Federal Reserve Bank of Minneapolis
Research Department Working Paper

ESTIMATING CONTINUOUS TIME RATIONAL EXPECTATIONS MODELS IN FREQUENCY DOMAIN: A CASE STUDY

Lawrence J. Christiano*

Working Paper 301

May 1987

ABSTRACT

This paper presents a completely worked example applying the frequency domain estimation strategy proposed by Hansen and Sargent [1980,1981a]. A bivariate, high order continuous time autoregressive moving average model is estimated subject to the restrictions implied by the rational expectations model of the term structure of interest rates. The estimation strategy takes into account the fact that one of the data series are point-in-time observations, while the other are time averaged. Alternative strategies are considered for taking into account nonstationarity in the data. Computing times reported in the paper demonstrate that estimation using the techniques of Hansen and Sargent is inexpensive.

*Federal Reserve Bank of Minneapolis. This paper represents an extension on sections of my doctoral dissertation. I am grateful to my advisors, Tom Sargent and John Taylor, for their encouragement and advice.

The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System. The material contained is of a preliminary nature, is circulated to stimulate discussion, and is not to be quoted without permission of the author.

1. Introduction

The purpose of this paper is to provide a completely worked example applying the frequency domain, continuous time estimation strategy advocated in Hansen and Sargent [1980,1981a]. The model we consider is a bivariate model of the term structure of interest rates. In this model, the short rate is the overnight federal funds rate and the long rate is the one year treasury bill rate. Since the federal funds rate is approximately a call rate, the rational expectations term structure hypothesis implies that the long rate is the integral of expected federal funds rates over one year. (A discussion of the rational expectations term structure hypothesis appears in Sargent [1979a,b] and Hansen and Sargent [1981b].)

Since the restrictions of the term structure hypothesis are very similar to those found in many other rational expectations models, the methodological lessons learned from the present study can be expected to apply generally. 1/ For example, in the estimation results reported below, we are forced to take into account the fact that the data on the short rate are temporally averaged, while those on the long rate are beginning of period, point-in-time observations. This mixture of averaged and point-in-time data is characteristic of settings in which some of the data are stocks (usually available point-in-time) and others are flows (usually temporally averaged). In addition, as in most empirical settings, we are forced to confront evident nonstationarity in the data. We report results from linearly detrending the data and from taking first differences. In both cases, we are careful to reconcile the detrending procedure with the underlying continuous time economic model.

What we find is that the data reject the term structure hypothesis according to the likelihood ratio criterion. A comparison of restricted and unrestricted moving average representations suggest that the reason for the rejection is principally because the restricted model fails to adequately describe the response of the averaged short rate to an innovation. Our negative findings are consistent with those of Hansen and Sargent [1981b], who test the term structure hypothesis using a different data set also drawn from the post war period.

The plan of the paper is as follows. Section 2 describes the term structure hypothesis and derives a class of restricted continuous time representations for the bivariate short rate/long rate process. Section 3 discussses the problem of parameter identification. Section 4 describes the estimation criterion used and summarizes the relevant asymptotic distribution theory. Section 5 describes calculations capable of producing a discrete time representation corresponding to each element of a class of continuous time representations. Section 6 presents and analyzes the empirical results. Conclusions appear in Section 7.

2. The Cross Equation Restrictions in Continuous Time

Let $R(t)$ denote the call rate and $R_n(t)$ the n-period long rate. Thus, according to the term structure hypothesis,

(1) $$R_n(t) = \frac{1}{n} E_t \int_0^n R(t+\tau)d\tau,$$

where E_t denotes the linear least squares projection on $\{R(t-\ell), R_n(t-\ell);\ \ell \geq 0\}$. Let $y(t) = (R(t)', R_n(t)')'$. We assume that $\{y(t)\}$ can be represented as a continuous time stochastic differential equation with moving average errors:

(2) $$\theta(D)y(t) = C(D)\varepsilon(t),$$

where $D = \frac{d}{dt}$, and $\{\varepsilon(t)\}$ is a continuous time, Gaussian white noise with

(3) $$E\varepsilon(t)\varepsilon(t-\tau)^T = \delta(\tau)I.$$

Here, $\delta(\tau)$ is the Dirac delta function, which can informally be thought of as a function with all weight concentrated at $\tau = 0$ and none elsewhere. 2/

In (2),

$$C(s) = \begin{bmatrix} \psi_1(s) & \psi_2(s) \\ \delta_1(s) & \delta_2(s) \end{bmatrix},$$

(4)
$$\begin{aligned}
\psi_i(s) &= \psi_0^i + \psi_1^i s + \cdots + \psi_{m-1}^i s^{m-1}, \qquad i = 1, 2 \\
\delta_i(s) &= \delta_0^i + \delta_1^i s + \cdots + \delta_{m-1}^i s^{m-1} \\
\theta(s) &= \theta_0 + \theta_1 s + \cdots + \theta_{m-1} s^{m-1} + s^m \\
&= (s-\rho_1) \cdots (s-\rho_m),
\end{aligned}$$

where m > 1, s is a complex variable and $\rho_i \neq \rho_j$ for i ≠ j. Also, we adopt the normalizations ψ_0^1, $\psi_1^2 \geq 0$, $\psi_0^2 \equiv 0$ and det(C(s)) = 0 implies that the real part of s is nonpositive. The ρ's in equation (4) are assumed to have negative real part. This guarantees that the {y(t)} process is covariance stationary. (Later, we will also consider the case where $\rho_i \to 0$ for some i.) Equation (1) places restrictions on the elements of (4). In appendix A, these are shown to be

(5a) $$\delta_i(D) = \sum_{k=1}^{m} P_k^i \alpha_k \prod_{\substack{j=1 \\ j \neq k}}^{m} (D-\rho_j),$$

i = 1, 2. Here,

(5b) $$P_k^i = \frac{\psi_i(\rho_k)}{\prod_{\substack{j=1 \\ j \neq k}}^{m} (\rho_k - \rho_j)},$$

(5c) $$\alpha_k = (e^{\rho_k n} - 1)/(n\rho_k),$$

k = 1, ..., m, and i = 1, 2. Evidently, P_k^i are the weights in the following partial fraction expansion:

(6) $$\frac{\psi_i(s)}{\theta(s)} = \sum_{k=1}^{m} \frac{P_k^i}{s - \rho_k},$$

i = 1, 2. It is easily verified from (5a) that, 3/

(7) $$\det\left(C(\rho_k)\right) = 0 \qquad k = 1, \ldots, m.$$

An alternative way of writing (5) is as follows. Let H = $[h_{ij}]$, F = $[f_{ij}]$, where $h_{ij} = \rho_i^{(j-1)}$, $f_{ij} = \alpha_i \rho_j^{(j-1)}$, i, j = 1, ..., m. Also, write

$$\delta = \begin{bmatrix} \delta_0 \\ \delta_1 \\ \vdots \\ \delta_{m-1} \end{bmatrix} \qquad \psi = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_{m-1} \end{bmatrix},$$

where $\delta_k = (\delta_k^1, \delta_k^2)$ and $\psi_k = (\psi_k^1, \psi_k^2)$, $k = 1, \ldots, m$. Using this notation, (5) is equivalent with

(8) $\qquad H\delta = F\psi.$

3. Parameter Identification

Results on consistency and asymptotic distribution require that the parameters of the econometric model be identified. In particular, they require that the mapping from the true structural parameters to the discrete time spectral density of the observed data possess a unique inverse in the space of admissible parameters. (See Kohn [1979] and Dunsmuir and Hannan [1976].) It is conceptually convenient to decompose the problem of parameter identification into two parts: (i) the uniqueness of the inverse mapping from the discrete spectral density to the continuous time spectrum, and (ii) the uniqueness of the inverse map from the continuous time spectrum to the space of admissible parameter values. The former problem is called the aliasing identification problem, while the latter can be called the "classical identification problem" because it arises in basically the same form in the context of discrete time estimation.

It is beyond the scope of the paper to provide an extensive discussion of the aliasing and classical identification problems. A discussion of the nature of the aliasing identification problem appears in Hansen and Sargent [1981c,1983] and in Phillips [1973], while the classical identification problem is studied in Hannan [1969]. Christiano [1982] proves rigorously that both the classical and aliasing identification problems are resolved in the model of this paper when m = 2 and observations on the variables are point-in-time. In fact, one of the variables of this model--the federal funds rate--is temporally averaged, so that the results of Christiano [1982] do not apply directly. However, results in Christiano [1985, forthcoming] show that averaging aids identification.

We limit the remaining remarks to verifying that there is no classical identification problem in the model of equation (2). The continuous time spectral density of $\{y(t)\}$ is

$$g_y(s) = \frac{C(s)C(-s)^T}{\theta(s)\theta(-s)},$$

where $s = i\omega$, $\omega \in (-\infty,+\infty)$. Consider a nonsingular matrix A, such that AA^T. Then, clearly

$$g_y(s) = \frac{\tilde{C}(s)\tilde{C}(-s)^T}{\theta(s)\theta(-s)},$$

where $\tilde{C}(s) = C(s)A$. A necessary condition for classical identification is that the only admissible A is A = I. Since equation (8) continues to hold after postmultiplication by any matrix A whatever, it follows that the cross-equation restrictions provide no help in restricting A to be the identity matrix. It is easily verified that the restrictions which do accomplish this are $\psi_0^2 = 0$, ψ_0^1, $\psi_1^2 \geq 0$.

Another potential source of underidentification in the classical sense arises from the ability of choosing an A matrix that is a nontrivial polynomial in s, with the property that $A(s)A(-s)^T = I = A(0)$ and $C(s)A(s)$ is of order m - 1. The effect of postmultiplying $C(s)$ by $A(s)$ is to change the sign of one or several zeroes of $\det C(s)$. If we restrict (2) to be an invertible representation; i.e., in which $\det C(s)$ has zeroes that are nonpositive in real part, then the only admissible $A(s)$ is $A(s) = I$. When $m = 2$, then the cross-equation restrictions force all the zeroes of $\det C(s)$ to be nonpositive in real part by (7). When $m > 2$ parameter values must be restricted so that the roots of $\det C(s)$ that are not zeroes of $\theta(s)$ are also nonpositive in real part.

4. The Likelihood Function

The available data are $\{\bar{y}(t);\ t = 1, \ldots, T\}$, where,

$$\bar{y}(t) = \begin{cases} \int_0^1 R(t+\tau)d\tau \\ R_n(t) \end{cases} \tag{9}$$

That is, we use beginning of the month point-in-time observations on $R_n(t)$ and monthly averages on $R(t)$. Below, we derive the frequency domain approximation to the likelihood for $\{\bar{y}(t);\ t = 1, \ldots, T\}$. The empirical results in section 6 suggest that the representation of $\{\bar{y}(t);\ t = 0, \pm 1, \ldots\}$ has a unit autoregressive root. This corresponds to a value of zero for one of the ρ's in (4). In anticipation of this we also derive the frequency domain likelihood for $\{\bar{y}(t+1)-\bar{y}(t);\ t = 1, 2, \ldots, T-1\}$

when one of the ρ's in (4) is zero. In formulating the frequency domain approximation to the likelihood, we follow the suggestion of Hansen and Sargent [1980,1981a].

4a. Likelihood for $\{\bar{y}(t);\ t = 1, \ldots, T\}$

The frequency domain approximation to the likelihood function requires the spectral density of $\{\bar{y}(t);\ t = 0, \pm 1, \pm 2, \ldots\}$. This in turn requires $R_{\bar{y}}(\tau) \equiv E\bar{y}(t)\bar{y}(t-\tau)^T$, $\tau = 0, \pm 1, \pm 2, \ldots$. Hansen and Sargent [1980] show that

$$R_{\bar{y}}(\tau) = \sum_{j=1}^{m} \hat{W}_j \mu_j^{\tau} \qquad \tau > 0$$

(10) $$R_{\bar{y}}(0) = \sum_{j=1}^{m} \tilde{\hat{W}}_j$$

$$R_{\bar{y}}(\tau) = R_{\bar{y}}(-\tau)^T \qquad \text{for } \tau \le 0.$$

Here,

(11a) $$\hat{W}_j = \begin{bmatrix} W_j^{11}\left(\frac{\mu_j - 1}{\rho_j}\right)\left(\frac{1-\mu_j^{-1}}{\rho_j}\right) & W_j^{12}\left(\frac{1-\mu_j^{-1}}{\rho_j}\right) \\ W_j^{21}\left(\frac{\mu_j - 1}{\rho_j}\right) & W_j^{22} \end{bmatrix}$$

(11b) $$\tilde{\hat{W}}_j = \begin{bmatrix} W_j^{11}\,\frac{2}{\rho_j}\left(\frac{\mu_j - 1}{\rho_j} - 1\right) & W_j^{21}\left(\frac{\mu_j - 1}{\rho_j}\right) \\ W_j^{21}\left(\frac{\mu_j - 1}{\rho_j}\right) & W_j^{22} \end{bmatrix}$$

$j = 1, \ldots, m$. In addition,

(11c)
$$\mu_j = e^{\rho_j}$$

$$W_j = \frac{C(\rho_j)C(-\rho_j^T)}{-2\rho_j \prod_{\substack{k=1 \\ k\neq j}}^{m} (\rho_j-\rho_k)(-\rho_j-\rho_k)} = \begin{bmatrix} W_j^{11} & W_j^{12} \\ W_j^{21} & W_j^{22} \end{bmatrix},$$

$j = 1, \ldots, m.$

The spectral density of $\{\bar{y}(t);\ t = 0, \pm 1, \pm 2, \ldots\}$ is defined as

(12a)
$$S_{\bar{y}}(e^{-i\omega}) \equiv \sum_{\tau=-\infty}^{+\infty} R_{\bar{y}}(\tau)e^{-i\omega\tau}$$
$$= M(e^{-i\omega}) + M(e^{i\omega})^T - K,$$

where

(12b)
$$M(e^{-i\omega}) = \sum_{j=1}^{m} \frac{\tilde{W}_j}{1 - \mu_j e^{-i\omega}}$$

(12c)
$$K = \sum_{j=1}^{m} (\tilde{W}_j+\tilde{W}_j^T) - R_{\bar{y}}(0).$$

The frequency domain approximation to the log likelihood is, neglecting an additive constant,

(13)
$$\pounds^F(\bar{y}(1),\ldots,\bar{y}(T);\phi) = -\frac{1}{2}\sum_{j=1}^{T} \log\det S_{\bar{y}}(e^{-i\omega_j})$$
$$-\frac{1}{2}\sum_{j=1}^{T} \operatorname{trace}\left[S_{\bar{y}}(e^{-i\omega_j})^{-1} I(\omega_j)\right],$$

where

$$(14) \qquad I(\omega_j) = \frac{1}{T}\left[\sum_{t=1}^{T} \bar{y}(t)e^{-i\omega_j t}\right]\left[\sum_{t=1}^{T} \bar{y}(t)'e^{i\omega_j t}\right],$$

$$\omega_j = \frac{2\pi j}{T}, \qquad j = 1, \ldots, T.$$

In (13), $\pounds^F$ is a function of the free parameters, ϕ, via $S_{\bar{y}}$. The periodogram matrix, $I(\cdot)$, is a function of the data and not of ϕ. Here,

$$(15) \qquad \phi = (\theta_0,\ldots,\theta_{m-1},\psi_1^1,\psi_2^1,\ldots,\psi_{m-1}^1,\psi_0^2,\psi_1^2,\ldots,\psi_{m-1}^2)$$

The dependence of $R_{\bar{y}},S_{\bar{y}},W_j,\tilde{W}_j,\tilde{\tilde{W}}_j,\mu_j$, $j = 1, \ldots, m$ on ϕ is not made explicit in order to keep from cluttering the notation. Results from Kohn [1979] or Dunsmuir and Hannan [1976] may be invoked to show that $\hat{\phi}_T$, the maximizer of (13), converges almost surely to its true value, which we denote ϕ_0. In addition, $\sqrt{T}(\hat{\phi}_T-\phi_0)$ is asymptotically normally distributed, with variance covariance matrix that is consistently estimated by

$$(16) \qquad -\left[\frac{1}{T} \frac{\partial^2 \pounds^F}{\partial\phi\partial\phi^T}\Big|_{\phi=\hat{\phi}_T}\right]^{-1}$$

4b. Likelihood for the First Differenced Data

In this section, we set $\theta_0 = \rho_m = 0$ and $\alpha_m = 1$ in (4) and (5c), and derive the frequency domain approximation to the likelihood of $\bar{y}(t+1) - \bar{y}(t)$, $t = 1, \ldots, T - 1$.

Let

$$(17a) \qquad \tilde{\theta}(D) = [\theta(D)/D] = \prod_{j=1}^{m-1} (D-\rho_j).$$

Then, (2) may be written

(17b) $$Dy(t) = [\frac{C(D)}{\tilde{\theta}(D)}]\varepsilon(t).$$

Since C(D) and $\tilde{\theta}(D)$ are both polynomials of order m - 1, $\{Dy(t), t\varepsilon(-\infty,+\infty)\}$ is a generalized stochastic process.

Notice that $\int_0^1 Dy(t+\tau)d\tau = y(t+1) - y(t)$. Also, the operator notation for unit averaging is $[(e^D-1)/D]$. Therefore, unit averaging (17) yields

(18) $$y(t+1) - y(t) = (\frac{e^D-1}{D})(\frac{C(D)}{\tilde{\theta}(D)})\varepsilon(t).$$

It is easy to verify that $\{y(t+1)-y(t), t\varepsilon(-\infty,+\infty)\}$ forms an ordinary, covariance stationary, stochastic process.

Define

(19) $$R_{\Delta y}(\tau) \equiv E\Delta y(t)\Delta y(t-\tau)^T, \ \tau\varepsilon(-\infty,+\infty)$$

where $\Delta y(t) \equiv y(t+1) - y(t)$. Appendix B shows how $R_{\Delta y}$ can be recovered from the spectral density of $\Delta y(t)$ implied by (18). What we require is

(20) $$R_{\Delta\bar{y}}(\tau) \equiv E\Delta\bar{y}(t)\Delta\bar{y}(t-\tau)^T, \ \tau = 0, \pm 1, \pm 2, \ldots$$

where $\Delta\bar{y}(t) = \bar{y}(t+1) - \bar{y}(t)$, and $\bar{y}(t)$ is defined in (8). The fact that $\Delta y(t)$ and $\Delta\bar{y}(t)$ are related by an averaging operator implies that $R_{\Delta\bar{y}}$ can be derived by suitably averaging $R_{\Delta y}$. Consider, for example, the 1,1 element of $R_{\Delta\bar{y}}$ and $R_{\Delta y}$, denoted by $R^{11}_{\Delta\bar{y}}$ and $R^{11}_{\Delta y}$, respectively:

(21)
$$R^{11}_{\Delta\bar{y}}(\tau) = E[\int_0^1 \{R(t+1+k)-R(t+k)\}dk][\int_0^1 \{R(t+1+\ell-\tau)-R(t+\ell-\tau)^T\}d\ell]$$

$$= \int_0^1 \int_0^1 \{E[R(t+1+k)-R(t+k)][R(t+1+\ell-\tau)-R(t+\ell-\tau)]^T\}dkd\ell$$

$$= \int_0^1 \int_0^1 R^{11}_{\Delta y}(k-\ell+\tau)dkd\ell.$$

Upon carrying out the required integration on $R_{\Delta y}$, we find,

(22)
$$R_{\Delta\bar{y}}(\tau) = \begin{cases} \sum_{j=1}^{m-1} \bar{W}_j & \tau = 0 \\ \sum_{j=1}^{m-1} \bar{\bar{W}}_j & \tau = 1 \\ \sum_{j=1}^{m-1} \bar{\bar{\bar{W}}}_j \, \mu_j^{\tau} & \tau > 1 \\ R_{\Delta\bar{y}}(-\tau)^T, & \tau \le 0 \end{cases}$$

where $\bar{W}_j$, $\bar{\bar{W}}_j$, and $\bar{\bar{\bar{W}}}_j$, j = 1, ..., m - 1 are given in Appendix B.

Define

$$M_{\Delta\bar{y}}(e^{-i\omega}) = \sum_{j=1}^{m-1} \frac{\bar{\bar{\bar{W}}}_j}{1 - \mu_j e^{-i\omega}}$$

(23)
$$D = R_{\Delta\bar{y}}(1) - \sum_{j=1}^{m-1} \bar{\bar{\bar{W}}}_j \mu_j$$

$$G = \sum_{j=1}^{m-1} (\bar{\bar{\bar{W}}}_j + \bar{\bar{\bar{W}}}_j^T) - R_{\Delta\bar{y}}(0).$$

Then, the spectral density of $\{\bar{y}(t+1)-\bar{y}(t);\ t = 0, \pm 1, \pm 2, \ldots\}$ is

$$\text{(24a)} \qquad S_{\Delta\bar{y}}(e^{-i\omega}) \equiv \sum_{t=-\infty}^{-\infty} R_{\Delta\bar{y}}(\tau)e^{-i\omega\tau}$$

$$= M_{\Delta\bar{y}}(e^{-i\omega}) + M_{\Delta\bar{y}}(e^{i\omega})^T + De^{-i\omega} + D^T e^{i\omega} - G$$

Finally, the frequency domain approximation to the likelihood is, ignoring an additive constant,

$$\text{(24b)} \quad \mathcal{L}^F(\bar{y}(T)-\bar{y}(T-1),\ldots,\bar{y}(2)-\bar{y}(1);\ \phi') = -\frac{1}{2}\sum_{j=1}^{T-1} \log\det S_{\bar{y}}(e^{-i\omega_j}) - \frac{1}{2}\sum_{j=1}^{T-1} \text{trace}\,[S_{\bar{y}}(e^{-i\omega_j})^{-1}\bar{I}(\omega_j)],$$

where

$$\text{(25)} \qquad \bar{I}(\omega_j) = \frac{1}{T}\Big[\sum_{t=1}^{T-1}(\bar{y}(t+1)-\bar{y}(t))e^{-i\omega_j t}\Big]\Big[\sum_{t=1}^{T-1}(\bar{y}(t+1)-\bar{y}(t))^T e^{i\omega_j t}\Big]$$

$$\omega_j = \frac{2\pi j}{T-1},\ j = 1, \ldots, T-1.$$

The parameter vector in the present case is

$$\text{(26)} \qquad \phi' = (\theta_1,\ldots,\theta_{m-1},\psi_1^1,\psi_2^1,\ldots,\psi_{m-1}^1,\psi_0^2,\psi_1^2,\ldots,\psi_{m-1}^2),$$

which coincides with ϕ in (15) except that θ_0 is not present in ϕ' since it is set to zero.

Let $\hat{\phi}_T'$ denote the admissible maximizer of (24). Then assuming $\{\Delta\bar{y}(t);\ t = 0, \pm 1, \pm 2, \ldots\}$ is covariance stationary, $\hat{\phi}_T' \xrightarrow{a.s.} \phi_0'$, where ϕ_0' denotes the true value of ϕ'. In addition, $\sqrt{T}\,(\hat{\phi}_T'-\phi_0')$ is asymptotically normal with mean zero and variance-covariance matrix consistently estimated by

(27) $$-\left[\frac{1}{T}\frac{\partial L^F}{\partial\phi'\partial(\phi')^T}\bigg|_{\phi'=\hat{\phi}'_T}\right]^{-1}.$$

These results can be justified by referring to Dunsmuir and Hannan [1976].

A factor that distinguishes the estimation problem posed in this section from that in section 4a is that $\det S_{\Delta\bar{y}}(1) = 0$. This is an implication of the facts $\det C(0) = 0$ (recall (7) and $\rho_m = 0$ and that $\Delta\bar{y}(t)$ is obtained by temporally averaging $y(t)$. 4/ A consequence of $\det S_{\Delta\bar{y}}(1) = 0$ is that the determinant of the moving average representation of sampled $\Delta\bar{y}(t)$ has a root at unity. 5,6/

The unit root in $\det S_{\Delta\bar{y}}(z)$ means that $\pounds^F$ is not well defined. We get around this by ignoring frequency zero in (24b). This has to be done in any case when mean adjusted data are used, since in this case the periodogram matrix at frequency zero is exactly zero so that the likelihood function is unbounded above. Therefore, frequency zero was ignored in (13) also.

5. Obtaining the Time Series Representation of the Sampled Data

In sections 6 and 7, use is made of the estimated time series representation of $\{\bar{y}(t), t = 0, \pm 1, \pm 2, \ldots\}$ when the latter is assumed to be covariance stationary, and of $\{\bar{y}(t+1)-\bar{y}(t), t = 0, \pm 1, \pm 2, \ldots\}$ when $\{\bar{y}(t)\}$ is assumed to have a unit autoregressive root. In this section we indicate the calculations required to obtain these representations.

5a. The Time Series Representation of $\{\bar{y}(t);\ t = 0, \pm 1, \pm 2, \ldots\}$

For simplicity in notation, let $z = e^{-i\omega}$. Then, the spectral density of $\{\bar{y}(t),\ t = 0, \pm 1, \pm 2, \ldots\}$ is, from (12),

$$(28)\qquad S_{\bar{y}}(z) = M(z) + M(z^{-1})^T - K$$

Write

$$(29)\qquad \theta^+(z) = \prod_{j=1}^{m} (1-\mu_j z).$$

Multiply (28) by $\theta^+(z)\theta^+(z^{-1})$, to get

$$(30)\qquad \theta^+(z)S_{\bar{y}}(z)\theta^+(z) = \sum_{j=1}^{m} \tilde{W}_j \prod_{\substack{k=1\\k\neq j}}^{m} (1-\mu_j z)\theta^+(z^{-1})$$

$$+ \sum_{j=1}^{m} \tilde{W}_j^T \prod_{\substack{k=1\\k\neq j}}^{m} (1-\mu_j z^{-1})\theta^+(z)$$

$$- \theta^+(z)\ K\ \theta^+(z^{-1})$$

$$= \Gamma(z),$$

say, where

$$(31)\qquad \Gamma(z) = \Gamma_0 + \Gamma_1 z + \cdots + \Gamma_m z^m + \Gamma_1^T z^{-1} + \ldots + \Gamma_m^T z^{-m}.$$

Using the algorithms in Whittle [1983, Chapter 9, section 3] or Rozanov [1967, Chapter I, section 10], (31) can be factored to give

$$(32a)\qquad \Gamma(z) = C^+(z)C^+(z^{-1})^T,$$

where

$$\text{(32b)}\quad \begin{aligned} &\text{dec } C^+(z) = 0 \text{ implies } |z| \geq 1 \\ &C^+(z) = C_0^+ + C_1^+ z + \cdots + C_m^+ z^m,\ C_0^+ \text{ lower triangular.} \end{aligned}$$

(Whittle's algorithm assumes $\det\Gamma(z) \neq 0$ for $|z| = 1$.) The time series representation of $\{\bar{y}(t),\ t = 0, \pm 1, \pm 2, \ldots\}$ is then

$$\theta^+(L)\bar{y}(t) = C^+(L)u(t+1),$$

where $\{u(t),\ t = 0, \pm 1, \pm 2, \ldots\}$ is white with variance equal to the identity matrix. (The dating on $u(t)$ is chosen to reflect the fact that the first element in $\bar{y}(t)$--the federal funds rate--is an average from t to t + 1.)

It may be verified that $C_m^+ \neq 0$ in (32b) because of the effects of averaging $y(t)$. In particular, the analogue of C_m^+ in the time series representation of $\{y(t),\ t = 0, \pm 1, \ldots\}$ is zero.

5b. The Time Series Representation of $\{\bar{y}(t+1)-\bar{y}(t);\ t = 0, \pm 1, \pm 2, \ldots\}$

We use the same approach described in the previous case. Set $z = e^{-i\omega}$. Then the spectral density of $\{\Delta\bar{y}(t),\ t = 0, \pm 1, \pm 2, \ldots\}$ is

$$\text{(33)}\quad S_{\Delta\bar{y}}(z) = M_{\Delta\bar{y}}(z) + M_{\Delta\bar{y}}(z^{-1})^T + Dz + D^T z^{-1} - G,$$

by (24a). Define

$$\text{(34)}\quad \tilde{\theta}^+(z) = \prod_{j=1}^{m-1} (1-\mu_j z).$$

Multiply (33) by $\tilde{\theta}^+(z)\tilde{\theta}^+(z^{-1})$ to get

(35)
$$\tilde{\theta}^{+}(z)S_{\Delta\bar{y}}(z)\,\tilde{\theta}^{+}(z^{-1}) = \sum_{j=1}^{m-1} \bar{\bar{W}}_j \prod_{\substack{k=1\\k\neq j}}^{m-1} (1-\mu_k z)\tilde{\theta}^{+}(z^{-1})$$
$$+ \sum_{j=1}^{m-1} \bar{\bar{W}}_j^T \prod_{\substack{k=1\\k\neq j}}^{m-1} (1-\mu_k z^{-1})\tilde{\theta}^{+}(z)$$
$$+ D\tilde{\theta}^{+}(z)\tilde{\theta}^{+}(z^{-1})z + D^T\tilde{\theta}^{+}(z)\tilde{\theta}^{+}(z^{-1})z^{-1}$$
$$- G\,\tilde{\theta}^{+}(z)\tilde{\theta}^{+}(z^{-1})$$
$$= \tilde{\Gamma}(z),$$

say, where

(36)
$$\tilde{\Gamma}(z) = \tilde{\Gamma}_0 + \tilde{\Gamma}_1 z + \cdots + \tilde{\Gamma}_m z^m + \tilde{\Gamma}_1^T z^{-1} + \ldots + \tilde{\Gamma}_m^T z^{-m}.$$

Here again, we may find a polynomial matrix $\tilde{C}(z)$, with the following properties:

(37)
$$\tilde{C}^{+}(z) = \tilde{C}_0^{+} + \tilde{C}_1^{+}z + \cdots + \tilde{C}_m^{+}z^m$$
$$\det \tilde{C}^{+}(z) = 0 \text{ implies } |z| \geq 1$$
$$\tilde{C}_0^{+} \text{ lower triangular,}$$
$$\tilde{C}^{+}(z)\tilde{C}^{+}(z^{-1})^T = \tilde{\Gamma}(z).$$

Since det $\tilde{\Gamma}(1) = 0$ in the present case, Whittle's factorization algorithm is inapplicable. For this reason, Rozanov's algorithm was used to solve (37).

The time series representation for $\{\Delta\bar{y}(t),\ t = 0, \pm 1, \pm 2, \ldots\}$ is given by

$$(38) \qquad \tilde{\theta}^{+}(L)\Delta\bar{y}(t) = C^{+}(L)u(t+2),$$

where $\{u(t),\ t = 0, \pm 1, \pm 2, \ldots\}$ is a discrete time white noise with variance matrix equal to the identity matrix. (The dating on $\tilde{u}$ reflects, first, that $\Delta\bar{y}(t) \equiv \bar{y}(t+1) - \bar{y}(t)$, and second, that the first element of $\bar{y}(t+1)$--the federal funds rate--is an average from $t + 1$ to $t + 2$.)

6. Empirical Results

In this section estimation results are presented using data on the term structure of interest rates. The data are monthly observations on the federal funds rate and the one year ($n = 12$) treasury bill rate. Observations from August 1954 to November 1977 ($T = 280$) were obtained from Salomon Brothers [1977] and are in percent terms.

The raw data on the treasury bill rate are middle-of-the-month observations from August 1954 to December 1958 (53 observations) and first-of-the month observations thereafter. The entire T-bill series represents point-in-time observations. The first 53 observations were converted to an approximate first-of-the month series by linear interpolation.

In the first part of this section, the results of analyzing the demeaned, linearly detrended data are presented. The time series representation of these data appear to have a unit root. Consequently, in the second part of this section, results are presented of analysis using first differenced data.

In each case, the likelihood ratio statistic is calculated by doubling the difference between the maximized values of the restricted and unrestricted log likelihood functions. The number of degrees of freedom is computed as the difference between the number of free parameters under the null hypothesis and the number of free parameters under the alternative hypothesis.

The alternative hypothesis is that the sampled data have a stationary representation. The order of this representation is the same as that of the discrete time representation corresponding to the model of the null hypothesis. It follows that the likelihood ratio statistic is a test of the following joint hypothesis: (a) that the data obey the term structure restrictions, (1); (b) that the continuous time representation of the short rate/long rate process has the rational spectral density implied by the specification in (2); and (c) that the averaging procedure actually applied to the data corresponds to the way we model it. That (b) implies restrictions on the discrete time representation is well known. Consider, for example, the scalar case in which $C(s) = C$ and $\theta(s) = s - \rho$. In this case, $y(t) = e^{\rho}y(t-1)+u(t)$, where $\{u(t), t = 0, \pm 1, \pm 2, \ldots\}$ is a white noise with variance $C^2[(e^{2\rho}-1)/(2\rho)]$. (See Phillips [1973].) Since $0 < e^{\rho} < 1$, the given continuous time specification implies a sampled representation which is AR(1) with autoregressive coefficient constrained to be positive.

6a. Analyzing the Demeaned, Detrended Data

The term structure theory discussed up to now applies to the raw data (or its first difference) and not to the demeaned and detrended data. Hence, before presenting the results of the analysis using the transformed data, an interpretation of the transformation needs to be given. Write

$$(39)\qquad \begin{bmatrix} R(t) \\ \\ R_n(t) \end{bmatrix} = \begin{bmatrix} c \\ \\ c_n \end{bmatrix} + \begin{bmatrix} \alpha \\ \\ \alpha_n \end{bmatrix} t + \begin{bmatrix} \tilde{R}(t) \\ \cdot \\ \tilde{R}_n(t) \end{bmatrix},$$

where, as before, $R(t)$ represents the call rate (approximated by the federal funds rate) and $R_n(t)$ represent the n-period long rate. According to the term structure hypothesis,

$$(1)\qquad R_n(t) = \frac{1}{n} \int_0^n E_t R(t+\tau)d\tau.$$

Substituting the first equation of (39) into (1),

$$(40)\qquad R_n(t) = c + \alpha t + \frac{\alpha}{n} \int_0^n \tau d\tau + \frac{1}{n} \int_0^n E_t \tilde{R}(t+\tau)d\tau$$

$$= c + \alpha t + \frac{\alpha}{2} n + \frac{1}{n} \int_0^n E_t \tilde{R}(t+\tau)d\tau.$$

Substituting the second equation of (39) into (40),

$$\tilde{R}_n(t) = \left[c - c_n + \frac{\alpha}{2} n\right] + (\alpha - \alpha_n)t + \frac{1}{n} \int_0^n E_t \tilde{R}(t+\tau)d\tau.$$

The latter equation implies that if,

$$(41)\qquad c_n = c + \frac{\alpha}{2} n$$

$$\alpha_n = \alpha,$$

then

(42) $$\bar{R}_n(t) = \frac{1}{n}\int_0^n E_t\bar{R}(t+\tau)d\tau.$$

Thus, proceeding as though the term structure theory (equation (1)) applies to the demeaned and detrended series $(\bar{R}_n(t),\bar{R}(t))$ is tantamount to imposing (41). Letting t vary from t = 1 in August 1954 to t = 280 in November 1977, OLS yields

(43) $$\hat{c}_{12} = .16, \qquad \hat{\alpha}_{12} = .0016$$
$$\hat{c} = .11, \qquad \hat{\alpha} = .0019.$$

Substituting $\hat{c}$ and $\hat{\alpha}$ from (43) into (41) to obtain $\tilde{c}_{12}$ and $\tilde{\alpha}_{12}$ we get,

(44) $$\tilde{c}_{12} = .12$$
$$\tilde{\alpha}_{12} = .0019.$$

Evidently, $\tilde{c}_{12}$ and $\hat{c}_{12}$, and $\tilde{\alpha}_{12}$ and $\hat{\alpha}_{12}$ deviate by 33 percent and 16 percent, respectively, from what is required by (41). Confidence intervals for the estimates in (43) were not calculated, hence it cannot be said whether the data reject (41). It seems clear that, at best, the data provide only weak evidence in favor of (41). We proceed now to formally test the remainder, (42) of the term structure restrictions, (1).

[Insert Tables 1 and 2]

The results of estimating the term structure model with $m = 2$ are presented in Tables 1 and 2. Table 1a shows the estimated values of the continuous time parameters. Note that one of the autoregressive roots of the continuous time system is very close to zero. The corresponding discrete time representation appears in Table 1b where the zero root of Table 1a shows up as a root at unity. The results of estimating an unrestricted model appear in Table 2. The Hessian of the log-likelihood function is singular at the indicated parameter values. The parameter values nevertheless do appear to be near a local maximum since the gradient is small. (The gradient is not reported here.) Because of the singularity, asymptotic standard errors could not be computed. According to the likelihood ratio statistic, the data appear to reject the hypothesis, (1), at any reasonable confidence level. Because the time series representation of the sampled data appears to have a unit autoregressive root, the function--equation (12)--that was used in estimation is probably not a good approximation to the exact likelihood function. This casts doubt on the results obtained using it. For this reason the analysis was redone using first-differenced data.

6b. Analyzing the First Differenced, Demeaned Data

Here, the results of analyzing the first differenced data are reported. The estimation criterion used was (24) without frequency zero and with $m = 3$.

[Insert Tables 3 and 4]

Parameter estimates and asymptotic errors of the restricted continuous time parameters appear in Table 3. The representation reported there is that of the continuous time derivative process $(DR(t),DR_{12}(t))$. As was explained below equation (17), this is a generalized stochastic process.

Results of estimating the unrestricted discrete-time model corresponding to the one in Table 3 are presented Table 4. Once again, note in Table 4 that the unit pole that appeared in Table 2 has been eliminated by first-differencing. Applying standard asymptotic theory, the likelihood ratio statistic indicates spectacular rejection of the null hypothesis. Because of the unit root, however, standard asymptotic theory does not apply in the present context. (See, for example, Sargan and Bhargava [1983].) Our conclusion, therefore, requires that the correct critical region not be drastically smaller that the critical region under normal theory.

The Chi-square statistics reported in Tables 2 and 4 are the same order of magnitude as those reported in Hansen and Sargent [1981b]. Working in discrete time they test the term structure hypothesis in which the short rate is a three-month treasury bill rate and the long rate is a five year treasury bond rate. They report a Chi-square statistic of 20 with 4 degrees of freedom under the null hypothesis. The area under the Chi-square density function with 4 or 10 degrees of freedom between 20 and 100 is on the order of .0005.

Hansen and Sargent [1981b] adopt an estimation strategy which avoids the unit root problem encountered in this paper. Consequently, they can appeal to the results of Kohn [1979] according to which the likelihood ratio statistic has an asymptotic Chi-square distribution. Their procedure is to model the bivariate process formed by the first difference of the short rate and the difference between the short and long rates. It is easy to show that, in the context of this paper, this process is covariance stationary. Moreover, the determinant of the moving average part of the resulting sampled representation need not have a root at unity. 7/

6c. Computing Time and Costs

All calculations were performed on the University Minnesota Cyber 172 computer. The nonlinear optimization routines used were taken from a package of routines called GQOPT. The package was obtained from S. M. Goldfeld of Princeton University.

Table 5 reports costs and computing times associated with the calculations reported in Tables 1 through 4. The information that seems especially relevant is that reported in the first column. There the amount of time needed to evaluate the likelihood function once is given. This quantity is independent of the optimization routine used or of starting values. The figures in this column indicate that the cost per function evaluation of estimating the parameters of the continuous time system are well within the feasible range. It is particularly interesting to note how very little computer time it takes to carry out the calculations that are unique to continuous time estimation. These are the calculations required to obtain the discrete time spectral density, given the structural parameters. The time required can be inferred by subtracting row two from row one, or row four from row three in column one of Table 5. Doing so, we deduce that the time required is between one and five one hundredth of a second, which is a small fraction of the total time required for one function evaluation.

It should be noted that in writing the computer program to execute the calculations, I was guided by the sole objective of making the code easy to read for debugging purposes. Presumably, a serious attempt to write efficient code would have resulted in a significant drop in computer time from the figures indicated in Table 5.

6d. Impulse Response Functions

A useful diagnostic device is to compare the model's restricted and unrestricted moving average representations. Taylor [1980] used this procedure as a complement to the standard likelihood ratio statistic to evaluate the goodness-of-fit of his staggered wage contract model. Sims [1980] has emphasized the usefulness of a model's moving average representation as a way of summarizing the dynamic properties of a time series model. In this section we follow the lead of Taylor and Sims and examine moving average representations in order to understand the reason for the high likelihood ratio statistic reported in Table 4.

Let

(45a) $$B^u(L) = \hat{\tilde{\theta}}^+(L)^{-1}\hat{C}^+(L)$$

(45b) $$B(L) = \tilde{\theta}^+(L)^{-1}C^+(L)$$

where $(\hat{\tilde{\theta}}^+,\hat{C}^+)$ are obtained from Table 4 and $(\tilde{\theta}^+,C^+)$ are obtained from Table 3b. Expanding (45) in nonnegative powers of L,

(46a) $$B^u(L) = \sum_{\tau=0}^{\infty} B^u_\tau L^\tau$$

(46b) $$B(L) = \sum_{\tau=0}^{\infty} B_\tau L^\tau.$$

Then, $[\bar{y}(t+1)-\bar{y}(t)] = B^u(L)u(t+2)$ according to the unrestricted estimation results, and $[\bar{y}(t+1)-\bar{y}(t)] = B(L)u(t+2)$ according to the restricted estimation results. Here, $\{u(t)\}$ is a discrete time white noise with variance normalized to be the identity matrix. (The matrices B^u_0 and B_0 are lower triangular.)

Figures 1-4 display B^u_τ and B_τ for $\tau = 0, 1, \ldots, 30$. Define $u(t) = (u_1(t),u_2(t))^T$. Following is a discussion of the dynamic effects of $u_1(t)$ and $u_2(t)$.

Figure 1 depicts the dynamic effect of a one standard deviation jump in $u_1(t+1)$ on $\bar{y}_1(t) - \bar{y}_1(t-1)$. (The normalization that I have adopted guarantees that $u_1(t+1)$ is proportional to the innovation in $\bar{y}_1(t) - \bar{y}_1(t-1)$.) Qualitatively, the response patterns to $u_1(t)$ in the restricted and unrestricted models are similar. There is a substantial quantitative difference, however. According to the unrestricted model, the initial effect of a jump in $u_1(t+1)$ is positive, followed by about two years of very strong negative effects. Figure 3 indicates that the dynamic effects of $u_1(t+1)$ on $\bar{y}_2(t) - \bar{y}_2(t-1)$ between restricted and unrestricted models are similar.

Figures 2 and 4 describe the dynamic effects of $u_2(t)$. (Under my normalization, $u_2(t+1)$ is not the innovation in $\bar{y}_2(t) - \bar{y}_2(t-1)$. The innovation is instead a linear combination of $u_1(t+1)$ and $u_2(t+1)$.) According to Figure 2, the qualitative nature of the dynamic effects of a unit jump in $u_2(t+1)$ on $\bar{y}_2(t) - \bar{y}_2(t-1)$ is very different between the restricted and unrestricted models. The unrestricted response pattern is initially positive, followed by a negative and then a positive again, whereupon it tapers off to zero. The restricted response pattern is exactly the opposite. The restricted and unrestricted response pattern in $\bar{y}_1(t) - \bar{y}_1(t-1)$ to $u_2(t+1)$ are similar.

In order to obtain a more accurate impression of the difference between B_τ^u and B_τ, it would be useful to include confidence intervals in Figures 1-4. I have not carried out the required computations in the interest of limiting the scope of this paper. [8/]

To summarize, the principle reason for the high likelihood ratio statistic reported in Table 4 appears to be that the constrained model does a poor job of capturing the dynamic response of the monthly averaged federal funds rate to an innovation. The constrained model also does a poor job of tracking the response in the three month treasury bill rate to a $u_2(t)$ disturbance. These statements must be regarded as suggestive, rather than definitive, since they are not accompanied by standard error statistics.

7. Conclusion and Suggestions for Further Research

The principle objective of this paper, to illustrate the application of an estimation procedure proposed by Hansen and Sargent [1980,1981a], has been accomplished. It was shown that the required calculations are not time consuming by ordinary standards. Some of the computations are complicated from a programming perspective. These are the calculations required to obtain the sampled representation of a continuous time process and to obtain the discrete time spectrum of (possibly) averaged data given a continuous time spectral density. However, these calculations are common across applications and so need to be programmed only once. The computations that are problem specific are no more difficult to program than what is now required to estimate discrete time rational expectations models. These are the calculations needed to obtain the continuous time spectral density given the structural parameters of the continuous time model.

Given the above remarks, it would be useful and practical to apply frequency domain estimation techniques to the estimation of other rational expectations models. In such an application it would be of great interest to also estimate the discrete time version of the model and compare results. This would provide empirical evidence on the extent to which the choice of timing interval in an econometric model "matters" in the sense of affecting the analyst's views about which economic theory best suits the data, or which economic policy is most likely to yield a desirable outcome.

Footnotes

[1]/Observations on the federal funds rate are available on a daily basis. Therefore, it would, in principle, be feasible to test the term structure hypothesis by estimating a restricted, daily, time series model. I used the time aggregated data instead in order to illustrate the use of continuous time estimation tools.

[2]/A stochastic process whose covariance function contains a delta function is said to be a generalized stochastic process. For a discussion of this, see Hannan [1970].

[3]/An implication of this is that when m = 2, (2) reduces to a first order, bivariate stochastic differential equation. To see this, note first that (7) implies $\det C(s) = \gamma\theta(s)$ when $m = 2$. Premultiply (2) by $C(D)^{-1} = C^a(D)/(\gamma\theta(D))$, where the superscript 'a' denotes the adjoint operator. The result is

$$C^a(D)y(t) = \gamma\varepsilon(t),$$

or

$$Dy(t) = Ay(t) + u(t),$$

where $A = -(C_1^a)^{-1}C_0^a$, $u(t) = \gamma(C_1^a)^{-1}\varepsilon(t)$. For $m > 2$, (2) fails to be a finite ordered differential equation, except in singular cases.

[4]/This fact can be shown by a simple application of the folding formula, which links continuous and discrete spectral densities. Let

$$S^c_{\Delta\bar{y}}(s) = \left(\frac{e^s-1}{s}\right)\left(\frac{e^{-s}-1}{-s}\right)\frac{G(s)C(s)C(-s)^T G(-s)^T}{\tilde{\theta}(s)\tilde{\theta}(-s)},$$

where

$$G(s) = \begin{bmatrix} \frac{e^s - 1}{s} & 0 \\ 0 & 1 \end{bmatrix}.$$

$S^c_{\Delta\bar{y}}(i\omega)$ is the spectral density of $\{\Delta\bar{y}(t),\ t\varepsilon(-\infty,+\infty)\}$ at frequency ω, $\omega\varepsilon(-\infty,+\infty)$. Let

$$\tilde{S}_{\Delta\bar{y}}(s) = \frac{G(s)C(s)C(-s)^T G(-s)^T}{\tilde{\theta}(s)\tilde{\theta}(-s)},$$

and note that det $\tilde{S}_{\Delta\bar{y}}(0) = 0$, since det $C(0) = 0$. In this notation,

$$S^c_{\Delta\bar{y}}(i\omega) = \frac{2(1-\cos\omega)}{\omega^2}\tilde{S}_{\Delta\bar{y}}(i\omega), \quad \omega \ \varepsilon \ (-\infty,+\infty).$$

Now, according to the folding formula (see Fishman [1968, p.50])

$$S_{\Delta\bar{y}}(e^{-i\omega}) = \sum_{k=-\infty}^{+\infty} \frac{2[1-\cos(\omega+2\pi k)]}{[\omega+2\pi k]^2}\tilde{S}_{\Delta\bar{y}}(i[\omega+2\pi k]),$$

$\omega \ \varepsilon \ (-\pi,\pi)$. Note:

$$\frac{2[1-\cos(0+2\pi k)]}{[0+2\pi k]^2} = \begin{cases} 0 & k \neq 0 \\ 1 & k = 0 \end{cases}.$$

The result for $k = 0$ can be derived as follows. First, note that

$$\frac{2[1-\cos(2\pi k)]}{(2\pi k)^2} = \left(\frac{e^s-1}{s}\right)\left(\frac{e^{-s}-1}{-s}\right)$$

$$= [1+ \tfrac{1}{2}s+ \tfrac{1}{3!}s^2+ \tfrac{1}{4!}s^3+\ldots](1+ \tfrac{1}{2}(-s)+ \tfrac{1}{3!}(-s)^2+\ldots),$$

for $s = 2\pi ki$. Then, set $s = 0$. Conclude that

$$S_{\Delta y}(1) = \tilde{S}_{\Delta \bar{y}}(0),$$

so that det $S_{\Delta \bar{y}}(1)$ = det $\tilde{S}_{\Delta \bar{y}}(0) = 0$.

5/A unit root in the determinant of the moving average representation is excluded in the results of Kohn [1979]. Dunsmuir and Hannan [1976], however, do not exclude a root at unity. Pham-Dinh [1978] also consider this case, although he limits his attention to the scalar case, and assumes there is a single unit root in the moving average. Sargan and Bhargava [1983] show--using a scalar first order moving average MA(1) model--that if estimation is carried out without imposing the unit root, then the MA(1) parameter is not asymptotically normally distributed. In the estimation proposed in section 4, the unit root is imposed.

6/It should be noted that the fact that we encounter a unit root has nothing to do with the fact that we are estimating in continuous time. It would occur for exactly the same reason if a discrete time rational expectations model were being estimated instead. To see this, suppose that we posited the following time series representation for y(t):

$$\theta(L)y(t) = C(L)\varepsilon(t),$$

with θ and C being scalar matrix polynomials, respectively. Suppose the rational expectations restrictions were of the following form:

$$y_2(t) = E_t \sum_{j=0}^{\infty} \beta^j y_1(t+j)$$

or

$$y_2(t) = E_t \sum_{j=0}^{N} y_1(t+j),$$

where $y(t)^T = (y_1(t), y_2(t))$. Under these restrictions, det $C(L)$ inherits all the zeroes of $\theta(L)$. In particular, if one of the zeroes of $\theta(L)$ were unity--as would be the case if the data required first differencing to induce stationarity--then det $C(1) = 0$. A way to avoid the unit root implication in this context is to model $(y_1(t)-y_1(t-1), y_2(t)-y_1(t))$ instead of $(1-L)y(t)$. The former does not necessarily have a unit root in the determinant of its moving average.

7/ We elaborate on these points in this footnote. Hansen and Sargent [1981b]'s strategy, adapted to the continuous time setup of this paper, is to model the bivariate process $z(t)^T = (\bar{R}(t+1)-\bar{R}(t), R_n(t)-\bar{R}(t))$, instead of $\Delta\bar{y}(t)^T = (\bar{R}(t+1)-\bar{R}(t), R_n(t+1)-R_n(t))$, as we do. Here, $\bar{R}(t) = (\frac{e^D-1}{D})R(t)$. It is easy to verify that, when $\theta(D) = D\tilde{\theta}(D)$, the continuous time representation of $z(t)$ is $z(t) = \frac{H(D)}{\tilde{\theta}(D)} C(D)\varepsilon(t)$, where $\theta(D)$ and $C(D)$ are defined in (4) and (5a) and

$$H(D) = \begin{bmatrix} (\frac{e^D-1}{D})^2 & 0 \\ -\frac{e^D-1}{D^2} & \frac{1}{D} \end{bmatrix}.$$

It may be verified that $[H(s)C(s)/\tilde{\theta}(s)]$ is analytic for all s with real part nonnegative, so that $\{z(t), t\varepsilon(-\infty,+\infty)\}$ is an ordinary, covariance stationary stochastic process. Moreover, the argument in footnote 4 does not apply in the present case, so that there is no reason to believe that a unit root necessarily appears in the determinant of the moving average part of sampled z(t).

The estimation strategy described in section 4 and appendix B can be modified to permit estimating the parameters of C and $\tilde{\theta}$ given observations on $\{z(t),\ t = 0, \pm 1, \pm 2, \ldots\}$. The heart of the strategy is to get the covariances of $\{z(t)\}$ by finding the inverse Laplace transform of its spectrum, which is

$$\frac{H(s)\tilde{C}(s)\tilde{C}(-s)^T H(-s)^T}{\tilde{\theta}(s)\tilde{\theta}(-s)},$$

for $s = i\omega$, $\omega\varepsilon(-\infty,+\infty)$.

8/These calculations are apparently nontrivial. One procedure for getting confidence bounds on the B's would start by obtaining a 95 percent confidence ellipsoid for the structural parameters, ϕ', defined in (26). Every point in a suitably fine grid in this ellipsoid could then be mapped into a set of B's. Graphing each of these sets of B's in Figures 1-4 would produce a shaded area which constitutes the 95 percent confidence bounds. A similar procedure could be used to get confidence bounds for the B^u's. The results in Runkle [1985] suggest the possibility that the confidence bounds could be quite large.

9/Following is a sketch of the procedure I used to solve (50). First, I obtained the following partial fractions expansion:

$$\frac{C(s)}{\tilde{\theta}(s)} = \sum_{j=1}^{m-1} \frac{A_j}{s-\rho_j} + C_{m-1},$$

so

$$Dy(t) = C_{m-1}\varepsilon(t) + \sum_{j=1}^{m-1} A_j \int_0^{\infty} e^{\rho_j \tau} \varepsilon(t-\tau)d\tau.$$

I then integrated both sides of the above equation from t to t + 1, giving an expression for y(t+1) - y(t). The last step was to integrate the first element in the latter vector from t to t + 1.

10/For the same reasons as those given in section 6.d, Figures 5-8 would be more informative if confidence intervals were also reported. These could be computed along the lines described in footnote 8.

References

Christiano, Lawrence J., 1982, "The Term Structure of Interest Rates and the Aliasing Identification problem," unpublished manuscript.

__________, forthcoming, "A Method for Estimating the Timing Interval in a Linear Econometric Model, With an Application to Taylor's Model of Staggered Contracts," Journal of Economic Dynamics and Control.

__________, 1985, "Formulating and Estimating a Continuous Time Version of Sargent and Wallace's Model of Hyperinflation," unpublished, January.

Dunsmuir, W., and E. J. Hannan, 1976, "Vector Linear Time Series Models," Advances in Applied Probability 8, 339-364.

Fishman, G., 1968, Spectral Methods in Econometrics, Rand Corporation, Santa Monica, California.

Goldfeld and Quandt, 1972, Nonlinear Methods in Econometrics, North Holland.

Hannan, E.J., 1970, Multiple Time Series, Wiley.

Hansen, Lars Peter, and Thomas J. Sargent, 1980, "Methods for Estimating Continuous Time Rational Expectations Models from Discrete Time Data," unpublished, March.

__________, 1981a, "Formulating and Estimating Continuous Time Rational Expectations Models," Federal Reserve Bank of Minneapolis Research Department Staff Report 75, October.

__________, 1981b, "Exact Linear Rational Expectations Models: Specification and Estimation," unpublished, January.

__________, 1981c, "Identification of Continuous Time Models from Discrete Data," unpublished manuscript.

__________, 1982, "Two Difficulties in Interpreting Vector Autoregressions," Federal Reserve Bank of Minneapolis Research Department Working Paper, November.

__________, 1983, "The Dimensionality of the Aliasing Problem in Models with Rational Spectral Densities," Econometrica, January.

Kohn, Robert, 1979, "Asymptotic Estimation and Hypothesis Testing Results for Vector Linear Time Series Models," Econometrica, vol. 47, no. 4, July.

Papoulis, Athanasios, 1962, The Fourier Integral and Its Applications, McGraw-Hill.

Phillips, P.C.B., 1973, "The Problem of Identification in Finite Parameter Continuous Time Models," Journal of Econometrics, Vol. 1, pp. 351-362.

Powell, M.J.D., 1971, "Recent Advances in Unconstrained Optimization," Mathematical Programming, 1, 26-57.

Rozanov, Yu. A., 1967, Stationary Random Processes, Holden-Day.

Runkle, David Edward, 1985, "Vector Autoregression Reality," Working Paper No. 85-8, Department of Economics, Brown University, March.

Salomon Brothers, 1977, An Analytical Record of Yields and Yield Spreads, Salomon, New York.

Sargan, J.D., and Alok Bhargava, " Maximum Likelihood Estimation of Regression Models with First Order Moving Average Errors

When the Root Lies on the Unit Circle," Econometrica, Vol. 51, No. 3, May.

Sargent, Thomas J., 1979a, "A Note on Maximum Likelihood Estimation of the Rational Expectation Model of the Term Structure," Journal of Monetary Economics, vol. 5.

___________, 1979b, Macroeconomic Theory, Academic Press.

Sims, Christopher A., 1980, "Macroeconomics and Reality," Econometrica, vol. 48, no. 1, pp. 1-48.

Taylor, John B., 1980, Output and Price Stability," Journal of Economic Dynamics and Control, 2, pp. 109-132.

Table 1a

Restricted Continuous Time Model [1/]

$$\hat{\theta}(D) = \underset{(.0012)^{2/}}{.000027} + \underset{(.0035)}{.176D} + D^2$$

$$\hat{C}(D) = \begin{bmatrix} \underset{(.0004)}{.008} & 0 \\ .008 & -.000005 \end{bmatrix} + \begin{bmatrix} \underset{(.0043)}{-.000062} & \underset{(.0023)}{.053} \\ .027 & .022 \end{bmatrix} D$$

zeroes of $\hat{\theta}(D)$: -.176, -.00016 $L_r = 1618.743$[3/]

zeroes of det $\hat{C}(d)$: -.176, -.00016 5 free parameters

Table 1b

Discrete Time Model Corresponding to Model in Table 1a [1/]

$$\theta^+(L) = 1 - 1.838L + .838L^2$$

$$C^+(L) = \begin{bmatrix} .035 & 0.00 \\ .013 & -.032 \end{bmatrix} + \begin{bmatrix} -.026 & -.017 \\ -.015 & .025 \end{bmatrix} L + \begin{bmatrix} -.011 & .0097 \\ 0.0 & 0.0 \end{bmatrix} L^2$$

zeroes of $\theta^+(L)$: 1.193, 1.000

zeroes of det $C^+(L)$: -7.217, 1.193, 1.000.

Table 2

Unrestricted Discrete Time Model [1/]

$$\hat{\theta}^{+}(L) = 1 - 1.572L + .572L^{2}$$

$$\hat{C}^{+}(L) = \begin{bmatrix} .034 & 0.00 \\ .013 & -.019 \end{bmatrix} + \begin{bmatrix} -.013 & -.002 \\ .006 & .025 \end{bmatrix} L + \begin{bmatrix} .004 & -.002 \\ -.003 & -.002 \end{bmatrix} L^{2}$$

zeroes of $\hat{C}^{+}(L)$: 1.749, 1.000 $L_u = 1699.965$ [3/]

zeroes of det $\tilde{C}(L)$: 6.304, 1.173 ± 2.879i, .746 13 free parameters

Likelihood ratio statistic = 162.44 with 8 degrees of freedom.

Notes To Tables 1 and 2:

[1/]The estimated AR and MA parameters correspond to the time series model of the demeaned, detrended data. Table 1a is a representation for point-in-time data. Tables 1b and 2 provide representations for the case where the first variable (the call rate) is averaged over the sampling interval and the second is point-in-time. The numbers in Table 1b solve (29) and (32), with m = 2.

[2/]Estimates of standard errors appear in parentheses. They are calculated by taking the square root of 280 (=T) times the appropriate element of the diagonal of (15). Naturally, standard error estimates do not appear beneath parameters that are not free.

[3/]L_r and L_u denote (up to an additive constant) the restricted and unrestricted values, respectively, of the log likelihood function evaluated at the optimum.

Table 3a

Restricted Continuous Time Model [1]

$$\tilde{\theta}(D) = \underset{(.017)^{2/}}{.087} + \underset{(.075)}{.267D} + D^2$$

$$\hat{C}(D) = \begin{bmatrix} \underset{(.007)}{.035} & 0.0 \\ .035 & 0.0 \end{bmatrix} + \begin{bmatrix} \underset{(.037)}{-.025} & \underset{(.059)}{.236} \\ .107 & .019 \end{bmatrix} D + \begin{bmatrix} \underset{(.065)}{.006} & \underset{(.030)}{.534} \\ .251 & .272 \end{bmatrix} D^2$$

zeroes of $\hat{\theta}(D)$: $-.134 \pm .263i$ $L_r = 256.403$[3]

zeroes of det $\hat{\tilde{C}}(D)$: $-.662, -.134 \pm .263i, 0.$ 7 free parameters

Table 3b

Discrete Time Representation of Averages

From the Model in Table 3a [5]

$$\tilde{\theta}^+(L) = 1 - 1.690L + .766L^2$$

$$C^+(L) = \begin{bmatrix} .451 & 0 \\ .131 & -.318 \end{bmatrix} + \begin{bmatrix} -.614 & .024 \\ -.095 & .566 \end{bmatrix} L + \begin{bmatrix} .086 & -.087 \\ -.170 & -.335 \end{bmatrix} L^2$$

$$+ \begin{bmatrix} .079 & .033 \\ .137 & .056 \end{bmatrix} L^3$$

zeroes of $\tilde{\theta}^+(L)$: $1.104 \pm .297i$

zeroes of det $C^+(L)$: $1.0, 1.104 \pm .297i, 1.970, -12.869, -.2\times 10^9$

Table 4

Unrestricted Discrete Time Model [4/]

$$\hat{\tilde{\theta}}^{+}(L) = 1.0 - \underset{(.130)}{1.665}L + \underset{(.122)}{.694}L^2$$

$$\hat{C}(L) = \begin{bmatrix} \underset{(.824)}{.220} & 0.0 \\ \underset{(.069)}{.201} & \underset{(.276)}{.025} \end{bmatrix} + \begin{bmatrix} \underset{(.696)}{-3.18} & \underset{(.599)}{.311} \\ \underset{(.511)}{-.351} & \underset{(.133)}{-.097} \end{bmatrix} L + \begin{bmatrix} \underset{(.254)}{.132} & \underset{(.477)}{-.456} \\ \underset{(.254)}{.095} & \underset{(.682)}{.235} \end{bmatrix} L^2 + \begin{bmatrix} \underset{(.177)}{-.034} & \underset{(.134)}{.144} \\ \underset{(.173)}{.053} & \underset{(.504)}{-.157} \end{bmatrix} L^3$$

zeroes of $\hat{\theta}^{+}(L)$: $1.199 \pm .045i$ $L_u = 343.802$ [3/]

zeroes of det $\hat{C}(L)$: -13.454, 3.028, 1.072, 1.0, .685 17 free parameters

Likelihood ratio statistic = 174.798 with 10 degrees of freedom.

Notes to Tables 3 and 4:

[1/]The estimated AR and MA representations constitute the parameters of the continuous time representation of $(DR(t), DR_{12}(t))$.

[2/]See Footnote 2 in "Notes to Tables 1 and 2."

[3/]L_r and L_u denote the restricted and unrestricted (up to an additive constant) values, respectively, of the log likelihood function evaluated at the optimum. The raw data is composed of 280 observations. One observation was lost in the first differencing process. Another observation was dropped so that the number of observations would be an even number--a matter of convenience in frequency domain estimation.

[4/]The estimated AR and MA parameters constitute a time series model of the demeaned, first differenced data.

[5/]The numbers in this table solve (34) and (37) with m = 3.

Table 5

Computation Costs and Times

	CPU Seconds per Function Evaluations 1/	Total Number of Function Evaluations	Total Cost 2/
Table 1a 3/	.18	186	$3.35
Table 2 3/	.17	783	$13.31
Table 3 4/	.25	501	$12.53
Table 4 5/	.20	5,598	$111.96

[1]Total central processing unit (CPU) seconds divided by the total number of function evaluations required to reach an optimum. The latter quantity appears in Column 2.

[2]Total cost is the product of the numbers in the first two columns and .10. Late night changes on the University Minnesota Cyber computer are approximately 10¢ per CPU second.

[3]Davidon-Fletcher-Powell nonlinear optimization algorithm used. (See Powell [1971].)

[4]GRADX nonlinear optimization algorithm used. For a description, see Goldfeld and Quandt [1972], pp. 5-9.

[5]Powell's [1964] conjugate gradient nonlinear optimization algorithm used.

Appendix A

The Cross Equation Restrictions

Following is a demonstration that (1) and (2) imply (5). Let

$$A_k = \frac{C(\rho_k)}{\prod_{\substack{i=1\\ i\neq k}}^{m} (\rho_k-\rho_i)} = \begin{bmatrix} A_k^1 \\ A_k^2 \end{bmatrix}, \quad k = 1, \ldots, m.$$

Here, A_k^i is a 1 x 2 row vector, i = 1, 2, k = 1, ..., m. Then, (2) is written

$$\text{(A1)} \qquad y(t) = \sum_{j=1}^{m} A_j \int_0^{\infty} e^{\rho_j \tau} \varepsilon(t-\tau)d\tau.$$

The first equation of (A1) is

$$\text{(A2)} \qquad R(t) = \sum_{j=1}^{m} A_j^1 \int_0^{\infty} e^{\rho_j \tau} \varepsilon(t-\tau)d\tau.$$

Integrating both sides of (A2),

$$\text{(A3)} \quad \frac{1}{n}\int_0^n R(t+s)ds = \sum_{j=1}^{m} A_j^1 \int_0^n \int_0^{\infty} e^{\rho_j \tau} \varepsilon(t-\tau+s)d\tau ds$$

$$= \frac{1}{n} \sum_{j=1}^{m} A_j^1 \int_0^n \{e^{\rho_j(\ell+s)} \varepsilon(t-\ell)d\ell + \int_0^{\infty} e^{\rho_j(\ell+s)} \varepsilon(t-\ell)d\ell\}ds$$

Applying E_t to both sides of (A3),

$$\text{(A4)} \qquad \frac{1}{n} E_t \int_0^n R(t+s)ds = \frac{1}{n} \sum_{j=1}^{m} A_j^1 \int_0^n [\int_0^{\infty} e^{\rho_j(\ell+s)} \varepsilon(t-\ell)d\ell]ds$$

$$= \sum_{j=1}^{m} A_j^1 \alpha_j \int_0^{\infty} e^{\rho_j \ell} \varepsilon(t-\ell)d\ell$$

$$= \sum_{j=1}^{n} \frac{A_j^1 \alpha_j}{D-\rho_j} \varepsilon(t),$$

where $\alpha_j = (e^{n\rho_j}-1)/(ne_j)$, $j = 1, \ldots, m$.

Now, (1) requires that

$$\frac{1}{n} E_t \int_0^n R(t+s)ds = \sum_{j=1}^{m} \frac{A_j^1 \alpha_j}{D-\rho_j} \varepsilon(t)$$

$$= R_n(t) = \left(\frac{\delta_1(D)}{\theta(D)} \frac{\delta_2(D)}{\theta(D)}\right)\varepsilon(t),$$

or

$$\text{(A5)} \qquad \sum_{j=1}^{m} \frac{A_j^1 \alpha_j}{D-\rho_j} = \left[\frac{\delta_1(D)}{\theta(D)}, \frac{\delta_2(D)}{\theta(D)}\right]$$

Equation (5a) follows upon noting that $A_j^1 = (P_j^1, P_j^2)$, where the P's are given in (5b), and upon multiplying both sides of (A5) by $\theta(D) = \prod_{j=1}^{m} (D-\rho_j)$.

Appendix B
Obtaining $R_{\Delta\bar{y}}$

First, we get $R_{\Delta y}(\tau)$, $\tau\varepsilon(-\infty,+\infty)$. $R_{\Delta\bar{y}}$ is obtained by averaging the latter. The spectral density of $\{\Delta y(t), t \; \varepsilon \; (-\infty,+\infty)\}$, given in (18), is ($s=i\omega$, $\omega \; \varepsilon \; (-\infty,+\infty)$):

$$g_{\Delta y}(s) = \left(\frac{e^{s}-1}{s}\right)\frac{C(s)C(-s)}{\tilde{\theta}(s)\tilde{\theta}(-s)}\left(\frac{e^{-s}-1}{-s}\right)$$

$$= \left(\frac{e^{s}-1}{s}\right)\left(\frac{e^{-s}-1}{-s}\right)\left[\frac{C(s)C(-s)^{T}-C_{m-1}C_{m-1}^{T}\tilde{\theta}(s)\tilde{\theta}(-s)+C_{m-1}C_{m-1}^{T}\tilde{\theta}(s)\tilde{\theta}(-s)}{\tilde{\theta}(s)\tilde{\theta}(-s)}\right]$$

$$= \left(\frac{e^{s}-1}{s}\right)\left(\frac{e^{-s}-1}{-s}\right)\left[C_{m-1}C_{m-1}^{T}+ \frac{C(s)C(-s)^{T}-C_{m-1}C_{m-1}^{T}\tilde{\theta}(s)\tilde{\theta}(-s)}{\tilde{\theta}(s)\tilde{\theta}(-s)}\right]$$

$$= \left(\frac{e^{s}-1}{s}\right)\left(\frac{e^{-s}-1}{-s}\right)C_{m-1}C_{m-1}^{T} + \left(\frac{e^{s}-1}{s}\right)\left(\frac{e^{-s}-1}{-s}\right)\left[\sum_{j=1}^{m-1}\frac{W_j}{s-\rho_j} + \sum_{j=1}^{m-1}\frac{W_j}{s-\rho_j}\right],$$

$$= S_1(s) + S_2(s),$$

say. Here, W_j, $j = 1, \ldots, m-1$ are defined in (11c). Now, $R_{\Delta y}$ is the unique function with the property that

$$g_{\Delta y}(s) = \int_{-\infty}^{+\infty} R_{\Delta y}(\tau)e^{-s\tau}d\tau,$$

for $s = i\omega$, $\omega \; \varepsilon \; (-\infty,+\infty)$. (Uniqueness of $R_{\Delta y}$ may be established using the Residue Theorem and Jordan's lemma. See Papoulis [1962].) Define $F_1(\tau)$ and $F_2(\tau)$ to be the unique functions such that

$$S_1(i\omega) = \int_{-\infty}^{+\infty} F_1(\tau)e^{-i\tau\omega}d\tau$$

$$S_2(i\omega) = \int_{-\infty}^{+\infty} F_2(\tau)e^{-i\tau\omega}d\tau,$$

for $\omega \in (-\infty,+\infty)$. Then, $R_{\Delta y}(\tau) = F_1(\tau) + F_2(\tau)$. It may be verified that

$$F_1(\tau) = \begin{cases} C_{m-1}C_{m-1}^T(1-|\tau|) & |\tau| \le 1 \\ 0 & |\tau| > 1 \end{cases}$$

$$F_2(\tau) = \begin{cases} \sum_{j=1}^{m} \frac{(e^{-\rho_j}-1)(e^{\rho_j}-1)}{-\rho_j^2} W_j e^{\rho_j\tau} & \tau \ge 1 \\ \sum_{j=1}^{m} \frac{1}{\rho_j^2}\{\rho_j(\tau-1)[W_j+W_j^T]+[W_j-W_j^T]-[2-e^{\rho_j}]e^{\rho_j\tau}W_j+e^{(1-\tau)\rho_j}W_j^T\} & 0 \le \tau \le 1 \end{cases}$$

$$F_2(\tau) = F_2(-\tau)^T \qquad \tau \le 0.$$

The function, $R_{\Delta\bar{y}}$, is obtained by integrating $R_{\Delta y}$ in the manner described in section 4b in the text. Formulas determining the matrices $\bar{W}_j$, $\bar{\bar{W}}_j$, $\bar{\bar{\bar{W}}}_j$, $j = 1, \ldots, m - 1$ appearing in (22) are given below. First, let

$$W_j = \begin{bmatrix} W_j^{11} & W_j^{12} \\ W_j^{21} & W_j^{22} \end{bmatrix}, \quad \bar{W}_j = \begin{bmatrix} \bar{W}_j^{11} & \bar{W}_j^{12} \\ \bar{W}_j^{21} & \bar{W}_j^{22} \end{bmatrix}$$

$$\bar{W}_j = \begin{bmatrix} \bar{W}_j^{11} & \bar{W}_j^{21} \\ \bar{W}_j^{21} & \bar{W}_j^{22} \end{bmatrix} \qquad \bar{\bar{W}}_j = \begin{bmatrix} \bar{\bar{W}}_j^{11} & \bar{\bar{W}}_j^{12} \\ \bar{\bar{W}}_j^{21} & \bar{\bar{W}}_j^{22} \end{bmatrix} ,$$

for j = 1, ..., m - 1. Then,

$$\bar{W}_j^{11} = -W_j^{11}\{\frac{4}{3}\frac{1}{\rho_j} + 2(2-e^{\rho_j})\frac{1}{\rho_j^3}[\frac{1}{\rho_j}(e^{\rho_j}-1)-1]$$

$$+ \frac{2}{\rho_j^3}[\frac{1}{\rho_j}(e^{\rho_j}-1)-e^{\rho_j}]\} + \frac{2}{3(m-1)}(\tilde{C}_{m-1}\tilde{C}_{m-1}^T)^{11}$$

$$\bar{W}_j^{12} = -(\frac{1}{2\rho_j})(W_j^{12}+W_j^{21}) + \frac{1}{\rho_j^2}(W_j^{12}-W_j^{21}) - \frac{1}{\rho_j^3}(2-e^{\rho_j})(e^{\rho_j}-1)W_j^{12}$$

$$+ \frac{1}{\rho_j^3} W_j^{21}(e^{\rho_j}-1) + \frac{1}{2(m-1)}(\tilde{C}_{m-1}\tilde{C}_{m-1}^T)^{12}(W_j^{12}-W_j^{21})$$

$$\bar{W}_j^{21} = \bar{W}_j^{12}$$

$$\bar{W}_j^{22} = W_j^{22}\frac{2}{\rho_j}[\frac{1}{\rho_j}(e^{\rho_j}-1)-1] + \frac{1}{m-1}(\tilde{C}_{m-1}\tilde{C}_{m-1}^T)^{22}$$

$$\bar{\bar{W}}_j = \begin{bmatrix} W_j^{11}[(\frac{e^{-\rho_j}-1}{-\rho_j})(\frac{e^{\rho_j}-1}{\rho_j})]^2 & W_j^{12}(\frac{e^{-\rho_j}-1}{-\rho_j})(\frac{e^{\rho_j}-1}{\rho_j})^2 \\ W_j^{21}(\frac{e^{-\rho_j}-1}{-\rho_j})^2(\frac{e^{\rho_j}-1}{\rho_j}) & W_j^{22}(\frac{e^{-\rho_j}-1}{-\rho_j})(\frac{e^{\rho_j}-1}{\rho_j}) \end{bmatrix}$$

$$\bar{\bar{W}}_j^{11} = W_j^{11}\frac{1}{\rho_j^3}\{[1-(e^{-\rho_j}-1)(e^{\rho_j}-1)e^{\rho_j}][\frac{1}{\rho_j}(e^{\rho_j}-1)-1]$$

$$+ (2-e^{\rho_j})[\frac{1}{\rho_j}(e^{\rho_j}-1)-e^{\rho_j}] - \frac{1}{3}\rho_j^2\} + \frac{1}{6(m-1)}(\tilde{C}_{m-1}\tilde{C}_{m-1}^T)^{11}$$

$$\bar{W}_j^{12} = \bar{\bar{W}}_j^{12}e^{\rho_j}$$

$$\bar{\bar{W}}_j^{21} = -(\frac{1}{2\rho_j})(W_j^{21}+W_j^{12}) + \frac{1}{\rho_j^2}(W_j^{21}-W_j^{12}) - \frac{1}{\rho_j^3}(2-e^{\rho_j})W_j^{21}(e^{\rho_j}-1)$$

$$+ \frac{1}{\rho_j^3} W_j^{12}(e^{\rho_j}-1) + \frac{1}{2(m-1)}(\tilde{C}_{m-1}\tilde{C}_{m-1}^T)^{21}$$

$$\bar{\bar{W}}_j^{22} = \bar{\bar{W}}_j^{22}e^{\rho_j}$$