Journal of Econometrics 9 (1979) 241–261. © North-Holland Publishing Company

THE SAMPLING DISTRIBUTION OF FORECASTS FROM A FIRST-ORDER AUTOREGRESSION*

Peter C.B. PHILLIPS

Yale University, New Haven, CT 06520, USA

University of Birmingham, Birmingham B15 2TT, UK

Received December 1977

Previous work on characterising the distribution of forecast errors in time series models by statistics such as the asymptotic mean square error has assumed that observations used in estimating parameters are statistically independent of those used to construct the forecasts themselves. This assumption is quite unrealistic in practical situations and the present paper is intended to tackle the question of how the statistical dependence between the parameter estimates and the final period observations used to generate forecasts affects the sampling distribution of the forecast errors. We concentrate on the first-order autoregression and, for this model, show that the conditional distribution of forecast errors given the final period observation is skewed towards the origin and that this skewness is accentuated in the majority of cases by the statistical dependence between the parameter estimates and the final period observation.

1. Introduction

As interest has grown in the use of finite parameter time series models for forecasting, some attention has also been given to the problem of characterizing the distribution of the error of forecasts derived from such models when the parameters have been estimated. Investigations to date have concentrated on the asymptotic distribution of forecasts and rely on the asymptotic distribution of the parameter estimates. Two of the most recent contributions are by Schmidt (1974) and Yamamoto (1976). Yamamoto considers a general autoregressive model and gives an expression up to order T^{-1}, where T is the sample size, of the mean square error of prediction in this model. Schmidt considers a more general multiple equation model with exogenous and lagged endogenous variables and obtains the limiting distribution of suitably standardised multi-step forecasts from this model.

*The research reported in this paper was supported by the SSRC under Grant No. HR 3432/1, and was completed while the author was a visitor at the Cowles Foundation for Research in Economics at Yale University. I am grateful to the referees for their helpful comments.

Such an asymptotic theory is undoubtedly of interest. But, in those applications where we can expect only a comparatively small sample of time series data we should be careful in our interpretation and use of these results. In the first place, it is well known [e.g. Hurwicz (1950), Shenton and Johnson (1966)] that parameter estimates in simple time series models can be badly biased in finite samples. This bias will carry over to the conditional distribution of forecasts given the observed values of the endogenous variable(s) used to initiate forecasts. Secondly, when we do forecast, conditional on certain values of the endogenous variables, then it is important to realise that the distribution of the parameter estimates will also be conditional and this will itself affect the distribution of forecast errors. This particular difficulty can be avoided by the assumption that the sample data used in parameter estimation is independent of the values of the endogenous variable which initiate the forecast. Although this assumption is commonly made [c.f. Yamamoto (1976)] it is hardly realistic in most practical forecasting situations. Finally, it is useful to distinguish between the distribution of forecast errors that is conditioned by observed values of the endogenous variable and the distribution which is not so conditioned.

In the case of the difference equation

$$y_t = \alpha y_{t-1} + u_t, \qquad t = \ldots, -1, 0, 1, \ldots, \tag{1}$$

where $|\alpha| < 1$ and the u_t are i.i.d. $(0, \sigma^2)$, we have the forecast error

$$\begin{aligned} \hat{y}_{T+1} - y_{T+1} &= (\hat{\alpha} - \alpha) y_T - u_{T+1} \\ &= \left[\sum_{t=1}^{T} u_t y_{t-1} \bigg/ \sum_{t=1}^{T} y_{t-1}^2 \right] y_T - u_{T+1}, \end{aligned} \tag{2}$$

where $\hat{\alpha}$ is the least squares estimator of α. As Malinvaud (1970, p. 554) has pointed out, the forecast error has zero mean when the distribution of the error u_t is symmetric. This result refers to the unconditional forecast error. On the other hand, when we consider the forecast error conditional on y_T, we can approximate the mean value of the error using the known expansion for the bias of $\hat{\alpha}$ in powers of T^{-1} [Hurwicz (1950), Shenton and Johnson (1966)], provided we assume that $\hat{\alpha}$ and y_T are independent. We have

$$E(\hat{y}_{T+1} - y_{T+1} | y_T) = -(2\alpha/T) y_T + 0(T^{-2}). \tag{3}$$

In fact, we can now go somewhat further than this. Under the same assumption about $\hat{\alpha}$ and y_T, we can obtain an Edgeworth-type expansion of the conditional distribution of $\hat{y}_{T+1}$ by using the following result I have given elsewhere [Phillips (1977), referred to below as P] for the expansion of the distribution function of $\sqrt{T}(\hat{\alpha}-\alpha)$:

$$P(\sqrt{T}(\hat{\alpha}-\alpha)\leqslant x)$$
$$=I\left(\frac{x}{(1-\alpha^2)^{1/2}}\right)+\frac{1}{\sqrt{T}}\frac{\alpha}{(1-\alpha^2)^{1/2}}i\left(\frac{x}{(1-\alpha^2)^{1/2}}\right)\left\{1+\frac{x^2}{1-\alpha^2}\right\}$$
$$+0(T^{-1}), \tag{4}$$

where $i(x)$ denotes the standard normal density and $I(x)=\int_{-\infty}^{x} i(y)\mathrm{d}y$.

We now consider the standardised variate $\sqrt{T}(\hat{y}_{T+1}-\bar{y}_{T+1})$ where $\bar{y}_{T+1}=E(y_T+1|y_T)=\alpha y_T$. We have

$$P(\sqrt{T}(\hat{y}_{T+1}-\bar{y}_{T+1})\leqslant x|y_T)$$

$$=P(\sqrt{T}(\hat{\alpha}-\alpha)y_T\leqslant x|y_T)=\begin{cases}P\left(\sqrt{T}(\hat{\alpha}-\alpha)\leqslant \dfrac{x}{y_T}\middle| y_T>0\right).\\ P\left(\sqrt{T}(\hat{\alpha}-\alpha)\geqslant \dfrac{x}{y_T}\middle| y_T<0\right)\end{cases}$$

When $y_T>0$, this becomes

$$I\left(\frac{x}{y_T(1-\alpha^2)^{1/2}}\right)+i\left(\frac{x}{y_T(1-\alpha^2)^{1/2}}\right)$$
$$\times\frac{1}{\sqrt{T}}\frac{\alpha}{(1-\alpha^2)^{1/2}}\left\{1+\frac{x^2}{y_T^2(1-\alpha^2)}\right\}+0(T^{-1}), \tag{5}$$

and when $y_T<0$,

$$I\left(\frac{-x}{y_T(1-\alpha^2)^{1/2}}\right)-i\left(\frac{x}{y_T(1-\alpha^2)^{1/2}}\right)$$
$$\times\frac{1}{\sqrt{T}}\frac{\alpha}{(1-\alpha^2)^{1/2}}\left\{1+\frac{x^2}{y_T^2(1-\alpha^2)}\right\}+0(T^{-1}). \tag{6}$$

Taking α to be positive we see that the distribution of forecasts is negatively skewed when $y_T>0$, positively skewed when $y_T<0$. When T is small, the correction term of $0(T^{-1/2})$ on the normal approximation can be substantial; and the normal approximation is less satisfactory as α increases in size.

These results suggest that we should be careful in using the normal distribution to characterise the sampling distribution of forecasts in dynamic models. Moreover, the fact that forecasts are systematically skewed towards the origin is suggestive in the light of the well known characteristic of dynamic simulations of econometric models to underestimate the amplitude of cycles.[1]

The question now arises to what extent these results are affected by the statistical dependence of y_T and $\hat{\alpha}$. The present paper is intended to tackle this question. We show that the correlation between y_T and $\hat{\alpha}$ leads to an additional correction term of $0(T^{-1/2})$ in the distribution of $\sqrt{T}(\hat{y}_{T+1} - \bar{y}_{T+1})$, and this additional term accentuates the skewness towards the origin in the majority of cases. We also look at the sampling distribution of forecasts for lead time $h>1$ and derive a similar Edgeworth-type expansion for this distribution.

2. One-period forecasts

We will work with the model given by (1) and $\hat{\alpha}$ will be defined as the ratio $\sum_{t=1}^{T} y_t y_{t-1}/\sum_{t=1}^{T} y_{t-1}^2$, so that $\hat{\alpha}$ and y_T are statistically dependent. In this section we give an expansion for the conditional distribution of $\sqrt{T}(\hat{y}_{T+1} - \bar{y}_{T+1})$, given y_T. We will leave the technical aspects of the derivation to the appendix; and in a later section we will consider the distribution of the forecast error $\hat{y}_{T+1} - y_{T+1}$ itself, rather than that of the standardised variate $\sqrt{T}(\hat{y}_{T+1} - \bar{y}_{T+1})$. It is useful to have an expansion for the conditional distribution of $\sqrt{T}(\hat{y}_{T+1} - \bar{y}_{T+1})$ given y_T. For, by comparing this expansion with (5) and (6) we can determine, at least in part, the small sample effect of the assumption that there are independent data available for the estimation of α.

From the appendix we have the following expansion of the distribution function of $\sqrt{T}(\hat{\alpha} - \alpha)$ conditional on y_T:

Theorem 1. In the model (1) and where $\hat{\alpha}$ is the least squares estimator $\sum_{t=1}^{T} y_t y_{t-1}/\sum_{t=1}^{T} y_{t-1}^2$ of α an approximation to the distribution function of $\sqrt{T}(\hat{\alpha} - \alpha)$ conditional on y_T is given by

$$P(\sqrt{T}(\hat{\alpha} - \alpha) \leq x | y_T)$$

$$= I\left(\frac{x}{(1-\alpha^2)^{1/2}}\right) + \frac{1}{\sqrt{T}} \frac{\alpha}{(1-\alpha^2)^{1/2}} i\left(\frac{x}{(1-\alpha^2)^{1/2}}\right) \times \left[1 + \frac{x^2}{1-\alpha^2} + \left\{1 - \left(\frac{y_T}{\sigma_y}\right)^2\right\}\right] + 0(T^{-1}), \tag{7}$$

[1] See, for instance, the simulation results of Green (1972) and Fromm et al. (1972).

where $\sigma_y^2 = \sigma^2/(1-\alpha^2)$.

We note immediately that the only difference between (7) and the expansion for the unconditional distribution of $\sqrt{T}(\hat{\alpha}-\alpha)$ [see (4) above] is the term in braces on the right-hand side of (7). We can see from (7) that, for the conditional distribution, the skewness of the unconditional distribution is accentuated when $|y_T| \leqslant \sigma_y$. Since y_T is normally distributed, this will be so in a clear majority of cases (68 %). But when y_T is an outlier the skewness of the conditional distribution is less marked.

From (7) we can now deduce an approximation to the conditional distribution of y_{T+1} given y_T. We have

$$P(\sqrt{T}(\hat{y}_{T+1}-\bar{y}_{T+1}) \leqslant x | y_T)$$

$$= I\left(\frac{x}{y_T(1-\alpha^2)^{1/2}}\right) + \frac{1}{\sqrt{T}} \frac{\alpha}{(1-\alpha^2)^{1/2}} i\left(\frac{x}{y_T(1-\alpha^2)^{1/2}}\right)$$
$$\times \left\{2 + \frac{x^2}{y_T^2(1-\alpha^2)} - \left(\frac{y_T}{\sigma_y}\right)^2\right\} + 0(T^{-1}), \tag{8}$$

when $y_T > 0$, and

$$= I\left(\frac{-x}{y_T(1-\alpha^2)^{1/2}}\right) - \frac{1}{\sqrt{T}} \frac{\alpha}{(1-\alpha^2)^{1/2}} i\left(\frac{x}{y_T(1-\alpha^2)^{1/2}}\right)$$
$$\times \left\{2 + \frac{x^2}{y_T^2(1-\alpha^2)} - \left(\frac{y_T}{\sigma_y}\right)^2\right\} + 0(T^{-1}), \tag{9}$$

when $y_T < 0$.

When we compare (8) and (9) with (5) and (6) we reach the same conclusion as in the case of the distribution of $\hat{\alpha}$. The effect of statistical dependence between $\hat{\alpha}$ and y_T on the distribution of the forecast $\hat{y}_{T+1}$ conditional on a given value of y_T is, in the majority of cases, to accentuate the skewness towards the origin which is present when $\hat{\alpha}$ is extimated from independent data.

It could be argued that we come closer to the assumption that $\hat{\alpha}$ can be estimated from independent data if we neglect several of the latest observations in the computation of $\hat{\alpha}$. To consider the effect of this procedure let us retain $\hat{\alpha}$ as before, estimated from the first $T+1$ observations $y_0, \ldots, y_T$, and suppose that we forecast using the latest observation y_{T+l} where $l>0$. Then $\hat{y}_{T+l+1}=\hat{\alpha}y_{T+l}$ and $\hat{y}_{T+l+1}-\bar{y}_{T+l+1}=(\hat{\alpha}-\alpha)y_{T+l}$. The expansion of the distribution function of $\sqrt{T}(\hat{\alpha}-\alpha)$ conditional on y_{T+l} proceeds in the same way as when $l=0$, and we find that

$$\begin{aligned}
&P(\sqrt{T}(\hat{\alpha}-\alpha)\leqq x|y_{T+l})\\
&=I\left(\frac{x}{(1-\alpha^2)^{1/2}}\right)+\frac{1}{\sqrt{T}}i\left(\frac{x}{(1-\alpha^2)^{1/2}}\right)\\
&\times\left[\frac{\alpha}{(1-\alpha^2)^{1/2}}\left\{1+\left(\frac{x}{(1-\alpha^2)}\right)^2\right\}+\frac{\alpha^{2l+1}}{(1-\alpha^2)^{1/2}}\left\{1-\left(\frac{y_T}{\sigma_y}\right)^2\right\}\right]\\
&+0(T^{-1}),
\end{aligned}$$

so that there is, indeed, a reduction in magnitude of the additional correction term of $0(T^{-1/2})$ as observations are neglected in the estimation of α. However, when T is small, some care needs to be taken in the selection of l. For, if all $T+l$ observations are used to estimate α, the primary coefficient of the correction term is, strictly speaking, $(T+l)^{-1/2}$; and the reduction in the overall importance of the correction term that results from the use of the extra l observations may well outweigh the effect of the reduction in the magnitude of the additional correction term resulting from neglect of the last l observations in the estimation of α.

3. Multi-period forecasts

When we develop forecasts h time periods ahead we use $y_{T+h}=\hat{\alpha}^h y_T$. If we now define $\bar{y}_{T+h}=E(y_{T+h}|y_T)=\alpha^h y_T$, we have $\hat{y}_{T+h}-\bar{y}_{T+h}=(\hat{\alpha}^h-\alpha^h)y_T$ and

$$P(\sqrt{T}(\hat{y}_{T+h}-\bar{y}_{T+h})\leqq x|y_T)=P(\sqrt{T}(\hat{\alpha}^h-\alpha^h)y_T\leqq x|y_T). \tag{10}$$

We can expand the probability on the right side of (10) in powers of $T^{-1/2}$ in much the same way as the previous section. The analysis is more complicated because $\hat{\alpha}^h$ is a function of a function of the more basic statistics (namely, the quadratic forms $y'C_1y$ and $y'C_2y$). An outline of the derivations is given in the appendix. We obtain the following expansion for the distribution of $\sqrt{T}(\hat{\alpha}^h-\alpha^h)$ conditional on y_T.

Theorem 2. In the model (1) an approximation to the distribution function of $\sqrt{T}(\hat{\alpha}^h-\alpha^h)$ conditional on y_T is given by

$$P(\sqrt{T}(\hat{\alpha}^h - \alpha^h) \leqq x | y_T)$$

$$= I\left(\frac{x}{\omega}\right) + \frac{1}{\sqrt{T}} \frac{\alpha}{(1-\alpha^2)^{1/2}} i\left(\frac{x}{\omega}\right)$$

$$\times \left[\left\{ 1 + \left(\frac{x}{\omega}\right)^2 \right\} - \left(\frac{h-1}{2}\right)\left(\frac{1-\alpha^2}{\alpha^2}\right)\left(\frac{x}{\omega}\right)^2 + \left\{ 1 - \left(\frac{y_T}{\sigma_y}\right)^2 \right\} \right]$$

$$+ 0(T^{-1}), \tag{11}$$

where $\omega = h\alpha^{h-1}(1-\alpha^2)^{1/2}$.

When $h=1$, (11) reduces to (7) above. We note from (11) that, as in the special case $h=1$, the skewness of the unconditional distribution of $\sqrt{T}(\hat{\alpha}^h - \alpha^h)$ is accentuated for the conditional distribution when $|y_T| \leqq \sigma_y$. We also see from (11) that the asymptotic distribution $I(x/\omega)$ becomes a more satisfactory approximation as h increases since the correction term of $0(T^{-1/2})$ in (11) involves an additional term [over and above those in (7)] which reduces the magnitude of the correction for $h > 1$.

We can now extract an expansion for the conditional distribution of $\hat{y}_{T+h}$ given y_T, as in (8) and (9) above. If $y_T > 0$, we have

$$P(\sqrt{T}(\hat{y}_{T+h} - \bar{y}_{T+h}) \leqq x | y_T)$$

$$= I\left(\frac{x}{y_T h\alpha^{h-1}(1-\alpha^2)^{1/2}}\right) + \frac{1}{\sqrt{T}} \frac{\alpha}{(1-\alpha^2)^{1/2}} i\left(\frac{x}{y_T h\alpha^{h-1}(1-\alpha^2)^{1/2}}\right)$$

$$\times \left[\left\{ 1 + \left(\frac{x}{y_T h\alpha^{h-1}(1-\alpha^2)^{1/2}}\right)^2 \right\} \right.$$

$$- \left(\frac{h-1}{2}\right)\left(\frac{1-\alpha^2}{\alpha^2}\right)\left(\frac{x}{y_T h\alpha^{h-1}(1-\alpha^2)^{1/2}}\right)^2$$

$$\left. + \left\{ 1 - \left(\frac{y_T}{\sigma_y}\right)^2 \right\} \right] + 0(T^{-1}),$$

and multi-period forecasts will be skewed towards the origin, provided h is not too large. As with (8) and (9), this skewness will in the majority of cases be accentuated by the dependence between $\hat{\alpha}$ and y_T.

4. Distribution of the forecast error

The above results can be used to obtain an approximation to the distribution of the forecast error. We illustrate for the case of the one-period forecast; and, as before, the derivations are given in the appendix.

Theorem 3. An approximation to the distribution function of the forecast error $\hat{y}_{T+1}-y_{T+1}$ conditional on y_T in the model (1) is given by

$$P(\hat{y}_{T+1}-y_{T+1}\leqq x|y_T)$$
$$=I\left(\frac{x}{\sigma}\right)+i\left(\frac{x}{\sigma}\right)\left[\frac{1}{T}\left(\frac{\alpha}{(1-\alpha^2)^{1/2}}\right)\left(\frac{y_T}{\sigma_y}\right)\left(3-\left(\frac{y_T}{\sigma_y}\right)^2\right)-\frac{1}{2T}\left(\frac{y_T}{\sigma_y}\right)^2\left(\frac{x}{\sigma}\right)\right]$$
$$+0(T^{-2}). \tag{12}$$

When $\hat{\alpha}$ and y_T are statistically independent the corresponding approximation to the distribution function of $\hat{y}_{T+1}-y_{T+1}$ conditional on y_T is given by

$$P(\hat{y}_{T+1}-y_{T+1}\leqq x|y_T)$$
$$=I\left(\frac{x}{\sigma}\right)+i\left(\frac{x}{\sigma}\right)\left[\frac{1}{T}\left(\frac{2\alpha}{(1-\alpha^2)^{1/2}}\right)\left(\frac{y_T}{\sigma_y}\right)-\frac{1}{2T}\left(\frac{y_T}{\sigma_y}\right)^2\left(\frac{x}{\sigma}\right)\right]$$
$$+0(T^{-2}). \tag{13}$$

It is clear from (12) that, up to $0(T^{-1})$, the conditional distribution of $\hat{y}_{T+1}-y_{T+1}$ is negatively skewed when $y_T>0$, positively skewed when $y_T<0$.The second term in the square brackets on the right side of (12) does not influence this skewness, since it involves an odd power of (x/σ). Comparing (12) and (13), we see that the skewness in the conditional distribution is accentuated when $|y_T|<\sigma_y$ and $\hat{\alpha}$ and y_T are statistically dependent, as in the conditional distribution of $\sqrt{T}(\hat{y}_{T+1}-\bar{y}_{T+1})$ considered in section 2. We conclude that, in the majority of cases and when y_T is not an outlier, the effect of the statistical dependence between $\hat{\alpha}$ and y_T is to magnify the skewness of the conditional distribution of the forecast error.

5. Final comments

As we emphasized in the introduction, a clear distinction must be drawn between the conditional and the unconditional distributions of forecasts in dynamic models. Our discussion has concentrated on the conditional distribution and it is this case which is of most interest since, in practice, we do forecast with given final period values of the endogenous variables. If these are observed without measurement error then this is information we should use in forecasting. But, in the evaluation of the success of a forecasting procedure, *on average* we might be interested in looking at the unconditional distribution. This is the approach taken in the derivation of characteristics such as the asymptotic mean square error by Yamamoto (1976) and Box and Jenkins (1970). Moreover, it is implicit in most of the sampling experiment analysis of forecasting performance in dynamic models. This explains why, in those sampling experiments [e.g. Orcutt and Winokur (1969), Malinvaud (1970, p. 554)] the sampling distribution of forecasts has appeared unbiased.

Note that a crude approximation for the unconditional distribution of the forecast error $\hat{y}_{T+1}-y_{T+1}$ can be obtained by multiplying (12) by the density of y_T, i.e., $(1/\sigma_y)\,i(y_T/\sigma_y)$, and integrating with respect to y_T. We find that

$$P(\hat{y}_{T+1}-y_{T+1}\leqq x)\sim I\left(\frac{x}{\sigma}\right)-\frac{1}{2T}\frac{x}{\sigma}i\left(\frac{x}{\sigma}\right)\sim I\left(\frac{x}{\sigma}\left(1-\frac{1}{2T}\right)\right),$$

and the approximate variance is given by $\sigma^2(1-(2T)^{-1})^{-2}\sim\sigma^2(1+T^{-1})$ which corresponds with the usual formula for the asymptotic mean square error of forecast [Box and Jenkins (1970, p. 269)]. The approximation is rather crude because although the limiting distribution of $\sqrt{T}(\hat{\alpha}-\alpha)$ as $T\to\infty$ is normal, the limiting distribution of $\sqrt{T}(\hat{y}_{T+1}-\bar{y}_{T+1})=\sqrt{T}(\hat{\alpha}-\alpha)y_T$ in the unconditional case is not and has a logarithmic discontinuity at the origin. It would seem advisable to use this known large sample behaviour in devising a suitable approximation for the small sample distribution.

The approach that has been used to derive expansions of the sampling distribution of forecasts in the simple model (1) can be used for higher-order autoregressive models and other time series models. In every case it will be necessary to write down the joint characteristic function of (i) the first and second sample moments of the data that appear in formula for the parameter estimates, and (ii) the final period values of the endogenous variable used to initiate forecasts. From this characteristic function the cumulants can be extracted and combined with the derivatives of the function representing the error in the parameter estimates in much the same way as in the proof of Theorem 2 to yield an approximation to the joint density of the parameter estimates and the final period values of the endogenous variable. The approximation to the conditional distribution of the parameter estimates given the final period values needed for forecasting then follows directly and this can be used to give the required approximation to the distribution of the forecast errors. However, in models more complicated than (1) it is necessary to carry out most of the heavy algebraic manipulation by computer. Some work on the development of the appropriate software is currently under way.

Appendix

Proof of Theorem 1

We start by considering the joint distribution of $\hat{\alpha}$ and y_T which we write as

$$P(\sqrt{T}(\hat{\alpha}-\alpha)\leqq x_1,\ y_T\leqq x_2)=P(\hat{\alpha}\leqq\alpha+x_1/\sqrt{T},\ y_T\leqq x_2)$$
$$=P(Q(r)\leqq 0,\ y_T\leqq x_2),$$

where $r=\alpha+x_1/\sqrt{T}$, $Q(r)=y'(C_1-rC_2)y$, $y'=(y_0,\ y_1,\ldots,y_T)$ and the matrices C_1 and C_2 are defined as

$$C_1=\tfrac{1}{2}\begin{bmatrix} 0 & 1 & \ldots & 0 & 0 \\ 1 & 0 & \ldots & 0 & 0 \\ . & . & \ldots & . & . \\ 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix},$$

$$C_2=\begin{bmatrix} 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ . & . & \ldots & . & . \\ 0 & 0 & \ldots & 1 & 0 \\ 0 & 0 & \ldots & 0 & 0 \end{bmatrix}.$$

Now the joint characteristic function of Q and y_T is given by[2]

$$\bar{\theta}(t_1,\ t_2)=|I-2it_1D\Omega|^{-1/2}$$
$$\times\ \exp\{\tfrac{1}{2}(it_2\Omega b)'\ (\Omega-2it_1\Omega D\Omega)^{-1}\ (it_2\Omega b)\},$$

where $D=C_1-rC_2$, $b'=(0,0,\ldots,0,1)$ and Ω is the $(T+1)\times(T+1)$ matrix with (i,j)th element $\sigma 2(1-\alpha^2)^{-1}\alpha^{|i-j|}$. The second characteristic (or cumulant generating function) is then

$$\bar{\lambda}(t_1,\ t_2)=\log(\bar{\theta}(t_1,t_2))$$
$$=-\tfrac{1}{2}\log\det(I-2it_1D\Omega)-(t_2^2/2)b'\Omega(\Omega-2it_1\Omega D\Omega)^{-1}\ \Omega b.$$

We now introduce the standardised variates

$$q_1=(Q-k_1)/\sqrt{k_2},\qquad q_2=y_T/\sigma_y,$$

where

$$k_1=E(y'(C_1-rC_2)y)=\mathrm{tr}((C_1-rC_2)\Omega),$$
$$k_2=2\mathrm{tr}((C_1-rC_2)\Omega)^2,$$

[2]This result follows readily from the result given by Lukacs and Laha (1964, p. 55).

and

$$\sigma_y^2=\sigma^2/(1-\alpha^2).$$

Then, the second characteristic of q_1 and q_2 is

$$\lambda(t_1,t_2)=\bar{\lambda}\left(\frac{t_1}{\sqrt{k_2}},\ \frac{t_2}{\sigma_y}\right)-\frac{ik_1t_1}{\sqrt{k_2}},$$

and, noting that $\log\det(I-2it_1k_2^{-1/2}D\Omega)=\sum_{j=1}^{T+1}\log(1-2it_1k_2^{-1/2}\delta_j)$ where δ_j is the jth eigenvalue of $D\Omega$, we have

$$\begin{aligned}\lambda(t_1,t_2)&=-\tfrac{1}{2}\sum_{j=1}^{T+1}\log\left(1-\frac{2it_1}{\sqrt{k_2}}\delta_j\right)\\&\quad-\frac{ik_1t_1}{\sqrt{k_2}}-\frac{t_2^2}{2\sigma_y^2}b'\left(I-\frac{2it_1}{\sqrt{k_2}}\Omega D\right)^{-1}\Omega b\\&=\sum_{m=1}^{\infty}\frac{2^{m-1}}{m}\left(\frac{it_1}{\sqrt{k_2}}\right)^m\sum_{j=1}^{T+1}\delta_j^m-\frac{ik_1t_1}{\sqrt{k_2}}\\&\quad-\frac{t_2^2}{2\sigma_y^2}\sum_{n=1}^{\infty}2^n\left(\frac{it_1}{\sqrt{k_2}}\right)^n b'(\Omega D)^n\Omega b\\&=-\frac{t_1^2}{2}-\frac{t_2^2}{2}+\sum_{m=3}^{\infty}\frac{(it_1)^m}{m!}\frac{k_m}{k_2^{m/2}}+(it_2)^2\sum_{n=1}^{\infty}\frac{(it_1)^nh_n}{k_2^{n/2}},\end{aligned}$$

where

$$k_m=(m-1)!2^{m-1}\sum_{j=1}^{T+1}\delta_j^m=(m-1)!2^{m-1}\operatorname{tr}(D\Omega)^m,$$

and

$$h_n=2^{n-1}b'(\Omega D)^n\Omega b/\sigma_y^2.$$

The characteristic function of q_1 and q_2 is now

$$\begin{aligned}\theta(t_1,t_2)&=\exp(\lambda(t_1,t_2))\\&=\exp\left(-\frac{t_1^2}{2}-\frac{t_2^2}{2}\right)\\&\quad\times\exp\left\{\sum_{m=3}^{\infty}\frac{(it_1)^m}{m!}\frac{k_m}{k_2^{m/2}}+(it_2)^2\sum_{n=1}^{\infty}\frac{(it_1)^nh_n}{k_2^{n/2}}\right\},\end{aligned}$$

and, since $k_m = 0(T)$ for all m (cf. P, p. 465) and $h_n = 0(1)$ for all n, we have the expansion

$$\begin{aligned}\theta(t_1,t_2) &= \exp\left(-\frac{t_1^2}{2}-\frac{t_2^2}{2}\right)\\ &\quad\times \exp\left\{\frac{k_3}{3!k_2^{3/2}}(it_1)^3+\frac{k_4}{4!k_2^2}(it_1)^4\right.\\ &\quad\left.+\frac{h_1}{k_2^{1/2}}(it_1)(it_2)^2+\frac{h_2}{k_2}(it_1)^2(it_2)^2+0(T^{-3/2})\right\}\\ &= \exp\left(-\frac{t_1^2}{2}-\frac{t_2^2}{2}\right)\left\{1+\frac{k_3}{3!k_2^{3/2}}(it_1)^3+\frac{k_4}{4!k_2^2}(it_1)^4\right.\\ &\quad+\frac{h_1}{k_2^{1/2}}(it_1)(it_2)^2+\frac{h_2}{k_2}(it_1)^2(it_2)^2+\tfrac{1}{2}\left(\frac{k_3}{3!k_2^{3/2}}\right)^2(it_1)^6\\ &\quad\left.+\tfrac{1}{2}\frac{h_1^2}{k_2}(it_1)^2(it_2)^4+\frac{k_3h_1}{3!k_2^2}(it_1)^4(it_2)^2\right\}+0(T^{-3/2}).\end{aligned}$$

Inverting this characteristic function term by term we obtain an expansion for the joint density of q_1 and q_2 as[3]

$$\begin{aligned}f(q_1,q_2) &= i(q_1)i(q_2)\left\{1+\frac{k_3}{3!k_2^{3/2}}H_3(q_1)+\frac{k_4}{4!k_2^2}H_4(q_1)\right.\\ &\quad+\tfrac{1}{2}\frac{k_3^2}{(3!)^2k_2^3}H_6(q_1)+\frac{h_1}{k_2^{1/2}}H_1(q_1)H_2(q_2)+\frac{h_2}{k_2}H_2(q_1)H_2(q_2)\\ &\quad\left.+\tfrac{1}{2}\frac{h_1^2}{k_2}H_2(q_1)H_4(q_2)+\frac{1}{3!}\frac{k_3h_1}{k_2^2}H_4(q_1)H_2(q_2)\right\}\\ &\quad+0(T^{-3/2}), \qquad\qquad (\text{A.1})\end{aligned}$$

where $H_n(\quad)$ denotes the Hermite polynomial of degree n.

[3]Note that term by term inversion of the characteristic function does not rigorously justify (A.1) as an asymptotic series, although it is the most convenient way of obtaining the explicit form of successive terms in the series. A similar comment holds for the integration needed to derive the corresponding expansion of the distribution function in (A.3) below. Verification of the asymptotic nature of these series can be obtained by appealing to an appropriate theorem on the validity of the Edgeworth series expansion. At present the most relevant theorem is given in Phillips (1977b). A complete verification of the expansions requires, of course, that the conditions of the theorem hold in the present case. Most of the conditions are not difficult to check out, but one side condition has not yet been validated for models with lagged endogenous variables as regressors. Some discussion of the problem is contained in Phillips (1977b, sect. 3).

Now the conditional density of q_1 given $q_2 = y_T/\sigma_y$ is the quotient

$$f(q_1|y_T/\sigma_y) = f(q_1, y_T/\sigma_y)/i(y_T/\sigma_y),$$

and

$$\begin{aligned} P(\sqrt{T}(\hat{\alpha}-\alpha) \leqq x_1|y_T) &= P(q_1 \leqq -k_1/\sqrt{k_2}|q_2 = y_T/\sigma_y) \\ &= \int_{-\infty}^{-k_1/\sqrt{k_2}} f(q_1|y_T/\sigma_y)\mathrm{d}q_1. \end{aligned} \tag{A.2}$$

From (A.1) and the fact that $H_n(x)i(x) = (-1)^n i^{(n)}(x)$ we obtain the expansion of the conditional density

$$\begin{aligned} f(q_1|y_T/\sigma_y) = i(q_1) &- \frac{k_3}{3!k_2^{3/2}} i^{(3)}(q_1) + \frac{k_4}{4!k_2^2} i^{(4)}(q_1) \\ &+ \tfrac{1}{2}\frac{k_3^2}{(3!)^2 k_2^3} i^{(6)}(q_1) - \frac{h_1}{k_2^{1/2}} i'(q_1) H_2\left(\frac{y_T}{\sigma_y}\right) \\ &+ \frac{h_2}{k_2} i^{(2)}(q_1) H_2\left(\frac{y_T}{\sigma_y}\right) + \tfrac{1}{2}\frac{h_1^2}{k_2} i^{(2)}(q_1) H_4\left(\frac{y_T}{\sigma_y}\right) \\ &+ \frac{1}{3!}\frac{k_3 h_1}{k_2^2} i^{(4)}(q_1) H_2\left(\frac{y_T}{\sigma_y}\right) + 0(T^{-3/2}). \end{aligned}$$

Integrating out, we have from (A.2)

$$\begin{aligned} &P(\sqrt{T}(\hat{\alpha}-\alpha) \leqq x_1|y_T) \\ &= I\left(-\frac{k_1}{\sqrt{k_2}}\right) - \frac{k_3}{3!k_2^{3/2}} I^{(3)}\left(-\frac{k_1}{\sqrt{k_2}}\right) + \frac{k_4}{4!k_2^2} I^{(4)}\left(-\frac{k_1}{\sqrt{k_2}}\right) \\ &\quad + \tfrac{1}{2}\frac{k_3^2}{(3!)^2 k_2^3} I^{(6)}\left(-\frac{k_1}{\sqrt{k_2}}\right) - \frac{h_1}{k_2^{1/2}} i\left(-\frac{k_1}{\sqrt{k_2}}\right) H_2\left(\frac{y_T}{\sigma_y}\right) \\ &\quad + \frac{h_2}{k_2} I^{(2)}\left(-\frac{k_1}{\sqrt{k_2}}\right) H_2\left(\frac{y_T}{\sigma_y}\right) + \tfrac{1}{2}\frac{h_1^2}{k_2} I^{(2)}\left(-\frac{k_1}{\sqrt{k_2}}\right) H_4\left(\frac{y_T}{\sigma_y}\right) \\ &\quad + \frac{1}{3!}\frac{k_3 h_1}{k_2^2} I^{(4)}\left(-\frac{k_1}{\sqrt{k_2}}\right) H_2\left(\frac{y_T}{\sigma_y}\right) + 0(T^{-3/2}). \end{aligned} \tag{A.3}$$

To obtain an explicit representation of the expansion we must take account of the fact that k_m is a function of $r=\alpha+x_1/\sqrt{T}$ for all values of m. From P (p. 466) we have

$$-\frac{k_1}{\sqrt{k_2}}=\frac{x_1}{(1-\alpha^2)^{1/2}}+\frac{1}{\sqrt{T}}\cdot\left(\frac{2\alpha}{(1-\alpha^2)^{1/2}}\right)\left(\frac{x_1^2}{1-\alpha^2}\right)+0(T^{-1}),$$

and, for our present purposes, it will be sufficient to consider only terms of $0(T^{-1/2})$ explicitly.[4] We, therefore, take the first, second and fifth terms of (A.3). Individually we have

$$I\left(-\frac{k_1}{\sqrt{k_2}}\right)=I\left(\frac{x_1}{(1-\alpha^2)^{1/2}}\right)$$

$$+\frac{1}{\sqrt{T}}i\left(\frac{x_1}{(1-\alpha^2)^{1/2}}\right)\frac{2\alpha}{(1-\alpha^2)^{1/2}}\left(\frac{x_1^2}{1-\alpha^2}\right)+0(T^{-1}),$$

$$I^{(3)}\left(-\frac{k_1}{\sqrt{k_2}}\right)=I^{(3)}\left(\frac{x_1}{(1-\alpha^2)^{1/2}}\right)+0(T^{-1/2})$$

$$=i\left(\frac{x_1}{(1-\alpha^2)^{1/2}}\right)\left\{\left(\frac{x_1^2}{1-\alpha^2}\right)-1\right\}+0(T^{-1/2}),$$

and from the calculations in P (p. 468, and appendix) we have

$$\frac{k_3}{k_2^{3/2}}=\frac{1}{\sqrt{T}}\left(\frac{6\alpha}{(1-\alpha^2)^{1/2}}\right)+0(T^{-1}),$$

and

$$k_2=2\operatorname{tr}((C_1-\alpha C_2)\Omega)^2+0(T^{1/2})$$

$$=\frac{T\sigma^4}{1-\alpha^2}[1+0(T^{-1/2})].$$

This leaves us with h_1. Now

$$h_1=b'\Omega D\Omega b/\sigma_y^2$$

$$=\left(\frac{1-\alpha^2}{\sigma^2}\right)b'\Omega(C_1-\alpha C_2)\Omega b+0(T^{-1/2}),$$

[4]Terms of $0(T^{-1})$ in the expansion of the distribution of $\sqrt{T}(\hat{\alpha}-\alpha)$ are obtained explicitly in Phillips (1976).

and after some algebra we find

$$b'\Omega C_1\Omega b=\left(\frac{\sigma^2}{1-\alpha^2}\right)^2\alpha\left(\frac{1-\alpha^{2T}}{1-\alpha^2}\right),$$

and

$$b'\Omega C_2\Omega b=\left(\frac{\sigma^2}{1-\alpha^2}\right)^2\alpha^2\left(\frac{1-\alpha^{2T}}{1-\alpha^2}\right).$$

Hence

$$\begin{aligned}h_1&=\left(\frac{1-\alpha^2}{\sigma^2}\right)\left(\frac{\sigma^2}{1-\alpha^2}\right)^2\left(\frac{1-\alpha^{2T}}{1-\alpha^2}\right)\alpha(1-\alpha^2)+0(T^{-1/2})\\&=\alpha\left(\frac{\sigma^2}{1-\alpha^2}\right)+0(T^{-1/2}),\end{aligned}$$

so that

$$\begin{aligned}\frac{h_1}{k_2^{1/2}}&=\alpha\left(\frac{\sigma^2}{1-\alpha^2}\right)\frac{1}{\sqrt{T}}\left(\frac{\sigma^2}{(1-\alpha^2)^{1/2}}\right)^{-1}\{1+0(T^{-1/2})\}\\&=\frac{1}{\sqrt{T}}\left(\frac{\alpha}{(1-\alpha^2)^{1/2}}\right)+0(T^{-1}).\end{aligned}$$

Using the above results and collecting terms in (A.3) we find that

$$\begin{aligned}&P(\sqrt{T}(\hat{\alpha}-\alpha)\leqq x_1|y_T)\\&=I\left(\frac{x}{(1-\alpha^2)^{1/2}}\right)+\frac{1}{\sqrt{T}}\frac{\alpha}{(1-\alpha^2)^{1/2}}i\left(\frac{x_1}{(1-\alpha^2)^{1/2}}\right)\\&\quad\left[1+\frac{x_1^2}{1-\alpha^2}+\left\{1-\left(\frac{y_T}{\sigma_y}\right)^2\right\}\right]+0(T^{-1}),\end{aligned}$$

as given in the theorem.

Proof of Theorem 2

We write $\eta_T(p)=\hat{\alpha}^h-\alpha^h$ and $e_T(p)=\hat{\alpha}-\alpha$ where $p'=(p_1,p_2)$, $p_1=T^{-1}\{y'C_1y-E(y'C_1y)\}$ and $p_2=T^{-1}\{y'C_2y-E(y'C_2y)\}$. Then

$$\eta_T(p)=((\hat{\alpha}-\alpha)+\alpha)^h=\sum_{j=1}^{h}\binom{h}{j}e_T(p)^j\alpha^{h-j}.$$

We note that $e_T(0)=0$ (cf. P, p. 44), so that $\eta_T(0)=0$ and both $e_T(p)$ and $\eta_T(p)$ have continuous derivatives up to the third order. We write derivatives of $\eta_T(p)$ and $e_T(p)$ evaluated at the origin as, for instance, $e_j=\partial e_T(0)/\partial p_j$ and, using the tensor summation convention of a repeated subscript, we have from the Taylor development of $\eta_T(0)$

$$\eta_T(p)=\eta_j p_j+\tfrac{1}{2}\eta_{jk}p_j p_k+0_p(T^{-3/2}),$$

since $p_j=0_p(T^{-1/2})$. Introducing the standardised variates $\bar{p}_j=\sqrt{T}p_j$, we have

$$\sqrt{T}\eta_T(p)=\eta_j\bar{p}_j+\frac{1}{2\sqrt{T}}\eta_{jk}\bar{p}_j\bar{p}_k+0_p(T^{-1}),$$

and we note that the derivatives of $\eta_T(q)$ and $e_T(q)$ at the origin are related as follows: $\eta_j=h\alpha^{h-1}e_j$, and $\eta_{jk}=h(h-1)\alpha^{h-2}e_je_k+h\alpha^{h-1}e_{jk}$. The joint distribution of $\hat{\alpha}^h$ and y_T is given by

$$P(\sqrt{T}(\hat{\alpha}^h-\alpha^h)\leqq x_1, y_T\leqq x_2)=P(\sqrt{T}\eta_T(p)\leqq x_1, y_T\leqq x_2),$$

and we now introduce the standardised statistics $q_1=\sqrt{T}\eta_T(p)/\omega$ and $q_2=y_T/\sigma_y$ where ω, which will be defined precisely later [by eq. (A.7)], tends as $T\to\infty$ to the variance of the limiting distribution of $\sqrt{T}\eta_T(p)$. The joint characteristic function of q_1 and q_2 is

$$\begin{aligned}\psi(t,s)&=E(\exp(itq_1+isq_2))=\int\exp(it\sqrt{T}\eta_T(p)/\omega+isp_2)\mathrm{d}F(\bar{p},y_T)\\&=\int\exp(isy_T/\sigma_y+it\eta_a\bar{p}_a/\omega)\\&\quad\times\left\{1+\frac{it}{2\omega\sqrt{T}}\eta_{ab}\bar{p}_a\bar{p}_b+0_p(T^{-1})\right\}\mathrm{d}F(\bar{p},y_T),\end{aligned}\tag{A.4}$$

where $F(\bar{p},y_T)$ denotes the joint distribution function of $(\bar{p},y_T)$.

Now the joint characteristic function of $(\bar{p},y_T)$ is given by

$$\begin{aligned}\mu(v,w)&=E(\exp(ivq_2+w'\bar{p}))\\&=\left|I-\frac{2i}{\sqrt{T}}F(w)\Omega\right|^{-1/2}\\&\quad\times\exp\left\{\tfrac{1}{2}(iv\Omega b)'\left(\Omega-\frac{2i}{\sqrt{T}}\Omega F(w)\Omega\right)^{-1}(iv\Omega b)\right\}\\&\quad\times\exp\left\{-\frac{iw_1}{\sqrt{T}}\mathrm{tr}(C_1\Omega)-\frac{iw_2}{\sqrt{T}}\mathrm{tr}(C_2\Omega)\right\},\end{aligned}$$

where $F(w)=w_1C_1+w_2C_2$. The second characteristic is

$$
\begin{aligned}
\chi(v,w) &= \log(\mu(v,w)) \\
&= -\tfrac{1}{2}\log\det\left(I-\frac{2i}{\sqrt{T}}F(w)\Omega\right) \\
&\quad -\frac{v^2}{2}b'\Omega\left(\Omega-\frac{2i}{\sqrt{T}}\Omega F(w)\Omega\right)^{-1}\Omega b \\
&\quad -\frac{iw_1}{\sqrt{T}}\operatorname{tr}(C_1\Omega)-\frac{iw_2}{\sqrt{T}}\operatorname{tr}(C_2\Omega).
\end{aligned} \tag{A.5}
$$

We will need to expand $\chi(v,w)$ in a Taylor series about the origin up to the third order so we introduce the following subscript notation for the derivatives of $\chi(v,w)$: the subscript v denotes differentiation with respect to v and the subscripts j, k and l indicate differentiation with respect to the components of w; and all derivatives are evaluated at the origin. Thus, we will have

$$\chi_{vv}=\partial^2\chi(0,\,0)/\partial v^2,$$

$$\chi_{vjk}=\partial^3\chi(0,\,0)/\partial v\partial w_j\partial w_k,$$

and

$$\chi_{jkl}=\partial^3\chi(0,\,0)/\partial w_j\partial w_k\partial w_l.$$

We note from the form of (A.5) that all first derivatives are zero and in addition $\chi_{vj}=0$, $\chi_{vjk}=0$ and $\chi_{vvv}=0$ for all j and k. Since derivatives of higher order than the third evaluated at the origin are of $0(T^{-1})$ we have the expansion

$$
\begin{aligned}
\chi(v,w) &= \tfrac{1}{2}\chi_{vv}v^2+\tfrac{1}{2}\chi_{jk}w_jw_k \\
&\quad +\tfrac{1}{6}(3\chi_{vvj}v^2w_j+\chi_{jkl}w_jw_kw_l)+0(T^{-1}),
\end{aligned}
$$

so that

$$
\begin{aligned}
\mu(v,w) &= \exp(\tfrac{1}{2}\chi_{vv}v^2+\tfrac{1}{2}\chi_{jk}w_jw_k) \\
&\quad \times\{1+\tfrac{1}{6}(3\chi_{vvj}v^2w_j+\chi_{jkl}w_jw_kw_l)\}+0(T^{-1}).
\end{aligned} \tag{A.6}
$$

From (A.4) we now have the representation

$$\psi(t,s)=\mu(s/\sigma_y, t\eta^0/\omega)$$
$$-\frac{it}{2\omega\sqrt{T}}\left(\frac{\partial^2\mu(s/\sigma_y, t\eta^0/\omega)}{\partial w_a \partial w_b}\right)\eta_{ab}+0(T^{-1}), \tag{A.7}$$

where $\eta^{0\prime}=(\eta_1,\eta_2)$, the vector of first derivatives of $\eta(\quad)$ at the origin.

We now define

$$\omega^2=-\chi_{jk}\eta_j\eta_k. \tag{A.8}$$

Then, using subscripts a and b to denote derivatives at the origin with respect to w_a and w_b, we find from (A.6) and (A.7) after some manipulation

$$\psi(t,s)=\exp\left(-\frac{s^2}{2}-\frac{t^2}{2}\right)\cdot\left[1-\frac{(it)^3}{6\omega^3}(i^3\chi_{jkl}\eta_j\eta_k\eta_l)\right.$$
$$+\frac{(it)^3}{2\omega^3\sqrt{T}}(\chi_{aj}\eta_j)(\chi_{bk}\eta_k)\eta_{ab}-\frac{(it)}{2\omega\sqrt{T}}(\chi_{ab}\eta_{ab})$$
$$\left.-\frac{(is)^2(it)}{2\sigma_y^2\omega}(i^3\chi_{vvj}\eta_j)\right]+0(T^{-1}).$$

Inverting, we find an expansion of the joint density of (q_1,q_2) given by

$$g(q_1,q_2)=i(q_1)i(q_2)\left[1-H_3(q_1)\left(\frac{1}{6\omega^3}\right)i^3\chi_{jkl}\eta_j\eta_k\eta_l\right.$$
$$+H_3(q_1)\left(\frac{1}{2\omega^3\sqrt{T}}\right)(\chi_{aj}\eta_j)(\chi_{bk}\eta_k)\eta_{ab}$$
$$-H_1(q_1)\left(\frac{1}{2\omega\sqrt{T}}\right)\chi_{ab}\eta_{ab}$$
$$\left.-H_1(q_1)H_2(q_2)\left(\frac{1}{2\sigma_y^2\omega}\right)i^3\chi_{vvj}\eta_j\right]+0(T^{-1}). \tag{A.9}$$

To obtain an explicit representation of (A.8) in terms of the parameters we need to evaluate the coefficients of the Hermite polynomials in the square brackets. Details of the derivations can be obtained from the author on request. We find that

$$\omega^2 = h^2\alpha^{2h-2}(1-\alpha^2),$$

$$i^3\chi_{jkl}\eta_j\eta_k\eta_l = \frac{-6}{\sqrt{T}} h^3\alpha^{3h-2}(1-\alpha^2) + 0(T^{-3/2}),$$

$$i^3\chi_{vvj}\eta_j = \frac{-2}{\sqrt{T}} h\alpha^h\sigma_y^2,$$

$$\chi_{ab}\eta_{ab} = -h(h-1)\alpha^{h-2}(1-\alpha^2) + 4h\alpha^h + 0(T^{-1}),$$

and

$$(\chi_{aj}\eta_j)(\chi_{bk}\eta_k)\eta_{ab} = h^3(h-1)\alpha^{3h-4}(1-\alpha^2)^2 - 4h^3\alpha 3^{h-2}(1-\alpha^2) + 0(T^{-1}).$$

Using these coefficients in (A.9) we obtain

$$\begin{aligned} g(q_1,q_2) = i(q_i)i(q_2)\Bigg[& 1 + H_3(q_1)\frac{1}{\sqrt{T}}\frac{\alpha}{(1-\alpha^2)^{1/2}} \\ & + \left\{ H_3(q_1) + H_1(q_1) \right\}\left(\frac{1}{2\sqrt{T}}\right) \\ & \times \left\{ (h-1)\frac{(1-\alpha^2)^{1/2}}{\alpha} - \frac{4\alpha}{(1-\alpha^2)^{1/2}} \right\} \\ & + H_1(q_1)H_2(q_2)\left(\frac{1}{\sqrt{T}}\right)\frac{\alpha}{(1-\alpha^2)^{1/2}} \Bigg] + 0(T^{-1}). \end{aligned}$$

From this joint density of (q_1, q_2) we can derive an expansion for the distribution function of $T(\hat{\alpha}^h - \alpha^h)$ given the value y_T. We have

$$\begin{aligned} P(\sqrt{T}(\hat{\alpha}^h - \alpha^h) \leqq x_1 | y_T) &= P(\sqrt{T}\eta_T(p) \leqq x_1 | y_T) \\ &= P(q_1 \leqq x/\omega | q_2 = y_T/\sigma_y) \\ &= \int_{-\infty}^{x_1/\omega} \frac{g(q_1, y_T/\sigma_y)}{i(y_T/\sigma_y)} \mathrm{d}q_1, \end{aligned}$$

and, upon integration, this reduces to

$$I\left(\frac{x_1}{\omega}\right) + i\left(\frac{x_1}{\omega}\right)\frac{1}{\sqrt{T}}\frac{\alpha}{(1-\alpha^2)^{1/2}}$$
$$\times \left[\left\{ 1 + \left(\frac{x_1}{\omega}\right)^2 \right\} - \left(\frac{h-1}{2}\right)\left(\frac{1-\alpha^2}{\alpha^2}\right)\left(\frac{x_1}{\omega}\right)^2 + \left\{ 1 - \left(\frac{y_T}{\sigma_y}\right)^2 \right\} \right] + 0(T^{-1}),$$

as given in the theorem.

Proof of Theorem 3

The characteristic function of $\hat{y}_{T+1}-y_{T+1}$ conditional on y_T is

$$\begin{aligned}
&E(\exp(it(\hat{y}_{T+1}-y_{T+1}))|y_T)\\
&=E(\exp(it(\hat{\alpha}-\alpha)y_T)|y_T)E(\exp(-itu_{T+1}))\\
&=\exp(-\tfrac{1}{2}t^2\sigma^2)E\left[\exp\left\{ity_T T^{-1}(1-\alpha^2)^{1/2}\left(\frac{\sqrt{T}(\hat{\alpha}-\alpha)}{(1-\alpha^2)^{1/2}}\right)\right\}\Bigg|\, y_T\right].
\end{aligned} \tag{A.10}$$

But, from the proof of Theorem 1 we have the following expansion of the characteristic function of $\sqrt{T}(\hat{\alpha}-\alpha)/(1-\alpha^2)^{1/2}$ conditional on y_T:

$$\begin{aligned}
&E[\exp(it\sqrt{T}(\hat{\alpha}-\alpha)/(1-\alpha^2)^{1/2})|y_T]\\
&=\exp(-\tfrac{1}{2}t^2)\\
&\quad\times\left[1-\frac{1}{\sqrt{T}}\frac{\alpha}{(1-\alpha^2)^{1/2}}\left\{2it+(it)^3+(it)\left(1-\left(\frac{y_T}{\sigma_y}\right)^2\right)\right\}\right]+0(T^{-1}).
\end{aligned} \tag{A.11}$$

Neglecting the remainder of $0(T^{-1})$ in this expansion, we then obtain from (A.10) and (A.11) an approximation to the characteristic function of $y_{T+1}-y_{T+1}$ conditional on y_T:

$$\begin{aligned}
&\exp(-\frac{t^2}{2}\{\sigma^2+(1-\alpha^2)y_T^2/T\})\\
&\quad\times\left[1-\frac{1}{T}2\alpha y_T(it)-\frac{1}{T^2}\alpha(1-\alpha^2)y_T^3(it)^3-\frac{1}{T}\alpha y_T\left(1-\left(\frac{y_T}{\sigma_y}\right)^2\right)(it)\right]\\
&=\exp(-t^2\sigma^2/2)\left[1-\frac{t^2}{2T}(1-\alpha^2)y_T^2+\frac{t^4}{8T^2}(1-\alpha^2)^2y_T^4\right]\\
&\quad\times\left[1-\frac{1}{T}2\alpha y_T(it)-\frac{1}{T^2}\alpha(1-\alpha^2)y_T^3(it)^3-\frac{1}{T}\alpha y_T\left(1-\left(\frac{y_T}{\sigma_y}\right)^2\right)(it)\right].
\end{aligned}$$

Inverting this expression term by term and neglecting those terms of $0(T^{-2})$

and smaller we find the following approximation to the density of $(\hat{y}_{T+1} - y_{T+1})/\sigma$ conditional on y_T:

$$i(x)+i'(x)\frac{1}{T}\left(\frac{2\alpha y_T}{\sigma}\right)+i^{(2)}(x)\frac{1}{2T}\frac{(1-\alpha^2)y_T^2}{\sigma^2}$$
$$+i'(x)\frac{1}{T}\frac{\alpha y_T}{\sigma}\left(1-\left(\frac{y_T}{\sigma_y}\right)^2\right)+0(T^{-2}).$$

Integrating this expression, we have the required expansion

$$P(\hat{y}_{T+1}-y_{T+1}\leqq x|y_T)$$
$$=I\left(\frac{x}{\sigma}\right)+i\left(\frac{x}{\sigma}\right)\left[\frac{1}{T}\left(\frac{2\alpha}{(1-\alpha^2)^{1/2}}\right)\left(\frac{y_T}{\sigma_y}\right)-\frac{1}{2T}\left(\frac{y_T}{\sigma_y}\right)^2\left(\frac{x}{\sigma}\right)\right.$$
$$\left.+\frac{1}{T}\left(\frac{\alpha}{(1-\alpha^2)^{1/2}}\right)\left(\frac{y_T}{\sigma_y}\right)\left(1-\left(\frac{y_T}{\sigma_y}\right)^2\right)\right]+0(T^{-2}).$$

References

Box, G.E.P. and G.M. Jenkins, 1970, Time series analysis: Forecasting and control (Holden Day, San Francisco, CA).

Fromm, G., L.R. Klein and G.R. Schink, 1972, Short- and long-term simulations with the Brookings model, in: B.G. Hickman, ed., Econometric models of cyclical behaviour (Columbia University Press, New York).

Green, G.R., 1972, Short- and long-term simulations with the O.B.E. econometric model, in: B.G. Hickman, ed., Econometric models of cyclical behaviour (Columbia University Press, New York).

Hurwicz, L., 1950, Least squares bias in time series, in: T.C. Koopmans, ed., Statistical inference in dynamic economic models (Wiley, New York).

Lukacs, E. and R.G. Laha, 1964, Applications of characteristic functions (Griffin, London).

Malinvaud, E., 1970, Statistical methods of econometrics (North-Holland, Amsterdam).

Orcutt, G.H. and H.S. Winokur, 1969, First order autoregression: Inference, estimation and prediction, Econometrica 37, 1–14.

Phillips, P.C.B., 1976, Large deviation expansions applied to tail probabilities in a first order stochastic difference equation, Discussion Paper Ser. A, no. 203 (University of Birmingham).

Phillips, P.C.B., 1977a, Approximation to some finite sample distributions associated with a first order stochastic difference equation, Econometrica 45, 463–485.

Phillips, P.C.B., 1977b, A general theorem in the theory of asymptotic expansions for approximations to the finite sample distribution of econometric estimators, Econometrica 45, 1517–1534.

Schmidt, P., 1974, The asymptotic distribution of forecasts in the dynamic simulation of an econometric model, Econometrica 42, 303–309.

Shenton, L.R. and W.L. Johnson, 1965, Moments of a serial correlation coefficient, Journal of the Royal Statistical Society B 27, 308–320.

Yamamoto, T., 1976, Asymptotic mean square prediction error for an autoregressive model with estimated coefficients, Applied Statistics 25, 123–127.