Econometrica, Vol. 44, No. 6 (November, 1976)

SOME EXPERIMENTAL RESULTS ON THE STATISTICAL PROPERTIES OF LEAST SQUARES ESTIMATES IN CONTROL PROBLEMS[1]

BY T. W. ANDERSON AND JOHN B. TAYLOR

The statistical properties of the certainty equivalence control rule and of the least squares estimates generated by this rule are examined experimentally in a linear model with two unknown parameters. It is found that the least squares certainty equivalence rule converges to its true value with probability one and is asymptotically efficient, having an asymptotic distribution with a variance as small as any other strongly consistent rule. However, while a linear combination of the parameter estimates is consistent, the evidence does not confirm that the individual estimates themselves are consistent. If these converge to their true values at all, they do so very slowly (on the order of $(\log t)^{-1}$).

1. INTRODUCTION

LET $y_1, y_2, \ldots$ be a sequence of random variables determined by the linear relation

$$(1.1) \qquad y_t = \alpha + \beta x_t + u_t \qquad (t = 1, 2, \ldots),$$

where $u_1, u_2, \ldots$ is a sequence of independent and identically distributed random variables with zero mean and finite variance σ^2, where $x_1, x_2, \ldots$ is a sequence of control variables, and where α and β are unknown parameters. A control problem within the context of (1.1) having applications in the fields of stabilization and regulation is to choose a control sequence (or rule) $x_1, x_2, \ldots$ so as to bring each element in the sequence $y_1, y_2, \ldots$ as close as possible to some target value y^*. Particular choices for a control sequence will differ both in the implied behavior of the dependent variable y_t and in the quality of the parameter estimates of α and β.

The purpose of this paper is to examine by Monte Carlo experimentation the properties of a control sequence in which x_1 and x_2 are distinct but otherwise arbitrary and

$$(1.2) \qquad x_t = \begin{cases} k, & \text{if} \quad r_t \geqslant k, \\ r_t, & \text{if} \quad -k \leqslant r_t \leqslant k, \\ -k, & \text{if} \quad r_t \leqslant -k, \end{cases} \qquad (t = 3, 4, \ldots),$$

where

$$(1.3) \qquad r_t = \frac{y^* - \hat{\alpha}_{t-1}}{\hat{\beta}_{t-1}} \qquad (t = 3, 4, \ldots),$$

[1] This research has been supported by National Science Foundation Grants SOC73-05453 at the Institute for Mathematical Studies in the Social Sciences, Stanford University and SOC74-1336 at Columbia University.

where $\hat{\alpha}_{t-1}$ and $\hat{\beta}_{t-1}$ are the least squares estimates of α and β based on $y_1, \ldots, y_{t-1}$ and $x_1, \ldots, x_{t-1}$, and where k is a pre-assigned positive number representing a bound on the control rule. (A bounded rule is considered because the moments of r_t for $t = 3$ do not exist when u_t is normally distributed.) This rule may be called the *least squares certainty equivalence* (LSCE) control rule, because it is obtained by treating the parameters as known with certainty and equal to their least squares estimates. The experiments reported here concern not only the performance of x_t, but also the joint distribution of $\hat{\alpha}_t$ and $\hat{\beta}_t$. The properties of these estimates have previously been given little treatment in the control literature despite their importance in practical problems.

Previous studies which have been concerned with the sampling properties of control rules and parameter estimates in this type of model are Taylor [**6**] and Aoki [**2**]. The former paper deals with the case where α is known and shows that the least squares certainty equivalence control rule is strongly consistent and has an asymptotic distribution with a variance as small as any other control rule in the class of strongly consistent rules. Aoki [**2**] considers the more general case where α and β are unknown but is only able to show that an approximation to (1.2) is consistent with no discussion of the asymptotic distribution.

There have also been numerous papers which have approached this problem from a Bayesian point of view. (See Aoki [**1**], Zellner [**8**], and Prescott [**5**].) However, because optimal Bayes control rules have been analytically as well as computationally difficult to find, most recent research has been devoted to finding approximations to these Bayes rules. (See Tse [**7**] and Chow [**3**], for example.) A distinctive feature of most of these studies is finding rules which give better control performance than certainty equivalence rules without regard to the behavior of the parameter estimates (although it is sometimes conjectured that better parameter estimates lead to better control performance).

The experimental findings of this paper suggest that the least squares certainty equivalence rule for (1.1) performs quite well. The sequence of values of the control variable is strongly consistent and has the same asymptotic normal distribution derived by Taylor [**6**] where α is known. This implies that the LSCE control rule cannot be improved on asymptotically by experimenting to obtain better parameter estimates. The results also suggest that the distribution of x_t converges to the normal surprisingly quickly. The asymptotic approximation is close even for sample sizes as small as $t = 10$. However, the distributions of the least squares parameter estimates $\hat{\alpha}_t$ and $\hat{\beta}_t$ indicate that these will usually be poor estimates of α and β. While plots of the joint distribution of $\hat{\alpha}_t$ and $\hat{\beta}_t$ indicate that a certain linear combination of the estimates is consistently estimated (with the usual rate of convergence), the parameters themselves converge to their true values so slowly that they are of questionable value. In addition, the usual estimates of the variances of $\hat{\alpha}_t$ and $\hat{\beta}_t$ are inaccurate. These results therefore indicate that control rules which are designed for experimentation are useful not so much in improving control performance, but in improving the parameter estimates themselves. However, improving the parameter estimates can be accomplished only by sacrificing control performance.

2. TRANSFORMATION TO CANONICAL FORM

The control rule and coefficient estimates were investigated for five different combinations of parameter values. In order to focus on parameter combinations which are really different in their effect on these estimates, the model was first transformed to a canonical form. This canonical form involves three free parameters and allows for easier interpretation of the Monte Carlo results.

This transformation is defined by

$$(2.1)\qquad y_t^0 = \frac{(y_t - y^*)}{\sigma},$$

$$(2.2)\qquad x_t^0 = \frac{(\alpha + \beta x_t - y^*)}{\sigma}, \quad \text{and}$$

$$(2.3)\qquad u_t^0 = \frac{u_t}{\sigma},$$

so that the process in equation (1.1) becomes

$$(2.4)\qquad y_t^0 = x_t^0 + u_t^0 \qquad (t = 1, 2, \ldots).$$

Thus, the target value for y_t^0 is 0, the true value of the intercept coefficient α^0 is 0, and the true value of the slope coefficient β^0 is 1. The least squares estimates of these coefficients can be written in terms of the least squares estimates of the coefficients of (1.1) as

$$(2.5)\qquad \hat{\alpha}_t^0 = \frac{1}{\sigma}\left[\hat{\alpha}_t + \hat{\beta}_t \frac{y^* - \alpha}{\beta} - y^*\right] \quad \text{and}$$

$$(2.6)\qquad \hat{\beta}_t^0 = \frac{\hat{\beta}_t}{\beta}.$$

From these a control sequence in canonical form is defined by x_1^0 and x_2^0 distinct and

$$(2.7)\qquad x_t^0 = \begin{cases} k^0, & \text{if} \quad r_t^0 \geqslant k^0, \\ r_t^0, & \text{if} \quad -k^0 \leqslant r_t^0 \leqslant k^0, \\ -k^0, & \text{if} \quad r_t^0 \leqslant -k^0, \end{cases} \qquad (t = 3, 4, \ldots),$$

where $r_t^0 = -\hat{\alpha}_{t-1}^0/\hat{\beta}_{t-1}^0$, $t = 3, 4, \ldots$, and $k^0 = (\alpha + \beta k - y^*)/\sigma$. All models of the form (2.4) with control rule (2.7) are therefore completely characterized by the initial values x_1^0 and x_2^0 and the bound k^0. Further, any properties of $\hat{\alpha}_t^0$, $\hat{\beta}_t^0$, and x_t^0 can be transformed into properties of $\hat{\alpha}_t$, $\hat{\beta}_t$, and x_t through equations (2.5), (2.6), and (2.7).

It is clear that equation (1.1) is in canonical form when $\alpha = 0$, $\beta = 1$, $\sigma^2 = 1$, and $y^* = 0$. Throughout the remainder of this paper these parameters are set at these values so that $\hat{\alpha}_t$, $\hat{\beta}_t$, and x_t are in canonical form. We can therefore dispense with superscript notation. All the following experiments were run in canonical form.

3. DESCRIPTION OF THE EXPERIMENTS

All results reported in this paper are based on 100 replications. The five different combinations of parameter values are listed in Table I. These represent a fairly wide variety of initial conditions. Since the primary focus of the investigation is on the asymptotic properties of x_t, $\hat{\alpha}_t$, and $\hat{\beta}_t$, the value of the bound k is not crucial

TABLE I

PARAMETER VALUES INVESTIGATED

Model	x_1	x_2	$x_1 + x_2$	$x_1 - x_2$	Var $\hat{\alpha}_2$	Var $\hat{\beta}_2$	Cov $(\hat{\alpha}_2, \hat{\beta}_2)$
I	0.5	−0.5	0	1	0.5	2	0
II	5	−5	0	10	0.5	0.02	0
III	1	0	1	1	1	2	−1
IV	10	0	10	10	1	0.02	−0.1
V	6	4	10	2	1.3	0.5	−0.4

TABLE II

ESTIMATED MOMENTS OF THE CONTROL RULE

Model	t	$E(x_{t+1})$	MSE (x_{t+1})	Var (x_{t+1})
I	10	5.92(−2)	1.18(−1)	1.15(−1)
	50	1.79(−2)	2.21(−2)	2.17(−2)
	100	3.95(−3)	1.08(−2)	1.08(−2)
	500	1.71(−3)	2.27(−3)	2.27(−3)
	1,000	2.03(−3)	1.23(−3)	1.23(−3)
	3,000	2.30(−3)	4.32(−4)	4.26(−4)
II	10	4.92(−2)	8.28(−2)	8.04(−2)
	50	9.42(−3)	2.22(−2)	2.21(−2)
	100	1.09(−3)	9.38(−3)	9.38(−3)
	500	3.02(−3)	1.76(−3)	1.75(−3)
	1,000	2.66(−3)	9.59(−4)	9.52(−4)
	3,000	1.65(−3)	3.59(−4)	3.56(−4)
III	10	1.16(−1)[a]	1.13(−1)	9.95(−2)
	50	2.88(−2)[a]	2.39(−2)	2.31(−2)
	100	1.17(−2)	1.17(−2)	1.16(−2)
	500	4.90(−3)	2.00(−3)	1.98(−3)
	1,000	3.72(−3)	1.11(−3)	1.10(−3)
	3,000	3.08(−3)	4.07(−4)	3.97(−4)
IV	10	5.54(−2)	9.02(−2)	8.72(−2)
	50	1.05(−2)	2.35(−2)	2.34(−2)
	100	2.62(−3)	9.52(−3)	9.52(−3)
	500	2.95(−3)	1.75(−3)	1.74(−3)
	1,000	2.52(−3)	9.52(−4)	9.46(−4)
	3,000	1.67(−3)	3.55(−4)	3.52(−4)
V	10	8.42(−2)[a]	9.79(−2)	9.08(−2)
	50	1.38(−2)	2.41(−2)	2.39(−2)
	100	4.41(−3)	9.21(−3)	9.20(−3)
	500	2.73(−3)	1.72(−3)	1.71(−3)
	1,000	2.27(−3)	9.53(−4)	9.48(−4)
	3,000	1.77(−3)	3.54(−4)	3.51(−4)

[a] Significant at the five per cent level.

and was chosen to be 10 for all experiments. This bound is rarely reached in the models considered and only for very small t.

All five models were run for 3,000 periods. The long runs were necessary to investigate the asymptotic distribution of the parameter estimates which evolves very slowly and to test for the strong convergence of the control. In all experiments the random variable u_t is symmetric and approximately normal. The same 300,000 random numbers were used for all experiments.

4. THE STATISTICAL PROPERTIES OF THE CONTROL RULE

It has been shown analytically by Taylor [**6**] that when α is *known*, x_t converges with probability one to $(y^* - \alpha)/\beta$ which equals 0 in canonical form, and $\sqrt{t}[x_{t+1} - (y^* - \alpha)/\beta]$ has a limiting normal distribution with mean 0 and variance σ^2/β^2 which equals one in canonical form. The conjecture that the control rule also has these properties when α is *unknown* is confirmed experimentally as described below.

The estimated moments of the control rule for the five models and six time periods are presented in Table II. The estimate of the variance of x_{t+1} (and the other estimated variances presented in this paper) are calculated by dividing the sum of deviations by 100 in order that the estimated mean square error (MSE) is the sum of the estimated variance and the square of the estimated bias. Since all models are in canonical form the mean square error of x_{t+1}, which is defined by $E[x_{t+1} - (y^* - \alpha)/\beta]^2$, is equal to the second moment Ex_t^2. Probability intervals for these estimated moments are found in Table III. Since x_{t+1} is a function of

TABLE III

95 PER CENT PROBABILITY INTERVALS FOR ESTIMATED MOMENTS OF x_{t+1}

t	Interval for $E(x_{t+1})$	Interval for MSE $(x_{t+1}$	
10	$\pm 6.20(-2)$	$7.42(-2)$	$1.30(-1)$
50	$\pm 2.76(-2)$	$1.48(-2)$	$2.59(-2)$
100	$\pm 1.96(-2)$	$7.42(-3)$	$1.30(-2)$
500	$\pm 8.77(-3)$	$1.48(-3)$	$2.59(-3)$
1,000	$\pm 6.20(-3)$	$7.42(-4)$	$1.30(-3)$
3,000	$\pm 3.58(-3)$	$2.47(-4)$	$4.32(-4)$

t observations, the intervals are computed assuming x_{t+1} is $N(0, t^{-1})$. Thus the estimate of Ex_{t+1} will have a normal distribution with mean 0 and variance $(100t)^{-1}$, and the estimate of $\mathrm{MSE}(x_{t+1})$ will be distributed as $\chi^2_{100}/(100t)^{-1}$ (with mean t^{-1} and variance $(50t^2)^{-1}$). As can be seen from Table II the estimated $\mathrm{MSE}(x_{t+1})$ are within the 95 per cent interval for all models and all reported time periods. The estimated var (x_{t+1}) is always within this interval. These results confirm the conjecture that the variance of the limiting distribution of $\sqrt{t}x_{t+1}$ is equal to one. They also indicate that the speed of convergence to this variance is quite rapid, being very close at $t = 10$ for all models. The estimate of Ex_{t+1} indicates that there is slight positive bias though this is statistically significant in only a few cases.

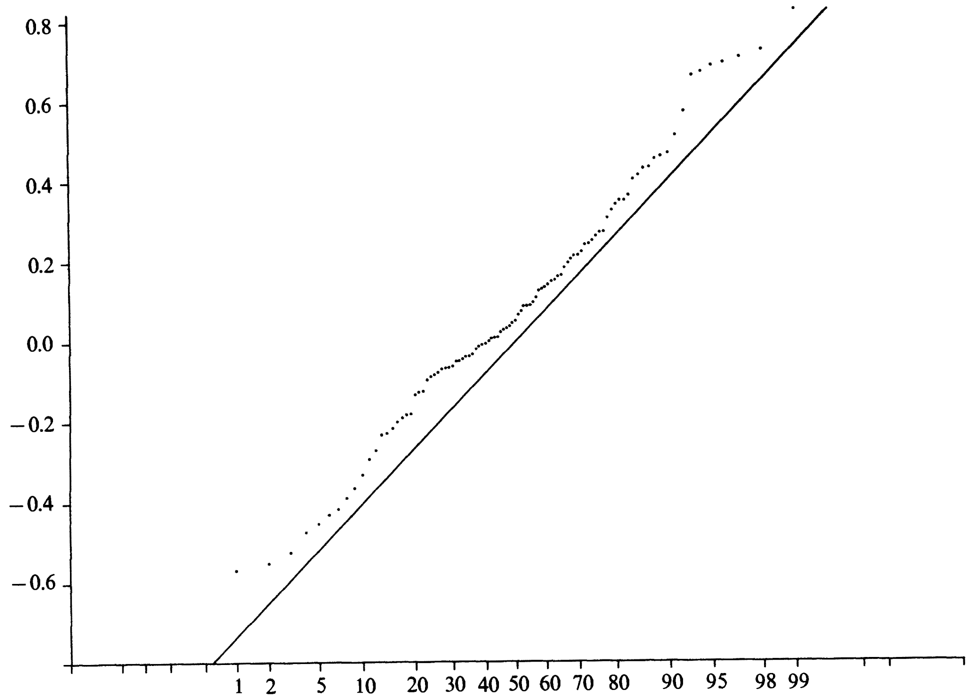

FIGURE 1.—Distribution of X_{11}, Model V.

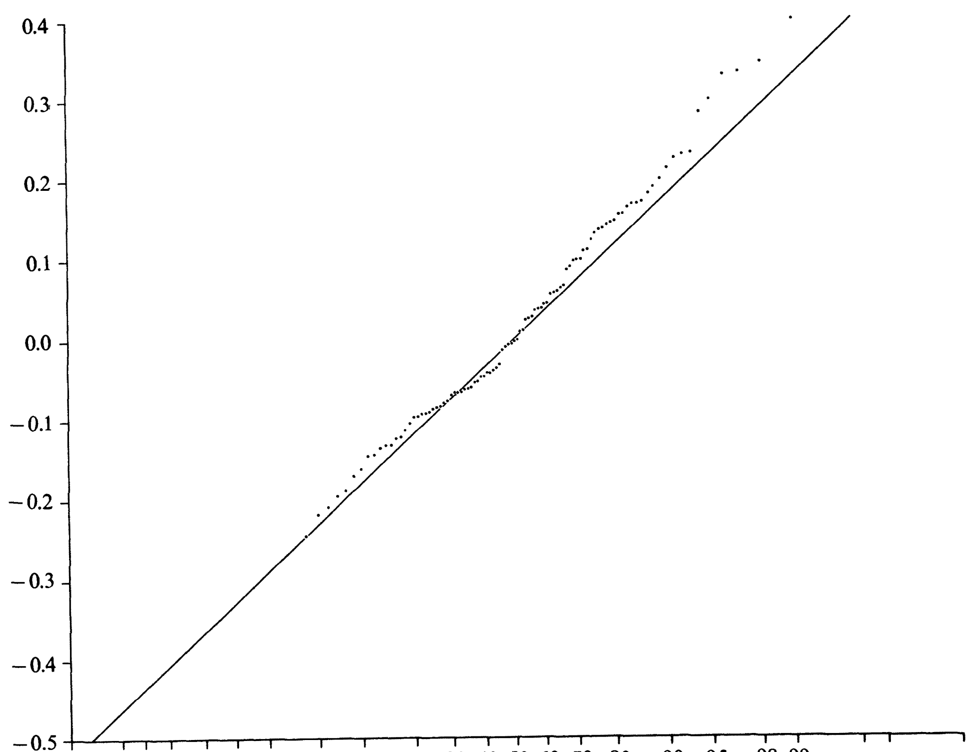

FIGURE 2.—Distribution of X_{51}, Model V.

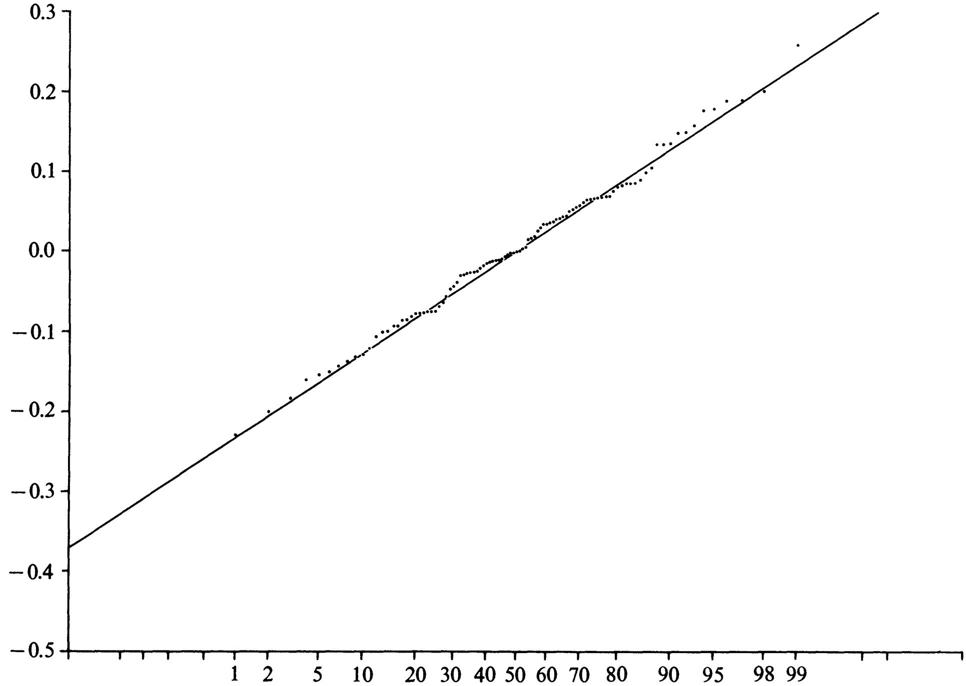

FIGURE 3.—Distribution of X_{101}, Model V.

The sample distribution for Model V is plotted on normal probability paper in Figures 1, 2, and 3 for $t = 10$, 50, and 100 along with the normal distribution with the appropriate variance.[2] For $t = 10$ the sample distribution is nearly normal but shifted slightly to the right, giving a smaller probability for negative values than the normal. The normal approximation for $t = 50$ and 100 is even closer. Thus, the conjecture of asymptotic normality is experimentally confirmed

Convergence in probability of the control rule is implied by the mean square convergence evident in Table II. To test for convergence with probability one $P[\max_{t>T}|x_{t+1}| < \varepsilon]$ was estimated for several ε and T. The results of this estimation for each model are presented in Table IV. The five values of ε were set to[3] $\sqrt{(2\log\log t)/t}$ for $t = 10$, 50, 100, 500, and 1,000. These values are the bounds appropriate for the mean of a random sample with the same expectation and variance as x_{t+1}, as given by the law of the iterated logarithm. The probability was not estimated for T larger than 1,000 so that the infinite range of the maximum is approximated by at least 2,000 periods. The results indicate that x_{t+1} does converge with probability one. For any fixed ε we can find a T to make the probability arbitrarily close to one. In addition, since all entries below the main diagonal of these tables are approximately one, the law of the iterated logarithm gives a good estimate of the upper bound. (There is one series in Model I for which $|x_{1408}|$ is greater than .2336, thus keeping the estimated probability at .99.)

[2] In each, 99 points are plotted; one value was deleted at random.

[3] The notation log refers to the natural logarithm.

TABLE IV

ESTIMATES OF $P[\max_{t>T} |x_{t+1}| < \varepsilon]$

Model	T	ε .4084	.2336	.1748	.0855	.0622
I	10	.76	.23	.17	.01	.01
	50	.99	.84	.66	.20	.10
	100	1.00	.87	.87	.38	.14
	500	1.00	.99	.99	.88	.62
	1,000	1.00	.99	.99	.95	.84
II	10	.72	.22	.03	.00	.00
	50	.99	.84	.62	.10	.02
	100	1.00	.98	.88	.33	.07
	500	1.00	1.00	1.00	.91	.68
	1,000	1.00	1.00	1.00	1.00	.90
III	10	.71	.26	.09	.01	.00
	50	.99	.80	.63	.15	.07
	100	1.00	.96	.84	.38	.12
	500	1.00	1.00	1.00	.88	.63
	1,000	1.00	1.00	1.00	.97	.87
IV	10	.77	.13	.04	.00	.00
	50	1.00	.85	.64	.13	.00
	100	1.00	.98	.88	.36	.07
	500	1.00	1.00	1.00	.90	.65
	1,000	1.00	1.00	1.00	1.00	.90
V	10	.75	.17	.06	.00	.00
	50	.99	.85	.62	.09	.01
	100	1.00	1.00	.87	.34	.06
	500	1.00	1.00	1.00	.91	.65
	1,000	1.00	1.00	1.00	1.00	.92

The implication of these experimental results is that the variance of the control rule in a model with two unknown parameters decreases as $1/t$ and is asymptotically no larger than the variance in a model with only one unknown parameter. This also implies that out of the class of strongly consistent controls, the least squares certainty equivalence control rule considered here is asymptotically efficient in the sense that its asymptotic variance is as small as any other control rule in this class.

If the measure of loss is $(y_s - y^*)^2$, then the expected loss in the first t periods is

$$\sum_{s=1}^{t} E(y_s - y^*)^2 = \beta^2 \sum_{s=1}^{t} \mathrm{MSE}\,(x_s) + t\sigma^2. \tag{4.1}$$

In canonical form it is $\Sigma_{s=1}^{t} E(x_s^2) + t$. Since the expected loss with knowledge of α and β is t, the "regret" is $\Sigma_{s=1}^{t} E(x_s^2)$ and can be considered as the loss due to the lack of knowledge of the parameters. The estimate of this quantity is given in Table VI.

The experimental results indicate that the limit of the normalized single period regret is given by

$$\lim_{t\to\infty} tE[(y_t - y^*)^2 - \sigma^2] = \sigma^2. \tag{4.2}$$

TABLE V

ESTIMATED MOMENTS OF THE PARAMETER ESTIMATES

Model	t	$E(\hat{\alpha}_t)$	Var $(\hat{\alpha}_t)$	$E(\hat{\beta}_t)$	Var $(\hat{\beta}_t)$	Cov $(\hat{\alpha}_t, \hat{\beta}_t)$	Cor $(\hat{\alpha}_t, \hat{\beta}_t)$
I	10	−7.20(−2)	2.11(−1)	1.62[a]	1.31	6.00(−2)	1.14(−1)
	50	−3.10(−2)	6.96(−2)	1.77[a]	7.44(−1)	−1.30(−2)	−5.73(−2)
	100	−1.15(−2)[a]	4.15(−2)	1.79[a]	8.13(−1)	−1.76(−2)	−9.58(−2)
	500	−1.15(−3)	8.47(−3)	1.69[a]	8.24(−1)	3.67(−4)	4.39(−3)
	1,000	−3.66(−3)	4.97(−3)	1.66[a]	8.05(−1)	−2.70(−3)	−4.26(−2)
	3,000	−3.47(−3)	1.53(−3)	1.54[a]	6.99(−1)	−2.38(−3)	−7.28(−2)
II	10	−4.88(−2)	8.29(−2)	1.00	1.52(−2)	7.46(−4)	−2.10(−2)
	50	−1.02(−2)	2.27(−2)	1.01	1.43(−2)	−1.42(−3)	−7.88(−2)
	100	−3.46(−3)	9.31(−3)	1.01	1.45(−2)	−2.66(−3)	−2.29(−1)[a]
	500	−3.27(−3)	1.77(−3)	1.01	1.42(−2)	−2.44(−4)	−4.86(−2)
	1,000	−2.73(−3)	1.01(−3)	1.01	1.43(−2)	−9.74(−5)	−2.56(−2)
	3,000	−2.10(−3)	3.79(−4)	1.02	1.37(−2)	−4.59(−4)	−2.01(−1)[a]
III	10	−1.99(−1)[a]	2.95(−1)	1.61[a]	7.51(−1)	−2.83(−1)	−6.01(−1)[a]
	50	−8.04(−2)[a]	8.78(−2)	1.57[a]	5.40(−1)	−1.22(−1)	−5.60(−1)[a]
	100	−4.43(−2)[a]	4.11(−2)	1.54[a]	4.99(−1)	−7.45(−2)	−5.20(−1)[a]
	500	−9.85(−3)	5.24(−3)	1.44[a]	3.67(−1)	−9.33(−3)	−2.13(−1)[a]
	1,000	−6.58(−3)	2.77(−3)	1.42[a]	3.37(−1)	−4.00(−3)	−1.31(−1)
	3,000	−4.82(−3)	9.50(−4)	1.38[a]	2.98(−1)	−2.41(−3)	−1.43(−1)
IV	10	−6.51(−2)[a]	9.26(−2)	1.01	8.94(−3)	−9.58(−3)	−3.33(−1)[a]
	50	−1.34(−2)	2.34(−2)	1.01	8.11(−3)	−3.06(−3)	−2.22(−1)[a]
	100	−4.90(−3)	9.28(−3)	1.01	8.08(−3)	−2.37(−3)	−2.73(−1)[a]
	500	−3.21(−3)	1.71(−3)	1.01	7.89(−3)	−2.93(−4)	−7.97(−2)
	1,000	−2.66(−3)	9.66(−4)	1.01	7.92(−3)	−1.87(−4)	−6.74(−2)
	3,000	−1.93(−3)	3.63(−4)	1.01	7.67(−3)	−2.83(−4)	−1.70(−1)
V	10	−9.99(−2)[a]	1.07(−1)	1.04[a]	1.43(−2)	−1.52(−2)	−3.86(−1)[a]
	50	−1.78(−2)	2.43(−2)	1.03[a]	1.24(−2)	−3.94(−3)	−2.27(−1)[a]
	100	−6.20(−3)	9.25(−3)	1.02[a]	1.18(−2)	−1.86(−3)	−1.78(−1)
	500	−3.07(−3)	1.77(−3)	1.02[a]	1.18(−2)	−3.66(−4)	−8.00(−2)
	1,000	−2.65(−3)	1.01(−3)	1.02[a]	1.19(−2)	−3.82(−4)	−1.11(−1)
	3,000	−2.03(−3)	3.78(−4)	1.03[a]	1.14(−2)	−2.87(−4)	−1.38(−1)

[a] Significant at the five per cent level.

Let $\hat{y}_t$ be the value of the dependent variable when a consistent control rule other than the LSCE rule is being used and let

$$\tau^2 = \lim_{t \to \infty} tE[(\hat{y}_t - y^*)^2 - \sigma^2]. \tag{4.3}$$

Then the asymptotic efficiency of the LSCE rule implies that $\sigma^2 \leqslant \tau^2$. This also implies that

$$\lim_{T \to \infty} \frac{\sum_{t=1}^{T} E[(y_t - y^*)^2 - \sigma^2]}{\sum_{t=1}^{T} E[(\hat{y}_t - y^*)^2 - \sigma^2]} = \frac{\sigma^2}{\tau^2} \leqslant 1. \tag{4.4}$$

This can be shown by dividing the numerator and denominator of the above expression by log T. (See Knopp [**4**, p. 32].)

5. THE STATISTICAL PROPERTIES OF THE PARAMETER ESTIMATES

The estimated moments of the parameter estimates in canonical form for all models are presented in Table V. The estimate of α shows a small negative bias for all models. The values of the estimate of $E\hat{\alpha}_t$ that are significantly different from zero at the 5 per cent level of significance (two-sided test) are marked in Table V; there are 6 out of 30 such values significant. The estimated variance of $\hat{\alpha}_t$ decreases approximately as $1/t$, though it is considerably larger than the estimated var (x_t). (Note that $x_{t+1} = -\hat{\alpha}_t/\hat{\beta}_t$ and $\hat{\beta}_t$ estimates one.) The estimate of β shows a positive bias which is quite large for some models; its variance decreases very slowly, as $1/\log t$ for large t. The correlation between $\hat{\alpha}_t$ and $\hat{\beta}_t$ is negative for all models and significantly different from zero for some values in four of the five models. The sequences do not show a tendency to converge to zero. The test of significance for the correlation is a t test on the basis that $(\hat{\alpha}_t, \hat{\beta}_t)$ are approximately normally distributed.

TABLE VI

ESTIMATED MOMENTS OF THE ELEMENTS OF MATRIX $\mathbf{A}_t$

Model	t	$E\left(\sum_1^t x_s\right)$	$\mathrm{Var}\left(\sum_1^t x_s\right)$	$E\left(\sum_1^t x_s^2\right)$	$\mathrm{Var}\left(\sum_1^t x_s^2\right)$	$E\left(\sum_1^t (x_s - \bar{x}_t)^2\right)$	$\mathrm{Var}\left(\sum_1^t (x_s - \bar{x}_t)^2\right)$
I	10	4.72(−1)	1.85(1)	1.43(1)	8.44(2)	1.25(1)	6.95(2)
	50	1.32	9.05(1)	1.82(1)	9.33(2)	1.63(1)	8.86(2)
	100	1.55	1.94(2)	1.89(1)	9.42(2)	1.69(1)	9.11(2)
	500	2.82	1.16(3)	2.05(1)	9.46(2)	1.82(1)	9.37(2)
	1,000	4.35	2.54(3)	2.14(1)	9.47(2)	1.89(1)	9.50(2)
	3,000	1.35(1)	7.73(3)	2.29(1)	9.38(2)	2.02(1)	9.48(2)
II	10	3.71(−1)	7.78	5.15(1)	2.84	5.07(1)	4.03(1)
	50	1.47	6.55(1)	5.30(1)	6.24	5.16(1)	1.51
	100	1.49	1.57(2)	5.37(1)	8.22	5.21(1)	2.35
	500	2.69	8.16(2)	5.50(1)	1.11(1)	5.34(1)	5.27
	1,000	4.56	1.79(3)	5.57(1)	1.22(1)	5.39(1)	7.04
	3,000	1.28(1)	5.60(3)	5.68(1)	1.37(1)	5.49(1)	1.04(1)
III	10	2.44	1.98(1)	1.66(1)	8.91(2)	1.40(1)	7.30(2)
	50	4.49	9.15(1)	1.95(1)	9.62(2)	1.72(1)	9.47(2)
	100	5.09	2.01(2)	2.03(1)	9.58(2)	1.80(1)	9.60(2)
	500	8.09	1.02(3)	2.19(1)	9.54(2)	1.97(1)	9.68(2)
	1,000	1.08(1)	2.16(3)	2.27(1)	9.56(2)	2.04(1)	9.64(2)
	3,000	2.13(1)	6.53(3)	2.39(1)	9.55(2)	2.16(1)	9.60(2)
IV	10	1.02(1)	9.54	1.02(2)	3.59	9.09(1)	4.23(1)
	50	1.14(1)	7.05(1)	1.04(2)	7.04	9.98(1)	1.45(1)
	100	1.15(1)	1.70(2)	1.05(2)	9.69	1.02(2)	9.47
	500	1.29(1)	8.38(2)	1.06(2)	1.24(1)	1.04(2)	6.21
	1,000	1.47(1)	1.79(3)	1.07(2)	1.33(1)	1.05(2)	7.59
	3,000	2.29(1)	5.61(3)	1.08(2)	1.50(1)	1.06(2)	1.13(1)
V	10	9.34	3.44(1)	7.98(1)	1.36(3)	6.76(1)	1.57(3)
	50	1.08(1)	1.01(2)	8.14(1)	1.34(3)	7.70(1)	1.40(3)
	100	1.11(1)	2.06(2)	8.21(1)	1.34(3)	7.88(1)	1.38(3)
	500	1.26(1)	8.65(2)	8.34(1)	1.35(3)	8.14(1)	1.36(3)
	1,000	1.43(1)	1.79(3)	8.41(1)	1.36(3)	8.21(1)	1.36(3)
	3,000	2.24(1)	5.60(3)	8.52(1)	1.37(3)	8.32(1)	1.35(3)

Scatter plots of $\hat{\alpha}_t$ and $\hat{\beta}_t$ are found in Figures 4, 5, and 6 for Model V for $t = 50$, 100, and 1,000. The plot for $t = 50$ is influenced by the initial conditions for this model which give a small var $\hat{\beta}_2$ and a large var $\hat{\alpha}_2$. At $t = 50$ the scatter shows a slight horizontal tendency, but as t grows the scatter becomes vertical, thus showing the quick convergence of $\hat{\alpha}_t$ relative to the slow convergence of $\hat{\beta}_t$.

The scatter is vertical because the model is in canonical form. Any transformation of this canonical form for which the transformed α is nonzero will lead to scatter which is tilted away from the vertical. Therefore, in any such model the estimates of α or of β will converge as slowly as the canonical $\hat{\beta}_t$ does for these results (though the linear combination (2.5) will converge as rapidly as the canonical $\hat{\alpha}_t$). The poor performance of these estimates could lead to serious mistakes in practical problems.

For further investigation of the asymptotic properties of these estimates we examine the elements of the random matrix $\mathbf{A}_t$ defined by

$$\mathbf{A}_t = \begin{pmatrix} t & \sum_{s=1}^{t} x_s \\ \sum_{s=1}^{t} x_s & \sum_{s=1}^{t} x_s^2 \end{pmatrix}, \tag{5.1}$$

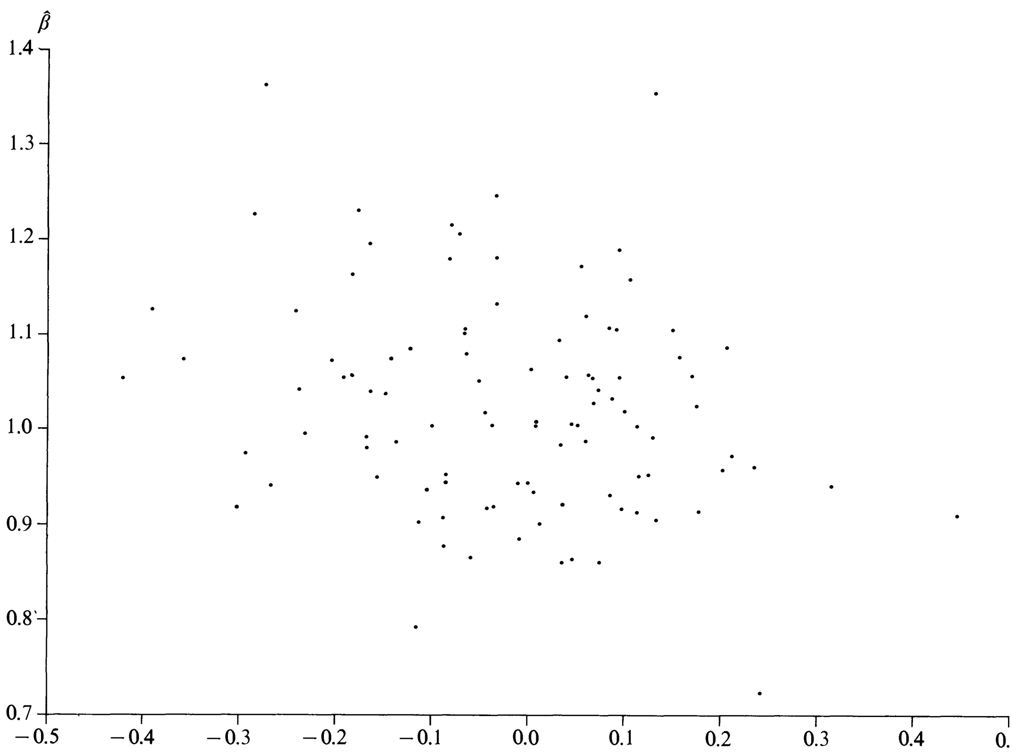

FIGURE 4.—Distribution of $\hat{\alpha}$ and $\hat{\beta}$, Model V, $t = 50$.

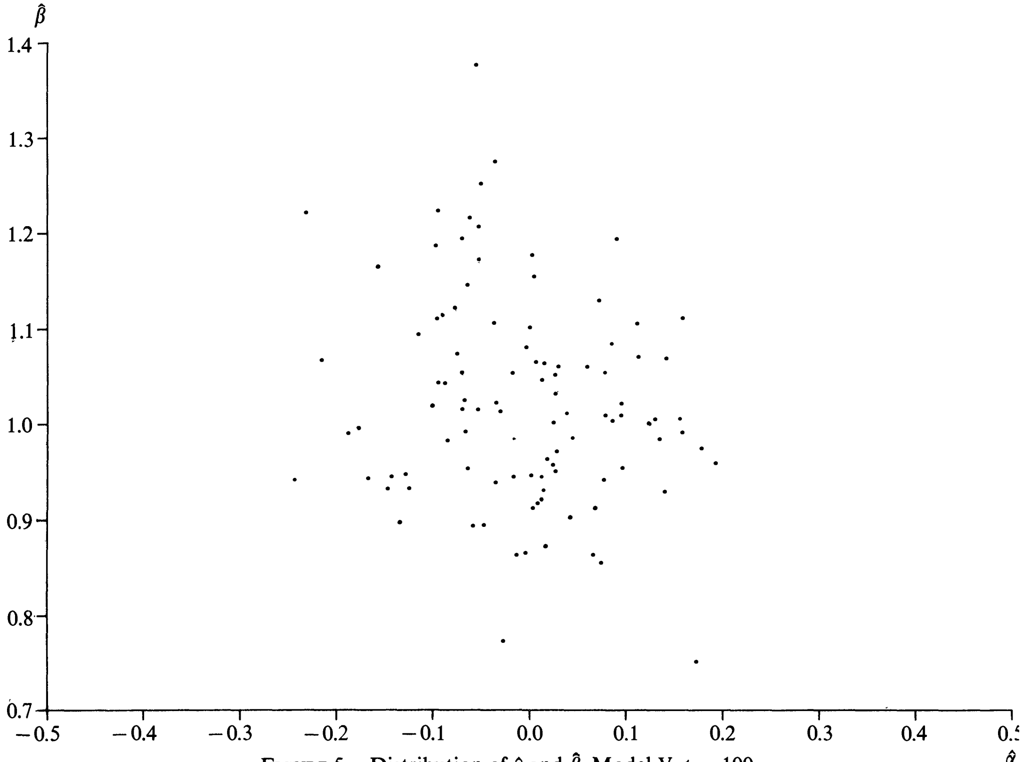

FIGURE 5.—Distribution of $\hat{\alpha}$ and $\hat{\beta}$, Model V, $t = 100$.

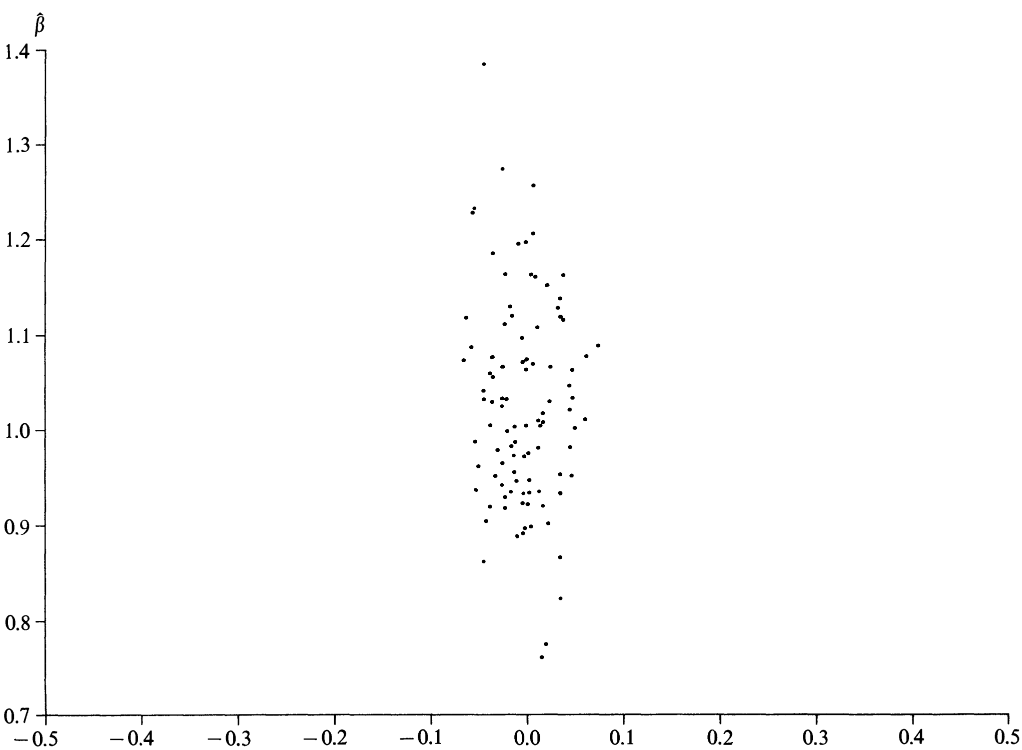

FIGURE 6.—Distribution of $\hat{\alpha}$ and $\hat{\beta}$, Model V, $t = 1{,}000$.

which is used in the calculation of the least squares estimates. If the x_t's were independent nonstochastic variables, the covariance matrix of $(\hat{\alpha}_t, \hat{\beta}_t)$ would be $\mathbf{A}_t^{-1}$ and the variance of $\hat{\beta}_t$ would be $1/\Sigma_{s=1}^t (x_s - \bar{x}_t)^2$. The estimated moments of the stochastic elements of $\mathbf{A}_t$ are contained in Table VI. The estimated values of

TABLE VII

ESTIMATED AVERAGE CORRELATION IN $\mathbf{A}_t[E(\rho_t)]$

t	Model I	Model II	Model III	Model IV	Model V
10	5.10(−2)	1.65(−2)	3.54(−1)[a]	3.19(−1)[a]	3.66(−1)[a]
50	5.43(−2)	2.74(−2)	2.38(−1)[a]	1.58(−1)[a]	1.88(−1)[a]
100	3.67(−2)	1.87(−2)	1.84(−1)[a]	1.12(−1)[a]	1.36(−1)
500	3.08(−2)	1.51(−2)	1.21(−1)[a]	5.56(−2)	7.09(−2)
1,000	3.05(−2)	1.85(−2)	1.03(−1)[a]	4.47(−2)	5.57(−2)
3,000	5.67(−2)	2.97(−2)	1.05(−1)[a]	3.99(−2)	4.91(−2)

[a] Significant at the five per cent level.

$E(\Sigma_{s=1}^t x_s^2)$ increase with t about as fast as log log t, which is very slow. Table VII contains estimates of the means of

$$\rho_t = \frac{\sum_{s=1}^{t} x_s}{\sqrt{t \sum_{s=1}^{t} x_s^2}}. \tag{5.2}$$

These estimates are positive; 11 out of 30 are significantly different from zero (5 per cent, two-sided test). The sequences do not exhibit a strong tendency to move towards zero.

6. EFFECTS OF STARTING VALUES

Since the first two values of the control variable x_1 and x_2 can be selected arbitrarily, it is of interest to study the effect of their choice on the properties of the rule. The estimates of $E(x_{t+1})$ and var (x_{t+1}) are somewhat larger for Models I and III than for Models II, IV, and V, particularly at $t = 3{,}000$. In Models I and III, $x_1 - x_2$ is small and var $(\hat{\beta}_2)$ is large.

In these two models (I and III) the estimates of var $(\hat{\alpha}_t)$ are about four times $1/t$ while in the other three models var $(\hat{\alpha}_t)$ is hardly larger than $1/t$. In the two models with large var $(\hat{\beta}_2)$, $\hat{\beta}_t$ exhibits a large bias and relatively large variance. In these cases var $(\hat{\beta}_t)$ tends to be much larger than $1/\Sigma_{s=1}^t (x_s - \bar{x}_t)^2$. The fact that the ratios of the variances $\hat{\alpha}_t$ and $\hat{\beta}_t$ to elements of the inverse of the averages of $\mathbf{A}_t$ varies, indicates that $\mathbf{A}_1^{-1}$ is not a good estimate of the covariance matrix of $\hat{\alpha}_t, \hat{\beta}_t$. The lower accuracy of $\hat{\alpha}_t, \hat{\beta}_t$ in Models I and III is reflected in the slower approach of x_t to 0.

The lower accuracy of the parameter estimates in Models I and III can be related to the facts that $\Sigma_{s=1}^t x_s^2$ and $\Sigma_{s=1}^t (x_s - \bar{x}_t)^2$ are considerably smaller for

TABLE VIII

ESTIMATES OF $E\left(\sum_{s=1}^{t} x_s^2\right)$ AND $E\left[\sum_{s=1}^{t} (x_s - \bar{x}_t)^2\right]$

t	I		II		III		IV		V	
2	0.5	0.5	50.0	50.0	1.0	0.5	100.0	50.0	52.0	2.0
10	14.3	12.5	51.5	50.7	16.6	14.0	102.2	90.9	79.8	67.6
3,000	22.9	20.2	56.8	54.9	23.9	21.6	107.6	107.6	105.6	83.2

these models than for the other three at every value of t. Table VIII gives the estimated expected value for $t = 2$, 10, and 3,000. The estimates of var $(\hat{\beta}_t)$ are roughly in the same order as the reciprocals of the estimates of $E\,\Sigma_{s=1}^{t} (x_s - \bar{x}_t)^2$.

In every model the increase in the estimate of $E\,\Sigma_{s=1}^{t} x_s^2$ from $t = 10$ to $t = 3{,}000$ (2,990 time periods) is small relative to the value at $t = 10$. In Models II and IV the increase from $t = 2$ to $t = 10$ (or $t = 3{,}000$) is small. In these cases where $\Sigma_{s=1}^{t} (x_s - \bar{x}_t)^2$ is large (that is, $(x_1 - x_2)$ is large) and, correspondingly, var $(\hat{\beta}_2)$ is small, almost all the information in estimating the slope is in the first two observations. Note that the accuracy in estimation is at the cost of a high expected loss.

One of the most interesting questions in this area is whether $\hat{\beta}_t \to 1$ with probability one (or in probability). A related question is whether $\Sigma_{s=1}^{t} (x_s - \bar{x}_t)^2 \to \infty$ with probability one (or in probability). Our evidence pertaining to these questions is inconclusive.

Stanford University and London School of Economics
and
Columbia University

Manuscript received May, 1975; revision received October, 1975.

REFERENCES

[1] AOKI, M.: *Optimization of Stochastic Systems*. New York: Academic Press, 1967.
[2] ———: "On Some Price Adjustment Schemes," *Annals of Economic and Social Measurement*, 3 (1974), 95–116.
[3] CHOW, G. C.: *Analysis and Control of Dynamic Economic Systems*. New York: Wiley, 1975.
[4] KNOPP, K.: *Infinite Sequences and Series*. New York: Dover, 1956.
[5] PRESCOTT, E. C.: "The Multiperiod Control Problem under Uncertainty," *Econometrica*, 40 (1972), 1043–1058.
[6] TAYLOR, J. B.: "Asymptotic Properties of Multiperiod Control Rules in the Linear Regression Model," *International Economic Review*, 15 (1974), 472–484.
[7] TSE, E.: "Adaptive Dual Control Methods," *Annals of Economic and Social Measurement*, 3 (1974), 65–83.
[8] ZELLNER, A.: *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley, 1971.