

THE EXACT MULTI-PERIOD MEAN-SQUARE FORECAST ERROR FOR THE FIRST-ORDER AUTOREGRESSIVE MODEL*

Asraul HOQUE

Monash University, Clayton, Vict. 3168, Australia

Jan R. MAGNUS

London School of Economics, London WC2A 2AE, England

Bahram PESARAN

National Institute of Economic & Social Research, London SW1P 3HE, England

Received May 1987, final version received November 1987

The finite-sample behaviour of the multi-period least-squares forecast is considered in the simple normal autoregressive model $y_t = \beta y_{t-1} + u_t$, where $|\beta| < 1$. Necessary and sufficient conditions are established for the existence of the forecast bias and the mean-square forecast error (MSFE) and an exact expression for the MSFE is given. Exact numerical results are obtained for both the stationary and the fixed start-up case. Our main conclusions are that for small values of β the MSFE is a decreasing function of the number of forecast periods, and that the behaviour of the MSFE in the stationary and the fixed start-up case is very similar, except for values of $|\beta|$ close to 1.

1. Introduction

The purpose of this paper is to study the behaviour in finite samples of the least-squares (LS) forecast in the simple autoregressive model

$$y_t = \beta y_{t-1} + u_t,$$

where $|\beta| < 1$ and $\{u_t\}$ is a sequence of independent and identically distributed (i.i.d.) $N(0, \sigma^2)$ random variables. The specification of the initial observation is important in finite samples and we shall distinguish between two cases: the stationary and the non-stationary (fixed start-up) case.

If $\hat{\beta}$ denotes the LS estimator of β based on n observations y_1, \dots, y_n , then the s -periods-ahead LS forecast is $\hat{y}_{n+s} = \hat{\beta}^s y_n$. It is well known [Malinvaud (1970)] that the forecast bias $E(\hat{y}_{n+s} - y_{n+s})$ is zero if it exists. We shall show that the forecast bias exists if and only if $s \leq n - 2$, and that the mean-square

*The first version of this paper was written in the Spring of 1986 while all three authors were at LSE. We are grateful to three referees and an associate editor for their careful and positive comments which improved the exposition.

forecast error (MSFE) of the forecast (that is, the variance of the forecast error) exists if and only if $s \leq [(n-2)/2]$ in which case we obtain an exact expression for the MSFE which can be calculated by numerical integration.

Our exact numerical results show, not surprisingly, that the MSFE is a decreasing function of n for all values of β and s . But they also show that it is *not* generally true that the MSFE is an increasing function of s ; indeed, for small values of β the MSFE appears to be a *decreasing* function of s . We further find that the behaviour of the MSFE in the stationary and the non-stationary case is very similar, *except for values of $|\beta|$ close to 1*.

There is an extensive literature on the properties of the least-squares estimator of β in this or related models [see, e.g., Hurwicz (1950), Anderson (1959), White (1961), Copas (1966), Thornber (1967), Phillips (1977), Sawa (1978), Dickey and Fuller (1979), Evans and Savin (1981), Tanaka (1983), Hoque (1985b), Hoque and Peters (1986), and Nankervis and Savin (1988)].

The literature on the properties of the LS forecast \hat{y}_{n+s} is somewhat less extensive. Davisson (1965) obtained an asymptotic approximation of the MSFE up to order n^{-1} for the one-period-ahead forecast and $|\beta| < 1$. This result was extended by Fuller and Hasza (1981) to processes with $|\beta| = 1$ and $|\beta| > 1$. See also Yamamoto (1976) and Baillie (1979). Phillips (1979) developed an Edgeworth type expansion for the distribution of the forecast error. His work was extended by Tanaka and Maekawa (1984) who considered the situation where the true model is ARMA(1, $1/\mu$) but misspecified as AR(1). Lütkepohl (1984) and Hoque (1985a) have recently considered more general ARMA and ARMAX models. Lütkepohl studied the precision of forecasts from macro-economic time series, while Hoque obtained the exact MSFE for the one-period-ahead forecast.

Monte Carlo studies of the forecast error were conducted by Orcutt and Winokur (1969), Lahiri (1975), Gonedes and Roberts (1977), and Fuller and Hasza (1980). Of these four studies only Lahiri's paper is directly relevant to our paper and we shall compare his Monte Carlo results to our exact results in section 4.

The plan of this paper is as follows. In sections 2 and 3 we present the model and state our two theorems concerning the mean and the variance of the s -period-ahead forecast error. Our numerical results are presented in sections 4 and 5. Three appendices, containing two new propositions and the proofs of the two theorems, conclude the paper.

2. The model

We shall consider the first-order autoregressive process $\{y_1, y_2, \dots\}$ defined by

$$y_t = \beta y_{t-1} + u_t, \quad t = 2, 3, \dots, \quad (1)$$

where $|\beta| < 1$ and $\{u_1, u_2, \dots\}$ is a sequence of i.i.d. $N(0, \sigma^2)$ random vari-

ables. Regarding the initial observation y_1 we postulate

$$y_1 = \delta u_1, \quad \delta > 0. \tag{2}$$

In order to obtain our exact numerical results we need to specify δ . Thus in sections 4 and 5 we shall assume either:

Assumption 1a (stationarity). $\delta = (1 - \beta^2)^{-1/2}$

in which case $\{y_1, y_2, \dots\}$ is a normal strictly stationary time series, or

Assumption 1b (non-stationarity). $\delta = 1$

in which case the series $\{y_1, y_2, \dots\}$ is not covariance stationary. Notice that Assumption 1b is equivalent to assuming $y_0 = 0$.

Let $y = (y_1, y_2, \dots, y_n)'$ be an $n \times 1$ vector of observations generated by (1) and (2). Then y is normally distributed $N(0, \sigma^2 LL')$, where

$$L = \begin{pmatrix} \delta & 0 & 0 & \dots & 0 & 0 \\ \delta\beta & 1 & 0 & \dots & 0 & 0 \\ \delta\beta^2 & \beta & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \delta\beta^{n-2} & \beta^{n-3} & \beta^{n-4} & \dots & 1 & 0 \\ \delta\beta^{n-1} & \beta^{n-2} & \beta^{n-3} & \dots & \beta & 1 \end{pmatrix}. \tag{3}$$

The least-squares estimator of β is

$$\hat{\beta} = \frac{\sum_{t=2}^n y_t y_{t-1}}{\sum_{t=2}^n y_{t-1}^2} = y' Ay / y' By, \tag{4}$$

where A and B are symmetric $n \times n$ matrices defined by

$$A = \frac{1}{2} \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 1 & 0 & 1 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}. \tag{5}$$

In the non-stationary case (Assumption 1b) the least-squares estimator is the maximum likelihood estimator.

The s -periods-ahead forecast is given by

$$\hat{y}_{n+s} = \hat{\beta}^s y_n, \quad s = 1, 2, \dots, \quad (6)$$

and the forecast error is therefore

$$\hat{y}_{n+s} - y_{n+s} = (\hat{\beta}^s - \beta^s) y_n - \sum_{j=0}^{s-1} \beta^j u_{n+s-j}. \quad (7)$$

From (7) we obtain the forecast bias,

$$E(\hat{y}_{n+s} - y_{n+s}) = E(\hat{\beta}^s y_n), \quad (8)$$

and the mean-square forecast error,

$$\begin{aligned} E(\hat{y}_{n+s} - y_{n+s})^2 &= \beta^{2s} \text{var}(y_n) + E(\hat{\beta}^{2s} y_n^2) \\ &\quad - 2\beta^s E(\hat{\beta}^s y_n^2) + \sigma^2 \sum_{j=0}^{s-1} \beta^{2j}, \end{aligned} \quad (9)$$

provided the expectations exist.

Malinvaud (1970, p. 554) was the first to point out that the expected value of the forecast error $\hat{y}_{n+s} - y_{n+s}$ is zero when the distribution of u_t is symmetric (as it is here), a result subsequently generalized by Fuller and Hasza (1980) and Dufour (1984, 1985). Theorem 1 gives necessary and sufficient conditions for the *existence* of the s -periods-ahead forecast bias.

Theorem 1. *The expectation of the forecast error of \hat{y}_{n+s} exists if and only if $1 \leq s \leq n - 2$, in which case*

$$E(\hat{y}_{n+s} - y_{n+s}) = 0.$$

Proof. Since it is well known that the forecast bias is zero when it exists, we only derive necessary and sufficient conditions for its existence; see appendix C.

3. The mean-square forecast error of \hat{y}_{n+s}

We shall now go one step further and determine conditions for the existence and obtain an exact expression for the mean-square forecast error (MSFE), $E(\hat{y}_{n+s} - y_{n+s})^2$. This is done in Theorem 2.

Theorem 2. Let A , B and L be the $n \times n$ matrices defined in (3) and (5). Let l' denote the last (n th) row of L . Let P be an orthogonal $n \times n$ matrix and Λ a diagonal $n \times n$ matrix, such that

$$P'L'BLP = \Lambda, \quad P'P = I_n,$$

and define

$$A^* = P'L'ALP, \quad l^* = P'l.$$

Then the mean-square forecast error of \hat{y}_{n+s} exists if and only if $1 \leq s \leq [(n - 2)/2]^1$ in which case

$$E(\hat{y}_{n+s} - y_{n+s})^2 = \sigma^2(c_{2s} - 2\beta^s c_s + (1 - K\beta^{2(n+s)})/(1 - \beta^2)), \quad (10)$$

where

$$K = \frac{1 - \delta^2(1 - \beta^2)}{\beta^2}. \quad (11)$$

Here c_k ($1 \leq k \leq n - 2$) is defined as

$$c_k = \frac{1}{(k - 1)!} \sum_{\nu} \gamma_k(\nu) \int_0^{\infty} t^{k-1} |\Delta| \left\{ \omega \theta' \theta + 2 \sum_{j=1}^k (jn_j r_j \omega_j) \theta' R^j \theta \right\} dt, \quad (12)$$

where the summation is over all $1 \times k$ vectors $\nu = (n_1, n_2, \dots, n_k)$ whose elements n_j are non-negative integers satisfying $\sum_{j=1}^k jn_j = k$,

$$\gamma_k(\nu) = k! 2^k \prod_{j=1}^k (n_j! (2j)^{n_j})^{-1}, \quad (13)$$

Δ is a diagonal positive definite $n \times n$ matrix, R a symmetric $n \times n$ matrix and θ an $n \times 1$ vector defined as

$$\Delta = (I_n + 2t\Lambda)^{-1/2}, \quad R = \Delta A^* \Delta, \quad \theta = \Delta l^*, \quad (14)$$

and the scalars ω , ω_j and r_j are defined by

$$\omega = \prod_{i=1}^k (\text{tr } R^i)^{n_i}, \quad \omega_j = \prod_{\substack{i=1 \\ i \neq j}}^k (\text{tr } R^i)^{n_i}, \quad (15)$$

¹The symbol $[x]$ denotes the integer part. Thus, $[x]$ is the largest integer $\leq x$.

and

$$\begin{aligned} r_j &= 1 && \text{if } n_j = 0, 1, \\ &= (\text{tr } R^j)^{n_j-1} && \text{if } n_j \geq 2. \end{aligned} \tag{16}$$

Remark 1. The matrix R^j denotes the j th power of the matrix R .

Remark 2. In the stationary case (Assumption 1a) we have $K = 0$; in the non-stationary case (Assumption 1b) we have $K = 1$.

Proof. See appendix C.

4. Exact results: The stationary case

Application of Theorem 2 involves numerical integration. We used the Numerical Algorithms Group (1984) (the so-called NAG) subroutine DO1AMF for this purpose. This subroutine also gives an estimate of the absolute error in the integration. For all results reported in this paper the absolute error was less than 10^{-5} . The eigenvalues and eigenvectors in A and P were calculated using the NAG subroutine FO2ABF. See Magnus (1986, sec. 7) for some further remarks on computation in a related problem.

The exact mean-square forecast error (MSFE) of the least-squares (LS) forecast \hat{y}_{n+s} was calculated for both the stationary and the non-stationary process for the following selected values of the autoregressive parameter (β), the number of observations (n) and the number of periods ahead (s):

$$\beta = 0.00, 0.10, 0.20, \dots, 0.90, 0.95, 0.99,$$

$$n = 10, 15, 20, 25,$$

$$s = 1, 2, 3, 4.$$

To consider larger values of n would have been computationally very costly, since computing time more than doubles for each additional five observations. Convergence of the integral was particularly slow for values of β close to one.

Before we discuss the numerical results for the stationary case (in this section) and the non-stationary case (in the next section) we note some simple facts which hold for both the stationary and the non-stationary process. First, the MSFE is proportional to σ^2 . Hence we may, without loss of generality, set $\sigma^2 = 1$. All numerical results reported below are for $\sigma^2 = 1$. Secondly, rewriting (9), we obtain

$$\text{MSFE} = E(\beta^2 - \hat{\beta}^s)^2 y_n^2 + \sigma^2 \sum_{j=0}^{s-1} \beta^{2j}. \tag{17}$$

Hence

$$\text{MSFE} \geq \sigma^2, \quad (18)$$

and, since $\hat{\beta}$ is a consistent estimator of β ,

$$\lim_{n \rightarrow \infty} \text{MSFE} = \sigma^2(1 - \beta^{2s}) / (1 - \beta^2). \quad (19)$$

Thirdly, since our model has no intercept term, the MSFE is an even function of β , that is,

$$\text{MSFE}(\beta) = \text{MSFE}(-\beta).^2 \quad (20)$$

Hence it suffices to calculate the MSFE's for non-negative values of β only.

In table 1 we present the numerical results for the stationary case. One would conjecture (at least we did conjecture) that the MSFE is a decreasing function of n , and an increasing function of $|\beta|$ and s . The first of these conjectures appears to be true, but the second and third are not generally true. Fig. 1 illustrates that the MSFE decreases with increasing n , thus confirming our first conjecture. It also illustrates that, at least for $s = 1$, the MSFE is *not* a monotone function of $|\beta|$ thus refuting our second conjecture. The maximum of the MSFE occurs at $\beta = 0.91$ (when $n = 10$), $\beta = 0.94$ (when $n = 15$), $\beta = 0.96$ (when $n = 20$) and $\beta = 0.97$ (when $n = 25$), and the MSFE drops substantially when β approaches 1.

For the one-period-ahead forecast the MSFE appears to be a very flat function of β , but for the two(or more)-periods-ahead forecast this is no longer the case. It appears that, although for $s \geq 2$ the MSFE is a monotone function of $|\beta|$, the increase in the MSFE is less for values of $|\beta|$ close to 1, especially if n is small.

Our third conjecture was that, for fixed n and β , the MSFE increases with s . Fig. 2 shows, for $n = 15$, that this conjecture is false for $\beta \leq 0.45$ (although it is true for $\beta \geq 0.50$). Indeed, for $\beta \leq 0.20$ the order is exactly reversed. Thus for β close to zero the one-period-ahead forecast is less precise (has larger MSFE) than the two-periods-ahead forecast, which in turn is less precise than the three-periods-ahead forecast, and so on. Exactly the same phenomenon occurs for $n = 20$ and $n = 25$. For $n = 10$ the situation is slightly different. For $n = 10$ and $\beta \geq 0.30$ the MSFE increases with s and for small values of β the one-period-ahead forecast is less precise than the two-, three- or four-periods-ahead forecasts, but the four-periods-ahead forecast is less precise than the two- or three-periods-ahead forecasts for all values of β . Such a surprising and

²See Cryer, Nankervis and Savin (1986). This paper shows that the distribution of the forecast error is the same for β and $-\beta$, and hence that all moments are the same for β and $-\beta$.

Table 1
Exact mean-square forecast error of least-squares forecast \hat{y}_{n+s} : Stationary case.

n	s	β											
		0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99
10	1	1.1284	1.1286	1.1293	1.1305	1.1323	1.1348	1.1384	1.1435	1.1504	1.1570	1.1544	1.1353
10	2	1.0498	1.0637	1.1056	1.1760	1.2757	1.4060	1.5693	1.7695	2.0124	2.2988	2.4412	2.5020
10	3	1.0379	1.0530	1.1004	1.1860	1.3202	1.5187	1.8033	2.2045	2.7647	3.5243	3.9506	4.1702
10	4	1.0826	1.0985	1.1490	1.2436	1.4006	1.6506	2.0433	2.6570	3.6118	5.0491	5.9051	6.2995
15	1	1.0759	1.0760	1.0763	1.0768	1.0776	1.0788	1.0806	1.0832	1.0874	1.0938	1.0958	1.0868
15	2	1.0163	1.0286	1.0656	1.1276	1.2148	1.3279	1.4682	1.6378	1.8416	2.0883	2.2250	2.3079
15	3	1.0057	1.0175	1.0547	1.1222	1.2285	1.3860	1.6116	1.9288	2.3722	2.9958	3.3864	3.6712
15	4	1.0029	1.0139	1.0491	1.1146	1.2225	1.3939	1.6638	2.0892	2.7646	3.8531	4.6090	5.2050
20	1	1.0544	1.0545	1.0546	1.0549	1.0554	1.0560	1.0571	1.0586	1.0613	1.0662	1.0692	1.0645
20	2	1.0083	1.0201	1.0553	1.1140	1.1966	1.3033	1.4348	1.5926	1.7798	2.0047	2.1339	2.2225
20	3	1.0020	1.0131	1.0478	1.1107	1.2098	1.3561	1.5649	1.8562	2.2585	2.8198	3.1813	3.4707
20	4	1.0007	1.0112	1.0443	1.1053	1.2049	1.3615	1.6054	1.9853	2.5810	3.5352	4.2187	4.8172
25	1	1.0425	1.0426	1.0427	1.0429	1.0432	1.0436	1.0442	1.0453	1.0471	1.0508	1.0538	1.0515
25	2	1.0051	1.0165	1.0507	1.1077	1.1877	1.2910	1.4179	1.5694	1.7477	1.9596	2.0827	2.1737
25	3	1.0010	1.0117	1.0453	1.1063	1.2021	1.3435	1.5445	1.8235	2.2055	2.7321	3.0731	3.3604
25	4	1.0003	1.0106	1.0430	1.1025	1.1991	1.3498	1.5831	1.9439	2.5040	3.3895	4.0267	4.6131
∞	1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
∞	2	1.0000	1.0100	1.0400	1.0900	1.1600	1.2500	1.3600	1.4900	1.6400	1.8100	1.9025	1.9801
∞	3	1.0000	1.0101	1.0416	1.0981	1.1856	1.3125	1.4896	1.7301	2.0496	2.4661	2.7170	2.9407
∞	4	1.0000	1.0101	1.0417	1.0988	1.1897	1.3281	1.5363	1.8477	2.3117	2.9975	3.4521	3.8822

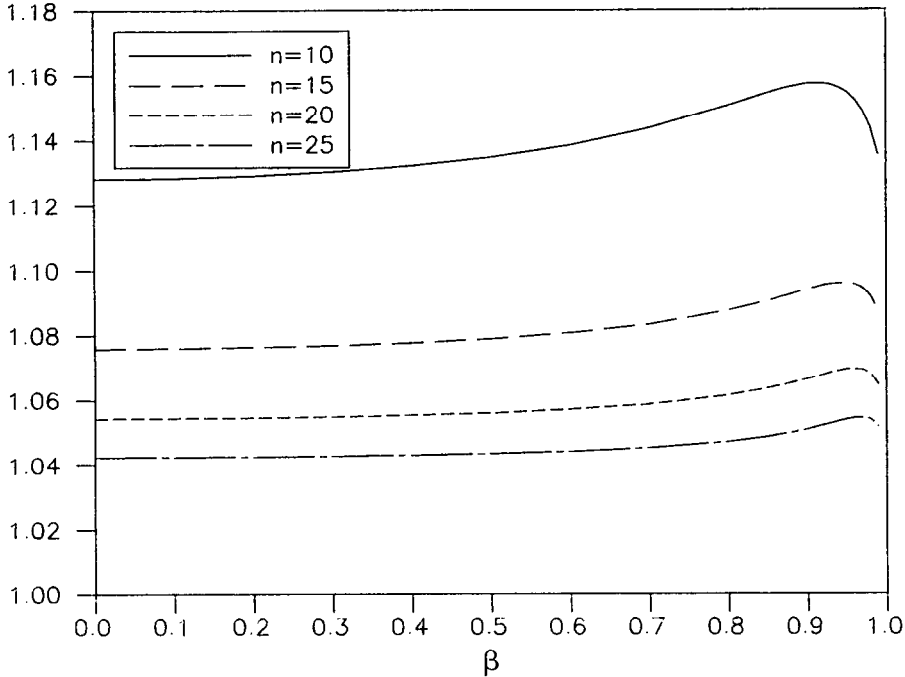


Fig. 1. MSFE of LS forecast for $s = 1$: Stationary case.

perhaps counterintuitive result merits further confirmation. In fact, supporting evidence that in certain cases the MSFE decreases as s increases can be found in Monte Carlo results reported by Fuller and Hasza (1980, table 1) for a model with an estimated intercept and Ansley and Newbold (1980, table 1).

For given values of β and s , we define the *relative efficiency* (RE) of the LS forecast as

$$RE = \frac{MSFE(\infty)}{MSFE(n)} \tag{21}$$

Since the MSFE is a decreasing function of n , we see that $0 < RE \leq 1$ for all values of β , s and n . Fig. 3 shows, for $n = 15$, that the RE is particularly high for small values of β and $s \geq 2$, and that the RE decreases with $|\beta|$ except, again, when $|\beta|$ is close to 1. The results for $n = 10, 20$ and 25 are similar. Hence, in the stationary case, the RE is at its lowest when $|\beta|$ is around 0.90 or 0.95.

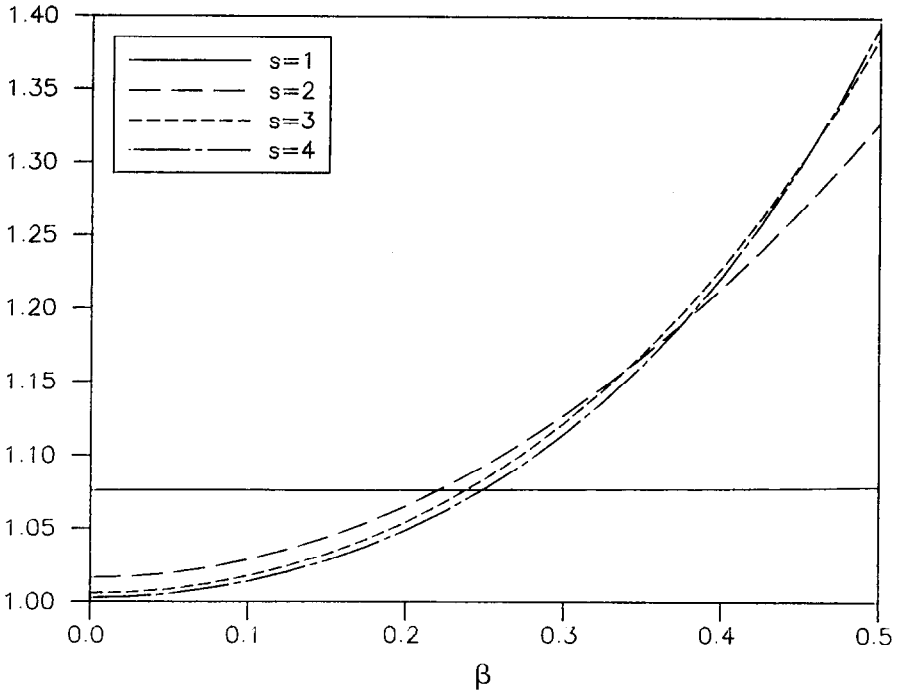


Fig. 2. MSFE of LS forecast for $n = 15$ and $\beta \leq 0.5$: Stationary case.

In fig. 2 we concluded that for small β and $n \geq 15$,

$$\text{MSFE}(s = 1) > \text{MSFE}(s = 2) > \text{MSFE}(s = 3) > \text{MSFE}(s = 4),$$

a rather striking result. Fig. 3 shows that we even have

$$\text{RE}(s = 1) > \text{RE}(s = 2) > \text{RE}(s = 3) > \text{RE}(s = 4),$$

due, of course, to the fact that for $\beta = 0$, $\text{MSFE}(n = \infty) = 1$ for every s . We also learned from fig. 2 that for $\beta \geq 0.50$ the precision of the forecast (as measured by the MSFE) decreases with increasing s . We now see from fig. 3 that for $\beta \geq 0.60$ or 0.70 not only the precision but also the RE of the forecast decreases with increasing s .

Finally, let us compare our numerical results with those reported in the literature. No exact results for this case are available, and the only Monte Carlo results that we could find are those obtained by Lahiri (1975). Lahiri, unaware of Malinvaud's (1970) result that the LS forecast is unbiased,³

³Actually, Malinvaud's remark already appeared in the first French edition of his book published in 1964.

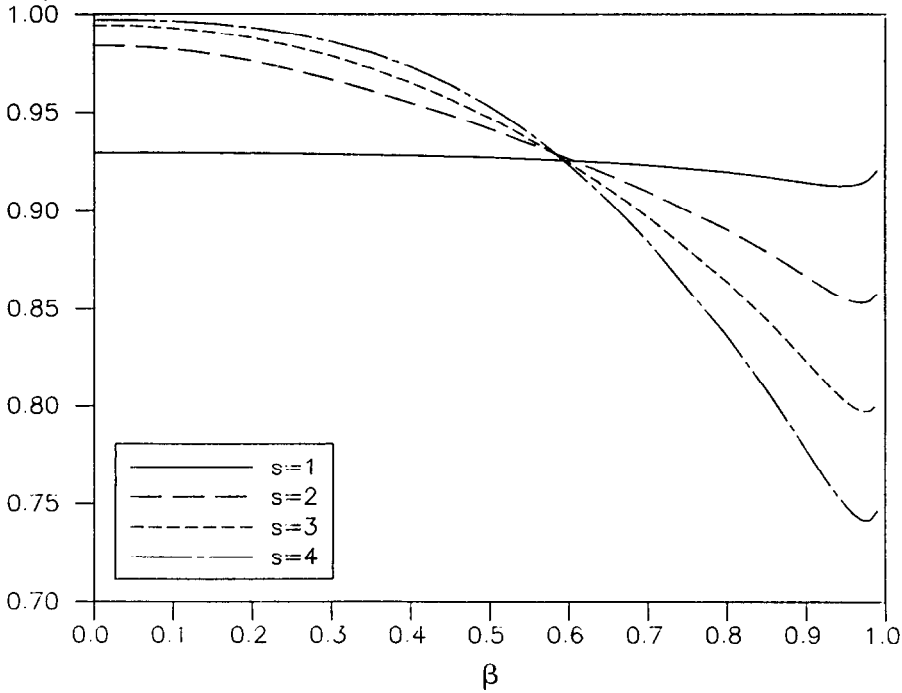


Fig. 3. Relative efficiency of LS forecast for $n = 15$: Stationary case.

calculated the sample bias and the sample MSFE of the LS forecast \hat{y}_{n+s} using one thousand replications for each combination of β and n . The combinations considered are $n = 10, 20, 40$, $\beta = 0.2, 0.4, 0.6, 0.8$, and $s = 1, 2, 3, 4, 5$. His results are quite poor.⁴ For example, he obtains an average absolute sample bias of 0.0374 whereas the true bias is 0. Similar deviations occur with the MSFE, particularly for $n = 10$. Lahiri also reports results based on Chow's (1973) modified Bayesian predictor; these results are even less accurate.

5. Exact results: The non-stationary case

The only difference between the stationary and the non-stationary case is in the specification of the initial observation y_1 . In section 4 we assumed $y_1 = (1 - \beta^2)^{-1/2} u_1$ (Assumption 1a), so that $\{y_t\}$ is a normal strictly sta-

⁴Lahiri was also unaware of the fact that for $n = 10$ and $s = 5$ the MSFE does not exist. See our Theorem 2 in section 3.

Table 2
Exact mean-square forecast error of least-squares forecast \hat{y}_{n+s} : Non-stationary case.

n	s	β											
		0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99
10	1	1.1284	1.1288	1.1299	1.1319	1.1350	1.1397	1.1467	1.1577	1.1752	1.2028	1.2216	1.2398
10	2	1.0498	1.0638	1.1061	1.1774	1.2790	1.4133	1.5847	1.8018	2.0808	2.4490	2.6787	2.8910
10	3	1.0379	1.0531	1.1010	1.1880	1.3258	1.5322	1.8346	2.2765	3.9260	4.6127	5.2871	5.8711
10	4	1.0826	1.0988	1.1507	1.2487	1.4133	1.6806	2.1126	2.8191	4.0017	6.0427	7.5917	9.2056
15	1	1.0759	1.0760	1.0765	1.0773	1.0786	1.0806	1.0837	1.0886	1.0975	1.1150	1.1294	1.1448
15	2	1.0163	1.0286	1.0658	1.1279	1.2156	1.3299	1.4727	1.6480	1.8658	2.1512	2.3342	2.5088
15	3	1.0057	1.0175	1.0548	1.1225	1.2294	1.3885	1.6181	1.9457	2.4175	3.1295	3.6327	4.1417
15	4	1.0029	1.0140	1.0492	1.1148	1.2234	1.3968	1.6724	2.1139	2.8395	4.1000	5.0893	6.1581
20	1	1.0544	1.0545	1.0547	1.0552	1.0559	1.0570	1.0587	1.0615	1.0667	1.0785	1.0900	1.1038
20	2	1.0083	1.0201	1.0553	1.1142	1.1970	1.3042	1.4370	1.5975	1.7920	2.0403	2.2009	2.3580
20	3	1.0020	1.0131	1.0478	1.1108	1.2101	1.3571	1.5676	1.8634	2.2790	2.8891	3.3211	3.7667
20	4	1.0007	1.0112	1.0443	1.1054	1.2052	1.3625	1.6084	1.9945	2.6113	3.6510	4.4672	5.3657
25	1	1.0425	1.0426	1.0427	1.0430	1.0435	1.0442	1.0452	1.0470	1.0505	1.0588	1.0682	1.0807
25	2	1.0051	1.0165	1.0507	1.1078	1.1880	1.2915	1.4192	1.5724	1.7551	1.9826	2.1289	2.2751
25	3	1.0010	1.0117	1.0453	1.1064	1.2023	1.3440	1.5459	1.8274	2.2172	2.7748	3.1662	3.5758
25	4	1.0003	1.0106	1.0430	1.1026	1.1992	1.3503	1.5846	1.9486	2.5200	3.4570	4.1846	4.9954
∞	1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
∞	2	1.0000	1.0100	1.0400	1.0900	1.1600	1.2500	1.3600	1.4900	1.6400	1.8100	1.9025	1.9801
∞	3	1.0000	1.0101	1.0416	1.0981	1.1856	1.3125	1.4896	1.7301	2.0496	2.4661	2.7170	2.9407
∞	4	1.0000	1.0101	1.0417	1.0988	1.1897	1.3281	1.5363	1.8477	2.3117	2.9975	3.4521	3.8822

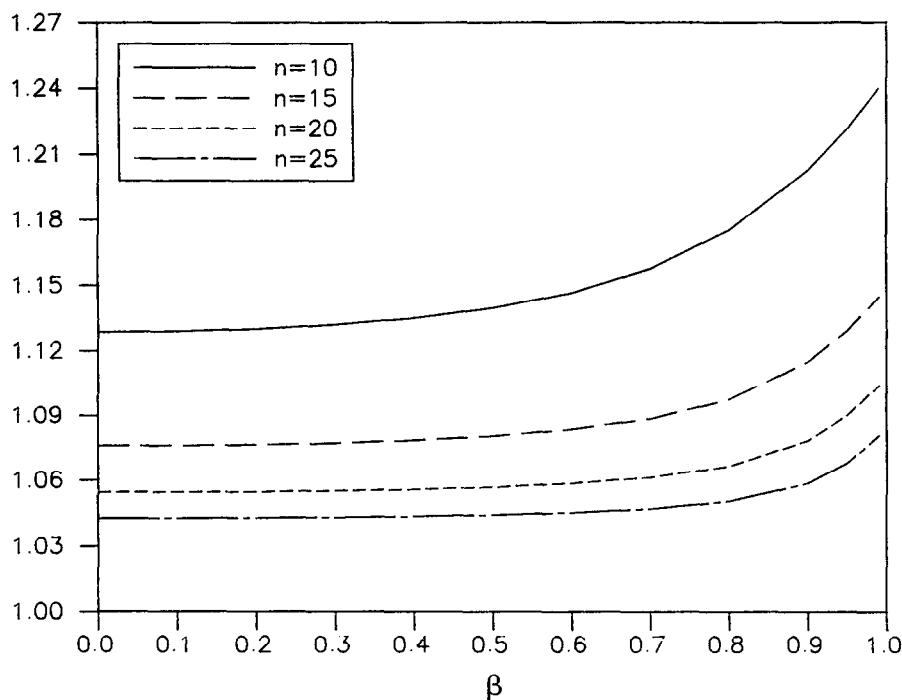


Fig. 4. MSFE of LS forecast for $s = 1$: Non-stationary case.

tionary time series; now we assume $y_1 = u_1$ or, equivalently, $y_0 = 0$ (Assumption 1b) which implies that $\{y_t\}$ is not covariance stationary. No results, either exact or Monte Carlo, for the MSFE of the LS forecast in this case are available in the literature. Our results are presented in table 2.

As in the stationary case we see that the MSFE is a decreasing function of n , and that it is not always an increasing function of s ; in fact, the behaviour of the MSFE in the non-stationary case (considered as a function of s) is identical to that in the stationary case.

There is one important difference, however, between the stationary and the non-stationary process: for every fixed n and s the MSFE in the non-stationary case is an increasing function of $|\beta|$, also for values of $|\beta|$ close to 1. A comparison of fig. 4 and fig. 1 illustrates this point. As to the relative efficiency of the LS forecast as defined in (21), the comments made for the stationary process remain valid for the non-stationary process, except that the RE now decreases with $|\beta|$ also when $|\beta|$ is close to 1. Apart from this one striking difference, the behaviour of the MSFE in the two cases is the same.

It appears that for given β , n and s the MSFE in the non-stationary case is always larger than or equal to the MSFE in the stationary case with equality if

and only if $\beta = 0$ or $n = \infty$. Thus if we denote, for given values of β , n and s , the MSFE in the stationary and the non-stationary case by $MSFE_1$ and $MSFE_2$, respectively, then

$$\phi \equiv MSFE_2/MSFE_1 > 1,$$

unless $\beta = 0$ or $n = \infty$. One would expect that ϕ decreases with n , and indeed this is the case. Also, by comparing tables 1 and 2, we see that ϕ increases with both $|\beta|$ and s . It appears that, for given n and s , ϕ is not only an increasing function of $|\beta|$ but that the increase is very rapid. Hence the difference between the MSFE's in the stationary and the non-stationary case is at its largest when $|\beta|$ is close to 1, s is large and n is small.

Appendix A: A result concerning the expectation of functions of normally distributed random variables

Below we obtain certain expectations using a technique developed by Ullah and Nagar (1974, app. B) and also used, in a different context, by Ullah and Ullah (1978, sec. 4).

Proposition 1. Let y_0 be a normally distributed $n \times 1$ vector with mean μ_0 and positive definite covariance matrix Ω_0 . For every μ in \mathbb{R}^n define

$$y(\mu) = y_0 + \mu - \mu_0,$$

so that $y(\mu_0) = y_0$ and $y(\mu) \sim N(\mu, \Omega_0)$. Let g be a real-valued function defined in an open neighbourhood $B(\mu_0)$ of μ_0 such that, for every $\mu \in B(\mu_0)$, $g(y(\mu))$ is a random variable with finite expectation, say $\theta(\mu)$. If θ is twice differentiable at μ_0 , then

$$Eg(y_0)y_0 = \Omega_0 h_0 + \theta_0 \mu_0,$$

and

$$Eg(y_0)y_0 y_0' = \theta_0(\Omega_0 + \mu_0 \mu_0') + \Omega_0 H_0 \Omega_0 + \Omega_0 h_0 \mu_0' + \mu_0 h_0' \Omega_0,$$

where

$$\theta_0 = \theta(\mu_0), \quad h_0 = \left. \frac{\partial \theta(\mu)}{\partial \mu} \right|_{\mu = \mu_0}, \quad H_0 = \left. \frac{\partial^2 \theta(\mu)}{\partial \mu \partial \mu'} \right|_{\mu = \mu_0}.$$

In particular, if $\mu_0 = 0$,

$$Eg(y_0)y_0 = \Omega_0 h_0,$$

and

$$Eg(y_0)y_0y_0' = \theta_0\Omega_0 + \Omega_0H_0\Omega_0.$$

Proof. Let

$$f(y, \mu) = (2\pi)^{-n/2}|\Omega_0|^{-1/2}\exp -\frac{1}{2}(y - \mu)'\Omega_0^{-1}(y - \mu).$$

Differentiating f twice with respect to μ gives

$$\frac{\partial f(y, \mu)}{\partial \mu} = f(y, \mu)\Omega_0^{-1}(y - \mu), \tag{A.1}$$

and

$$\frac{\partial^2 f(y, \mu)}{\partial \mu \partial \mu'} = f(y, \mu)(\Omega_0^{-1}(y - \mu)(y - \mu)'\Omega_0^{-1} - \Omega_0^{-1}). \tag{A.2}$$

Using (A.1) and the fact that differentiation under the integral sign is permitted we obtain

$$\begin{aligned} Eg(y_0)(y_0 - \mu_0) &= \int_{\mathbf{R}^n} g(y)(y - \mu_0)f(y, \mu_0) \, dy \\ &= \Omega_0 \int g(y) \left[\frac{\partial f(y, \mu)}{\partial \mu} \right]_{\mu=\mu_0} \, dy \\ &= \Omega_0 \left[\frac{\partial}{\partial \mu} \int g(y)f(y, \mu) \, dy \right]_{\mu=\mu_0} \\ &= \Omega_0 \left[\frac{\partial}{\partial \mu} Eg(y) \right]_{\mu=\mu_0} = \Omega_0 h_0. \end{aligned} \tag{A.3}$$

Similarly, using (A.2), we obtain

$$Eg(y_0)(y_0 - \mu_0)(y_0 - \mu_0)' = \theta_0\Omega_0 + \Omega_0H_0\Omega_0. \tag{A.4}$$

The result now follows easily from (A.3), (A.4) and the fact that $Eg(y_0) = \theta_0$. ■

Appendix B: The first and second derivatives (with respect to μ) of $E(y'Ay/y'By)^s$ at $\mu = \mathbf{0}$ when $y \sim N(\mu, \Omega)$

Let $y \sim N_n(\mu, \Omega)$. Proposition 2 gives the first and second derivatives (with respect to μ) of $E(y'Ay/y'By)^s$ at $\mu = \mathbf{0}$ for an arbitrary positive integer s .

Hoque (1985a, app. B) obtained the special cases $s = 1$ and $s = 2$ at an arbitrary point μ in \mathbb{R}^n . The general result for arbitrary s and μ is given in Magnus (1988).

Proposition 2. Let y_0 be a normally distributed $n \times 1$ vector with mean 0 and positive definite covariance matrix $\Omega_0 = LL'$. Let A be a symmetric $n \times n$ matrix, B a positive semidefinite $n \times n$ matrix, $B \neq 0$. For every μ in \mathbb{R}^n define

$$y = y(\mu) = y_0 + \mu,$$

so that $y(0) = y_0$ and $y \sim N(\mu, \Omega_0)$. Let s be a positive integer and g_s a real-valued function defined by

$$g_s(x) = \begin{cases} (x'Ax/x'Bx)^s & \text{if } x \in \mathbb{R}^n - \mathcal{N}(B), \\ = 0 & \text{if } x \in \mathcal{N}(B), \end{cases}$$

where $\mathcal{N}(B)$ denotes the nullspace of B , i.e., the set $\{x \in \mathbb{R}^n, Bx = 0\}$. If, for every μ sufficiently close to zero, $g_s(y(\mu))$ is a random variable with finite expectation, say $\theta_s(\mu)$, then

- (a) θ_s is ∞ times continuously differentiable at $\mu = 0$, and
- (b) the first two derivatives of θ_s at $\mu = 0$ are

$$\left. \frac{\partial \theta_s(\mu)}{\partial \mu} \right|_{\mu=0} = 0,$$

and

$$\left. \frac{\partial^2 \theta_s(\mu)}{\partial \mu \partial \mu'} \right|_{\mu=0} = -\theta_s(0)\Omega_0^{-1} + L'^{-1}PQ_sP'L^{-1},$$

where

$$Q_s = \frac{1}{(s-1)!} \sum_{\nu} \gamma_s(\nu) \int_0^\infty t^{s-1} |\Delta| \left\{ \omega \Delta^2 + 2 \sum_{j=1}^s j n_j r_j \omega_j \Delta R^j \Delta \right\} dt, \tag{B.1}$$

P is an orthogonal $n \times n$ matrix, and Λ a diagonal $n \times n$ matrix such that

$$P'L'BLP = \Lambda, \quad P'P = I_n,$$

the summation in (B.1) is over all $1 \times s$ vectors $\nu = (n_1, n_2, \dots, n_s)$ whose

elements n_j are non-negative integers satisfying $n_1 + 2n_2 + \dots + sn_s = s$,

$$\gamma_s(\nu) = s!2^s \prod_{j=1}^s (n_j!(2j)^{n_j})^{-1},$$

$$\Delta = (I_n + 2tA)^{-1/2}, \quad R = \Delta A^* \Delta,$$

$$\omega = \prod_{i=1}^s (\text{tr } R^i)^{n_i}, \quad \omega_j = \prod_{\substack{i=1 \\ i \neq j}}^s (\text{tr } R^i)^{n_i},$$

and

$$r_j = 1 \quad \text{if } n_j = 0, 1,$$

$$= (\text{tr } R^j)^{n_j-1} \quad \text{if } n_j \geq 2.$$

Remark. The matrix R^j denotes the j th power of the matrix R .

Proof. It follows from Theorem 6 of Magnus (1986) that

$$\theta_s(\mu) = \frac{1}{(s-1)!} \sum_{\nu} \gamma_s(\nu) \int_0^{\infty} t^{s-1} |\Delta| \psi_1(\mu) \psi_2(\mu) dt, \tag{B.2}$$

where

$$\psi_1(\mu) = \exp(-\frac{1}{2}\mu'V\mu), \quad V = \Omega_0^{-1} - L'^{-1}P\Delta^2P'L^{-1},$$

$$\psi_2(\mu) = \prod_{j=1}^s [\phi_j(\mu)]^{n_j}, \quad \phi_j(\mu) = \text{tr } R^j + j\mu'W_j\mu,$$

and

$$W_j = L'^{-1}P\Delta R^j\Delta P'L^{-1}.$$

[Notice that $\psi_2(0) = \omega$.] Since ψ_1 and ψ_2 are ∞ times continuously differentiable at $\mu = 0$, so is θ_s . This proves (a). To prove (b), we differentiate ψ_1 and ψ_2 twice at $\mu = 0$ and use (B.2).

Appendix C: Proof of Theorems 1 and 2

Existence. Let $y = (y_1, \dots, y_n)'$ be an $n \times 1$ vector of observations generated by (1) under initial condition (2). Let A and B be defined as in (5) and note

that $\hat{\beta} = y'Ay/y'By$, as given in (4). Now, let $e = (0, \dots, 0, 1)'$ of order $n \times 1$. Then

$$r(B) = n - 1, \quad Be = 0, \quad e'Ae = 0, \quad Ae \neq 0, \quad e'e \neq 0, \quad (C.1)$$

and

$$y_n = e'y, \quad y_n^2 = y'(ee')y.$$

Following the procedure of Magnus (1986, theorem 7) and using theorem 1 of Kinal (1980), we can then prove that $E(\hat{\beta}^s y_n)$ and $E(\hat{\beta}^s y_n^2)$ exist if and only if $0 \leq s < n - 1$. [See also Magnus (1988, theorems 2 and 3).]

From (8) we see that the expectation of the forecast error exists if and only if $E(\hat{\beta}^s y_n)$ exists, that is, if and only if $0 \leq s < n - 1$. Similarly, from (9), we see that the MSFE exists if and only if $E(\hat{\beta}^{2s} y_n^2)$ and $E(\hat{\beta}^s y_n^2)$ exist, that is, if and only if $0 \leq s < (n - 1)/2$.

Derivation of the MSFE. Since the forecast error, given in (7), is proportional to σ , the MSFE is proportional to σ^2 . It suffices therefore to prove Theorem 2 for $\sigma^2 = 1$. From Magnus (1986, theorem 7) we know that

$$E\left(\frac{(y + \mu)'A(y + \mu)}{(y + \mu)'B(y + \mu)}\right)^k, \quad 1 \leq k \leq n - 2,$$

exists for all $\mu \in \mathbb{R}^n$. Hence we obtain from Propositions 1 and 2,

$$E\hat{\beta}^k y y' = LPQ_k P'l', \quad 1 \leq k \leq n - 2, \quad (C.2)$$

where

$$Q_k = \frac{1}{(k - 1)!} \sum_{\nu} \gamma_k(\nu) \int_0^{\infty} t^{k-1} |\Delta| \left\{ \omega \Delta^2 + 2 \sum_{j=1}^k j n_j r_j \omega_j \Delta R^j \Delta \right\} dt.$$

In particular we obtain

$$E\hat{\beta}^k y_n^2 = l'PQ_k P'l = l^{*'}Q_k l^* = c_k,$$

where c_k is defined in (12). This, together with (9), yields

$$\begin{aligned} E(\hat{y}_{n+s} - y_{n+s})^2 &= \beta^{2s} \text{var}(y_n) + c_{2s} - 2\beta^s c_s + \sum_{j=0}^{s-1} \beta^{2j} \\ &= c_{2s} - 2\beta^s c_s + (1 - K\beta^{2(n+s)})/(1 - \beta^2), \end{aligned}$$

since

$$\text{var}(y_n) = l'l = \sum_{j=0}^{n-2} \beta^{2j} + \delta^2 \beta^{2(n-1)},$$

and K is defined in (11).

References

- Anderson, T.W., 1959, On asymptotic distributions of estimates of parameters of stochastic difference equations, *Annals of Mathematical Statistics* 30, 676–687.
- Ansley, C.F. and P. Newbold, 1980, Finite sample properties of estimators for autoregressive moving average models, *Journal of Econometrics* 13, 159–183.
- Baillie, R.T., 1979, The asymptotic mean squared error of multistep prediction from the regression model with autoregressive errors, *Journal of the American Statistical Association* 74, 175–184.
- Chow, G.C., 1973, Multiperiod predictions from stochastic difference equations by Bayesian methods, *Econometrica* 41, 109–118.
- Copas, J.B., 1966, Monte Carlo results for estimation in a stable Markov time series, *Journal of the Royal Statistical Society A* 129, 110–116.
- Cryer, J.D., J.C. Nankervis and N.E. Savin, 1986, Mirror-image and invariant distributions in ARMA models, Submitted for publication.
- Davissou, L.D., 1965, The prediction error of stationary Gaussian time series of unknown covariance, *IEEE Transactions on Information Theory* 11, 527–532.
- De Gooijer, J.G., 1980, Exact moments of the sample autocorrelations from series generated by general ARIMA processes of order (p, d, q) , $d = 0$ or 1, *Journal of Econometrics* 14, 365–379.
- Dickey, D.A. and W.A. Fuller, 1979, Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association* 74, 427–431.
- Dufour, J.-M., 1984, Unbiasedness of predictions from estimated autoregressions when the true order is unknown, *Econometrica* 52, 209–215.
- Dufour, J.-M., 1985, Unbiasedness of predictions from estimated vector autoregressions, *Econometric Theory* 1, 387–402.
- Evans, G.B.A. and N.E. Savin, 1981, Testing for unit roots: 1, *Econometrica* 49, 753–779.
- Fuller, W.A. and D.P. Hasza, 1980, Predictors for the first-order autoregressive process, *Journal of Econometrics* 13, 139–157.
- Fuller, W.A. and D.P. Hasza, 1981, Properties of predictors for autoregressive time series, *Journal of the American Statistical Association* 76, 155–161.
- Gonedes, N.J. and H.V. Roberts, 1977, Differencing of random walks and near random walks, *Journal of Econometrics* 6, 289–308.
- Hoque, A., 1985a, The exact moments of forecast error in the general dynamic model, *Sankhya* 47 B, 128–143.
- Hoque, A., 1985b, Finite sample analysis of the first order autoregressive model, *Calcutta Statistical Association Bulletin* 34, 51–63.
- Hoque, A. and T.A. Peters, 1986, Finite sample analysis of the ARMAX models, *Sankhya* 48 B, 266–283.
- Hurwicz, L., 1950, Least squares bias in time series, in: T.C. Koopmans, ed., *Statistical inference in dynamic economic models*, Cowles Commission monograph 10 (Wiley, New York).
- Kinal, T.W., 1980, The existence of moments of k -class estimators, *Econometrica* 48, 241–249.
- Lahiri, K., 1975, Multiperiod predictions in dynamic models, *International Economic Review* 16, 699–711.
- Lütkepohl, H., 1984, Forecasting contemporaneously aggregated vector ARMA processes, *Journal of Business and Economic Statistics* 2, 201–214.
- Magnus, J.R., 1986, The exact moments of a ratio of quadratic forms in normal variables, *Annales d'Économie et de Statistique* 4, 95–109.

- Magnus, J.R., 1988, On certain moments relating to ratios of quadratic forms in normal variables: Further results, Submitted for publication.
- Malinvaud, E., 1970, *Statistical methods of econometrics*, 2nd ed. (North-Holland, Amsterdam).
- Nankervis, J.C. and N.E. Savin, 1988, The exact moments of the least-squares estimator for the autoregressive model: Corrections and extensions, *Journal of Econometrics* 37, 381–388.
- Numerical Algorithms Group, 1984, *NAG Fortran Library Manual*, Mark 11, Oxford, NAGLIB.
- Orcutt, G.H. and H.S. Winokur, 1969, First order autoregression: Inference, estimation, and prediction, *Econometrica* 37, 1–14.
- Phillips, P.C.B., 1977, Approximations to some finite sample distributions associated with a first-order stochastic difference equation, *Econometrica* 45, 463–485.
- Phillips, P.C.B., 1979, The sampling distribution of forecasts from a first-order autoregression, *Journal of Econometrics* 9, 241–261.
- Sawa, T., 1978, The exact moments of the least squares estimator for the autoregressive model, *Journal of Econometrics* 8, 159–172.
- Tanaka, K., 1983, Asymptotic expansions associated with the AR(1) model with unknown mean, *Econometrica* 51, 1221–1231.
- Tanaka, K. and K. Maekawa, 1984, The sampling distributions of the predictor for an autoregressive model under misspecifications, *Journal of Econometrics* 25, 327–351.
- Thorner, H., 1967, Finite sample Monte Carlo studies: An autoregressive illustration, *Journal of the American Statistical Association* 62, 801–818.
- Ullah, A. and A.L. Nagar, 1974, The exact mean of the two-stage least squares estimator of the structural parameters in an equation having three endogenous variables, *Econometrica* 42, 749–758.
- Ullah, A. and S. Ullah, 1978, Double k -class estimators of coefficients in linear regression, *Econometrica* 46, 705–722.
- White, J.S., 1961, Asymptotic expansions for the mean and variance of the serial correlation coefficient, *Biometrika* 48, 85–94.
- Yamamoto, T., 1976, Asymptotic mean square prediction error for an autoregressive model, *Applied Statistics* 25, 123–127.