Journal of Econometrics 42 (1989) 157–179. North-Holland

THE EXACT MULTI-PERIOD MEAN-SQUARE FORECAST ERROR FOR THE FIRST-ORDER AUTOREGRESSIVE MODEL WITH AN INTERCEPT

Jan R. MAGNUS

London School of Economics, London WC2A 2AE, England

Bahram PESARAN

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

Received April 1988, final version received October 1988

We study the exact finite-sample behaviour of the mean-square forecast error (*MSFE*) of multi-period least-squares forecasts in the normal autoregressive model $y_t = \alpha + \beta y_{t-1} + u_t$. We obtain necessary and sufficient conditions for the existence of the *MSFE* and give an exact expression which we use to obtain numerical results for both the stationary and the fixed start-up model. We conclude, inter alia, that the behaviour of the *MSFE* in the model with intercept can be very different from that in the model without intercept, especially when β is close to unity.

1. Introduction

This paper is a further attempt to analyze the mean-square forecast error (*MSFE*) of multi-period least-squares forecasts in dynamic models. While Hoque, Magnus, and Pesaran (1988) considered the *MSFE* arising from a first-order autoregression without intercept,

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

where $|\beta| < 1$ and $\{u_t\}$ is a sequence of independent and identically distributed $\mathrm{N}(0, \sigma^2)$ random variables, we now go one step further and study the model with an intercept,

$$y_t = \alpha + \beta y_{t-1} + u_t. \tag{2}$$

We shall see that the behaviour of the *MSFE* in the AR(1) model with an intercept can be very different from that in the model without an intercept, and that this is particularly true for β close to unity.

In the no-intercept model (1), let $\hat{\beta}$ denote the least-squares (LS) estimator of β based on n observations $y_1, \dots, y_n$. The s-periods-ahead LS forecast is then $\hat{y}_{n+s} = \hat{\beta}^s y_n$. Malinvaud (1970) showed that the forecast bias $\mathrm{E}(\hat{y}_{n+s} - y_{n+s})$ vanishes, if it exists. Hoque, Magnus, and Pesaran (1988) showed that

0304-4076/89/$3.50©1989, Elsevier Science Publishers B.V. (North-Holland)

the forecast bias exists if and only if $s \leq n-2$. They also showed that the *MSFE* exists if and only if $s \leq [(n-2)/2]$, and they obtained an exact expression for the *MSFE* which they used to obtain exact results by numerical integration. Their two main conclusions were (i) that for $|\beta|$ close to 1 the *MSFE* is very sensitive to the specification of the initial observation and (ii) that it is not generally true that the *MSFE* is an increasing function of s and that, indeed, for β close to zero the *MSFE* is *decreasing* in s.

The latter result is counterintuitive, so it may be useful to provide some analytical evidence in addition to the exact results. First, we may write the asymptotic approximation up to order n^{-1} as

$$MSFE = \sigma^2\left(\frac{1-\beta^{2s}}{1-\beta^2} + \frac{1}{n-1}s^2\beta^{2(s-1)}\right)$$

[see, e.g., Maekawa (1987)], and one verifies easily that the same counterintuitive phenomenon occurs here. Secondly, at $\beta = 0$, we have

$$MSFE = \sigma^2\left(1 + \mathrm{E}\left(\hat{\beta}^s y_n\right)^2\right). \tag{3}$$

If we consider (3) for a moment as a continuous function of s, then

$$\frac{\mathrm{d}\,MSFE}{\mathrm{d}s} = \sigma^2 \mathrm{E}\left(\hat{\beta}^s y_n\right)^2 \log \hat{\beta}^2,$$

and this expectation is negative because $\log\hat{\beta}^2$ is negative with probability almost equal to one.[1]

In the model (2) with an intercept, Fuller and Hasza (1980) proved that the forecast bias is zero if it exists and if the process is mean-stationary. Magnus and Pesaran (1987) showed that the forecast bias in this case exists if and only if $s \leq n-3$ and they provided an exact expression for the forecast bias which is applicable whether the process is mean-stationary or not. The results for the not mean-stationary case show that the forecast bias is not, in general, a monotone function of either β, n, or s, and that the covariance-stationarity of the process is only relevant when β is close to -1.

In this paper we study the *MSFE* in the model with an intercept (2). We show that the *MSFE* exists if and only if $s \leq [(n-3)/2]$ and provide an exact expression for the *MSFE* in this case. The exact results are compared with the asymptotic approximation and with Monte Carlo results obtained by Orcutt and Winokur (1969) and Fuller and Hasza (1980). The paper highlights the importance of the specification of the initial observation. If $\beta > 0$, then we find

[1] There is a small but positive probability that $|\hat{\beta}| > 1$ even though $|\beta| < 1$. This probability is particularly small when the true β is zero. The exception ($n = 10$, $s = 4$) to the monotonic behaviour of the *MSFE* at $\beta = 0$ noted in Hoque, Magnus, and Pesaran (1988, section 4 and table 1) is explained by the fact that this case lies on the boundary of the existence region of the *MSFE*.

it is important to know whether the process is mean-stationary or not, but almost irrelevant to have information about the covariance-stationarity of the process. If $\beta < 0$, the opposite is true. Then it is important to know whether the process is covariance-stationary or not, but almost useless to know whether the process is mean-stationary or not.

In section 2 we present the model and the precise assumptions concerning the initial observation. The main theorem is stated in section 3. Sections 4 and 5 discuss the exact results. An appendix containing the proof of the theorem concludes the paper.

2. The model

We shall be exclusively concerned with the first-order autoregressive process $\{y_1, y_2, \ldots\}$ with an intercept term,

$$y_t = \alpha + \beta y_{t-1} + u_t, \qquad t = 2, 3, \ldots, \tag{4}$$

where both α and β are unknown and $\{u_2, u_3, \ldots\}$ is a sequence of i.i.d. $N(0, \sigma^2)$ random variables. Regarding the initial observation y_1 we postulate

$$y_1 = \mu_1 + \delta u_1, \tag{5}$$

where $u_1 \sim N(0, \sigma^2)$ is independent of $u_2, u_3, \ldots$, and $\delta > 0$. In finite samples the actual values of μ_1 and δ are important and we shall return to their specification shortly.

Let $y = (y_1, y_2, \ldots, y_n)'$ be an $n \times 1$ vector of observations generated by (4) and (5). Then y is normally distributed with mean

$$\mu = \frac{\alpha}{1-\beta}\begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + \left(\mu_1 - \frac{\alpha}{1-\beta}\right)\begin{pmatrix} 1 \\ \beta \\ \beta^2 \\ \vdots \\ \beta^{n-1} \end{pmatrix} \tag{6}$$

and positive definite covariance matrix LL', where

$$L = \sigma \begin{pmatrix} \delta & 0 & 0 & \cdots & 0 & 0 \\ \delta\beta & 1 & 0 & \cdots & 0 & 0 \\ \delta\beta^2 & \beta & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \delta\beta^{n-2} & \beta^{n-3} & \beta^{n-4} & \cdots & 1 & 0 \\ \delta\beta^{n-1} & \beta^{n-2} & \beta^{n-3} & \cdots & \beta & 1 \end{pmatrix}. \tag{7}$$

Given the vector of observations $y=(y_1, y_2, \ldots, y_n)'$, the least-squares estimators of α and β, obtained by minimizing $\sum_{t=2}^{n}(y_t-\alpha-\beta y_{t-1})^2$, are given by

$$\hat{\alpha}=\bar{y}_{**}-\hat{\beta}\bar{y}_{*} \tag{8}$$

and

$$\hat{\beta}=\sum_{t=2}^{n}(y_t-\bar{y}_{**})(y_{t-1}-\bar{y}_{*})\Big/\sum_{t=2}^{n}(y_{t-1}-\bar{y}_{*})^2, \tag{9}$$

where

$$\bar{y}_{*}=(1/(n-1))\sum_{t=2}^{n}y_{t-1}, \qquad \bar{y}_{**}=(1/(n-1))\sum_{t=2}^{n}y_t.$$

Defining the $n\times n$ matrices

$$A=\tfrac{1}{2}(D_{*}'MD_{**}+D_{**}'MD_{*}), \qquad B=D_{*}'MD_{*}, \tag{10}$$

where M is the idempotent $(n-1)\times(n-1)$ matrix

$$M=I_{n-1}-(1/(n-1))ii', \qquad i=(1,1,\ldots,1)',$$

and D_{*} and D_{**} are the $(n-1)\times n$ selection matrices

$$D_{*}=(I_{n-1}:0), \qquad D_{**}=(0:I_{n-1}),$$

we can write the least-squares estimator $\hat{\beta}$ in (9) as a ratio of two quadratic forms in normal variables,

$$\hat{\beta}=y'Ay/y'By.$$

The s-periods-ahead forecast is defined recursively as

$$\hat{y}_{n+1}=\hat{\alpha}+\hat{\beta}y_n,$$

$$\hat{y}_{n+s}=\hat{\alpha}+\hat{\beta}\hat{y}_{n+s-1}, \qquad s=2,3,\ldots,$$

so that

$$\hat{y}_{n+s}=\hat{\alpha}\sum_{j=0}^{s-1}\hat{\beta}^j+\hat{\beta}^s y_n, \qquad s=1,2,\ldots. \tag{11}$$

From (4) and (11) we obtain the forecast error

$$\hat{y}_{n+s}-y_{n+s}=\hat{\alpha}\sum_{j=0}^{s-1}\hat{\beta}^j+(\hat{\beta}^s-\beta^s)y_n-\alpha\sum_{j=0}^{s-1}\beta^j-\sum_{j=0}^{s-1}\beta^j u_{n+s-j}, \tag{12}$$

and the expected value of $(\hat{y}_{n+s} - y_{n+s})^2$, if it exists, is the mean-square forecast error of the forecast $\hat{y}_{n+s}$. In the next section we shall obtain an exact expression for the *MSFE* and show that it exists if and only if $s \leq [(n-3)/2]$.

Let us now return to the specification of μ_1 and δ. The following choices seem the most natural ones to consider. As to the mean component μ_1 we assume either

Assumption 1a. $\mu_1 = \alpha/(1-\beta)$,

in which case $\{y_1, y_2, \ldots\}$ is mean-stationary ($\mathrm{E}y_t = \mu_1$, $t = 1, 2, \ldots$), or

Assumption 1b. $\mu_1 = \alpha$,

which, if we assume that (4) also holds for $t = 1$, is equivalent to assuming that y_0 is distributed symmetrically about zero, but implies that the process is not mean-stationary (except when $\beta = 0$).

As to the variance component δ, we assume either

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

in which case the series $\{y_1, y_2, \ldots\}$ is covariance-stationary ($\mathrm{cov}(y_{s+t}, y_s)$ is independent of s), or

Assumption 2b. $\delta = 1$,

which is equivalent to assuming that y_0 is a nonrandom constant, but implies of course that the process is not covariance-stationary (except again when $\beta = 0$).

If Assumptions 1a and 2a are both satisfied, then $\{y_1, y_2, \ldots\}$ is a normal strictly stationary time series; if Assumptions 1b and 2b are both satisfied, then $y_0 = 0$.

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

In order to obtain an exact expression for the mean-square forecast error (*MSFE*), $\mathrm{E}(\hat{y}_{n+s} - y_{n+s})^2$, we need some additional notation.

Let A and B be the $n \times n$ matrices ($n \geq 5$) defined in (10) and let μ and L be the $n \times 1$ vector and $n \times n$ matrix defined in (6) and (7), respectively. Let P be an orthogonal $n \times n$ matrix and Λ a diagonal $n \times n$ matrix such that

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

and define the $n \times n$ matrix and $n \times 1$ vectors

$$A^* = P'L'ALP, \qquad \mu^* = P'L^{-1}\mu,$$

$$w_1 = \frac{1}{n-1}(-1,0,\dots,0,1)',$$

$$w_2 = \frac{1}{n-1}(-1,-1,\dots,-1,n-1)',$$

$$w_3 = \frac{1}{n-1}(1,1,\dots,1,0)'.$$

For $i, j = 1,2,3$, let $W_{ij} = \frac{1}{2}(w_i w_j' + w_j w_i')$ and denote by ε_s the bias of the forecast $\hat{y}_{n+s}$, that is, $\varepsilon_s = \mathrm{E}(\hat{y}_{n+s} - y_{n+s})$.

Let $\{C_{k,s}, 1 \le k \le 2s\}$ be symmetric $n \times n$ matrices defined by

$$C_{k,s} = \begin{cases} (k+1)W_{11} - 2\beta^s W_{12} + 2(1-\beta^s)W_{13} & \text{if } k = 1,\dots,s-1, \\ (s-1)W_{11} + 2W_{12} - 2\beta^s W_{22} + 2(1-\beta^s)W_{23} & \text{if } k = s, \\ (2s-k-1)W_{11} + 2W_{12} & \text{if } k = s+1,\dots,2s-1, \\ W_{22} & \text{if } k = 2s, \end{cases}$$

and, for any symmetric $n \times n$ matrix C, let

$$\zeta_k[C] = \frac{\exp(-\frac{1}{2}\mu^{*\prime}\mu^*)}{(k-1)!} \sum_{\nu} \gamma_k(\nu) \int_0^\infty t^{k-1} |\Delta| \exp(\tfrac{1}{2}\xi'\xi) g_k(t)\,\mathrm{d}t, \tag{13}$$

where

$$\begin{aligned} g_k(t) = {} & \omega(\operatorname{tr}\Gamma + \xi'\Gamma\xi) + 2\sum_{j=1}^{k} j n_j r_j \omega_j (\operatorname{tr} R^j\Gamma + 2\xi' R^j \Gamma \xi) \\ & + 4\sum_{j=1}^{k} j^2 n_j (n_j - 1) r_j^* \omega_j (\xi' R^j \Gamma R^j \xi) \\ & + 8\sum_{i<j} \omega_{ij} (i n_i r_i)(j n_j r_j)(\xi' R^i \Gamma R^j \xi). \end{aligned}$$

The summation in (13) is over all $1 \times k$ vectors $\nu = (n_1, n_2, \dots, n_k)$ whose elements n_j are nonnegative integers satisfying $\sum_{j=1}^{k} j n_j = k$. Further,

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

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

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

$$\Gamma = \Delta P' L' C L P \Delta, \qquad \xi = \Delta \mu^*,$$

and the scalars ω, ω_j, ω_{ij}, r_j, and r_j^* are defined by

$$\omega = \prod_{j=1}^{k} \varphi_j^{n_j}, \qquad \varphi_j = \operatorname{tr} R^j + j\xi' R^j \xi,$$

$$\omega_j = \begin{cases} 1 & \text{if} \quad k = 1, \\ \prod_{\substack{i=1 \\ i \neq j}}^{k} \varphi_i^{n_i} & \text{if} \quad k \geq 2, \end{cases}$$

$$\omega_{ij} = \begin{cases} 1 & \text{if} \quad k = 1, \\ \prod_{\substack{h=1 \\ h \notin \{i,j\}}}^{k} \varphi_h^{n_h} & \text{if} \quad k \geq 2, \end{cases}$$

$$r_j = \begin{cases} 1 & \text{if} \quad n_j = 0, 1, \\ \varphi_j^{n_j - 1} & \text{if} \quad n_j \geq 2, \end{cases}$$

$$r_j^* = \begin{cases} 1 & \text{if} \quad n_j = 0, 1, 2, \\ \varphi_j^{n_j - 2} & \text{if} \quad n_j \geq 3. \end{cases}$$

With all notation explained we can now state Theorem 1.

Theorem 1. The mean-square forecast error of $\hat{y}_{n+s}$ exists if and only if $1 \leq s \leq [(n-3)/2]$, in which case

$$\begin{aligned} \mathrm{E}(\hat{y}_{n+s} - y_{n+s})^2 = {} & \sigma^2 \sum_{k=0}^{s-1} \beta^{2k} - \alpha^2 \left(\sum_{k=0}^{s-1} \beta^k \right)^2 - 2\alpha\varepsilon_s \sum_{k=0}^{s-1} \beta^k \\ & + \left(w_1 - \beta^s w_2 + (1 - \beta^s) w_3\right)' \\ & \times (LL' + \mu\mu')\left(w_1 - \beta^s w_2 + (1 - \beta^s) w_3\right) \\ & + \sum_{k=1}^{2s} \zeta_k [C_{k,s}]. \end{aligned} \tag{14}$$

Proof. See appendix.

4. Exact results: The mean-stationary case

Let us begin by noticing that, as in the model without an intercept [see Hoque, Magnus, and Pesaran (1988)], the mean-square forecast error (*MSFE*) satisfies

$$MSFE \geq \sigma^2 \quad \text{and} \quad \lim_{n \to \infty} MSFE = \sigma^2(1-\beta^{2s})/(1-\beta^2),$$

and that these results are independent of assumptions regarding the initial observation, that is, independent of μ_1, α, and δ. However, unlike the model without intercept we now have that the *MSFE* is *not* proportional to σ^2 (unless the process is mean-stationary) and that the *MSFE* is *not* an even function of β.

The exact *MSFE* of the least-squares (LS) forecast $\hat{y}_{n+s}$ was calculated[2] using Theorem 1 under Assumptions 1a or 1b and 2a or 2b for the following selected values of β, n (the number of observations), and s (the number of periods ahead):

$$\beta = -0.99, -0.98, -0.95, -0.90, -0.80, \ldots,$$
$$\times 0.80, 0.90, 0.95, 0.98, 0.99$$
$$n = 10, 15, 20,$$
$$s = 1, 2, 3, 4.$$

In this section we shall discuss the results for the mean-stationary case (Assumption 1a) where $\mu_1 = \alpha/(1-\beta)$. The results for the not mean-stationary case (Assumption 1b) are discussed in the following section. If the process is mean-stationary we know that (i) the LS forecast is unbiased, (ii) the *MSFE* is independent of the values of α and μ_1, and (iii) the *MSFE* is proportional to σ^2. The first property follows from Fuller and Hasza (1980) and Magnus and Pesaran (1987), while properties (ii) and (iii) follow from the fact that, under Assumption 1a, eq. (4) can be written as

$$y_t - \mathrm{E}y_t = \beta(y_{t-1} - \mathrm{E}y_{t-1}) + u_t.$$

Hence in the mean-stationary case there is no loss of generality by assuming $\alpha = \mu_1 = 0$ and $\sigma^2 = 1$. Tables 1 and 2 contain the exact numerical results for this case under Assumptions 2a (table 1) and 2b (table 2), respectively.

Let us first discuss the results for the *strictly stationary* case (table 1) where $\mu_1 = \alpha/(1-\beta)$ and $\delta = (1-\beta^2)^{-1/2}$. Fig. 1 illustrates as expected that the

[2] Application of Theorem 1 involves numerical integration. We used the Numerical Algorithms Group (1984) (the so-called NAG) subroutine D01AMF 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 Λ and P were calculated using the NAG subroutine F02ABF.

Table 1

Exact *MSFE* of least-squares forecast $\hat{y}_{n+s}$: $\alpha = \mu_1 = 0$, $\delta = (1-\beta^2)^{-1/2}$, $\sigma = 1$.

		β												
n	s	−0.99	−0.90	−0.80	−0.60	−0.40	−0.20	0.00	0.20	0.40	0.60	0.80	0.90	0.99
10	1	1.2701	1.2848	1.2774	1.2682	1.2664	1.2698	1.2781	1.2922	1.3144	1.3468	1.3781	1.3740	1.3370
10	2	2.5232	2.3386	2.0616	1.6353	1.3601	1.2122	1.1845	1.2800	1.5107	1.8977	2.4297	2.6861	2.8174
10	3	4.4330	3.8255	3.0456	2.0246	1.4936	1.2541	1.2093	1.3359	1.6846	2.3918	3.6201	4.3631	4.8912
15	1	1.1616	1.1703	1.1629	1.1562	1.1550	1.1564	1.1599	1.1661	1.1763	1.1939	1.2224	1.2318	1.2102
15	2	2.3102	2.0936	1.8514	1.4903	1.2518	1.1198	1.0908	1.1658	1.3505	1.6599	2.1202	2.3835	2.5349
15	3	3.7687	3.1030	2.4672	1.6856	1.2895	1.1152	1.0822	1.1674	1.4032	1.8909	2.8233	3.4794	3.9781
15	4	5.2487	3.9279	2.8337	1.7172	1.2701	1.1033	1.0770	1.1641	1.4174	2.0172	3.4152	4.5789	5.6157
20	1	1.1208	1.1200	1.1149	1.1111	1.1105	1.1113	1.1133	1.1169	1.1228	1.1336	1.1551	1.1683	1.1560
20	2	2.2222	2.0045	1.7829	1.4474	1.2208	1.0929	1.0619	1.1285	1.2962	1.5746	1.9914	2.2526	2.4229
20	3	3.5393	2.8797	2.3132	1.6098	1.2487	1.0888	1.0563	1.1290	1.3357	1.7642	2.5950	3.2276	3.7503
20	4	4.8320	3.5610	2.6089	1.6329	1.2346	1.0818	1.0539	1.1270	1.3439	1.8515	3.0467	4.1204	5.1518
∞	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	1.0000
∞	2	1.9801	1.8100	1.6400	1.3600	1.1600	1.0400	1.0000	1.0400	1.1600	1.3600	1.6400	1.8100	1.9801
∞	3	2.9407	2.4661	2.0496	1.4896	1.1856	1.0416	1.0000	1.0416	1.1856	1.4896	2.0496	2.4661	2.9407
∞	4	3.8822	2.9975	2.3117	1.5363	1.1897	1.0417	1.0000	1.0417	1.1897	1.5363	2.3117	2.9975	3.8822

Table 2

Exact *MSFE* of least-squares forecast $\hat{y}_{n+s}$: $\alpha = \mu_1 = 0$, $\delta = \sigma = 1$.

		β												
n	s	−0.99	−0.90	−0.80	−0.60	−0.40	−0.20	0.00	0.20	0.40	0.60	0.80	0.90	0.99
10	1	1.3699	1.3311	1.3035	1.2774	1.2695	1.2705	1.2781	1.2929	1.3170	1.3516	1.3783	1.3636	1.3319
10	2	2.9611	2.5159	2.1465	1.6566	1.3653	1.2129	1.1845	1.2804	1.5110	1.8916	2.3835	2.6022	2.7904
10	3	5.7928	4.3458	3.2766	2.0752	1.5041	1.2555	1.2093	1.3368	1.6870	2.3818	3.4974	4.1334	4.8183
15	1	1.2263	1.1933	1.1735	1.1594	1.1560	1.1566	1.1599	1.1663	1.1774	1.1967	1.2259	1.2298	1.2060
15	2	2.5189	2.1608	1.8781	1.4958	1.2530	1.1200	1.0908	1.1658	1.3506	1.6596	2.1096	2.3490	2.5132
15	3	4.3082	3.2606	2.5229	1.6945	1.2909	1.1153	1.0822	1.1675	1.4034	1.8895	2.7928	3.3871	3.9244
15	4	6.3215	4.2160	2.9251	1.7287	1.2715	1.1034	1.0770	1.1641	1.4176	2.0152	3.3592	4.4004	5.5120
20	1	1.1578	1.1320	1.1203	1.1128	1.1110	1.1114	1.1133	1.1170	1.1234	1.1353	1.1584	1.1692	1.1523
20	2	2.3566	2.0409	1.7958	1.4499	1.2213	1.0930	1.0619	1.1286	1.2963	1.5750	1.9898	2.2380	2.4042
20	3	3.8419	2.9544	2.3365	1.6132	1.2492	1.0888	1.0563	1.1290	1.3357	1.7640	2.5854	3.1828	3.7041
20	4	5.4091	3.6879	2.6437	1.6367	1.2350	1.0818	1.0539	1.1270	1.3439	1.8510	3.0279	4.0328	5.0647

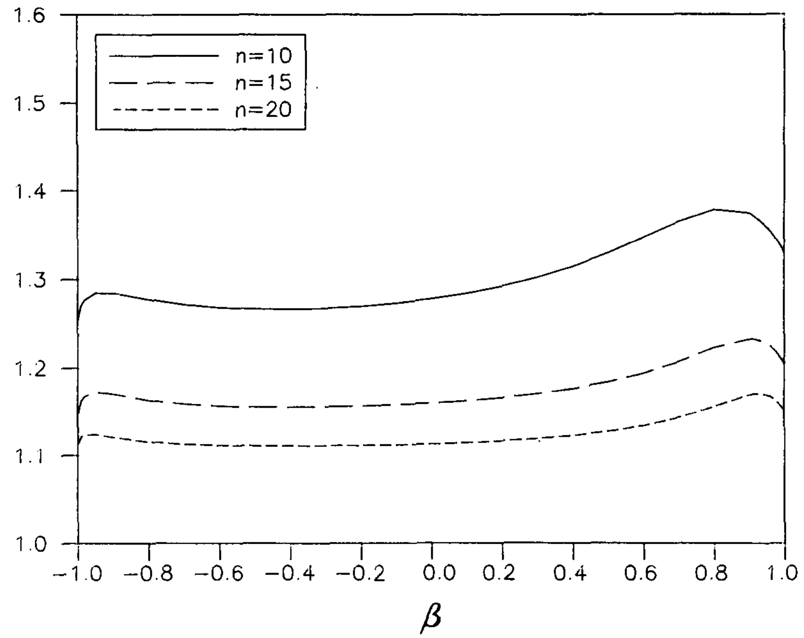

Fig. 1. *MSFE* of LS forecast for $s = 1$: strictly stationary case.

MSFE decreases with increasing n. (This is true for all $n \geq 2s + 3$; if $n < 2s + 3$, then the *MSFE* does not exist, see Theorem 1.) It also shows that the *MSFE* is not a monotone function of $|\beta|$, at least when $s = 1$, since the *MSFE* falls when β approaches ± 1. The drop is particularly significant for β close to $+1$. Finally, fig. 1 shows that the *MSFE* is not symmetric about $\beta = 0$. In the strictly stationary case we always have

$$MSFE(|\beta|) \geq MSFE(-|\beta|),$$

but we shall see that this inequality no longer holds if different assumptions are made regarding the initial observation.

For $s \geq 2$ these conclusions remain unaltered except that the *MSFE* is now monotone in $|\beta|$. However, as in the model without intercept, the increase in the *MSFE* is much less for values of $|\beta|$ close to 1, especially when n is small.

The asymptotic approximation of the *MSFE* up to order n^{-1} is [Fuller and Hasza (1981, p. 157)]

$$MSFE = \sigma^2 \left\{ \frac{1 - \beta^{2s}}{1 - \beta^2} + \frac{1}{n-1} \left(s^2 \beta^{2(s-1)} + \left(\frac{1 - \beta^s}{1 - \beta} \right)^2 \right) \right\}.$$

In particular, for $s = 1$,

$$MSFE = \sigma^2\left(1 + \frac{2}{n-1}\right).$$

The agreement between our exact results and the asymptotic approximation is quite good, but of course the interesting behaviour of the *MSFE* for $|\beta|$ close to 1 is not captured by the approximation.

In case the true value of α is zero *and we know this*, then we estimate of course the no-intercept model $y_t = \beta y_{t-1} + u_t$. But if $\alpha = 0$ *and we don't know this*, then we estimate the model with an intercept. This ignorance about the truth will lead to a higher *MSFE*. A comparison of table 1 with the corresponding table in Hoque, Magnus, and Pesaran (1988) confirms this. The asymptotic approximations show that the increase in the *MSFE* up to order n^{-1} will be

$$\frac{1}{n-1}\left(\frac{1-\beta^s}{1-\beta}\right)^2,$$

and this agrees quite well with our exact results.

The most striking result found in Hoque, Magnus, and Pesaran (1988) was that in the no-intercept model and for β close to zero the *MSFE* decreases when s increases. Some analytic evidence for this counterintuitive result was given in section 1 of this paper. Fig. 2 illustrates that, in the model with an intercept, the situation is *almost* the same; almost, because for $n = 10$, *MSFE* $(s = 3) >$ *MSFE* $(s = 2)$ when β is close to zero. It is easy to see analytically why the *MSFE* is not now strictly decreasing in s for all n, since, at $\alpha = \beta = \mu_1 = 0$,

$$\frac{\mathrm{d}MSFE}{\mathrm{d}s} = \mathrm{E}\left[\left(\hat{\beta}^s y_n + \frac{\hat{\alpha}(1-2\hat{\beta}^s)}{2(1-\hat{\beta})}\right)^2 - \left(\frac{\hat{\alpha}}{2(1-\hat{\beta})}\right)^2\right]\log\hat{\beta}^2,$$

and the sign of the derivative is not unambiguous.

Some Monte Carlo results for the strictly stationary case were obtained by Fuller and Hasza (1980, table 1). These results are reasonably accurate for values of $|\beta|$ not too close to 1. For $|\beta|$ close to 1, however, the Monte Carlo results are quite poor. (This is a consequence of their method, where the same number of replications was used for each β, while for $|\beta|$ close to 1 many more replications are needed to obtain the same accuracy.) For example, for $n = 10$, $\beta = 0.99$, and $s = 1, 2, 3$, Fuller and Hasza find 1.31, 2.70, and 4.46, whereas the exact results are 1.34, 2.82, and 4.89.

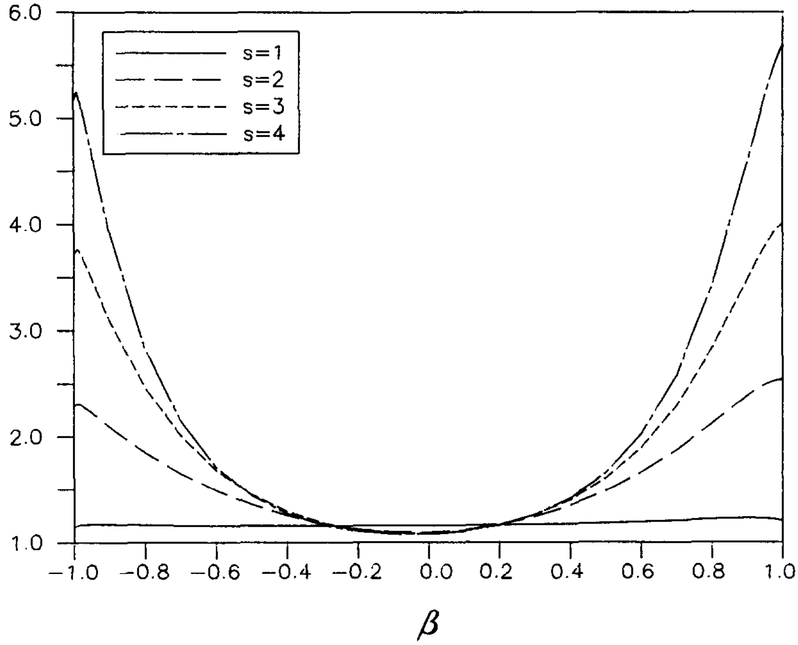

Fig. 2. *MSFE* of LS forecast for $n = 15$: strictly stationary case.

The strictly stationary case is based on Assumptions 1a and 2a. Let us now consider what happens if Assumptions 1a and 2b hold. This is a fixed start-up case with $y_0 = \alpha/(1-\beta)$ so that the process is mean-stationary but not covariance-stationary. The results are given in table 2. Comparing tables 1 and 2 we see that the *MSFE* does not seem to be affected by the choice of δ if β is positive, but that there is a big difference in *MSFE* if β is close to -1. Fig. 3 illustrates this. For $\beta < 0$ the *MSFE* is now a monotone function of β and it is larger than in the covariance-stationary case. It is interesting that exactly the same kind of result was found in Magnus and Pesaran (1987) for the bias of the forecast error in the not mean-stationary process. For a detailed analysis and explanation of this phenomenon, see Magnus and Rothenberg (1988).

Some Monte Carlo results are available for this case in Orcutt and Winokur (1969, table 7), but unfortunately only for $\beta \geq 0$ and based on only one thousand replications. The Monte Carlo results are not very trustworthy, but they are in general agreement with our exact results.

5. Exact results: The not mean-stationary case

The forecast error of the least-squares forecast $\hat{y}_{n+s}$ depends on seven parameters: α, β, σ, μ_1, δ, n, s. It is easy to see from (12) that the forecast

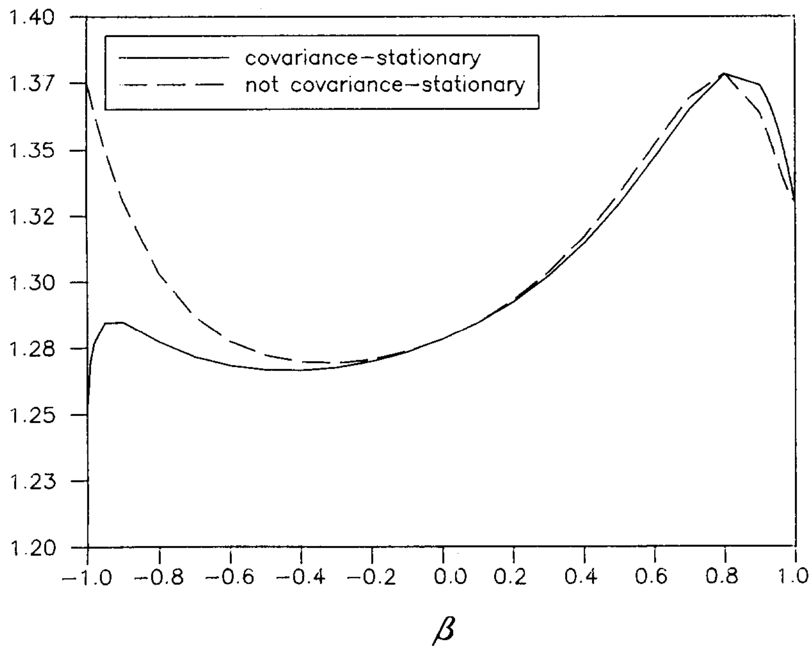

Fig. 3. *MSFE* of LS forecast for $n = 10$, $s = 1$: mean-stationary case.

error is linearly homogeneous in α, μ_1, and σ, so that

$$\frac{MSFE(\alpha, \mu_1, \sigma)}{\sigma^2} = MSFE\left(\frac{\alpha}{\sigma}, \frac{\mu_1}{\sigma}, 1\right). \tag{15}$$

In the mean-stationary case discussed in the previous section we have

$$\frac{MSFE(\alpha, \mu_1, \sigma)}{\sigma^2} = MSFE(0, 0, 1),$$

which depends only on β, δ, n, and s. If, however, the process is not mean-stationary, then the *MSFE* will depend on all seven parameters.[3] In particular, and in contrast to the mean-stationary case, (i) the LS forecast will be biased, (ii) the *MSFE* will depend on the values of α and μ_1, and (iii) the *MSFE* will not be proportional to σ^2.

[3]In fact, $MSFE/\sigma^2$ depends on α, μ_1, and σ only through $[\alpha - \mu_1(1-\beta)]/\sigma$. See Magnus and Rothenberg (1988).

It is clear from (15) that the mean-stationary process is a limiting case of the not mean-stationary process, since

$$\lim_{\sigma \to \infty} \frac{MSFE(\alpha, \mu_1, \sigma)}{\sigma^2} = MSFE(0,0,1). \tag{16}$$

Furthermore, the *MSFE* is an even function of α and μ_1:

$$MSFE(\alpha, \mu_1, \sigma) = MSFE(-\alpha, -\mu_1, \sigma).$$

In this section we only consider the case $\alpha = \mu_1$ which implies that y_0 is distributed symmetrically about zero. Then $MSFE/\sigma^2$ depends only on β, δ, n, s, and the ratio α/σ. From (16) we know the behaviour of $MSFE/\sigma^2$ when σ is large relative to α. When σ is small relative to α one can show that

$$\frac{MSFE}{\sigma^2} = O(1) + O(\sigma^2). \tag{17}$$

[Compare Magnus and Pesaran (1987, theorem 4).] In this section we study the case where α and σ are of the same order of magnitude. More specifically we set

$$\alpha = \mu_1 = \sigma,$$

so that

$$\frac{MSFE(\alpha, \mu_1, \sigma)}{\sigma^2} = MSFE(1,1,1).$$

No results from the literature are available for this case, either exact or Monte Carlo.

Table 3 presents the results for the case where the process is covariance-stationary but not mean-stationary. If we compare these results with the strictly stationary case (table 1), we find that the *MSFE* is hardly affected by the choice of μ_1 if β is negative, but that there is a big difference in *MSFE* for β close to $+1$. For $\beta > 0$ the *MSFE* is now a monotone function of β, and it is larger than in the mean-stationary case. Thus, we obtain exactly the opposite results as in table 2. Apparently for $\beta > 0$ the specification of μ_1 is important, but the specification of δ not, whereas for $\beta < 0$ the specification of δ is important, but not the specification of μ_1!

In table 4 we present the zero start-up case, where $y_0 = 0$ and the process is neither mean-stationary nor covariance-stationary. We know from our previ-

Table 3

Exact *MSFE* of least-squares forecast $\hat{y}_{n+s}$: $\alpha = \mu_1 = 1$, $\delta = (1 - \beta^2)^{-1/2}$, $\sigma = 1$.[a]

		β												
n	s	−0.99	−0.90	−0.80	−0.60	−0.40	−0.20	0.00	0.20	0.40	0.60	0.80	0.90	0.99
10	1	1.2700	1.2834	1.2753	1.2662	1.2651	1.2693	1.2781	1.2913	1.3086	1.3298	1.3645	1.4256	1.5470
10	2	2.5228	2.3339	2.0553	1.6308	1.3581	1.2117	1.1845	1.2796	1.5099	1.9137	2.6378	3.3292	4.3843
10	3	4.4319	3.8125	3.0294	2.0144	1.4895	1.2532	1.2093	1.3347	1.6791	2.4104	4.1150	6.0352	9.2824
15	1	1.1628	1.1696	1.1619	1.1555	1.1545	1.1562	1.1599	1.1657	1.1739	1.1843	1.1995	1.2305	1.3199
15	2	2.2977	2.0915	1.8491	1.4891	1.2513	1.1197	1.0908	1.1657	1.3502	1.6603	2.1622	2.6042	3.3401
15	3	3.7069	3.0983	2.4627	1.6836	1.2889	1.1151	1.0822	1.1674	1.4030	1.8946	2.9558	4.1069	6.2305
15	4	****	3.9195	2.8265	1.7147	1.2695	1.1032	1.0770	1.1640	1.4171	2.0223	3.6496	5.7760	10.2338
20	1	1.1208	1.1196	1.1144	1.1108	1.1103	1.1112	1.1133	1.1167	1.1215	1.1276	1.1351	1.1518	1.2238
20	2	2.2220	2.0032	1.7818	1.4469	1.2205	1.0929	1.0619	1.1285	1.2959	1.5728	1.9926	2.3285	2.9171
20	3	3.5390	2.8772	2.3111	1.6090	1.2485	1.0888	1.0563	1.1290	1.3356	1.7644	2.6350	3.5028	5.1545
20	4	****	3.5568	2.6060	1.6321	1.2345	1.0818	1.0539	1.1270	1.3439	1.8529	3.1294	4.6759	8.0278

[a] For $n = 15$ and $n = 20$ the integral did not converge for $\beta = -0.99$ and $s = 4$.

Table 4

Exact *MSFE* of least-squares forecast $\hat{y}_{n+s}$: $\alpha = \mu_1 = \delta = \sigma = 1$.

		β												
n	s	−0.99	−0.90	−0.80	−0.60	−0.40	−0.20	0.00	0.20	0.40	0.60	0.80	0.90	0.99
10	1	1.3645	1.3262	1.2993	1.2747	1.2681	1.2700	1.2781	1.2919	1.3108	1.3326	1.3667	1.4296	1.5475
10	2	2.9347	2.4956	2.1322	1.6504	1.3629	1.2124	1.1845	1.2799	1.5103	1.9143	2.6589	3.3655	4.3905
10	3	5.7007	4.2816	3.2355	2.0600	1.4993	1.2545	1.2093	1.3355	1.6813	2.4162	4.1782	6.1364	9.3016
15	1	1.2237	1.1912	1.1722	1.1585	1.1554	1.1566	1.1599	1.1663	1.1743	1.1860	1.1997	1.2308	1.3196
15	2	2.5102	2.1540	1.8737	1.4943	1.2526	1.1194	1.0908	1.1659	1.3504	1.6606	2.1674	2.6162	3.3411
15	3	4.2834	3.2448	2.5139	1.6925	1.2907	1.1153	1.0822	1.1674	1.4029	1.8950	2.9691	4.1893	6.2350
15	4	6.2678	4.1868	2.9111	1.7256	1.2708	1.1034	1.0770	1.1643	1.4172	2.0227	3.6745	5.8392	10.2451
20	1	1.1566	1.1311	1.1195	1.1123	1.1108	1.1113	1.1133	1.1168	1.1220	1.1286	1.1348	1.1512	1.2237
20	2	2.3521	2.0378	1.7940	1.4492	1.2210	1.0929	1.0619	1.1285	1.2960	1.5732	1.9946	2.3326	2.9171
20	3	3.8311	2.9479	2.3332	1.6123	1.2490	1.0888	1.0563	1.1290	1.3356	1.7645	2.6396	3.5166	5.1556
20	4	5.3874	3.6766	2.6386	1.6356	1.2348	1.0818	1.0539	1.1270	1.3439	1.8528	3.1372	4.7032	8.0310

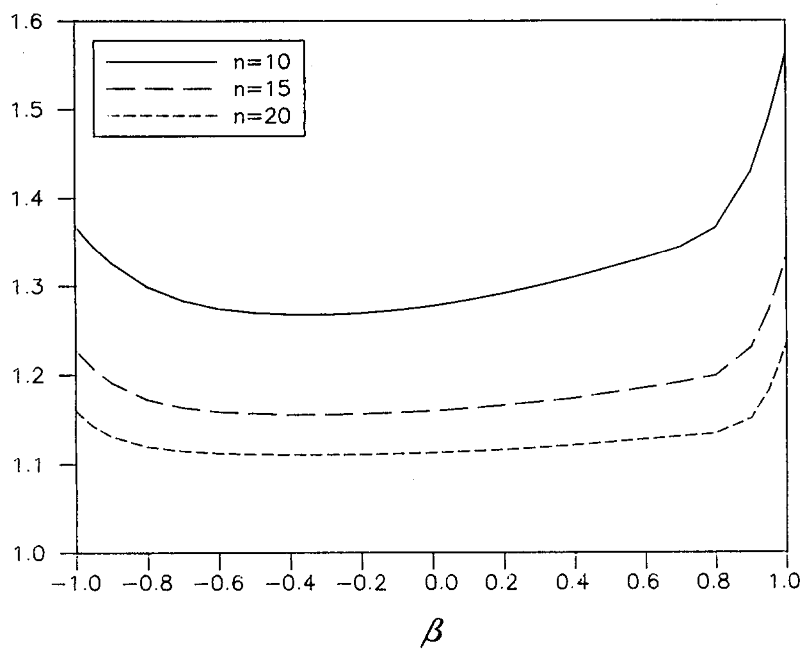

Fig. 4. *MSFE* of LS forecast for $s = 1$: zero start-up case.

ous analysis that going from mean-stationary to not mean-stationary affects the values of the *MSFE* for $\beta > 0$, while going from covariance-stationary to not covariance-stationary affects the values for $\beta < 0$. The zero start-up case, being a combination of the two, gives a combination of the two results. This is clearly illustrated by comparing fig. 4 with fig. 1. The *MSFE* is now monotone in $|\beta|$, but still not symmetric about $\beta = 0$.

Further results (not reported here) show, for all cases considered, that the *MSFE* is strictly decreasing in n, also for small values of n, in spite of the fact that the absolute value of the forecast bias is *not* strictly decreasing in n [see Magnus and Pesaran (1987)]. The reason is that the bias is small relative to the variance of the forecast error and that the latter is strictly decreasing in n.

Appendix: Proof of Theorem 1

The proof is in two parts. In the first part we show that the mean-square forecast error (*MSFE*), if finite, is given by (14). In the second part we show that the *MSFE* is finite if and only if $s \leq [(n-3)/2]$.

From (12) we obtain the *MSFE*:

$$\begin{aligned}
\mathrm{E}(\hat{y}_{n+s}-y_{n+s})^2 &= \mathrm{E}\left(\hat{\alpha}\sum_{j=0}^{s-1}\hat{\beta}^j + (\hat{\beta}^s-\beta^s)y_n\right)^2 + \alpha^2\left(\sum_{j=0}^{s-1}\beta^j\right)^2 \\
&\quad + \mathrm{E}\left[\sum_{j=0}^{s-1}\beta^j u_{n+s-j}\right]^2 \\
&\quad - 2\alpha\left[\sum_{j=0}^{s-1}\beta^j\right]\mathrm{E}\left[\hat{\alpha}\sum_{j=0}^{s-1}\hat{\beta}^j + (\hat{\beta}^s-\beta^s)y_n\right] \\
&= \mathrm{E}\hat{\alpha}^2\left(\sum_{j=0}^{s-1}\hat{\beta}^j\right)^2 + \mathrm{E}(\hat{\beta}^s-\beta^s)^2 y_n^2 \\
&\quad + 2\mathrm{E}\hat{\alpha}y_n(\hat{\beta}^s-\beta^s)\sum_{j=0}^{s-1}\hat{\beta}^j + \alpha^2\left(\sum_{j=0}^{s-1}\beta^j\right)^2 \\
&\quad + \sigma^2\sum_{j=0}^{s-1}\beta^{2j} - 2\alpha\left(\sum_{j=0}^{s-1}\beta^j\right)\left(\varepsilon_s + \alpha\sum_{j=0}^{s-1}\beta^j\right) \\
&= K_1 + \beta^{2s}\mathrm{E}y_n^2 + \mathrm{E}\hat{\alpha}^2\left(\sum_{j=0}^{s-1}\hat{\beta}^j\right)^2 + \mathrm{E}\hat{\beta}^{2s}y_n^2 \\
&\quad - 2\beta^s\mathrm{E}\hat{\beta}^s y_n^2 + 2\mathrm{E}\hat{\alpha}y_n(\hat{\beta}^s-\beta^s)\sum_{j=0}^{s-1}\hat{\beta}^j,
\end{aligned}$$

where

$$K_1 = \sigma^2\sum_{j=0}^{s-1}\beta^{2j} - \alpha^2\left(\sum_{j=0}^{s-1}\beta^j\right)^2 - 2\alpha\varepsilon_s\sum_{j=0}^{s-1}\beta^j. \tag{A.1}$$

We now define the $n \times 1$ vectors

$$v_1 = \frac{1}{n-1}(0,1,\ldots,1,1)',$$

$$v_2 = \frac{1}{n-1}(1,1,\ldots,1,0)',$$

$$v_3 = (0,0,\ldots,0,1)',$$

so that

$$\hat{\alpha} = v_1'y - \hat{\beta}v_2'y \quad \text{and} \quad y_n = v_3'y.$$

Then, using the relationship

$$\left(\sum_{j=0}^{s-1} \hat{\beta}^j\right)^2 = \sum_{j=0}^{s-1} (j+1)\hat{\beta}^j + \sum_{j=s}^{2s-2} (2s-j-1)\hat{\beta}^j, \tag{A.2}$$

we get

$$\begin{aligned}
\mathrm{E}(\hat{y}_{n+s} - y_{n+s})^2 &= K_1 + \beta^{2s}\mathrm{E}(v_3'y)^2 + \mathrm{E}(v_1'y - \hat{\beta}v_2'y)^2 \\
&\quad \times \left(\sum_{j=0}^{s-1} (j+1)\hat{\beta}^j + \sum_{j=s}^{2s-2} (2s-j-1)\hat{\beta}^j\right) \\
&\quad + \mathrm{E}(v_3'y)^2\hat{\beta}^{2s} - 2\beta^s\mathrm{E}(v_3'y)^2\hat{\beta}^s \\
&\quad + 2\mathrm{E}(v_1'y - \hat{\beta}v_2'y)(v_3'y)(\hat{\beta}^s - \beta^s)\sum_{j=0}^{s-1}\hat{\beta}^j \\
&= K_1 + \beta^{2s}\mathrm{E}(v_3'y)^2 + \mathrm{E}(v_1'y)^2 \\
&\quad - 2\beta^s\mathrm{E}(v_1'y)(v_3'y) + \mathrm{E}\sum_{j=1}^{2s}\eta_j\hat{\beta}^j,
\end{aligned} \tag{A.3}$$

where the random variables $\eta_1, \ldots, \eta_{2s}$ are quadratic forms in y defined as follows:

for $j = 1, \ldots, s-1$,

$$\begin{aligned}
\eta_j &= (j+1)(v_1'y)^2 + (j-1)(v_2'y)^2 - 2j(v_1'y)(v_2'y) \\
&\quad - 2\beta^s((v_1 - v_2)'y)v_3'y,
\end{aligned}$$

for $j = s$,

$$\eta_s = (s-1)(v_1'y)^2 + (s-1)(v_2'y)^2 - 2s(v_1'y)(v_2'y)$$
$$+2(v_1'y)(v_3'y) - 2\beta^s((v_3 - v_2)'y)v_3'y,$$

for $j = s+1, \dots, 2s-1$,

$$\eta_j = (2s-j-1)(v_1'y)^2 + (2s-j+1)(v_2'y)^2 - 2(2s-j)(v_1'y)(v_2'y)$$
$$+2((v_1 - v_2)'y)v_3'y,$$

and for $j = 2s$,

$$\eta_{2s} = ((v_3 - v_2)'y)^2.$$

Now notice that

$$v_1 = w_1 + w_3, \qquad v_2 = w_3, \qquad v_3 = w_2 + w_3.$$

Thus the η_j variables can be expressed in terms of w_1, w_2, and w_3 as follows: for $j = 1, \dots, s-1$,

$$\eta_j = (j+1)(w_1'y)^2 - 2\beta^s(w_1'y)(w_2'y) + 2(1-\beta^s)(w_1'y)(w_3'y),$$

for $j = s$,

$$\eta_s = (s-1)(w_1'y)^2 + 2(w_1'y)(w_2'y) - 2\beta^s(w_2'y)^2$$
$$+2(1-\beta^s)(w_2'y)(w_3'y),$$

for $j = s+1, \dots, 2s-1$,

$$\eta_j = (2s-j-1)(w_1'y)^2 + 2(w_1'y)(w_2'y),$$

and for $j = 2s$,

$$\eta_{2s} = (w_2'y)^2.$$

As a result we obtain

$$\mathrm{E}\sum_{j=1}^{2s} \eta_j \hat{\beta}^j = \mathrm{E}\sum_{j=1}^{2s} (y'C_{j,s}y)\hat{\beta}^j = \sum_{j=1}^{2s} \zeta_j[C_{j,s}], \tag{A.4}$$

where the matrices $C_{j,s}$ $(1 \leq j \leq 2s)$ are defined in section 3, and the second equality follows from Theorem 5(b) of Magnus (1988).

Next,

$$\begin{aligned}
&\beta^{2s}\mathrm{E}(v_3'y)^2 + \mathrm{E}(v_1'y)^2 - 2\beta^s\mathrm{E}(v_1'y)(v_3'y) \\
&= \mathrm{E}\{(v_1 - \beta^s v_3)'y\}^2 \\
&= \mathrm{E}\{(w_1 - \beta^s w_2 + (1-\beta^s)w_3)'y\}^2 \\
&= (w_1 - \beta^s w_2 + (1-\beta^s)w_3)'(LL' + \mu\mu')(w_1 - \beta^s w_2 + (1-\beta^s)w_3).
\end{aligned} \tag{A.5}$$

Inserting (A.1), (A.5), and (A.4) in (A.3) completes the first part of the proof.

To prove the second part we need to show that the *MSFE* exists if and only if $s \leq [(n-3)/2]$. It is clear that the *MSFE* exists if and only if ε_s and

$$\mathrm{E}(y'Ay/y'By)^k(y'C_{k,s}y), \qquad k = 1,\ldots,2s,$$

both exist. Now, by Magnus and Pesaran (1987, theorem 1), ε_s exists if and only if $s < n-2$. Further, since the matrix B, given in (10), has rank $n-2$, there exists an $n \times 2$ matrix Q, namely $Q = (v_2, v_3)$, with full column-rank such that $BQ = 0$. We have

$$Q'AQ = 0, \quad AQ \neq 0, \quad Q'C_{k,s}Q \neq 0,$$

so that, by Theorem 3(iv) of Magnus (1988), $\mathrm{E}(y'Ay/y'By)^k(y'C_{k,s}y)$ exists for all $1 \leq k \leq 2s$ if and only if $2s < n-2$. Hence the *MSFE* exists for $1 \leq s \leq [(n-3)/2]$ and does not exist for $s \geq [(n-1)/2]$. This concludes the proof of Theorem 1.

References

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.

Hoque, A., J.R. Magnus, and B. Pesaran, 1988, The exact multi-period mean-square forecast error for the first-order autoregressive model, Journal of Econometrics 39, 327–346.
Maekawa, K., 1987, Finite sample properties of several predictors from an autoregressive model, Econometric Theory 3, 359–370.
Magnus, J.R., 1986, The exact moments of a ratio of quadratic forms in normal variables, Annales d'Economie 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.
Magnus, J.R. and B. Pesaran, 1987, The bias of forecasts from a first-order autoregression, submitted for publication.
Magnus, J.R. and T.J. Rothenberg, 1988, Least-squares autoregression with near-unit root, submitted for publication.
Malinvaud, E., 1970, Statistical methods of econometrics, 2nd ed. (North-Holland, Amsterdam).
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.