UNIVERSITY OF NOTTINGHAM

Discussion Papers in Economics

Discussion Paper
No. 00/13

ASYMPTOTIC MEAN SQUARED FORECAST ERROR WHEN AN AUTOREGRESSION WITH LINEAR TREND IS FITTED TO DATA GENERATED WITH AN I(0) OR I(1) PROCESS

by Tae-Hwan Kim, Stephen Leybourne and Paul Newbold

July 2000 DP 00/13

ISSN 1360-2438

UNIVERSITY OF NOTTINGHAM

Discussion Papers in Economics

Discussion Paper
No. 00/13

ASYMPTOTIC MEAN SQUARED FORECAST ERROR WHEN AN AUTOREGRESSION WITH LINEAR TREND IS FITTED TO DATA GENERATED WITH AN I(0) OR I(1) PROCESS

by Tae-Hwan Kim, Stephen Leybourne and Paul Newbold

Tae-Hwan Kim is Lecturer, and Stephen, Leybourne and Paul Newbold are Professors, School of Economics, University of Nottingham

July 2000

ASYMPTOTIC MEAN SQUARED FORECAST ERROR WHEN AN AUTOREGRESSION WITH LINEAR TREND IS FITTED TO DATA GENERATED BY AN I(0) OR I(1) PROCESS

Tae-Hwan Kim, Stephen J. Leybourne and Paul Newbold

University of Nottingham

May 2000

Abstract

Assume that a time series is generated by an autoregression which has at most one unit root. A correctly specified model, including linear time trend, is estimated by ordinary least squares, but no allowance is made for any unit root in the generating process. We investigate the impact of estimation error on the mean squared error of forecasts calculated from the fitted model.

1 INTRODUCTION

Consider a time series whose generating process is a finite order autoregression, possibly incorporating a single unit root or a linear time trend. Suppose that a model with linear trend and correctly specified autoregressive order, but not any unit autoregressive root, is fitted by ordinary least squares to a series of T observations. This fitted model is then employed in the usual way to forecast h steps ahead. The forecast error variance is then the sum of two components - one whose value is that which would occur if the parameters of the generating process were known, and a second of order T^{-1} following from sampling variability in the parameter estimators. Our purpose is to assess the second term, as T becomes large but the forecast horizon h remains fixed.

Our analysis follows a sequence of investigations, beginning with Yamamoto (1976), who derived the asymptotic mean squared error (a.m.s.e.) of forecasts obtained from a correctly specified model fitted to data generated by a finite order stationary autoregression with zero mean. Subsequently, this was extended by Yamamoto (1981) to the case of a mixed autoregressive-moving average generating process. Baillie (1979) extended the results of Yamamoto (1976) to the case of a linear regression model with fixed or stochastic exogenous variables and autoregressive errors. Fuller and Hasza (1981) generalised Yamamoto (1976) to the case where the stationary autoregressive generating process has unknown mean. Following some preliminary definitions and analysis in Section 2, in Section 3 of the paper we extend this line of research to the case where the generating process is a stationary autoregression around an unknown linear trend.

Fuller and Hasza (1981) also derived forecast a.m.s.e. when an autoregressive model is fitted to data generated by a unit root process, but provided an explicit expression only for the pure prediction error component, simply proving that the estimation error component is $O_p(T^{-1/2})$. In Section 4 of the paper we analyse this latter case in more detail, allowing a linear trend in the fitted model, as might be done when the generating process incorporates drift. It is straightforward to infer from our analysis corresponding results for the case of a driftless generating process and a fitted model including intercept but no trend. Our results apply to a finite forecast horizon h, and in particular it is interesting to compare them with the corresponding results of Section 3 for the trend-stationary generating model for one-step-ahead prediction, $h = 1$. Because the general case is algebraically complex, and therefore difficult to interpret, we also explicitly discuss the comparison for general h when the fitted model is a first order autoregression.

Several other authors have analysed this topic, including extensions to the multivariate case. However, either the correct order of integration is assumed known, as for example in Reinsel and Lewis (1987) and Sampson (1991), the forecast horizon h and sample size T are allowed to grow simultaneously, as in Stock (1996), Phillips (1998) and Kemp (1999), or, as in Clements and Hendry (1999), where the impact of an incorrect choice of the order of integration is assessed, no autoregressive terms are included in the fitted model when the data-generating process is difference-stationary.

2 MULTI-STEP PREDICTION ERROR

Let the series y_t be generated by an autoregressive process of order p, AR(p), with a linear time trend

$$y_t = \sum_{j=1}^{p} \phi_j y_{t-j} + \mu + \delta\left(\frac{t-1}{T}\right) + \varepsilon_t, \qquad t = 1, ..., T, \tag{1}$$

where the particular normalization of the time trend function has been chosen, without loss of generality, for clarity of presentation. We impose the following assumption on ε_t.

Assumption 1. *ε_t is distributed as $IID(0, \sigma^2)$ and $0 < \sigma^2 < \infty$.*

Define a vector of the true parameters $\phi \equiv (\phi_1, ..., \phi_p, \mu, \delta)'$ and a selection matrix $S_{nm} \equiv \begin{bmatrix} I_n & 0_{n\times(m-n)} \end{bmatrix}$ which selects the first n elements in an $m \times 1$ vector. To specify the $h-$steps ahead forecast of y_t, it is convenient to transform the DGP in (1) into a vector AR(1) process

$$Y_t = AY_{t-1} + E_t$$

where $Y_t \equiv (y_t, y_{t-1}, ..., y_{t-p+1}, 1, t/T)', E_t \equiv (\varepsilon_t, 0, ..., 0)'$ and

$$A \equiv \begin{bmatrix} A_1 & A_f \\ 0 & A_2 \end{bmatrix} \tag{2}$$

with $A_1 \equiv \begin{bmatrix} (S_{p(p+2)}\phi)' \\ I_{p-1} \quad 0_{(p-1)\times 1} \end{bmatrix}$, $A_2 \equiv \begin{bmatrix} 1 & 0 \\ T^{-1} & 1 \end{bmatrix}$, and $A_f \equiv \begin{bmatrix} \mu \quad \delta \\ 0_{(p-1)\times 2} \end{bmatrix}$.

Then, Y_{T+h} is given by

$$Y_{T+h} = A^h Y_T + \sum_{j=0}^{h-1} A^j E_{T+h-j}.$$

Assume that the regression model for (1) is correctly specified. Under this assumption, the OLS estimator $\hat{\phi} \equiv (\hat{\phi}_1, ..., \hat{\phi}_p, \hat{\mu}, \hat{\delta})'$ satisfies

$$y_t = \sum_{j=1}^{p} \hat{\phi}_j y_{t-j} + \hat{\mu} + \hat{\delta}\left(\frac{t-1}{T}\right) + e_t \tag{3}$$

where e_t is the regression residual. Given $\hat{\phi}$, A is estimated by $\hat{A} \equiv A(\hat{\phi})$. The $h-$steps ahead forecast of Y_t, based on $\hat{A}$, standing at time T is

$$\hat{Y}_{T+h} \equiv \hat{A}^h Y_T$$

and the forecast error is

$$\hat{Y}_{T+h} - Y_{T+h} = -\sum_{j=0}^{h-1} A^j E_{T+h-j} + (\hat{A}^h - A^h) Y_T.$$

The $h-$steps ahead forecast error of y_t is the first element of $\hat{Y}_{T+h} - Y_{T+h}$, and is given by

$$\hat{y}_{T+h} - y_{T+h} = -\sum_{j=0}^{h-1} e_1' A^j E_{T+h-j} + (e_1'\hat{A}^h - e_1' A^h) Y_T \tag{4}$$

where e_1 is a vector with one in the first element and zeros elsewhere. Assuming that $T^{1/2}(e_1'\hat{A}^h - e_1'A^h)Y_T \xrightarrow{d} \mathcal{F}$, we define the asymptotic mean squared error of the $h-$steps ahead forecast $\hat{y}_{T+h}$, denoted $a.m.s.e.\{\hat{y}_{T+h}\}$,

$$a.m.s.e.\{\hat{y}_{T+h}\} \equiv \sigma_h^2 + v_h^2$$

where $\sigma_h^2 \equiv var\left[-\sum_{j=0}^{h-1} e_1' A^j E_{T+h-j}\right]$ and $v_h^2 \equiv T^{-1} var(\mathcal{F})$. Three comments should be noted. First, the convergence condition $T^{1/2}(e_1'\hat{A}^h - e_1'A^h)Y_T \xrightarrow{d} \mathcal{F}$ is satisfied for most estimators of ϕ, including the OLS estimator $\hat{\phi}$ which will be used in this paper. Second, since it can be easily shown that

$$\sigma_h^2 = \sigma^2 \sum_{j=0}^{h-1} (e_1' A^j e_1)^2, \tag{5}$$

we will focus on the second component v_h^2. Finally, since it is easily shown that the $h-$steps forecast error is invariant to μ and δ, and hence so is the asymptotic mean squared error, we without loss of generality set $\mu = \delta = 0$ in all the derivations.

To obtain the asymptotic mean square error of $\hat{y}_{T+h}$ using the expression in (4), it is convenient to expand $e_1'\hat{A}^h$, viewed as a function of $\hat{\phi}$, as a Taylor series around the true value ϕ, which can be shown to give

$$e_1'\hat{A}^h = e_1'A^h + (\hat{\phi} - \phi)' \frac{\partial(e_1'A^h)}{\partial\phi} + o_p(T^{-1/2}). \tag{6}$$

One can show by using the methods in Yamamoto (1976) and Fuller and Hasza (1981) that

$$\frac{\partial(e_1'A^h)}{\partial\phi} = \sum_{j=0}^{h-1} (e_1' A'^j e_1) A^{h-1-j} \equiv M_h. \tag{7}$$

Note that, for $\mu = \delta = 0$, M_h is a 2×2 block diagonal matrix. We will denote the two matrices on the diagonal as M_{h1} and M_{h2}, with respective dimensions $p \times p$ and 2×2. Collecting (4), (6) and (7) yields for the $h-$steps prediction error

$$\hat{y}_{T+h} - y_{T+h} = -\sum_{j=0}^{h-1} e_1' A^j E_{T+h-j} + (\hat{\phi} - \phi)' M_h Y_T + o_p(T^{-1/2})$$

which will be the basis for our subsequent discussion.

3 ASYMPTOTIC MSE OF PREDICTION FOR TREND STATIONARY PROCESSES

In this section the results of Yamamoto (1976) and Fuller and Hasza (1981) are extended to the case where a linear trend term is included in the fitted model. We impose the following assumption on the AR(p) generating process for y_t in (1) to render it trend stationary

Assumption 2. *All the roots of the equation*

$$1 - \phi_1 z - \phi_2 z^2 - ... - \phi_p z^p = 0$$

lie outside the unit circle.

It is well known that under Assumptions 1 and 2 and the assumption $\mu = \delta = 0$, which as we have noted implies no loss of generality in the asymptotic mean squared forecast error,

$$T^{1/2}(\hat{\phi} - \phi) \xrightarrow{d} N(0, \sigma^2 \Gamma^{-1})$$

where $\Gamma \equiv \begin{bmatrix} \Gamma_1 & 0 \\ 0 & \Gamma_2 \end{bmatrix}$ with $\Gamma_1 \equiv E(X_t X_t')$, $X_t \equiv (y_t, y_{t-1}, ..., y_{t-p+1})'$ where now y_t is generated by a zero-mean autoregression, and $\Gamma_2 \equiv \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}$. Then, the following theorem immediately follows.

Theorem 1 *Suppose that Assumptions 1 and 2 hold. Then,*

$$a.m.s.e.\{\hat{y}_{T+h}\} = \sigma_h^2 + \upsilon_h^2$$

where σ_h^2 is given in (5) and

$$\upsilon_h^2 = \sigma^2 T^{-1} \left\{ tr(M_{h1}' \Gamma_1^{-1} M_{h1} \Gamma_1) + 4 \left[\sum_{j=0}^{h-1} (e_1' A_1'^j e_1) \right]^2 \right\}$$

where $M_{h1} = \sum_{j=0}^{h-1} (e_1' A_1'^j e_1) A_1^{h-1-j}$ which is the first $p \times p$ submatrix of M_h.

The general result of Theorem 1 is algebraically quite involved. In the special cases where either $p = 1$ or $h = 1$ considerable simplification results, yielding findings that are more easily interpretable and directly comparable with previous research. These two cases are covered by the following corollaries.

Corollary 1 *Suppose that Assumptions 1 and 2 hold and that* $p = 1$. *Then,*

$$a.m.s.e.\{\hat{y}_{T+h}\} = \sigma_h^2 + v_h^2$$

where

$$\sigma_h^2 = \sigma^2 \sum_{j=0}^{h-1} \phi_1^{2j}$$

and

$$v_h^2 = \sigma^2 T^{-1} \left\{ h^2 \phi_1^{2(h-1)} + 4 \left[(1-\phi_1)^{-1}(1-\phi_1^h) \right]^2 \right\}.$$

The result of Corollary 1 can be compared with findings of Yamamoto (1976) and Fuller and Hasza (1981), both of which obtain results of the form

$$v_h^2 = \sigma^2 T^{-1} \left\{ h^2 \phi_1^{2(h-1)} + k \left[(1-\phi_1)^{-1}(1-\phi_1^h) \right]^2 \right\}.$$

The former shows that, in the case where the intercept and slope terms are (correctly) both omitted from the regression (3), $k = 0$, while it follows from Corollary 2.1 of the latter that when only the slope term is omitted, $k = 1$. Hence, including a redundant trend term in the regression would add $3\sigma^2 T^{-1} \left[(1-\phi_1)^{-1}(1-\phi_1^h) \right]^2$ to forecast a.m.s.e.

Corollary 2 *Suppose that Assumptions 1 and 2 hold and that* $h = 1$. *Then,*

$$a.m.s.e.\{\hat{y}_{T+1}\} = \sigma_1^2 + v_1^2$$

where

$$\sigma_1^2 = \sigma^2$$

and

$$v_1^2 = \sigma^2 T^{-1}(p+4).$$

Again this result is easily related to those of Yamamoto (1976) and Fuller and Hasza (1981). For their respective assumed generating processes they find $k = 0$ and $k = 1$ in

$$v_1^2 = \sigma^2 T^{-1}(p+k).$$

Thus, including a redundant trend term in the regression would add $3\sigma^2 T^{-1}$ to the asymptotic mean squared error of one-step-ahead forecasts.

4 ASYMPTOTIC MSE OF PREDICTION FOR DIFFERENCE STATIONARY PROCESSES

In this section, although a model of the form (1) is fitted and projected forward for forecasting, the true generating process is difference-stationary, so that the following assumption is imposed.

Assumption 3. *The characteristic equation for the DGP in (1) can be factorised as*

$$1 - \phi_1 z - \phi_2 z^2 - ... - \phi_p z^p = (1 - z)(1 - a_1 z - a_2 z^2 - ... - a_{p-1} z^{p-1})$$

and all the roots of the equation

$$1 - a_1 z - a_2 z^2 - ... - a_{p-1} z^{p-1} = 0$$

lie outside the unit circle.

In the presence of difference-stationarity, the assumption $\delta = 0$ in the generating process (1) is standard. In what follows, we further assume, again without loss of generality, that $\mu = 0$ in that process. Choi (1993) derived the asymptotic distribution of the OLS estimators of the autoregressive parameters under Assumption 3 when constant and trend are not included in the model. Here, we follow his approach for the OLS estimator $\hat{\phi}$ in (3). Define a $(p+2) \times (p+2)$ square matrix B as follows

$$B \equiv \begin{bmatrix} B_1 & B_2 & 0_{p\times 2} \\ 0_{2\times(p-1)} & 0_{2\times 1} & I_2 \end{bmatrix}$$

where B_1 is a $p \times (p-1)$ matrix with (i, j) element 1 for $i = j$ $(i = 1, ..., p-1)$, -1 for $i = j+1$ $(i = 2, ..., p)$, and 0 elsewhere, and $B_2 \equiv (1, -a_1, -a_2, ..., -a_{p-1})'$. Specifically, recalling that $Y_t = (y_t, y_{t-1}, ..., y_{t-p+1}, 1, t/T)'$, the matrix B has the property

$$B'Y_t = (\Delta Y_t', \tilde{z}_t')' \tag{8}$$

where $\Delta Y_t \equiv (\Delta y_t, \Delta y_{t-1}, ..., \Delta y_{t-p+2})'$ is a vector with $p-1$ elements from which the deterministic components are excluded, and $\tilde{z}_t \equiv (z_t, 1, t/T)'$with $z_t = z_{t-1} + \varepsilon_t$. It can be easily verified that B is nonsingular. Next define a scaling matrix D_T as

$$D_T \equiv \begin{bmatrix} D_{T1} & 0 \\ 0 & D_{T2} \end{bmatrix}$$

where $D_{T1} \equiv T^{1/2} I_{p-1}$, and $D_{T2} \equiv \begin{bmatrix} T & 0 \\ 0 & T^{1/2} I_2 \end{bmatrix}$.

Lemma 1 *Let $\hat{\phi} = (\hat{\phi}_1, ..., \hat{\phi}_p, \hat{\mu}, \hat{\delta})'$ denote the OLS estimator in (3). Then, under Assumptions 1 and 3*

$$\hat{\phi} - \phi = BD_T^{-1} \begin{bmatrix} F_1^* \\ F_2^* \end{bmatrix} + B \begin{bmatrix} O_p(T^{-1}) \\ O_p(T^{-3/2}) \end{bmatrix}$$

where

$$\begin{aligned} F_1^* &\equiv (T^{-1} \sum_{t=p+1}^{T} \Delta Y_{t-1} \Delta Y_{t-1}')^{-1} T^{-1/2} \sum_{t=p+1}^{T} \Delta Y_{t-1} \varepsilon_t \\ F_2^* &\equiv (D_{T2}^{-1} \sum_{t=p+1}^{T} \tilde{z}_{t-1} \tilde{z}_{t-1}' D_{T2}^{-1})^{-1} D_{T2}^{-1} \sum_{t=p+1}^{T} \tilde{z}_{t-1} \varepsilon_t. \end{aligned}$$

Further,

$$F_1^* \xrightarrow{d} N(0, \sigma^2 \Gamma_a^{-1})$$

where $\Gamma_a \equiv E(\Delta Y_t \Delta Y_t')$, and

$$F_2^* \Rightarrow \Sigma \Im^{-1} \mathcal{H}$$

where

$$\Sigma \equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & \sigma \end{bmatrix}, \quad \Im \equiv \begin{bmatrix} \int W(r)^2 dr & \int W(r) dr & \int r W(r) dr \\ \int W(r) dr & 1 & 1/2 \\ \int r W(r) dr & 1/2 & 1/3 \end{bmatrix}$$

and

$$\mathcal{H} \equiv \begin{bmatrix} 1/2\{W(1)^2 - 1\} \\ W(1) \\ W(1) - \int W(r) dr \end{bmatrix}$$

where $W(r)$ is standard Brownian motion.

The following lemma provides an expression for the model estimation error component, scaled by $T^{1/2}$, of the error for an $h-$steps ahead forecast.

Lemma 2 *Suppose that Assumptions 1 and 3 hold. Then,*

$$T^{1/2} Y_T' M_h' (\hat{\phi} - \phi) = C_1^* + C_2^* + O_p(T^{-1/2})$$

where

$$\begin{aligned} C_1^* &\equiv \Delta Y_T' \sum_{j=0}^{h-1} \sum_{k=0}^{j} (e_1' A^{*\prime k} e_1) A^{*\prime (h-1-j)} F_1^* \\ C_2^* &\equiv \sum_{j=0}^{h-1} \sum_{k=0}^{j} (e_1' A^{*\prime k} e_1)(T^{-1/2} z_T, 1, 1) F_2^* \\ A^* &\equiv \begin{bmatrix} & a' & \\ I_{p-2} & & 0_{(p-2)\times 1} \end{bmatrix} \\ a &\equiv (a_1, a_2, ..., a_{p-1})'. \end{aligned}$$

In general, the terms C_1^* and C_2^* will not simplify further, but to provide some insight we rearrange them as

$$
\begin{aligned}
C_1^* &= \Delta Y_T' \left[\sum_{j=0}^{h-1} (e_1' A^{*\prime k} e_1) A^{*\prime (h-1-j)} + \sum_{j=0}^{h-1} \sum_{k=0}^{j-1} (e_1' A^{*\prime k} e_1) A^{*\prime (h-1-j)} \right] F_1^* \\
C_2^* &= \left[h + \sum_{j=0}^{h-1} (e_1' A^{*\prime k} e_1) + \sum_{j=0}^{h-1} \sum_{k=1}^{j-1} (e_1' A^{*\prime k} e_1) \right] (T^{-1/2} z_T, 1, 1) F_2^*.
\end{aligned}
$$

Written in this way, the double summation terms are essentially interaction terms that arise when a unit root and AR components are simultaneously present. For example, if the DGP of Δy_t is a stationary $AR(p-1)$ with constant and trend and there is no unit root present, then fitting a correctly specified model to this series would yield

$$
\begin{aligned}
C_1^* &= \Delta Y_T' \left[\sum_{j=0}^{h-1} (e_1' A^{*\prime k} e_1) A^{*\prime (h-1-j)} \right] F_1^* \\
C_2^* &= \left[\sum_{j=0}^{h-1} (e_1' A^{*\prime k} e_1) \right] (1, 1) F_2^*
\end{aligned}
\tag{9}
$$

where F_2^* is redefined, excluding the terms corresponding to z_t. The asymptotic mean squared error of the $h-$steps forecast calculated from C_1^* and C_2^* in (9) is essentially the same as the one given in Theorem 1. The only difference is that the dimension of the AR parameters is p rather than $p-1$.

The asymptotic mean squared error of the $h-$steps ahead forecast $\hat{y}_{T+h}$ for difference stationary processes immediately follows from Lemma 2.

Theorem 2 *Suppose that Assumptions 1 and 3 hold. Then,*

$$a.m.s.e.\{\hat{y}_{T+h}\} = \sigma_h^2 + v_h^2$$

where σ_h^2 *is given in (5) and*

$$v_h^2 = \sigma^2 T^{-1} \left\{ tr(\Gamma_a \Pi_h \Gamma_a^{-1} \Pi_h') + n_1 \pi_h^2 \right\}$$

where $\Pi_h \equiv \sum_{j=0}^{h-1} \sum_{k=0}^{j} (e_1' A^{*\prime k} e_1) A^{*\prime (h-1-j)}$, $\pi_h \equiv \sum_{j=0}^{h-1} \sum_{k=0}^{j} (e_1' A^{*\prime k} e_1)$, $n_1 \equiv var(\aleph' \Im^{-1} \mathcal{H})$ *and* $\aleph \equiv (W(1), 1, 1)'$.

As in the previous section, the general result is algebraically very complex, and it is useful to consider separately the special cases where $p = 1$ and where $h = 1$. The following corollaries set out the result for these cases.

Corollary 3 *Suppose that Assumptions 1 and 3 hold and that $p = 1$. Then,*

$$a.m.s.e.\{\hat{y}_{T+h}\} = \sigma_h^2 + \upsilon_h^2$$

where

$$\sigma_h^2 = h\sigma^2$$

and

$$\upsilon_h^2 = \sigma^2 T^{-1}(h^2 n_1) = 6T^{-1}h^2\sigma^2$$

where n_1 is given in Theorem 2.

The final conclusion of the corollary uses $n_1 = 6$, a result obtained through simulation, with a series of 5,000 independent standard normal variates employed to generate a random walk and the appropriately normalised sums to approximate the corresponding functionals.

Of course, the conclusion of Corollary 3, which gives the asymptotic mean squared error of forecasts obtained when a first order autoregression with linear trend is fitted to a random walk, can be obtained more directly using the result given by Clements and Hendry (1996) on the limiting distribution of $T(\hat{\rho}^h - 1)$, where $\hat{\rho}$ is the least squares estimator of the autoregressive parameter.

Corollary 4 *Suppose that Assumptions 1 and 3 hold and that $h = 1$. Then,*

$$a.m.s.e.\{\hat{y}_{T+1}\} = \sigma_1^2 + \upsilon_1^2$$

where

$$\sigma_1^2 = \sigma^2$$

and

$$\upsilon_1^2 = \sigma^2 T^{-1}\{(p-1) + n_1\} = \sigma^2 T^{-1}(p+5).$$

It is interesting to compare the conclusions of Corollaries 2 and 4. In particular, it emerges that it would be misleading to extend the result for a trend stationary generating process to the case where an autoregression with linear trend is fitted to data generated by a difference stationary process. Although the same number of parameters is estimated in each case, the a.m.s.e is higher by $\sigma^2 T^{-1}$ in the latter. Similarly it can be concluded that, if the analyst knows the correct order of integration in that case, and so fits to the first differences an autoregression of order $(p-1)$ with intercept but no trend, using the result of Fuller and Hasza (1981) a gain of $\sigma^2 T^{-1}(p+5) - \sigma^2 T^{-1}p = 5\sigma^2 T^{-1}$ in a.m.s.e. would result.

For completeness we consider two further simple cases. First, suppose that the true generating process is difference stationary with no drift, and that the trend term is excluded from the fitted model, which is now

$$y_t = \sum_{j=1}^{p} \hat{\phi}_j y_{t-j} + \hat{\mu} + e_t. \tag{10}$$

In this case, the asymptotic mean square error of the $h-$steps forecast of y_t is provided in the following theorem.

Theorem 3 *Suppose that Assumptions 1 and 3 hold and the $h-$steps ahead forecast is constructed based on the fitted model in (10). Then,*

$$a.m.s.e.\{\hat{y}_{T+h}\} = \sigma_h^2 + \upsilon_h^2$$

where σ_h^2 is given in (5) and

$$\upsilon_h^2 = \sigma^2 T^{-1} \left\{ tr(\Gamma_a \Pi_h \Gamma_a^{-1} \Pi_h') + n_2 \pi_h^2 \right\}$$

where $n_2 \equiv var\left[\aleph' S_{23}'(S_{23} \Im S_{23}')^{-1} S_{23} \mathcal{H}\right]$, and $S_{23} \equiv \left[\begin{array}{cc} I_2 & 0_{2\times 1} \end{array}\right]$.

Proofs of Theorem 3 and the corollaries that follow from it are along precisely the same line as those of Theorem 2 and Corollaries 3 and 4, and are therefore omitted. When $p = 1$ or $h = 1$, we have the following corollaries, the value $n_2 = 3$ having again been obtained by simulation.

Corollary 5 *Suppose that Assumptions 1 and 3 hold, $p = 1$ and the $h-$steps ahead forecast is constructed based on the fitted model in (10). Then,*

$$a.m.s.e.\{\hat{y}_{T+h}\} = \sigma_h^2 + \upsilon_h^2$$

where

$$\sigma_h^2 = h\sigma^2$$

and

$$\upsilon_h^2 = \sigma^2 T^{-1}(h^2 n_2) = 3T^{-1} h^2 \sigma^2$$

where n_2 is given in Theorem 3.

Corollary 6 *Suppose that Assumptions 1 and 3 hold, $h = 1$ and the $h-$steps ahead forecast is constructed based on the fitted model in (10). Then,*

$$a.m.s.e.\{\hat{y}_{T+1}\} = \sigma_1^2 + \upsilon_1^2$$

where

$$\sigma_1^2 = \sigma^2$$

and

$$\upsilon_1^2 = \sigma^2 T^{-1}\{(p-1) + n_2\} = \sigma^2 T^{-1}(p+2).$$

Comparing with Yamamoto (1976), notice that an analyst who knew the correct order of integration and fitted a zero-mean autoregression of order $(p-1)$ to the series of first differences would achieve a gain of $3\sigma^2T^{-1}$ in a.m.s.e. of one-step ahead forecasts.

Finally, suppose that the true process is driftless difference stationary and an autoregression of order p with no intercept or trend is fitted; that is

$$y_t = \sum_{j=1}^{p} \hat{\phi}_j y_{t-j} + e_t.$$

Then, it can be easily shown that the model estimation error component υ_h^2 in the a.m.s.e of the $h-$steps forecast is given by

$$\upsilon_h^2 = \sigma^2 T^{-1} \left\{ tr(\Gamma_a \Pi_h \Gamma_a^{-1} \Pi_h') + n_3 \pi_h^2 \right\}$$

where $n_3 \equiv var\left[\aleph' S_{13}'(S_{13}\Im S_{13}')^{-1} S_{13}\mathcal{H}\right]$, with $S_{13} \equiv \left[\begin{array}{cc} 1 & 0_{1\times 2} \end{array}\right]$. The expression for υ_h^2 is in the usual way specialised to

$$\upsilon_h^2 = \sigma^2 T^{-1}(h^2 n_3) = 2T^{-1}h^2\sigma^2$$

when $p = 1$ and

$$\upsilon_1^2 = \sigma^2 T^{-1} \left\{ (p-1) + n_3 \right\} = \sigma^2 T^{-1}(p+1)$$

when $h = 1$. The value $n_3 = 2$ has again been obtained from simulation.

5 SUMMARY

We have analysed the impact on prediction mean squared error when an autoregression with linear trend is estimated and the fitted model is projected forward to derive forecasts. By contrast with some previous work on this topic, although we let the number of observations available for estimation grow, the forecast horizon is kept fixed, so that our results are applicable to short-term forecasting. We assume that the fitted model is correctly specified, although it is not assumed that account is taken of any unit autoregressive root in the generating process.

Sections 3 and 4 of the paper consider respectively the cases of trend stationary and difference stationary processes. The results of the former provide an extension to the linear trend case of well known conclusions relating to the case of a fixed mean, which is either assumed known or must be estimated. In Section 4 we demonstrated how the analysis can be modified if the time series is difference stationary. In both sections, we found it useful to provide insight by specialising our general results to the cases of one step ahead prediction and of a fitted first order autoregression.

It is interesting to note that the results for the difference stationary case are somewhat different than would be obtained by taking the trend stationary results and allowing an autoregressive parameter to approach to unity.

References

[1] Baillie, R.T. (1979) The Asymptotic Mean Squared Error of Multi-step Prediction from the Regression Model with Autoregressive Errors. *Journal of the American Statistical Association* 74, 175-184.

[2] Choi, I. (1993) Asymptotic Normality of the Least-Squares Estimates for Higher Order Autoregressive Integrated Processes with Some Applications. *Econometric Theory* 9, 263-281.

[3] Clements, M.P. and D.F. Hendry (1996) Multi-step Estimation for Forecasting. *Oxford Bulletin of Economics and Statistics* 58, 657-684.

[4] Clements, M.P. and D.F. Hendry (1999) Forecasting with Difference-Stationary and Trend-Stationary Models. Manuscript, Department of Economics, Warwick University.

[5] Fuller, W.A. and D.P. Hasza (1981) Properties of Predictors for Autoregressive Time Series. *Journal of the American Statistical Association* 76, 155-161.

[6] Hamilton, J.D. (1994) Time Series Analysis. Princeton, New Jersey: Princeton University Press.

[7] Kemp, G.C.R. (1999) The Behavior of Forecast Errors from a Nearly Integrated AR(1) Model as Both Sample Size and Forecast Horizon Become Large. *Econometric Theory* 15, 238-256.

[8] Phillips, P.C.B. (1998) Impulse Response and Forecast Error Variance Asymptotics in Nonstationary VARs. *Journal of Econometrics* 83, 21-56.

[9] Reinsel, G.C. and A.L. Lewis (1987) Prediction Mean Square Error for Non-stationary Multivariate Time Series Using Estimated Parameters. *Economics Letters* 24, 57-61.

[10] Sampson, M. (1991) The Effect of Parameter Uncertainty on Forecast Variances and Confidence Intervals for Unit Root and Trend Stationary Time-series Models. *Journal of Applied Econometrics* 6, 67-76.

[11] Stock, J.H. (1996) VAR, Error Correction and Pretest Forecasts at Long Horizons. *Oxford Bulletin of Economics and Statistics* 58, 685-701.

[12] Yamamoto, T. (1976) Asymptotic Mean Square Prediction Error for an Autoregressive Model with Estimated Coefficients. *Journal of Applied Statistics* 25, 123-127.

[13] Yamamoto, T. (1981) Predictions of Multivariate Autoregressive-Moving Average Models. *Biometrika* 68, 485-492.

6 APPENDIX: PROOFS

6.1 Proof of Theorem 1

Since $\sigma_h^2 = E\left[-\sum_{j=0}^{h-1} e_1' A^j E_{T+h-j}\right]^2$, we have

$$\sigma_h^2 = \sum_{j=0}^{h-1} e_1' A^j E(E_{T+h-j}E_{T+h-j}')A^{j\prime} e_1 = \sigma^2 \sum_{j=0}^{h-1} (e_1' A^j e_1)^2$$

as given in equation (5).

The OLS estimator $\hat{\phi} = (\hat{\phi}_1, ..., \hat{\phi}_p, \hat{\mu}, \hat{\delta})'$ can be expressed as

$$\begin{aligned} T^{1/2}(\hat{\phi} - \phi) &= \left[T^{-1} \sum_{t=p+1}^{T} Y_{t-1}Y_{t-1}{}'\right]^{-1} T^{-1/2} \sum_{t=p+1}^{T} Y_{t-1}\varepsilon_t \\ &= \left[\begin{array}{c} F_1 \\ F_2 \end{array}\right] + O_p(T^{-1/2}) \end{aligned}$$

in the case $\mu = \delta = 0$, where

$$\begin{aligned} F_1 &\equiv \left[T^{-1} \sum_{t=p+1}^{T} X_{t-1}X_{t-1}{}'\right]^{-1} T^{-1/2} \sum_{t=p+1}^{T} X_{t-1}\varepsilon_t, \\ F_2 &\equiv \left[T^{-1} \sum_{t=p+1}^{T} x_{t-1}x_{t-1}{}'\right]^{-1} T^{-1/2} \sum_{t=p+1}^{T} x_{t-1}\varepsilon_t \end{aligned}$$

and $x_t \equiv (1, t/T)'$. It is straightforward to show that

$$F_i \xrightarrow{d} N(0, \sigma^2\Gamma_i^{-1})$$

where $i = 1, 2$ and F_1 and F_2 are asymptotically independent since they are jointly normal and the asymptotic covariance matrix is zero.

Hence, the model estimation component of the error for an $h-$steps ahead forecast, scaled up by $T^{1/2}$, is given by

$$T^{1/2}Y_T' M_h'(\hat{\phi} - \phi) = Y_T' M_h' \left[\begin{array}{c} F_1 \\ F_2 \end{array}\right] + O_p(T^{-1/2}).$$

It can be shown that

$$\begin{aligned} e_1' A'^j e_1 &= e_1' A_1'^j e_1 \\ Y_T' A'^j &= (X_T' A_1'^j, 1, 1) + O(T^{-1}). \end{aligned}$$

Hence, we have

$$\begin{aligned} Y_T' M_h' &= Y_T' \sum_{j=0}^{h-1} (e_1' A_1'^j e_1) A'^{(h-1-j)} \\ &= \left[\begin{array}{cc} X_T' \sum_{j=0}^{h-1} (e_1' A_1'^j e_1) A_1'^{(h-1-j)} & \sum_{j=0}^{h-1} (e_1' A'^j e_1)(1,1) + O(T^{-1}) \end{array} \right]. \end{aligned}$$

Therefore, the model estimation component of the error for the $h-$steps ahead forecast is

$$\begin{aligned} & T^{1/2} Y_T' M_h' (\hat{\phi} - \phi) \\ = & \; X_T' \sum_{j=0}^{h-1} (e_1' A_1'^j e_1) A_1'^{(h-1-j)} F_1 + \sum_{j=0}^{h-1} (e_1' A_1'^j e_1)(1,1) F_2 + O_p(T^{-1/2}) \\ = & \; X_T' M_{h1} F_1 + \sum_{j=0}^{h-1} (e_1' A_1'^j e_1)(1,1) F_2 + O_p(T^{-1/2}) \\ \equiv & \; C_1 + C_2 + O_p(T^{-1/2}) \end{aligned}$$

where the last line defines C_1 and C_2. Since $F_i \xrightarrow{d} N(0, \sigma^2 \Gamma_i^{-1})$ for $i = 1, 2$, the asymptotic variances of C_1 and C_2 are given by

$$\begin{aligned} asyvar\{C_1\} &= \sigma^2 tr(M_{h1}' \Gamma_1^{-1} M_{h1} \Gamma_1) \\ asyvar\{C_2\} &= \sigma^2 \left[\sum_{j=0}^{h-1} (e_1' A'^j e_1) \right]^2 (1,1) \Gamma_2^{-1} (1,1)' \\ &= 4\sigma^2 \left[\sum_{j=0}^{h-1} (e_1' A_1'^j e_1) \right]^2 . \end{aligned}$$

Using the well known fact that X_T is asymptotically uncorrelated with $\hat{\phi}$ (hence with F_1 and F_2), we have

$$v_h^2 = \sigma^2 T^{-1} \left\{ tr(M_{h1}' \Gamma_1^{-1} M_{h1} \Gamma_1) + 4 \left[\sum_{j=0}^{h-1} (e_1' A_1'^j e_1) \right]^2 \right\}.$$

6.2 Proof of Corollary 1

When $p = 1$, $e_1' A^j e_1 = \phi_1^j$. Hence, $\sigma_h^2 = \sigma^2 \sum_{j=0}^{h-1} \phi_1^{2j}$. Also note that the terms in the expression for v_h^2 in Theorem 1 are simplified as follows.

$$\begin{aligned} M_{h1} &= h \phi_1^{(h-1)}. \\ \Gamma_1 &= E(y_t^2). \\ e_1' A_1'^j e_1 &= \phi_1. \end{aligned}$$

Hence, $tr(M_{h1}'\Gamma_1^{-1}M_{h1}\Gamma_1) = h^2\phi_1^{2(h-1)}$ and $\sum_{j=0}^{h-1}(e_1'A_1'^j e_1) = (1-\phi_1)^{-1}(1-\phi_1^h)$, which delivers the desired result.

6.3 Proof of Corollary 2

Note that $M_{h1} = I_{p+2}$ when $h = 1$. Hence, $tr(M_{11}'\Gamma_1^{-1}M_{11}\Gamma_1) = tr(I_p) = p$.

Since $A^0 = I$, $\sum_{j=0}^{h-1}(e_1'A_1'^j e_1) = 1$ which completes the proof.

6.4 Proof of Lemma 1

The OLS estimator $\hat{\phi} = (\hat{\phi}_1, ..., \hat{\phi}_p, \hat{\mu}, \hat{\delta})'$ can be expressed as

$$\begin{aligned}
\hat{\phi} - \phi &= B\left[\sum_{t=p+1}^{T} B'Y_{t-1}(B'Y_{t-1})'\right]^{-1}\sum_{t=p+1}^{T} B'Y_{t-1}\varepsilon_t \\
&= BD_T^{-1}\left[D_T^{-1}\sum_{t=p+1}^{T} B'Y_{t-1}(B'Y_{t-1})'D_T^{-1}\right]^{-1} D_T^{-1}\sum_{t=p+1}^{T} B'Y_{t-1}\varepsilon_t.
\end{aligned}$$

It is straightforward to show using the property in (8) that

$$\begin{aligned}
& D_T^{-1}\sum_{t=p+1}^{T} B'Y_{t-1}(BY_{t-1})'D_T^{-1} \\
= & \begin{bmatrix} T^{-1}\sum_{t=p+1}^{T}\Delta Y_{t-1}\Delta Y_{t-1}' & O_p(T^{-1/2}) \\ O_p(T^{-1/2}) & D_{T2}^{-1}\sum_{t=p+1}^{T}\tilde{z}_{t-1}\tilde{z}_{t-1}'D_{T2}^{-1} \end{bmatrix}
\end{aligned}$$

so that

$$\begin{aligned}
& \left[D_T^{-1}\sum_{t=p+1}^{T} B'Y_{t-1}(B'Y_{t-1})'D_T^{-1}\right]^{-1} \qquad (11) \\
= & \begin{bmatrix} (T^{-1}\sum_{t=p+1}^{T}\Delta Y_{t-1}\Delta Y_{t-1}')^{-1} & 0 \\ 0 & (D_{T2}^{-1}\sum_{t=p+1}^{T}\tilde{z}_{t-1}\tilde{z}_{t-1}'D_{T2}^{-1})^{-1} \end{bmatrix} + O_p(T^{-1/2}).
\end{aligned}$$

Further, since

$$D_T^{-1}\sum_{t=p+1}^{T} B'Y_{t-1}\varepsilon_t = \begin{bmatrix} T^{-1/2}\sum_{t=p+1}^{T}\Delta Y_{t-1}\varepsilon_t \\ D_{T2}^{-1}\sum_{t=p+1}^{T}\tilde{z}_{t-1}\varepsilon_t \end{bmatrix}, \qquad (12)$$

it follows from (11) and (12) that

$$\hat{\phi} - \phi = BD_T^{-1}\begin{bmatrix} F_1^* \\ F_2^* \end{bmatrix} + B\begin{bmatrix} O_p(T^{-1}) \\ O_p(T^{-3/2}) \end{bmatrix} \qquad (13)$$

where F_1^* and F_2^* are as given in the statement of the lemma. Using standard results in Hamilton (1994), it can then be shown that

$$F_1^* \xrightarrow{d} N(0, \sigma^2 \Gamma_a^{-1})$$

where $\Gamma_a \equiv E(\Delta Y_t \Delta Y_t')$, and

$$F_2^* \Rightarrow \Sigma \Im^{-1} \mathcal{H}.$$

6.5 Proof of Lemma 2

It can be shown using (13) that the model estimation component of the error for an $h-$steps ahead forecast, scaled up by $T^{1/2}$, is given by

$$T^{1/2} Y_T' M_h' (\hat{\phi} - \phi) = T^{1/2} Y_T' M_h' B \left(D_T^{-1} \begin{bmatrix} F_1^* \\ F_2^* \end{bmatrix} + \begin{bmatrix} O_p(T^{-1}) \\ O_p(T^{-3/2}) \end{bmatrix} \right).$$

Recalling the expression for M_h

$$M_h = \sum_{j=0}^{h-1} (e_1' A'^j e_1) A^{h-1-j},$$

it can be shown that

$$\begin{aligned} e_1' A'^j e_1 &= \sum_{k=0}^{j} (e_1' A^{*\prime k} e_1) \\ Y_T' A'^j B &= \begin{bmatrix} \Delta Y_T' A^{*\prime j} & \tilde{z}_T' + O(T^{-1}) \end{bmatrix} \end{aligned}$$

(in fact $\tilde{z}_T' + O(T^{-1}) = (z_T, 1, 1 + O(T^{-1}))'$) where A^* is given in Lemma 2. Hence, we have

$$\begin{aligned} T^{1/2} Y_T' M_h' B &= T^{1/2} Y_T' \sum_{j=0}^{h-1} (e_1' A'^j e_1) A'^{(h-1-j)} B \\ &= T^{1/2} \sum_{j=0}^{h-1} (e_1' A'^j e_1) Y_T' A'^{(h-1-j)} B \\ &= T^{1/2} \sum_{j=0}^{h-1} (e_1' A'^j e_1) \begin{bmatrix} \Delta Y_T' A^{*\prime(h-1-j)} & \tilde{z}_T' + O(T^{-1}) \end{bmatrix} \\ &= T^{1/2} \begin{bmatrix} \Delta Y_T' \sum_{j=0}^{h-1} (e_1' A'^j e_1) A^{*\prime(h-1-j)} & \sum_{j=0}^{h-1} (e_1' A'^j e_1) \tilde{z}_T' + O(T^{-1}) \end{bmatrix}. \end{aligned}$$

Therefore, the model estimation component of the error for the $h-$steps ahead forecast is

$$
\begin{aligned}
& T^{1/2} Y_T' M_h' (\hat{\phi} - \phi) \\
= \; & T^{1/2} \left[\; \Delta Y_T' \sum_{j=0}^{h-1} (e_1' A'^j e_1) A^{*\prime(h-1-j)} \quad \sum_{j=0}^{h-1} (e_1' A'^j e_1) \tilde{z}_T' + O(T^{-1}) \; \right] \\
& \times \left[D_T^{-1} \left[\begin{array}{c} F_1^* \\ F_2^* \end{array} \right] + \left[\begin{array}{c} O_p(T^{-1}) \\ O_p(T^{-3/2}) \end{array} \right] \right] \\
= \; & \Delta Y_T' \sum_{j=0}^{h-1} (e_1' A'^j e_1) A^{*\prime(h-1-j)} F_1^* + \sum_{j=0}^{h-1} (e_1' A'^j e_1)(T^{-1/2} z_T, 1, 1) F_2^* + O_p(T^{-1/2}).
\end{aligned}
$$

On substituting in for $e_1' A'^j e_1$ we can write this as

$$
T^{1/2} Y_T' M_h' (\hat{\phi} - \phi) = C_1^* + C_2^* + O_p(T^{-1/2})
$$

where

$$
\begin{aligned}
C_1^* &= \Delta Y_T' \sum_{j=0}^{h-1} \sum_{k=0}^{j} (e_1' A^{*\prime k} e_1) A^{*\prime(h-1-j)} F_1^* \\
C_2^* &= \sum_{j=0}^{h-1} \sum_{k=0}^{j} (e_1' A^{*\prime k} e_1)(T^{-1/2} z_T, 1, 1) F_2^*.
\end{aligned}
$$

6.6 Proof of Theorem 2

To obtain the asymptotic variance of $T^{1/2} Y_T' M_h' (\hat{\phi} - \phi)$, we will show later that C_1^* and C_2^* are asymptotically uncorrelated. Since $F_1^* \xrightarrow{d} N(0, \sigma^2 \Gamma_a^{-1})$, the asymptotic variance of C_1^* is given by

$$
asyvar\{C_1^*\} = \sigma^2 tr(\Gamma_a \Pi_h \Gamma_a^{-1} \Pi_h')
$$

where

$$
\Pi_h \equiv \sum_{j=0}^{h-1} \sum_{k=0}^{j} (e_1' A^{*\prime k} e_1) A^{*\prime(h-1-j)}
$$

Noting that $(T^{-1/2} z_T, 1, 1) F_2^* \Rightarrow \sigma \aleph' \Im^{-1} \mathcal{H}$, the asymptotic variance of C_2^* is given by:

$$
asyvar\{C_2^*\} = \sigma^2 n_1 \pi_h^2
$$

where

$$
\begin{aligned}
n_1 &= var(\aleph' \Im^{-1} \mathcal{H}) \\
\pi_h &= \sum_{j=0}^{h-1} \sum_{k=0}^{j} (e_1' A^{*\prime k} e_1)
\end{aligned}
$$

To show asymptotic uncorrelatedness between C_1^* and C_2^*, we focus on the simple case where no intercept and trend are included in the regression model for clarity of presentation. The proof goes through in the general case. We write C_1^* and C_2^* as

$$
\begin{aligned}
C_1^* &= \Delta Y_T' Q_1 (T^{-1} \sum_{t=p+1}^{T} \Delta Y_{t-1} \Delta Y_{t-1}')^{-1} T^{-1/2} \sum_{t=p+1}^{T} \Delta Y_{t-1} \varepsilon_t \\
&= \Delta Y_T' Q_1 \Gamma^{-1} T^{-1/2} \sum_{t=p+1}^{T} \Delta Y_{t-1} \varepsilon_t + O_p(T^{-1/2}), \\
C_2^* &= q_2 T^{-1/2} z_T (T^{-2} \sum_{t=p+1}^{T} z_{t-1}^2)^{-1} T^{-1} \sum_{t=p+1}^{T} z_{t-1} \varepsilon_t
\end{aligned}
$$

where Q_1 and q_2 are constants. We now simplify each term in C_1^* and C_2^* as follows.

$$
\begin{aligned}
T^{-1/2} \sum_{t=p+1}^{T} \Delta Y_{t-1} \varepsilon_t &= T^{-1/2} \sum_{t=p+1}^{T-k} \Delta Y_{t-1} \varepsilon_t + T^{-1/2} \sum_{t=T-k+1}^{T} \Delta Y_{t-1} \varepsilon_t \\
&= T^{-1/2} \sum_{t=p+1}^{T-k} \Delta Y_{t-1} \varepsilon_t + O_p\{(k/T)^{1/2}\}, \\
T^{-1/2} z_T &= T^{-1/2} z_{T-k} + T^{-1/2} \sum_{t=T-k+1}^{T} \varepsilon_t \\
&= T^{-1/2} z_{T-k} + O_p\{(k/T)^{1/2}\}, \\
T^{-2} \sum_{t=p+1}^{T} z_{t-1}^2 &= T^{-2} \sum_{t=p+1}^{T-k} z_{t-1}^2 + T^{-2} \sum_{t=T-k+1}^{T} z_{t-1}^2 \\
&= T^{-2} \sum_{t=p+1}^{T-k} z_{t-1}^2 + O_p\{(k/T)^2\}, \\
T^{-1} \sum_{t=p+1}^{T} z_{t-1} \varepsilon_t &= T^{-1} \sum_{t=p+1}^{T-k} z_{t-1} \varepsilon_t + T^{-1} \sum_{t=T-k+1}^{T} z_{t-1} \varepsilon_t \\
&= T^{-1} \sum_{t=p+1}^{T-k} z_{t-1} \varepsilon_t + O_p(k/T).
\end{aligned}
$$

Hence, we have

$$
\begin{aligned}
C_1^* &= \Delta Y_T' Q_1 \Gamma^{-1} T^{-1/2} \sum_{t=p+1}^{T-k} \Delta Y_{t-1} \varepsilon_t + O_p\{(k/T)^{1/2}\}, \\
C_2^* &= q_2 T^{-1/2} z_{T-k} (T^{-2} \sum_{t=p+1}^{T-k} z_{t-1}^2)^{-1} T^{-1} \sum_{t=p+1}^{T-k} z_{t-1} \varepsilon_t + O_p\{(k/T)^{1/2}\}.
\end{aligned}
$$

We assume that $k = O(T^{1-\delta})$, where $0 < \delta < 1$. Then

$$\begin{aligned} C_1^* &= \Delta Y_T' Q_1 \Gamma^{-1} T^{-1/2} \sum_{t=p+1}^{T-k} \Delta Y_{t-1} \varepsilon_t + O_p(T^{-\delta/2}), \\ C_2^* &= q_2 T^{-1/2} z_{T-k} (T^{-2} \sum_{t=p+1}^{T-k} z_{t-1}^2)^{-1} T^{-1} \sum_{t=p+1}^{T-k} z_{t-1} \varepsilon_t + O_p(T^{-\delta/2}). \end{aligned}$$

Since $\Delta Y_T'$ is a zero-mean stationary $AR(p-1)$ process, this term is asymptotically uncorrelated with all the other random variables in C_1^* and C_2^* which completes the proof.

6.7 Proof of Corollary 3

When $p = 1$, $\Delta Y_T, A^*$ and F_1^* do not exist. Hence, C_1^* does not exist. On the other hand,

$$C_2^* = h(T^{-1/2} z_T, 1, 1) F_2^*$$

which implies that

$$\upsilon_h^2 = \sigma^2 T^{-1} (h^2 n_1)$$

where $n_1 = var(\aleph' \Im^{-1} \mathcal{H})$.

6.8 Proof of Corollary 4

When $h = 1$, we have

$$\begin{aligned} C_1^* &= \Delta Y_T' F_1^* \\ C_2^* &= (T^{-1/2} z_T, 1, 1) F_2^* \end{aligned}$$

which implies that

$$\upsilon_1^2 = \sigma^2 T^{-1} \left[(p-1) + n_1 \right].$$