

A Note on Maximum Likelihood Estimation
of the Rational Expectations Model
of the Term Structure*

Thomas J. Sargent

January 1978

Staff Report #: 26

University of Minnesota
and
Federal Reserve Bank
of Minneapolis

The views expressed herein are solely those of the author and do not necessarily represent the views of the Federal Reserve Bank of Minneapolis or the Federal Reserve System

*Research for this paper was supported by the Federal Reserve Bank of Minneapolis, which is not responsible for the conclusions. Robert Litterman ably performed the calculations. Comments from a referee are gratefully acknowledged.

The key implications of the rational expectations theory of the term structure of interest rates are that certain sequences of forward interest rates can be described as martingales. These implications are ones for which the most convenient and powerful tests of the theory can be made.¹ However, as Modigliani, Sutch and Shiller have emphasized, from the point of view of implementing the theory in the context of a macroeconomic model, it is not sufficient to represent the theory simply by its implications that those sequences of forward yields are martingales. To get the theory in a form that can be used in a macroeconomic model, Modigliani, Shiller, and Sutch in effect characterized the theory by its implications for the regression of long rates on current and past short term rates. In addition to delivering something that might be used in a macroeconomic model, this approach can also be justified purely on the grounds that it provides a way of testing the theory on the basis of a much sparser data set than is required in order to test that the appropriate sequences of forward rates are martingales. That is, to test the model using the procedures to be discussed below, all that are required are suitable time series on a single long-term rate and a single short term rate²; but to test some of the martingale implications directly requires time series over the entire term structure of rates.

This note is written by way of pursuing the general Modigliani, Sutch and Shiller approach. However, rather than follow Modigliani, Sutch, and Shiller in focusing on the projection of long rates on current and past short rates, I will proceed by estimating the vector autoregression of long and short rates. This is a convenient representation for extracting predictions from the model, and also conserves

all of the information required to compute the projection of one interest rate series on current and past (and maybe future) values of the other series. In particular, a compact formula is given for the restriction on the bivariate vector autoregression of the long term rate and the short term rate that is implied by the rational expectations theory. Then two procedures are given for estimating the vector autoregression under that restriction: the first being a two-step procedure that gives consistent but not fully efficient estimates under the restriction; the second being a maximum likelihood estimator. Some sample calculations are carried out and the pertinent likelihood ratio statistic is reported. The maximum likelihood algorithm used here would be a convenient one to use for estimating and testing rational expectations models of other relationships. The main purpose of this paper is to illustrate the feasibility of maximum likelihood estimation in the face of the complicated nonlinear restrictions implied by rational expectations in multi-period horizon models. To my knowledge, applications of this approach are not available in the literature.

Let R_{1t} be the one period rate and R_{nt} be the n-period rate. I assume that the process of first differences $(\Delta R_{1t}, \Delta R_{nt})$ is a second-order jointly stationary, indeterministic stochastic process. Among other things, this means that the covariances between ΔR_{1t} and ΔR_{nt-s} exist and are independent of time t ; it also means that the variances of ΔR_{1t} and ΔR_{nt} exist and don't depend on t . I will work with the m^{th} -order bivariate autoregressive representation for the $(\Delta R_{1t}, \Delta R_{nt})$ process, the existence of which is implied by the preceding assumptions.^{3/}

The theory imposes restrictions on this vector autoregression so long as agents have at least as much information as is contained in m lagged ΔR_{1t} 's and ΔR_{nt} 's, as will be proved by applying a variant of an argument of Shiller [1972].

I will represent the rational expectations theory of the term structure in the form^{4/}

$$(1) \quad R_{nt} = \frac{1}{n}(R_{1t} + E_t R_{1t+1} + \dots + E_t R_{1t+n-1})$$

where I will interpret $E_t x$ to mean the linear least squares forecast of x based on information available at time t . I will assume that this information set includes at least (but possibly more than) current and all lagged values of both R_{1t} and R_{nt} . Let Ω_t be the

information set that agents use at time t , so that $E_t x \equiv Ex | \Omega_t$. I assume that $\Omega_t \supset \Omega_{t-1} \supset \Omega_{t-2} \dots$. First, differencing (1) gives

$$(R_{nt} - R_{nt-1}) = \frac{1}{n}[(R_{1t} - R_{1t-1}) + (ER_{1t+1} | \Omega_t - ER_{1t} | \Omega_{t-1}) + \dots + (ER_{1t+n-1} | \Omega_t - ER_{1t+n-2} | \Omega_{t-1})].$$

Let θ_{t-1} be any subset of Ω_{t-1} . Then use the law of iterated projections^{5/} to project both sides of the above equation on θ_{t-1} to get

$$(2) \quad E \Delta R_{nt} | \theta_{t-1} = \frac{1}{n}[E \Delta R_{1t} | \theta_{t-1} + E \Delta R_{1t+1} | \theta_{t-1} + \dots + E \Delta R_{1t+n-1} | \theta_{t-1}].$$

If we let $\theta_{t-1} = \{\Delta R_{1t-1}, \Delta R_{1t-2}, \dots, \Delta R_{1t-m}, \Delta R_{nt-1}, \dots, \Delta R_{nt-m}\}$, equation (2) implies a restriction across the systematic parts of the m^{th} -order vector autoregression for $(\Delta R_{1t}, \Delta R_{nt})$. Let the m^{th} -order vector autoregression for $\Delta R_{1t}, \Delta R_{nt}$ be

$$(3) \quad \begin{aligned} \Delta R_{1t} &= \sum_{i=1}^m \alpha_i \Delta R_{1t-i} + \sum_{i=1}^m \beta_i \Delta R_{nt-i} + a_{1t} \\ \Delta R_{nt} &= \sum_{i=1}^m \gamma_i \Delta R_{1t-i} + \sum_{i=1}^m \delta_i \Delta R_{nt-i} + a_{nt} \end{aligned}$$

where $\mathcal{E} a_{jt} \cdot \Delta R_{lt-i} = \mathcal{E} a_{jt} \Delta R_{nt-i} = 0$ for $j=1, n$ and $i=1, \dots, m$, where \mathcal{E} is the mathematical expectation operator.

The random variables a_{lt}, a_{nt} are the innovations in the $\Delta R_{lt}, \Delta R_{nt}$ processes, and are the one step-ahead prediction errors in ^{linearly}predicting ΔR_{lt} and ΔR_{nt} , respectively, on the basis of m observations of past ΔR_{lt} 's and ΔR_{nt} 's. Equation (3) can be written compactly as

$$(4) \quad x_t = Ax_{t-1} + a_t$$

where

$$x_t = \begin{bmatrix} \Delta R_{lt} \\ \Delta R_{lt-1} \\ \vdots \\ \Delta R_{lt-m+1} \\ \Delta R_{nt} \\ \Delta R_{nt-1} \\ \vdots \\ \Delta R_{nt-m+1} \end{bmatrix} \quad a_t = \begin{bmatrix} a_{lt} \\ 0 \\ 0 \\ \vdots \\ 0 \\ a_{nt} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$A = \begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_m & \beta_1 & \beta_2 & \dots & \beta_m \\ 1 & 0 & & 0 & 0 & 0 & & 0 \\ 0 & 1 & & 0 & 0 & 0 & & 0 \\ \vdots & \vdots \\ 0 & 0 & & 1 & 0 & 0 & & 0 \\ \gamma_1 & \gamma_2 & & \gamma_m & \delta_1 & \delta_2 & \dots & \delta_m \\ 0 & 0 & & 0 & 1 & 0 & & 0 \\ \vdots & \vdots \\ 0 & 0 & & 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix} .$$

Letting c be the $(1 \times 2m)$ row vector with one in the first column, zeroes elsewhere and letting d be the $(1 \times 2m)$ row vector with one in the $(m + 1)$ st column, zeroes elsewhere, we have

$$\Delta R_{1t} = cx_t$$

$$\Delta R_{nt} = dx_t \quad .$$

Using (4), we can write

$$x_{t+1} = A^2 x_{t-1} + Aa_t + a_{t+1}$$

$$x_{t+2} = A^3 x_{t-1} + A^2 a_t + Aa_{t+1} + a_{t+2}$$

⋮
⋮
⋮

$$(5) \quad x_{t+j} = A^{j+1} x_{t-1} + A^j a_t + A^{j-1} a_{t+1} + \dots + a_{t+j} \quad .$$

Since a_{t+j} satisfies ^{6/} $Ea_{t+k} | \theta_{t-1} = 0$ for $k=0, 1, 2, \dots$, we have from (5) that

$$(6) \quad Ex_{t+j} | \theta_{t-1} = A^{j+1} x_{t-1} \quad .$$

Now (4) implies that

$$\Delta R_{1t} = cAx_{t-1} + a_{1t}$$

(7)

$$\Delta R_{nt} = dAx_{t-1} + a_{nt} \quad .$$

But restriction (2) on the systematic part of the vector autoregression, together with (6), implies

$$(8) \quad \Delta R_{nt} = \frac{1}{n} c[A + A^2 + \dots + A^n]x_{t-1} + a_{n,t} \quad .$$

Comparing (7) with (8), we see that the rational expectations theory imposes the following restriction across the nontrivial rows of A :

$$(9) \quad d A = \frac{1}{n} c[A + A^2 + \dots + A^n].$$

Equation (9) is a compact representation of the restrictions that the rational expectations theory of the term structure imposes on the m^{th} -order bivariate vector autoregression of the $(\Delta R_{1t}, \Delta R_{nt})$ process.^{7/}

I propose the following methods for estimating the vector autoregression for $\Delta R_{1t}, \Delta R_{nt}$ under restriction (9). It is instructive first to consider a two-step procedure which potentially yields consistent, though not fully efficient estimates under the restriction. First, estimate by least squares the first row of A, i.e., estimate the first of equations (7). Then pursue the following iterative scheme for calculating the $(m+1)$ st row of A implied by this choice of the first row. First set the $(m+1)$ st row of A (i.e., the one corresponding to the autoregression for ΔR_{nt}) to a row of zeroes. Set the other rows of A at their known values. Call this preliminary estimate A_0 . Then form a revised estimate of the $(m+1)$ st row of A according to

$$d A_1 = \frac{1}{n} c[A_0 + A_0^2 + \dots + A_0^n].$$

In forming the other rows of A_1 leave the other rows of A at their initial values. Then recalculate A again, iterating on

$$(10) \quad d A_{i+1} = \frac{1}{n} c[A_i + A_i^2 + \dots + A_i^n]$$

where A_i is the estimate of A on the i th iteration. At each step in forming A_{i+1} , all rows of A except the $(m+1)$ st are kept equal to the corresponding rows of A_0 . If it converges, this algorithm will find an A that satisfies restriction (9). Experience indicates that this scheme often converges, especially where the eigenvalues of A are well below unity. The elements of the first row of A are consistently estimated by least squares. The

preceding algorithm, since it calculates the $(m+1)^{st}$ row of A as a function of the first row of A, will produce consistent estimates of that $(m+1)^{st}$ row under the usual regularity conditions.

The preceding algorithm in effect computes the γ 's and δ 's of (3) that satisfy (9) as functions of the α 's and β 's. Let us denote the solution to the iteration on (10) as the (set) function

$$(11) \quad (\gamma, \delta) = \phi(\alpha, \beta) ;$$

ϕ maps the α 's and β 's into a set of γ 's and δ 's that satisfy restriction (9). Our first estimator of the γ 's and δ 's is then simply $\phi(\alpha, \beta)$ evaluated with α and β being set at their least squares estimates. Call this the "two-step estimator."

Under the hypothesis that (a_{1t}, a_{nt}) is bivariate normal, the likelihood function of a sample of (a_{1t}, a_{nt}) for $t=1, \dots, T$ is

$$(12) \quad L(\alpha, \beta, \gamma, \delta, V | \{\Delta R_{1t}\}, \{\Delta R_{nt}\}) = (2\pi)^{-T} |V|^{-T/2} \exp\left(-\frac{1}{2} \sum_{t=1}^T e_t' V^{-1} e_t\right)$$

where

$$e_t = \begin{pmatrix} a_{1t} \\ a_{nt} \end{pmatrix}, \quad V = E e_t e_t'$$

Maximizing (12) subject to (3) without any restrictions on the coefficients, i.e., taking the m α_i , β_i , γ_i , and δ_i 's all as free parameters, is equivalent with estimating each equation of (3) by least squares.

Under the restriction (9), or equivalently (11), the likelihood function (12) becomes a function only of the α 's and β 's. As Wilson [1973] has noted, maximum likelihood estimates with an unknown V are obtained by

minimizing with respect to the α 's and β 's the criterion

$$(13) \quad |\hat{V}| = \left| \frac{1}{T} \sum_{t=1}^T e_t(\alpha, \beta) e_t(\alpha, \beta)' \right|$$

where the e_t 's (i.e. the a_{1t} and a_{nt} 's) are functions of the α 's and β 's (as well as the ΔR_{1t} 's and ΔR_{nt} 's) by virtue of their being calculated from (3) with (11) being imposed. A standard derivative free nonlinear minimization routine is capable of minimizing (13) numerically. The least squares estimates of α and β would seem to be good starting values from which to pursue the nonlinear minimization. The maximum likelihood estimator of V turns out to be

$$\hat{V} = \frac{1}{T} \sum_{t=1}^T \hat{e}_t \hat{e}_t' \quad ,$$

where the \hat{e}_t 's are the estimated vectors of residuals.

Let L_u be the value of (12) at its unrestricted maximum while L_r is the value of (12) under the restriction (9). Then under the null hypothesis that the rational expectations model is correct,

$$(14) \quad -2 \log_e \left(\frac{L_r}{L_u} \right)$$

is asymptotically distributed as $\chi^2(q)$ where q is the number of restrictions imposed. In our case $q = 2m$, where m is the number of lags in the autoregression (3). High values of the likelihood ratio (14) lead to rejection of the restrictions (11) that are implied by the rational expectations theory of the term structure.^{8/}

Table 1 reports three sets of estimates of equation 3 for $m=4$ where R_{nt} is taken as the rate on five-year government bonds while R_{1t} is taken as the three-month Treasury bill rate. The data are quarterly and point-in-time, first of month data for the first month of each quarter.^{9/} The data on the left-hand side variables of (3) span the period 1953II-1971IV. There are thus 71 observations on the disturbances, so that $T=71$. The table reports estimates of (3) unconstrained (i.e., least squares estimates of each equation of (3)), the two-step estimates which impose the rational expectations restrictions (11), and the maximum likelihood estimates that impose (11).

The likelihood ratio statistic pertinent for testing the null hypothesis that the rational expectations restrictions are correct is 8.58. Since the likelihood ratio statistic is distributed as chi-square with eight degrees of freedom, the marginal significance level is .3788. The likelihood ratio test thus does not provide any strong evidence for rejecting the rational expectations restrictions.

As indicated by the $|V|$ statistic, it is interesting that the two-step estimates provide a considerably poorer fit than do the maximum likelihood estimates.

Notice that the γ 's and δ 's estimated under the restriction (9) by both the two-step estimator and the maximum likelihood estimator are close to zero, so that with respect to the information in four lagged R_{1t} 's and R_{nt} 's, the long rate seems approximately described by a "weak" martingale.^{10/} That the restrictions given by the rational expectations theory of the term structure imply such an approximation for long rates under suitable regularity conditions was exploited earlier by Sargent [1976].

Modigliani and Sutch [1965] worked with a version of the theory in which only lagged short interest rates were included in the information set carried along in the model. As the argument leading to equation (2) shows, the rational expectations restrictions (2) are predicted to hold with θ_{t-1} being any subset of Ω_{t-1} , and in particular with θ_{t-1} being chosen in the fashion of Modigliani and Sutch, namely, $\theta_{t-1} = \{\Delta R_{1t-1}, \Delta R_{1t-2}, \dots, \Delta R_{1t-m}\}$. This specification of θ_{t-1} leads to the restriction on (3) that $\beta_i = \delta_i = 0$, $i=1, \dots, m$, where now the least squares orthogonality conditions become $\sum_{j=1}^n \Delta R_{1t-i} = 0$ for $j=1, n$, and $i=1, \dots, m$. With this restriction on the β 's and δ 's, (9) continues to represent the rational expectations restrictions across the α 's and γ 's. In fact, with the β_i 's being zero, iterations on (10) are guaranteed to converge in one step. All of the estimation theory goes through as before.

Table 2 reports three sets of estimates with $\beta_i = \delta_i = 0$, $i=1, \dots, m$, with θ_{t-1} specified as $\{\Delta R_{1t-1}, \dots, \Delta R_{1t-4}\}$. The likelihood ratio statistic pertinent for testing the null hypothesis that the rational expectations restrictions (9) are correct is only 3.0788. Since this statistic is distributed as chi-square with four degrees of freedom under the null hypothesis, the marginal significance level is .5447, which once again provides no strong evidence for rejecting the rational expectations restrictions.

It is interesting to test whether lagged ΔR_{nt} 's are usefully included in the information set θ_{t-1} . Comparing the unrestricted estimates in Tables 1 and 2, i.e., the first sets of estimates, we note that the Table 2 estimates are computed under a restriction on the Table 1 specification. A likelihood ratio statistic for testing the null hypothesis that $\beta=\delta=0$ can be computed as $T\{\log|V_r| - \log|V_u|\}$ where

$|V_r|$ is the determinant of V estimated in Table 2, while $|V_u|$ is the determinant in Table 1. This statistic is distributed as chi-square with eight degrees of freedom. The value of the test statistic turns out to be 12.94, which has a marginal significance level of .114. Computing the analogous test on the maximum likelihood restricted estimates (the third sets of estimates in Tables 1 and 2) gives a likelihood ratio statistic of 7.438, which is distributed as chi-square with four degrees of freedom and so has a marginal significance level of .114. I would interpret these significance levels as being mildly though not spectacularly supportive of Modigliani and Sutch's choice of θ_{t-1} .

It should be emphasized that the theory predicts that none of the representations estimated in this paper will be invariant with respect to an intervention that alters the stochastic processes facing agents and thereby alters the second-order characteristics of the distributions of yields. For example, despite the moderate success of results that choose θ_{t-1} to be $\{\Delta R_{1t-1}, \Delta R_{1t-2}, \dots\}$, it would not be appropriate to impose arbitrary alternative stochastic processes for the short rate (arguing that it is the monetary authority's instrument) and expect such term-structure relations to remain invariant. 11/

Footnotes

¹ Some of these implications were spelled out by Roll [1970] and tested against data by Roll [1970] and Sargent [1972].

² Modigliani and Shiller [1973] made this point but did not formulate a formal econometric test.

³ A nonrigorous discussion of vector autoregressions, vector stochastic processes, and some of their applications in macroeconomics is contained in Sargent [1978].

⁴ Equation (1) is only an approximation to the correct formula linking long with expected short rates. Shiller and Modigliani [1973] Shiller [1972] recommend the alternative approximation

$$R_{nt} = (1 - \gamma) \sum_{j=0}^{\infty} \gamma^j E_t R_{1t+j}$$

where $\gamma = 1/(1 + r_0)$, r_0 being a "representative short term rate", which Modigliani and Shiller recommend taking as the mean long term rate. Modigliani and Shiller recommend that this approximation be used for very long term rates.

⁵ Use of the law of iterated projections in this way is the argument of Shiller [1972] referred to earlier. The law of iterated projections states that $E(y|z) = E(E(y|x, z)|z)$, where x, y, z are random variables and E is the linear least squares projection operator. The law is easily proved as a consequence of the fact that least squares residuals are orthogonal to least squares predictions.

⁶Technically, this holds only if $E a_{jt} \Delta R_{1t-i} = E a_{jt} \Delta R_{nt-i} = 0$ for $j=1, n$, and $i=1, 2, \dots$. This amounts to the condition that the m^{th} -order vector autoregression equals the infinite-order vector autoregression, so that coefficients on ΔR_1 and ΔR_n lagged more than m periods would be zero if they had been included in the population representation (3). Practically, the requirement amounts to choosing m large enough to account for the serial correlation and cross-serial correlation in the $(\Delta R_1, \Delta R_n)$ process.

⁷If we had used Modigliani and Shiller's formula (see footnote (4)) restriction (9) would become

$$dA = (1 - \gamma)c A \sum_{j=0}^{\infty} \gamma^j A^j .$$

Assume that the eigenvalues of A are distinct, so that A can be written $A = P \Lambda P^{-1}$ where the columns of P are the eigenvectors of A while Λ is a diagonal matrix of eigenvalues of A . Then the above restriction can be written in the compact form

$$dA = (1 - \gamma)c A P \left\{ \frac{1}{1 - \gamma \lambda} \right\} P^{-1}$$

assuming $\max |\gamma \lambda_i| < 1$, $|\gamma \lambda_i| < 1$

where $\left\{ \frac{1}{1 - \gamma \lambda} \right\}$ is the diagonal matrix with $\frac{1}{1 - \gamma \lambda_i}$ in the (i, i) position. By making use of this formula, the algorithm proposed in the text can easily be modified for Modigliani and Shiller's formula.

⁸Using the calculations of Wilson [1973, p. 80], it is possible to show that the likelihood ratio (14) could be calculated from

$$T\{\log_e |V_r| - \log_e |V_u|\}$$

where V_r and V_u are the restricted and unrestricted estimates of V , respectively. In our case $T = 71$.

⁹The data were obtained from the Salomon Brothers publication An Analytical Record. Those data are monthly but are mid month until 1959, at which time they are first of month. I linearly interpolated the earlier mid month data in order to obtain approximate first of month series for the years 1953-1958.

¹⁰The martingale property is a characteristic of conditional mathematical expectations. By a "weak" martingale I mean to denote a condition analogous to the martingale property ($\mathcal{E}_t x_{t+1} = x_t$ where \mathcal{E}_t is mathematical expectation conditioned on some information set including at least x_t) holding for linear least squares projections (i.e. the condition $E_t x_{t+1} = x_t$ where $E_t x_{t+1}$ is the linear least squares projection of x_{t+1} based on information available at time t).

¹¹This term structure example can thus be added to the consumption, investment, and labor supply examples provided by Lucas [1976].

References

- [1] An Analytical Record of Yields and Yield Spreads, Salomon Brothers, New York.
- [2] Granger, C. W. J., "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," Econometrica, (July 1969), pp. 424-438.
-
- [3] Lucas, Robert E., "Econometric Policy Evaluation: A Critique," in The Phillips Curve and the Labor Market edited by Karl Brunner and Allan Meltzer. Vol. I of the Carnegie-Rochester Conferences on Public Policy, a supplementary series to the Journal of Monetary Economics Amsterdam: North Holland (1976).
-
- [4] Modigliani, Franco and Robert J. Shiller, "Inflation, Rational Expectations, and the Term Structure of Interest Rates," Economica, N.S., Vol. 40 (February 1973), pp. 12-43.
- [5] Modigliani, Franco and Richard Sutch, "Innovations in Interest Rate Policy," American Economic Review, Papers and Proceedings (May 1966).
- [6] Roll, Richard, The Behavior of Interest Rates. Basic Books, 1970.
- [7] Sargent, Thomas J., "A Classical Macroeconometric Model for the United States," Journal of Political Economy, April 1976.
- [8] Sargent, Thomas, "Rational Expectations and the Term Structure of Interest Rates," Journal of Money, Credit and Banking, February 1972.

[9] Sargent, Thomas J., "Macroeconomic Theory," manuscript (1977).
To be published by Academic Press, 1978.

[10] Shiller, Robert J., "Rational Expectations and the Structure of
Interest Rates," (PhD thesis, M.I.T., 1972).

[11] Sims, Christopher A., "Money, Income, and Causality," American
Economic Review, (September 1972).

[12] Wilson, G. T., "The Estimation of Parameters in Multivariate
Time Series Models," Journal of the Royal Statistical Societies,
B, 1973(1), pp. 76-85.

Table 1: Estimates for a Five-Year Government
Bond Rate and 91-day Treasury Bill Rate,
1953II-1971IV

Unrestricted Estimates

j	1	2	3	4
α_j	-.3663	-.3235	.1234	-.0694
β_j	.6373	.4322	-.3286	.1703
γ_j	-.2962	.0203	.2480	-.1047
δ_j	.2812	.1200	-.3934	.0765

$$V = \begin{pmatrix} .3080 & .2072 \\ .2072 & .1924 \end{pmatrix}, |V| = .0163082085$$

Two-Step Estimates*

j	1	2	3	4
γ_j	-.0199	-.0083	.0016	-.0021
δ_j	.0285	.0085	-.0047	.0053

$$V = \begin{pmatrix} .3080 & .2072 \\ .2072 & .2162 \end{pmatrix}, |V| = .0236367$$

* α 's and β 's are the same as unrestricted estimates.

Maximum Likelihood Estimates

j	1	2	3	4
α_j	-.0717	-.3660	-.1465	.0433
β_j	.3700	.3270	.0995	.0900
γ_j	-.0183	-.0154	-.0033	.0014
δ_j	.0298	.0172	.0063	.0029

$$V = \begin{pmatrix} .3362 & .2336 \\ .2336 & .2171 \end{pmatrix}, |V| = .0184034$$

Likelihood ratio statistic = 8.5816

Marginal significance level = .3788

Let X be a chi-square distributed random variable and let x be the test statistic. Then the marginal significance level is defined as $\text{Prob}\{X > x\}$ under the null hypothesis.

Table 2: Estimates for a Five-Year Government Bond
and 91-day Treasury Bill Rate, 1953II-1971IV,
($\beta=\delta=0$)

Unrestricted Estimates

j	1	2	3	4
α_j	.0847	-.2229	.0267	.1492
γ_j	-.1173	-.0207	.0808	.0011

$$V = \begin{pmatrix} .3494 & .2311 \\ .2311 & .2089 \end{pmatrix}, |V| = .0195689615$$

Two-Step Estimates*

j	1	2	3	4
γ_j	.0020	-.0024	.0091	.0077

$$V = \begin{pmatrix} .3494 & .2311 \\ .2311 & .2176 \end{pmatrix}, |V| = .02258$$

* α 's are the same as the unrestricted estimates.

Maximum Likelihood Estimates

j	1	2	3	4
α_j	.2165	-.2117	-.0592	.1540
γ_j	.0055	-.0065	.0053	.0086

$$V = \begin{pmatrix} .3603 & .2411 \\ .2412 & .2182 \end{pmatrix}, |V| = .020436$$

Likelihood ratio test statistic = 3.0788
Marginal significance level = .5447

Let X be a chi-square distributed random variable and let x be the test statistic. Then the marginal significance level is defined as $\text{Prob}\{X>x\}$ under the null hypothesis.