1} + \gamma_{42}\Delta c_{t-2} + \gamma_{43}\Delta c_{t-3} + e_{4t}, \tag{3.133}$$

in order to account for the possibility that the employment variable, ΔY_{4t}, might be slightly lagging. The state-space representation of the model then is given by

Measurement Equation

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \\ \Delta y_{3t} \\ \Delta y_{4t} \end{bmatrix} = \begin{bmatrix} \gamma_1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \gamma_2 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \gamma_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \gamma_{40} & \gamma_{41} & \gamma_{42} & \gamma_{43} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \Delta c_t \\ \Delta c_{t-1} \\ \Delta c_{t-2} \\ \Delta c_{t-3} \\ e_{1t} \\ e_{1,t-1} \\ e_{2t} \\ e_{2,t-1} \\ e_{3t} \\ e_{3,t-1} \\ e_{4t} \\ e_{4,t-1} \end{bmatrix}, \tag{3.134}$$

Transition Equation

$$
\begin{bmatrix} \Delta c_t \\ \Delta c_{t-1} \\ \Delta c_{t-2} \\ \Delta c_{t-3} \\ e_{1t} \\ e_{1,t-1} \\ e_{2t} \\ e_{2,t-1} \\ e_{3t} \\ e_{3,t-1} \\ e_{4t} \\ e_{4,t-1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \psi_{11} & \psi_{12} & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \psi_{41} & \psi_{42} \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \Delta c_{t-1} \\ \Delta c_{t-2} \\ \Delta c_{t-3} \\ \Delta c_{t-4} \\ e_{1,t-1} \\ e_{1,t-2} \\ e_{2,t-1} \\ e_{2,t-2} \\ e_{3,t-1} \\ e_{3,t-2} \\ e_{4,t-1} \\ e_{4,t-2} \end{bmatrix} + \begin{bmatrix} w_t \\ 0 \\ 0 \\ 0 \\ \epsilon_{1t} \\ 0 \\ \epsilon_{2t} \\ 0 \\ \epsilon_{3t} \\ 0 \\ \epsilon_{4t} \\ 0 \end{bmatrix}. \tag{3.135}
$$

Parameter estimates of the model are reported in table 3.3. For various issues related to the specification tests, refer to Stock and Watson 1991. Given these parameter estimates, we get $\Delta c_{t|t}$, $t = 1, 2, \ldots, T$, by running the Kalman filter again. Then, following the steps in section 3.5.2, we get an estimate of δ, the mean of the first differences of C_t. Notice that the new coincident index is identified only up to an arbitrary choice of the initial value. Thus, with an arbitrary starting values for $C_{0|0}$, we calculate $C_{t|t}$, $t = 1, 2, \ldots, T$ using equation (3.121):

$$C_{t|t} = C_{t|t-1} + \Delta c_{t|t} + \hat{\delta}.$$

For direct comparison of the new coincident index to the DOC coincident index, the following adjustments are made to the new coincident index:

$$C^*_{t|t} = C_{t|t} * SD(\Delta C_t^{\mathrm{DOC}})/SD(\Delta c_{t|t}), \tag{3.136}$$

where $SD(\Delta C_t^{\mathrm{DOC}})$ and $SD(\Delta c_{t|t})$ refer to standard deviations of the first differences of the DOC index and of $\Delta c_{t|t}$, respectively. Finally, $C^*_{t|t}$ is adjusted again so that it is equal to the value of the DOC index in January 1970. The resulting new index of coincident indicators is depicted in figure 3.15 against the DOC coincident index. The new index reveals higher growth in the 1970s and lower growth in the 1980s and 1990s than the DOC index.

Table 3.3
Parameter estimates of stock and Watson's (1991) Dynamic Facto Model of Coincident Indicators

Variables	Parameters	Estimates	
Δc_t	ϕ_1	0.5505	(0.0816)
	ϕ_2	0.0079	(0.0759)
	γ_1	0.6412	(0.0414)
Δy_{1t}	ψ_{11}	−0.0903	(0.0872)
	ψ_{12}	−0.0020	(0.0039)
	σ_1^2	0.2439	(0.0327)
	γ_2	0.2420	(0.0234)
Δy_{2t}	ψ_{21}	−0.3156	(0.0509)
	ψ_{22}	−0.0249	(0.0080)
	σ_2^2	0.3079	(0.0223)
	γ_3	0.5071	(0.0404)
Δy_{3t}	ψ_{31}	−0.3703	(0.0541)
	ψ_{32}	−0.0343	(0.0100)
	σ_3^2	0.6412	(0.0497)
	γ_{40}	0.1402	(0.0110)
	γ_{41}	0.0081	(0.0131)
Δy_{4t}	γ_{42}	0.0105	(0.0129)
	γ_{43}	0.0438	(0.0100)
	ψ_{41}	−0.0261	(0.0671)
	ψ_{42}	0.3435	(0.0751)
	σ_4^2	0.0191	(0.0021)
Log likelihood		305.19	

Note: Standard errors are in parentheses.

Appendix: GAUSS Programs to Accompany Cha ɔter 3

1. TVP.OPT: A time-varying-parameter model of U.S. monetary growth function (based on Kim and Nelson 1989).

2. UC_UNI.OPT: A univariate unobserved compo ıents model of U.S. real GDP (based on Clark 1987).

3. UC_BI.OPT: A bivariate unobserved component: model of U.S. real GDP and unemployment rate (based on Clark 1989).

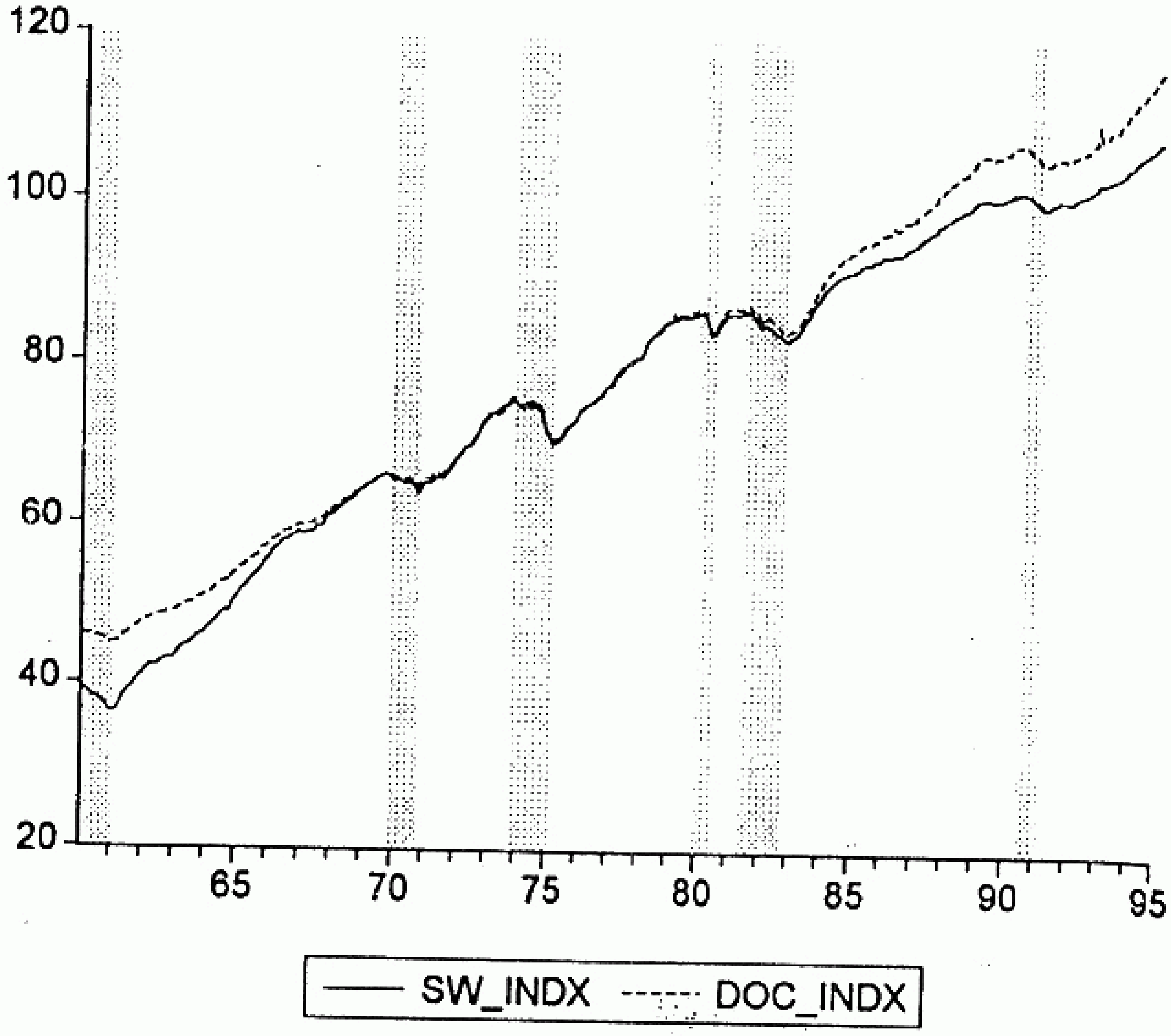

Figure 3.15
The Stock and Watson index from dynamic factor model vs. DOC coincident index

4. S&W.OPT: A dynamic factor model of four coincident economic indicators: an experimental coincident index (based on Stock and Watson 1991).

References

Antoncic, Madelyn. 1986. "High and Volatile Real Interest Rates: Where Does the Fed Fit In?" *Journal of Money, Credit, and Banking,* 18(1), 18–27.

Bera, Anil K., and Sangkyu Lee. 1993. "Information Matrix Test, Parameter Heterogeneity and ARCH: A Synthesis." *Review of Economic Studies,* 60, 229–240.

Burmeister, Edwin, and Kent D. Wall. 1982. "Kalman Filtering Estimation of Unobserved Rational Expectations with an Application to German Hyperinflation." *Journal of Econometrics,* 20, 255–284.

Burmeister, Edwin, Kent D. Wall, and James Hamilton. (1986). "Estimation of Unobserved Expected Monthly Inflation Using Kalman Filtering." *Journal of Business and Economic Statistics,* 4, 147–160.

Burns, A. F., and W. C. Mitchell. 1946. *Measuring Business Cycles.* New York: National Bureau of Economic Research.

Clark, Peter K. 1987. "The Cyclical Component of U.S. Economic Activity." *Quarterly Journal of Economics,* 102, 797–814.

Clark, Peter K. 1989. "Trend Reversion in Real Output and Unemployment." *Journal of Econometrics,* 40, 15–32.

Engle, R. F. 1982. "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation." *Econometrica,* 46 (50) 987–1007.

Engle, Robert F., and Mark Watson. 1981. "A One-Factor Multivariate Time Series Model of Metropolitan Wage Rates." *Journal of the American Statistical Association,* 76(376), Applications Section, 774–781.

Engle, Robert F., and Mark Watson. 1987. "Kalman Filter: Applications to Forecasting and Rational-Expectations Models." In *Advances in Econometrics,* Fifth World Congress, of the Econometric Society, ed. Truman Bewley, vol. 1, 245–281.

Garbade, Kenneth D., and Paul Wachtel. (1978). "Time Variation in the Relationship between Inflation and Interest Rates." *Journal of Monetary Economics,* 4 775–765.

Hai, Weike, Nelson C. Mark, and Yangru Wu. 1996. "Understanding Spot and Forward Exchange Rate Regressions." Working paper no. 96-03, Ohio State University, Columbus.

Hamilton, James D. 1994a. *Time Series Analysis.* Princeton, NJ: Princeton University Press.

Hamilton, James D. 1994b. "State-Space Models." In *Handbook of Econometrics,* ed. R. F. Engel and D. L. McFadden, vol. 4, chap. 50, 3014–3077.

Harrison, P. J., and C. F. Stevens. 1976. "Bayesian Forecasting." *Journal of the Royal Statistical Society,* Series B, 38, 205–247.

Harvey, A. C. 1987. "Applications of the Kalman Filter in Econometrics." In *Advances in Econometrics,* Fifth World Congress of the Econometric Society, ed. by Truman Bewley, vol. 1., 285–313.

Harvey, Andrew C. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge: Cambridge University Press.

Harvey, Andrew C. 1990. *The Econometric Analysis of Time Series.* 2d ed. Cambridge: MIT Press.

Kalman, R. E. 1960. "A New Approach to Linear Filtering and Prediction Problems." *Transactions ASME Journal of Basic Engineering,* D82, 35–45.

Kim, Chang-Jin, and Charles R. Nelson. 1989. "The Time-Varying-Parameter Model for Modeling Changing Conditional Variance: The Case of the Lucas Hypothesis." *Journal of Business and Economic Statistics,* 7(4) 433–440.

Kim, Chang-Jin, and Charles R. Nelson. 1997. "Friedman's Plucking Model of Business Fluctuations: Tests and Estimates of Permanent and Transitory Components." Working paper, University of Washington, Seattle.

McNees, S. K. 1986. "Modeling the Fed: A Forward-Looking Monetary Policy Reaction Function." *New England Economic Review,* Nov./Dec., 3–8.

Nelson, Charles R. 1988. "Spurious Trend and Cycle in the State Space Decomposition of a Time Series with a Unit Root." *Journal of Economic Dynamics and Control,* 12, 475–488.

Nelson, C. R., and C. I. Plosser. 1982. "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications." *Journal of Monetary Economics,* 10, 139–162.

Perron, P. 1990. "Testing for a Unit Root in a Time Series with a Changing Mean." *Journal of Business and Economic Statistics,* 8, 153–162.

Stock, James H., and Mark W. Watson. 1991. "A Probability Model of the Coincident Economic Indicators." In *Leading Economic Indicators: New Approaches and Forecasting Records,* ed. K. Lahiri and G. H. Moore. Cambridge: Cambridge University Press 63–89.

Tsay, Ruey S. 1987. "Conditional Heteroscedastic Time Series Models." *Journal of the American Statistical Association,* (82)(398) Theory and Methods, 590–604.