The Chi-Square Approximation of the Restricted Likelihood Ratio Test for the Sum of Autoregressive Coefficients with Interval Estimation

Willa W. Chen *   Rohit S. Deo †

November 4, 2007

Abstract: The restricted likelihood (RL) of an autoregressive (AR) process of order one with intercept/trend possesses enormous advantages, such as yielding estimates with significantly reduced bias, powerful unit root tests, small curvature and a well-behaved likelihood ratio test ($RLRT$) near the unit root. We consider the likelihood ratio test based on the Restricted Likelihood ($RLRT$) for the sum of the coefficients in $AR(p)$ processes with intercept/trend. The limit of the leading error term in the chi-square approximation to the $RLRT$ distribution is shown to be finite as the unit root is approached, suggesting a good approximation over the entire parameter space and well-behaved interval inference for nearly integrated processes. Our result is stronger than the pointwise first order result currently available for competing confidence interval procedures in the intercept/trend model. We extend the correspondence between the AR coefficients and the partial autocorrelations from the stationary to the unit root case. The resulting parameter space is the bounded p-dimensional hypercube $(-1, 1] \times (-1, 1)^{p-1}$, which greatly simplifies both computation and optimisation of the Restricted Likelihood (RL) as well as the computation of confidence intervals for the sum of the AR coefficients. In simulations, we show that the $RLRT$ intervals have almost exact coverage and also shorter lengths and significantly higher power against stationary alternatives than other competing intervals. An application to the Nelson-Plosser data yields intervals that can be markedly different from those in the literature.

Keywords: near unit root, Restricted Maximum Likelihood Estimator, autoregressive, Bartlett correction

*Department of Statistics, Texas A&M University, College Station, Texas 77843, USA

†University of Texas-Austin, Texas. Part of this work was completed when Deo was at New York University.

1 Introduction:

We will assume that we observe $(X_1, ..., X_n)$ from the general $AR(p)$ model with intercept and/or trend given by

$$X_t = \beta_0 + \beta_1 t + U_t, \tag{1}$$

where

$$U_t = \sum_{i=1}^{p} \alpha_i U_{t-i} + v_t$$

and $v_t \sim i.i.d.\ N\left(0, \sigma^2\right)$. When the roots $\{z_i\}_{i=1}^p$ of the polynomial $g(z) = z^p - \sum_{i=1}^p \alpha_i z^{p-i}$ are strictly less then one in absolute value, we will assume that the initial values $(U_0, ..., U_{1-p})$ are drawn from their stationary distribution, thus ensuring that $(X_1, ..., X_n)$ are trend stationary. In the case of a unit root $\sum_{i=1}^p \alpha_i = 1$, we will assume that the initial first differences $(U_1 - U_0, U_0 - U_{-1}, ..., U_{3-p} - U_{2-p})$ are drawn from their stationary distribution, ensuring that $(X_2 - X_1, ..., X_n - X_{n-1})$ are strictly stationary. In many applications, one is interested in generating confidence intervals for the sum of the autoregressive coefficients $\sum_{i=1}^p \alpha_i$ which is a measure of how close the process is to being non-stationary (Andrews and Chen 1994). However, inference in such models is known to be particularly difficult since the usual asymptotic normal approximation to the distribution of the appropriate t-statistic is very poor near the unit root in finite samples. Phillips (1977) demonstrated this theoretically for the $AR(1)$ process with known mean by obtaining the Edgeworth expansion of the t-statistic for the AR coefficient α_1. Phillips (1977) showed that

$$P\left(t_{\hat{\alpha}_1} \leq x\right) = F_Z(x) + \frac{\alpha_1}{\sqrt{1-\alpha_1^2}} \left\{ \frac{1+3\alpha_1^2}{1-\alpha_1^2} + 2\left[\frac{1+\alpha_1^2}{1-\alpha_1^2}\right]^2 x^2 \right\} \frac{f_z(x)}{\sqrt{n}} + O\left(n^{-1}\right), \tag{2}$$

where $t_{\hat{\alpha}_1}$ is the t-statistic for α_1 based on the OLS estimator, while F_Z and f_Z are the c.d.f and the p.d.f. of a standard normal variable, respectively. It is obvious from the leading error term in this expansion that the quality of the asymptotic normal approximation deteriorates markedly as α_1 approaches unity. This can be seen very effectively in the plot on the left in Figure 1,where we plot the empirical densities of the t-statistic for various values of α_1 based on a sample of size 100, as well as the limiting standard normal density. It is seen that the density of the t-statistic deviates further and further from that of the standard normal as the unit root is approached. As a consequence, the standard method of constructing confidence intervals for α_1

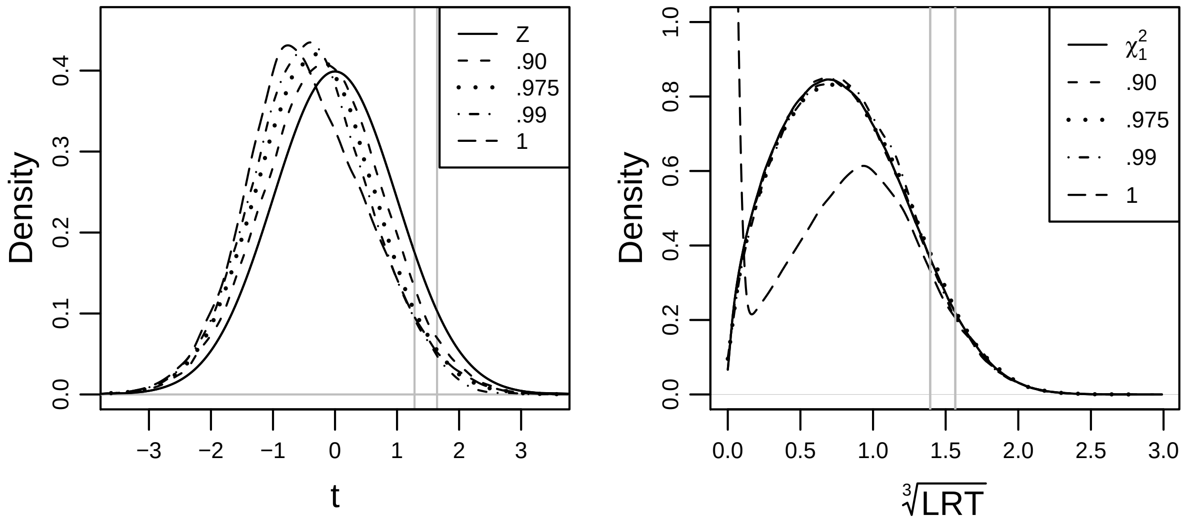

Figure 1: Empirical densities of t statistics and cubic root transformed LRT statistics of AR(1) processes without intercept. The vertical lines are 90th and 95th percentiles. Both plots are based on 100,000 repetitions of an AR(1) with sample size $n = 100$ and AR coefficient $\alpha = .9, .975, .99$ and 1.

by inverting the t-statistic does not work well in finite samples for autoregressive processes. This problem persists when using the normal approximation for the distribution of the t-statistic for the sum of the $AR(p)$ coefficients, $\sum_{i=1}^{p} \alpha_i$, in higher order AR processes. The various methods that have been suggested in the literature to circumvent this problem include using a parametric grid bootstrap (Andrews, 1991 and Andrews and Chen 1994), a non-parametric grid-bootstrap (Hansen, 1999) and using a local-to-unity formulation (Phillips, Moon and Xiao, 2001, Stock, 1991 and Elliott and Stock, 2001). Hansen (1999) proved the pointwise (over the parameter space) first-order convergence of the grid bootstrap distribution to the distribution of the t-statistic for $\sum \alpha_i$ in the $AR(p)$ process with intercept/trend in (1). Under the assumption of zero intercept and trend (i.e. $\beta_0 = \beta_1 = 0$ in (1)) Mikusheva (2007) proved that this first-order convergence is actually uniform over the parameter space for higher order AR processes, regardless of how close the largest root may be to unity. However, as we will see below, it is precisely the presence of the intercept/trend that causes problems for inference in AR processes near the unit root and if the intercept may be assumed to be zero, the standard LRT provides well behaved inference. Furthermore, it is unrealistic in practice to assume that the intercept/trend is known and most empirical economic series, such as those considered in Nelson and Plosser (1982), require the inclusion of a trend in addition to an intercept. Mikusheva (2007), however, does not provide any result for the higher order $AR(p)$ model with intercept/trend, nor does her proof extend to this case. Hence, for $AR(p)$ models with intercept/trend, which is the only case that is both problematic and empirically relevant, the strongest current result for the grid bootstrap procedure is the pointwise first-order convergence result of Hansen (1999).

As stated above, it is the presence of the intercept/trend in the AR model that causes problems for inference near the unit root. To see this, we first note that for the Likelihood Ratio Test (LRT) for α_1 in $AR(1)$ models with zero mean/intercept, Chen and Deo (2006a) show that

$$P\left(LRT \leq x\right) = P\left(\chi_1^2 \leq x\right) - \frac{0.25}{n}\left\{G_3\left(x\right) - G_1\left(x\right)\right\} + O\left(n^{-2}\right), \tag{3}$$

where $G_s\left(\cdot\right)$ is the c.d.f. of a χ_s^2 variable. This expansion implies that the LRT distribution is pointwise well approximated by the χ^2 distribution in finite samples. Simulation results in van Giersbergen (2006a) suggest that this continues to be true close to the unit root. This can also be seen in the empirical densities of the LRT for α_1 in a zero-mean model shown in the plot on the right in Figure 1, which are plotted together with the limiting χ_1^2 density. (Since a χ_1^2 density is very right skewed, we plot the density of the cube root of the LRT to ensure a density that looks more symmetric in order to make the comparisons in the right tail clearer). These empirical densities are seen to be remarkably well-approximated by the limiting distribution (in this case a χ_1^2), both when α_1 is far from the unit root as well as when α_1 is close to unity, in stark contrast to the situation in Figure 1 for the t-statistic. The good quality of the χ_1^2 approximation close to the unit root can be explained by the fact that the right tail of the distribution of the LRT of a zero intercept AR model under the unit root has been theoretically shown (equation 1.3 of Larsson, 1999) to behave like that of a χ_1^2 distribution. This can also be seen in the plot on the right in Figure 1, where the empirical density of the LRT is plotted for the unit root process. Hence, if a zero mean AR model may be reasonably assumed, the simple LRT in conjunction with the χ^2 distribution will yield confidence intervals with good coverage properties. It may seem puzzling at first that the LRT is so well approximated by its limiting distribution for a zero-mean process even though the t-statistic fails so miserably. However, Chen and Deo (2006a) pointed out that this puzzling superior performance of the LRT could be explained by the properties of invariance and curvature, as laid out in a series of fundamental papers by Sprott (1973, 1975, 1980, 1984, 1990).

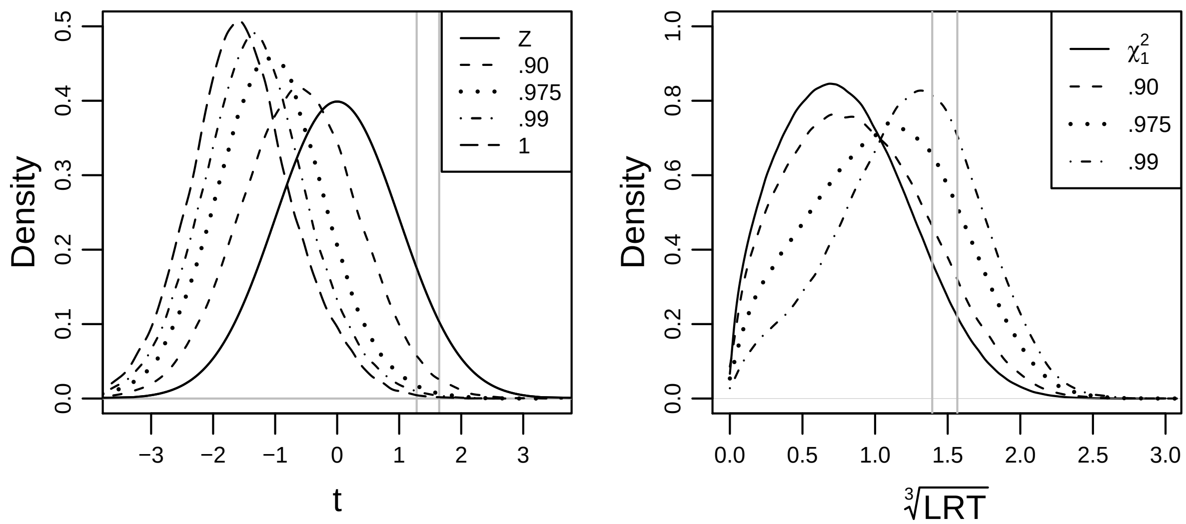

Figure 2: Empirical densities of t statistics and cubic root transformed LRT statistics of AR(1) processes with unknown mean. The vertical lines are 90th and 95th percentiles. Both plots are based on 100,000 repetitions of an AR(1) with sample size $n = 100$ and AR coefficient $\alpha = .9, .975, .99$ and 1.

Though it is heartening to see that the standard LRT will provide inference with good coverage properties in zero-mean models, from an empirical perspective it is important to allow for an intercept/trend. Using results from van Giersbergen (2006a), Chen and Deo (2006a) showed that the distribution of the likelihood ratio test (LRT) for the AR coefficient in the $AR(1)$ model with intercept has an expansion of the form

$$P\left(LRT \le x\right) = P\left(\chi_1^2 \le x\right) + \frac{0.25}{n}\frac{1+7\alpha_1}{1-\alpha_1}\left\{G_3\left(x\right) - G_1\left(x\right)\right\} + O\left(n^{-2}\right), \tag{4}$$

where $G_s\left(\cdot\right)$ is the c.d.f. of a χ_s^2 variable. The leading error term in this expansion explodes near the unit root, implying that the χ^2 approximation to the LRT distribution will be poor in that region. Just how poor the χ_1^2 approximation to the LRT distribution in the intercept model can be, is seen in the plot on the right in Figure 2, where the finite sample distribution is very different from the limiting χ_1^2, even for $\alpha_1 = 0.9$. The plot on the left in Figure 2 shows that the situation is no better for the distribution of the t-statistic in the model with an intercept. On comparing the expansions in (3) and (4) as well as the empirical densities on the right in Figures 1 and 2, it immediately becomes clear that it is the presence of the nuisance intercept parameter (and in general, a deterministic component) that causes the χ^2 approximation to the LRT distribution to be poor.

This observation prompted Chen and Deo (2006a) to consider the restricted likelihood (RL), which eliminates the nuisance deterministic component parameters through an appropriate linear transformation. Chen and Deo (2006a) showed that the RL of an AR process with intercept/trend can be expressed as the LRT of a zero mean AR process plus a smaller order term. As a consequence, the likelihood ratio test based on the restricted likelihood ($RLRT$) for a process with deterministic components behaves approximately like the LRT for a zero mean process. From the entire discussion above, one would then expect the distribution of the $RLRT$ for $AR(p)$ processes with intercept/trend to be well-approximated by that of the limiting χ^2 distribution. In this paper, we study this question by considering the finite sample behaviour of the restricted likelihood ratio test ($RLRT$) for the sum of the AR coefficients of $AR(p)$ processes of arbitrary order p with intercept/trend. The expansion of the distribution of the $RLRT$ for the sum of the AR coefficients is considered and the limit of the leading error term in the expansion is shown to be finite as the unit root is approached. This generalises the result for the $RLRT$ for the $AR(1)$ with intercept case given in Chen and Deo (2006a), in which the leading error term is derived exactly and shown to be $-0.25n^{-1}$, and suggests that the χ^2 approximation to the distribution of the $RLRT$ does not deteriorate for nearly integrated $AR(p)$ processes of arbitrary order p. Hence, the resulting confidence intervals for the sum of the AR coefficients based on inverting the $RLRT$ are expected to have good coverage properties. We provide simulation results in the paper that support this belief. It should be noted that for $AR(p)$ models with intercept/trend, the result that we obtain in this paper is a second-order pointwise result and hence stronger than the first-order pointwise result for the bootstrap based procedures given in Hansen (1999), which is currently the strongest known result for those procedures when the intercept/trend is unknown.

Though the $RLRT$ may provide intervals with good coverage properties, it is not immediately clear whether these intervals are superior to those that can be generated by other competing methods, such as the Andrews and Chen (1994) and Hansen (1999) method. In this paper, we carry out such a comparison on the basis of three criteria: (i) the length of the interval (ii) the power of the interval to reject the unit root under a stationary alternative and (iii) ease of computation. We find in our simulation study that the $RLRT$ based intervals are uniformly superior to the intervals based on the competing procedures on each of these counts. An empirical analysis of the Nelson-Plosser data yields $RLRT$ based intervals that can be substantially different from those computed by the Andrews and Chen (1994) and Hansen (1999) procedure.

A third contribution of this paper is to provide an alternative reparametrisation of the $AR(p)$ coefficients $\{\alpha_i\}_{i=1}^{p}$ in terms of the first p partial autocorrelations. This reparametrisation has the advantage that it yields a simple way of computing and optimising the Restricted Likelihood for $AR(p)$ processes and it is also of independent interest since it generalises an earlier result of Barndorff-Nielsen and Schou (1973) from the stationary case to the unit root case. The need for reparametrisation is motivated by the fact that the leading error term in the expansion of the distribution of the $RLRT$ is a function of the vector of AR coefficients $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_p)'$. A major obstacle in studying the behaviour of this error term in the vicinity of the unit root is that in a stationary $AR(p)$ process the parameter space covered by $\boldsymbol{\alpha}$ is a complicated subset of $\mathbb{R}^p$, defined implicitly through the restrictions imposed on the roots $\{z_i\}_{i=1}^{p}$ of the polynomial $g(z)$. More specifically, while we can assert that $\sum_{i=1}^{p} \alpha_i \to 1$ as $\boldsymbol{\alpha}$ approaches the unit root case, it does not seem possible to similarly describe the limiting behaviour of the individual α_i. However, in stationary $AR(p)$ models, Barndorff-Nielsen and Schou (1973) have proved the one-to-one onto correspondence between the $AR(p)$ coefficients $\boldsymbol{\alpha}$ and the first p partial autocorrelations $\boldsymbol{\phi} = (\phi_{11}, ..., \phi_{pp})'$, where the s^{th} partial autocorrelation ϕ_{ss} is defined as the conditional correlation between U_t and U_{t+s} given the intervening $(U_{t+1}, ..., U_{t+s-1})$. The advantage of the partial autocorrelations is that each of them lies in the open interval $(-1, 1)$ in the stationary case. In this paper we extend this one-to-one correspondence between $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ to the unit root case and prove that at the unit root, the $\boldsymbol{\alpha}$ vector corresponds to $\boldsymbol{\phi} \in [1] \times (-1, 1)^{p-1}$. Thus, as the unit root is approached, only the first partial autocorrelation $\phi_{11} \to 1$, whereas the remaining $p-1$ partial autocorrelations stay strictly within $(-1, 1)$. Using this result and reparametrising $\boldsymbol{\alpha}$ in terms of $\boldsymbol{\phi}$ allows us to study the behaviour of the $RLRT$ distribution near the unit root more easily and provides more insight into the problem.

The extension of the one-to-one correspondence between $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ that we prove here has an additional benefit in terms of computing and optimising the restricted likelihood. Using this correspondence, we provide a unified expression for the restricted likelihood for the model (1) under both stationarity and the unit root. This expression does not possess a singularity at the unit root and thus is numerically stable as it transitions smoothly from the stationary case to the unit root case. Furthermore, the expression is also easy to compute and to optimise since it is in terms of $\boldsymbol{\phi}$, which lies in the bounded p-dimensional hypercube $(-1,1]\times(-1,1)^{p-1}$. Finally, we show that the sum of the AR coefficients $\sum_{i=1}^{p}\alpha_i$ is a simple bounded function of the partial autocorrelations, providing a feasible method of computing confidence intervals for $\sum_{i=1}^{p}\alpha_i$ by inverting the restricted likelihood ratio test over a grid of values in a bounded interval.

In the next Section we describe the restricted likelihood and review some of the results in the literature that show its various advantages over ordinary maximum likelihood as well as ordinary least squares for both estimation and inference. We then provide our theoretical result on the finite sample behaviour of the $RLRT$ for the sum of the AR coefficients. In Section 3 we describe the correspondence between $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ in the stationary case and then provide our result on the extension to the unit root case. Section 4 provides the findings from a detailed simulation study in which we consider the performance of not only the $RLRT$ based procedure but also compare it to the intervals of Hansen (1999) and Andrews and Chen (1994). In Section 6 we provide a simple algorithm for the computation and optimisation of the restricted likelihood and for computing confidence intervals for $\sum_{i=1}^{p}\alpha_i$. Section 5 provides our empirical results on the Nelson-Plosser data using the algorithm described in Section 6. All proofs are in the Appendix at the end.

2 The Restricted Likelihood:

The model in (1) can be expressed in a linear model framework as

$$\mathbf{X} = \mathbf{W}_r\theta_r + \mathbf{U},$$

where $\mathbf{X} = (X_1, ..., X_n)'$, $\mathbf{U} = (U_1, ..., U_n)'$, $\mathbf{W}_r$ and θ_r are the design matrix and parameter vector respectively corresponding to the intercept model for $r = 1$ (i.e. $\theta_1 = \beta_0$ and $\mathbf{W}_1 = (1, ..., 1)'$) and intercept and linear trend model for $r = 2$ (i.e. $\theta_2 = (\beta_0, \beta_1)'$ and $\mathbf{W}_2 = \big(\mathbf{W}_1, \mathbf{t}\big)$ where $\mathbf{t}' = (1, 2, ..., n)'\big)$. The restricted likelihood was proposed by Kalbfleisch and

Sprott (1970) to estimate the parameters of the error covariance matrix, $Var(\mathbf{U})$, for such linear models where the regression coefficients θ_r are nuisance parameters. Kalbflesich and Sprott (1970) defined the restricted likelihood to be the exact likelihood of the linearly transformed data $\mathbf{T}_r\mathbf{X}$, where $\mathbf{T}_r$ is any matrix of full row-rank such that $\mathbf{T}_r\mathbf{W}_r = \mathbf{0}$. Thus, the likelihood of the transformation $\mathbf{T}_r\mathbf{X}$ does not depend on the nuisance regression coefficient parameters (here, the intercept and trend parameters). The particular choice of the matrix $\mathbf{T}_r$ is irrelevant since the likelihood of $\mathbf{T}_r\mathbf{X}$ will change only by a multiplicative constant for different choices of $\mathbf{T}_r$ (Harville, 1974) and hence will have no effect on either estimation or testing of hypothesis. As a result, the restricted likelihood for the process (1) based on the data $(X_1, ..., X_n)$ is the exact likelihood of the second differences $\{X_t - X_{t-2} + 2X_{t-1}\}_{t=3}^{n}$ (If the trend coefficient β is known to be zero, the restricted likelihood based on $(X_1, ..., X_n)$ is the exact likelihood of the first differences $\{X_t - X_{t-1}\}_{t=2}^{n}$). Using Harville's (1974) formula, the restricted log-likelihood (up to an additive constant) for the process (1) based on $\mathbf{X} = (X_1, ..., X_n)'$ is given by

$$\begin{aligned} L\left(\mathbf{X},\boldsymbol{\alpha},\sigma^2\right) = &-\frac{n-r}{2}\log\sigma^2 + \frac{1}{2}\log\frac{\left|\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha})\right|}{\left|\mathbf{W}_r'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha})\mathbf{W}_r\right|} \\ &-\frac{1}{2\sigma^2}\left\{\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha})\mathbf{X} - \mathbf{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha})\mathbf{W}_r(\mathbf{W}_r'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha})W_r)^{-1}\mathbf{W}_r'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha})\mathbf{X}\right\}, \end{aligned} \tag{5}$$

where $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_p)'$, $Var(\mathbf{U}) = \sigma^2\boldsymbol{\Sigma}(\boldsymbol{\alpha})$ and $r = 1$ for the intercept model and $r = 2$ for the trend model. The restricted likelihood estimates of $\boldsymbol{\alpha}$ can be obtained by maximising the concentrated restricted log-likelihood

$$\begin{aligned} L_1(\mathbf{X},\boldsymbol{\alpha}) = &-\frac{n-r}{2}\log\left[\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha})\mathbf{X} - \mathbf{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha})\mathbf{W}_r(\mathbf{W}_r'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha})W_r)^{-1}\mathbf{W}_r'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha})\mathbf{X}\right] \\ &+\frac{1}{2}\log\frac{\left|\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha})\right|}{\left|\mathbf{W}_r'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha})\mathbf{W}_r\right|}. \end{aligned} \tag{6}$$

Harville (1977) showed that the restricted maximum likelihood (REML) estimates of the error covariance matrix did not lose any efficiency. In a time series context, Cooper and Thompson (1977), Levenbach (1972) and Tunnicliffe Wilson (1989) found through simulations that REML estimates had smaller bias than the usual maximum likelihood estimates, while Rahman and King (1997) found that score tests based on the restricted likelihood had good size and power properties. Cheang and Reinsel (2000) derived expressions for the bias of REML estimates in the model (1) with no trend ($\beta_1 = 0$) and showed that this bias is smaller than that of ML estimates

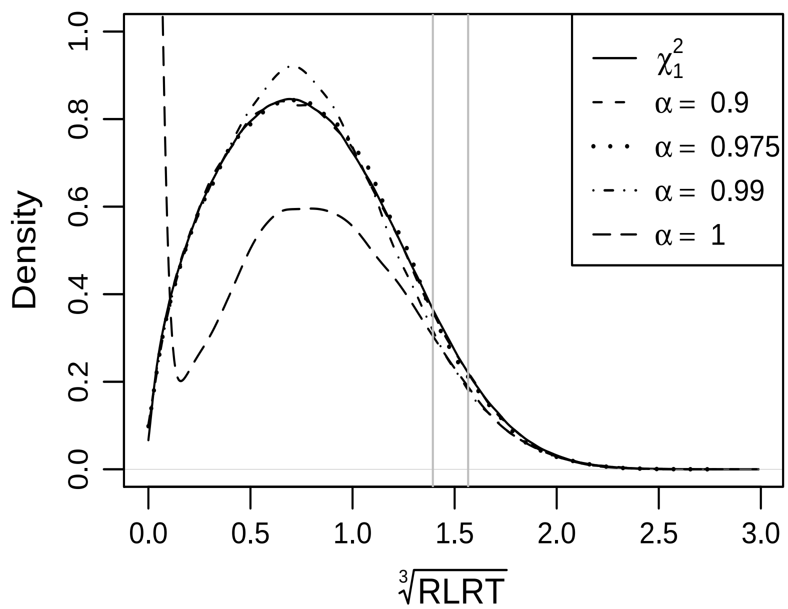

Figure 3: Empirical densities of cubic root transformed Restricted LRT statistics of AR(1) processes with intercept. The vertical lines are 90th and 95th percentiles. The plot is based on 100,000 repetitions of an AR(1) with sample size $n = 100$ and AR coefficient $\alpha = .9, .975, .99$ and 1.

in nearly integrated AR models, while Kang, Shin and Lee (2003) extended this result to the case of linear trend. Francke and de Vos (2005) considered the use of the restricted likelihood in unit root tests. They found that the resulting tests almost attained the power envelope in $AR(1)$ models and provided an intriguing heuristic explanation for this finding. Chen and Deo (2006a) showed that the restricted likelihood for stationary $AR(1)$ models with intercept had both small Efron (1975) curvature and well behaved likelihood ratio tests for the autoregressive coefficient.

More specifically, Chen and Deo (2006a) obtained the expansion of the likelihood ratio test based on the restricted likelihood ($RLRT$) for the AR coefficient of a stationary $AR(1)$ model with intercept and showed that

$$P\left(RLRT \leq x\right) = P\left(\chi_1^2 \leq x\right) - \frac{0.25}{n}\left\{G_3\left(x\right) - G_1\left(x\right)\right\} + O\left(n^{-2}\right). \tag{7}$$

Thus, the $RLRT$ distribution for the $AR(1)$ model with intercept has an expansion that is identical up to the second order to that of the LRT distribution for the zero mean $AR(1)$ model, stated in (3). The leading error term of -0.25 in this expansion is small, does not depend on α_1 and does not explode as the unit root is approached. Figure 3 shows the empirical densities of the $RLRT$ for various values of α_1 and it is observed that the χ_1^2 approximation work very well for values of α_1 both near and far from the unit root. Chen and Deo (2006b) also obtained the limiting distribution of the $RLRT$ under the unit root and showed by numerical evaluation that the right tail of this unit root distribution behaves like that of a χ_1^2 variable. This result is similar to the finding of Larsson (1995), mentioned in the Introduction above, for the distribution of the standard LRT at the unit root for a no-intercept process. This finding can also be seen in the density plot of the $RLRT$ for the unit root in Figure 3. Thus, the result above suggests that unlike the distribution of the t-statistic and the standard LRT for an $AR(1)$ model with intercept, the $RLRT$ distribution will be well approximated by the χ^2 distribution and will thus provide interval inference for the AR coefficient with good coverage properties. The following Theorem extends the result of Chen and Deo (2006a) for an $AR(1)$ process by considering the finite sample behaviour of the $RLRT$ for testing a composite null hypothesis $H_0 : \sum_{i=1}^{p} \alpha_i = a$ for AR processes of any order p.

Theorem 1 *Let* $\mathbf{X} = (X_1, ..., X_n)'$ *follow the process in (1) where the vector of AR coefficients* $\boldsymbol{\alpha}$ *is such that the polynomial* $g(z) = z^p - \sum_{i=1}^{p} \alpha_i z^{p-i}$ *has roots* $|z_p| \leq ... \leq |z_1| < 1$. *For* $a = \sum_{i=1}^{p} \alpha_i$, *let*

$$R_n(a) = -2 \left\{ \max_{\{\boldsymbol{\alpha}: \sum_{i=1}^{p} \alpha_i = a, |z_1| \leq 1\}} L_1(\mathbf{X}, \boldsymbol{\alpha}) - \max_{\{\boldsymbol{\alpha}: |z| \leq 1\}} L_1(\mathbf{X}, \boldsymbol{\alpha}) \right\}$$

where $L_1(\mathbf{X}, \alpha)$ *is given in (6), be the restricted likelihood ratio test for testing the composite null hypothesis* $H_0 : \sum_{i=1}^{p} \alpha_i = a$. *Then, we have*

$$P(R_n(a) \leq x) - P\left(\chi_1^2 \leq x\right) = n^{-1} C(\boldsymbol{\alpha}) \left\{ P\left(\chi_3^2 \leq x\right) - P\left(\chi_1^2 \leq x\right) \right\} + O\left(n^{-2}\right)$$

where $C(\boldsymbol{\alpha})$ *is a continuous function of* $\boldsymbol{\alpha}$ *such that*

$$\lim_{\boldsymbol{\alpha} \to \widetilde{\boldsymbol{\alpha}}} C(\boldsymbol{\alpha}) < \infty$$

for any vector $\widetilde{\boldsymbol{\alpha}} \equiv (\widetilde{\alpha}_1, ..., \widetilde{\alpha}_p)'$, *where* $\widetilde{\boldsymbol{\alpha}}$ *is such that the polynomial* $z^p - \sum_{i=1}^{p} \widetilde{\alpha}_i z^{p-i}$ *has roots* $|\widetilde{z}_p| \leq ... \leq |\widetilde{z}_2| < 1$ *and* $\widetilde{z}_1 = 1$.

Remark 1 *The result in the above Theorem is a generalisation of that stated in (7) for the* $AR(1)$ *case in which* $C(\boldsymbol{\alpha}) = -0.25$, *due to which it trivially follows that* $\lim_{\alpha_1 \to 1} C(\boldsymbol{\alpha}) < \infty$. *From Theorem 3 of van Giersbergen (2006b), arguing as in Chen and Deo (2006a), it can be shown that for an* $AR(2)$ *process we can identify* $C(\boldsymbol{\alpha})$ *as* $C(\boldsymbol{\alpha}) \equiv 0.25$.

Remark 2 *Close inspection of the proof of Theorem 1 shows that the term* $C(\boldsymbol{\alpha})$ *is of the form*

$$C(\boldsymbol{\alpha}) = \frac{\mathcal{P}_1(\boldsymbol{\phi})}{\mathcal{P}_2(\boldsymbol{\phi}) \prod_{i=2}^{p} (1 - \phi_{ii})^{b_i} \prod_{j=1}^{p} (1 + \phi_{jj})^{d_j}},$$

where b_i *and* d_j *are non-negative integers,* $\boldsymbol{\phi}$ *is the vector of partial autocorrelations corresponding to* $\boldsymbol{\alpha}$ *as shown in Theorem 2 and* $\mathcal{P}_i(\boldsymbol{\phi})$ *are multivariate polynomials in the elements of* $\boldsymbol{\phi}$ *such that* $\mathcal{P}_2(\boldsymbol{\phi})$ *is bounded as* $\phi_{11} \to 1$. *Since the other term in the denominator is free of a* $1 - \phi_{11}$ *term, it follows from Theorem 2 that* $C(\boldsymbol{\alpha})$ *is bounded as the unit root is approached.*

Theorem 1 suggests that the distribution of the $RLRT$ will be well-approximated by the χ_1^2 distribution even for nearly integrated $AR(p)$ processes. Simulation results that we report in Section 4 support this suggestion. A useful consequence of Theorem 1 is that confidence intervals for $\sum_{i=1}^{p} \alpha_i$ obtained by inverting the $RLRT$ acceptance region will have good coverage properties. In order for this to be practically convenient, we need to have efficient methods for computing and optimising the restricted likelihood, as well as of computing confidence intervals for $\sum_{i=1}^{p} \alpha_i$. In Section 6 we describe in detail a computationally efficient algorithm for these calculations. The algorithm is based on the results in the next Section, in which we reparametrise the AR coefficients $\boldsymbol{\alpha}$ through the partial autocorrelations $\boldsymbol{\phi} = (\phi_{11}, ..., \phi_{pp})'$.

3 The correspondence between $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ ::

Let $\mathbf{A}_1 = \{\boldsymbol{\alpha} : \max_{1 \le i \le p} |z_i| < 1\}$, where z_i are the roots of the polynomial $g(z)$, denote the parameter space of the coefficients of a stationary $AR(p)$ process. Now, let $\boldsymbol{\phi} = (\phi_{11}, ..., \phi_{pp})'$ denote the vector of the first p partial autocorrelations of a stationary $AR(p)$ process and let $\boldsymbol{\Phi}_1 = \{\boldsymbol{\phi} : (\phi_{11}, ..., \phi_{pp}) \in (-1, 1)^p\}$ denote their parameter space. Defining $(\phi_{p1}, \phi_{p2}, ..., \phi_{pp}) \equiv \boldsymbol{\alpha} = (\alpha_1, ..., \alpha_p)'$, any $\boldsymbol{\phi} \in \boldsymbol{\Phi}_1$ can be transformed to $\boldsymbol{\alpha} \in \mathbf{A}_1$ by first computing for $k = 2, ..., p$ the

Durbin-Levinson recursion

$$\phi_{ki} = \phi_{k-1,i} - \phi_{kk}\phi_{k-1,k-i} \qquad i = 1, 2, ..., k-1 \tag{8}$$

and then setting

$$\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_p)' = (\phi_{p1}, \phi_{p2}, ..., \phi_{p,p-1}, \phi_{pp})'. \tag{9}$$

Conversely, any vector $\boldsymbol{\alpha} \in \mathbf{A}_1$ can be transformed to $\boldsymbol{\phi} \in \boldsymbol{\Phi}_1$ by using the recursion

$$\phi_{k-1,i} = \phi_{k,i} + \phi_{kk}\phi_{k,k-i} \qquad i = 1, ..., k-1. \tag{10}$$

Theorem 2 of Barndorff-Nielsen and Schou (1973) states that the mapping that transforms $\boldsymbol{\alpha} \in \boldsymbol{A}_1$ to $\boldsymbol{\phi} \in \boldsymbol{\Phi}_1$ is one-to-one and onto $(-1, 1)^p$. Furthermore, both this mapping and its inverse mapping are continuously differentiable.

Since we are interested in such a correspondence at the unit root, we define the parameter space of the coefficients of a unit root $AR(p)$ process as $\mathbf{A}_2 = \{\boldsymbol{\alpha} : \max_{2\leq i\leq p} |z_i| < 1,\ z_1 = 1\}$, where $|z_p| \leq ... \leq |z_2| \leq |z_1|$ are the absolute values of the roots of $g(z)$. The following Lemma shows that $\mathbf{A}_2$ is in one-to-one correspondence with the space $\boldsymbol{\Phi}_2 = \{\boldsymbol{\phi} : \phi_{11} = 1, |\phi_{ii}| < 1,\ i = 2, ..., p\}$.

Lemma 1 *For the recursions in (10) and (8), $\boldsymbol{\alpha} \in \mathbf{A}_2$ if and only if $\boldsymbol{\phi} \in \boldsymbol{\Phi}_2$. Furthermore, the mapping from $\mathbf{A}_2 \rightarrow \boldsymbol{\Phi}_2$ is one-to-one and onto $[1] \times (-1, 1)^{p-1}$.*

Defining the combined parameter spaces $\mathbf{A} = \mathbf{A}_1 \cup \mathbf{A}_2$ and $\boldsymbol{\Phi} = \boldsymbol{\Phi}_1 \cup \boldsymbol{\Phi}_2$, we get the following Theorem for the recursions in (10) and (8) from Lemma 1, Theorem 2 of Barndorff-Nelson and Schou (1973) and the fact that the mapping from $\boldsymbol{\alpha}$ to $\boldsymbol{\phi}$ is continuous.

Theorem 2 *The mapping $\mathbf{A} \rightarrow \boldsymbol{\Phi}$ is one-to-one and onto $(-1, 1] \times (-1, 1)^{p-1}$ and can be partitioned as $\mathbf{A}_1 \rightarrow \boldsymbol{\Phi}_1$ onto $(-1, 1)^p$ and $\mathbf{A}_2 \rightarrow \boldsymbol{\Phi}_2$ onto $[1] \times (-1, 1)^{p-1}$. Furthermore, if $\boldsymbol{\alpha}_s \in \mathbf{A}_1$ is a sequence of vectors such that $\boldsymbol{\alpha}_s \rightarrow \tilde{\boldsymbol{\alpha}} \in \mathbf{A}_2$, then the corresponding sequence of partial autocorrelation vectors $\boldsymbol{\phi}_s \in \boldsymbol{\Phi}_1$ is such that $\boldsymbol{\phi}_s \rightarrow \tilde{\boldsymbol{\phi}} \in \boldsymbol{\Phi}_2$.*

Theorem 2 is of use not only in studying the finite sample behaviour of the $RLRT$ for $\sum_{i=1}^{p} \alpha_i$ near the unit root but also in providing a feasible way of computing and optimising the restricted likelihood and computing confidence intervals for $\sum_{i=1}^{p} \alpha_i$, as we show in Section 6.

In the next Section we report the findings from a detailed simulation study.

4 Simulation study:

In this section we study the finite sample behaviour of the $RLRT$ based confidence intervals and compare these intervals to those constructed from the procedure of Andrews and Chen (1994) and Hansen (1999). In the first part of the simulation study, the data was generated from the $AR(1)$ process $X_t = \alpha X_{t-1} + v_t$, where $v_t \sim i.i.d.N\left(0, \sigma_v^2\right)$ and $X_0 \sim N\left(\mu, \left(1-\alpha^2\right)^{-1}\right)$ when $|\alpha| < 1$ and $X_0 = 0$ when $\alpha = 1$. The sample size was set to $n = 100$ and the number of replications was 20,000. We constructed 90% and 95% confidence intervals based on the $RLRT$ by inverting the acceptance region of the $RLRT$, where the acceptance region is always defined using the χ_1^2 distribution. Confidence intervals were also constructed by the methods of Andrews and Chen (1994) and Hansen (1999). The parametric grid bootstrap was used for these procedure since the actual data was normally distributed. Also, the upper end point of the Andrews and Hansen intervals was truncated at 1, so that the intervals always lay entirely in the parameter space. There were replications where the Andrews and Hansen intervals were empty and in such cases we defined them as the singleton [1]. The $RLRT$ based intervals, however, always lie in the parameter space by construction since the restricted likelihood is optimised over the admissible parameter space and also always have positive Lebesgue measure since the RL is continuous in the parameters. The intervals for all three procedures were computed for both models with intercept as well as for models with trend.

The simulation coverage rates of the $RLRT$ intervals is reported in Table I for α_1 belonging to $\{0.95, 0.99, 0.995, 1\}$ and it is seen that the coverage rates are very close to the nominal ones, for both the intercept case as well as the intercept and trend case. The coverage is a little conservative in the neighbourhood of the unit root but not by an enormous amount. We do not report coverage rates for the Andrews and Hansen procedures since these intervals are designed to have exact coverage in an $AR(1)$ model.

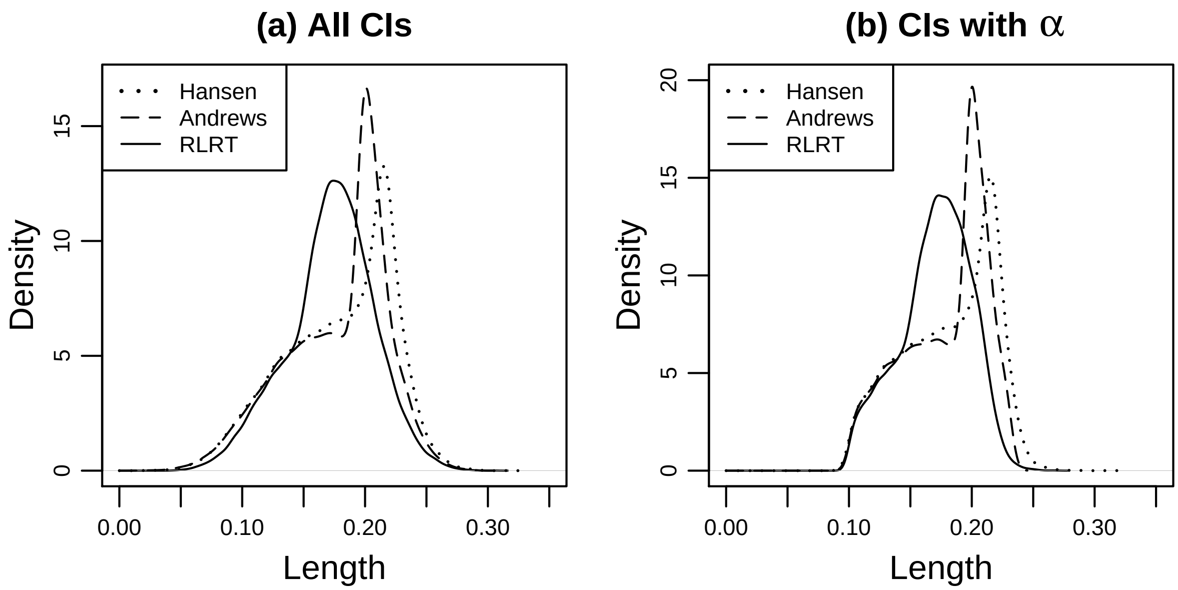

Figure 4: Empirical densities of 90% Confidence Interval lengths based on 20,000 replications of an AR(1) with AR coefficient $\alpha = 0.9$ and sample size $n = 100$. Plot (a) is for all intervals. Plot (b) is for those intervals that cover the true α.

In Table II we report the percentage of times the intervals did not contain the the value of unity when the true value of α_1 was either 0.9 or 0.95. Thus, this percentage is the power of the confidence interval procedure to reject the unit root in favour of the true stationary process. It is seen that for both values of α_1 as well as both the levels of confidence, the RLRT based intervals have significantly higher power for rejecting the unit root compared to the Andrews and Hansen intervals. The power of the *RLRT* interval is approximately twice that of the Hansen interval and approximately 1.5 times that of the Andrews interval. Thus, the *RLRT* procedure provides intervals with correct coverage for α_1 while simultaneously yielding a more powerful unit root test. The next comparison that we make between the three interval procedures is in terms of the distribution of their lengths. In Figure 4 we plot the empirical densities of the lengths of the 90% confidence intervals when $\alpha_1 = 0.9$ for each of the procedures for the model with intercept. We plot both the densities of the lengths of all the intervals, as well as the densities of only those intervals that actually contain the true value of the parameter in them. The plots for the other levels of confidence, the other value of α_1 and also for the model with intercept and trend were similar qualitatively to this plot and so we do not report them here. It is clear from the empirical densities that the lengths of the $RLRT$ based intervals are generally stochastically smaller than the lengths of the Andrews and Hansen intervals. Furthermore, the spread of the distribution of the lengths is much smaller for the $RLRT$ intervals than for the other intervals. The superiority of the $RLRT$ intervals is even more pronounced when one focuses on the intervals that contain the true value of the parameter. In Tables III and IV we report the mean and the standard deviation of the lengths of the 90% intervals for all three procedures for the intercept and trend models respectively. The pattern of the results is similar for the 95% intervals. It is seen from the table that the mean of the length of all the $RLRT$ intervals is smaller than those of the other intervals except when $\alpha_1 = 0.95$ and this exception can be explained by the fact that the Andrews and Hansen intervals get truncated at 1 and this truncation happens more often when α_1 is close to unity (We recall that as stated above, both the Andrews and Hansen intervals can be empty). Furthermore, when one considers the mean lengths of only those intervals that contained the true value of the parameter, we see that the $RLRT$ interval uniformly dominates the other procedures at all values of α_1. It should be noted that for the $RLRT$ intervals, the mean length of the intervals that contain the true value of the parameter is uniformly smaller than the mean length of all the intervals. In contrast, for the Andrews and Hansen procedure the mean lengths of the intervals that contain the true parameter value are larger than the mean lengths of all the intervals. Finally, the standard deviation of the $RLRT$ interval lengths is uniformly smaller than that of the lengths of the other procedures, with the reduction in standard deviation being as much as 20%.

Finally, we also note that the $RLRT$ based intervals are much simpler to compute than the bootstrap based ones, since the $RLRT$ statistic value needs to be compared just to the χ_1^2 critical value, which is easily available. The actual computation of the $RLRT$ statistic itself is very simple, since according to our reparametrisation through the partial autocorrelations, the optimisation is over the simple bounded set $(-1, 1] \times (-1, 1)^{p-1}$ and can be done with any reasonable statistical software. On the other hand, the bootstrap based procedures require the user to generate a different bootstrap distribution for every candidate value of the parameter when computing the confidence interval.

We next turn our attention to examining the performance of $RLRT$ based confidence intervals for $AR(2)$ processes. The data is generated from the $AR(2)$ process given by $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + v_t$, where $v_t \sim i.i.d.N\left(0,1\right)$. In Table V we report the coverage rates of $RLRT$ intervals for the sum of the coefficients $\alpha_1 + \alpha_2$ for models with intercept, where the sample size is $n = 100$ and the number of replications is 20,000. Four different parameter configurations of (α_1, α_2) for the stationary case, viz. $(1.3, -0.4)$, $(1.55, -0.6)$, $(1.75, -0.76)$ and $(1.775, -0.78)$ were considered. The values of $\alpha_1 + \alpha_2$ for these four configurations are 0.9, 0.95, 0.99 and 0.995 respectively while the roots of the associated polynomial $z^2 - \alpha_1 z - \alpha_2 = 0$ are (0.5, 0.8), (0.75, 0.8), (0.95, 0.8) and (0.975, 0.8) respectively. To evaluate the performance of the restricted likelihood under a unit root, we generated $n = 100$ values from the series with $\alpha_1 = 1.8$ and $\alpha_2 = -0.8$, for which the two roots of the associated polynomial are 1 and 0.8. To generate the non-stationary series, we first simulated ΔX_t for $t = 1, ..., 100$ from the stationary process $\Delta X_t = -\alpha_2 \Delta X_{t-1} + v_t$ and then integrated the differenced series ΔX_t setting $X_0 = 0$ (We note again that setting $X_0 = 0$ is irrelevant to our calculations since the restricted likelihood uses only ΔX_t for $t = 2, ..., 100$ and hence X_0 could be set to any arbitrary value). It is seen from the table that the $RLRT$ intervals provide excellent coverage of the true value, even when the process is close to a unit root one.

In the next Section, we report our findings from our empirical study in which we compute $RLRT$ based confidence intervals for the sum of the AR coefficients for the well-known Nelson-Plosser data set.

5 Empirical Application:

We now proceed to compute confidence intervals for $\sum_{i=1}^{p} \alpha_i$ for autoregressive models of the form (1) with time trend estimated for the original Nelson-Plosser data (1982) that runs up to 1970, as well as for the extended Nelson-Plosser data that runs up to 1988. In keeping with the literature that has analysed the Nelson-Plosser data, we use the same autoregressive order p as Nelson-Plosser, and natural logs are used for all the series except bond yield. Tables I and II report the series that we analysed, the order p of the AR model used for each series, the maximum RL point estimate of $\sum_{i=1}^{p} \alpha_i$ as well as the 90% and 95% CI for $\sum_{i=1}^{p} \alpha_i$ computed by inverting the acceptance region of the $RLRT$ using a χ_1^2 distribution. The χ_1^2 critical value was also used to decide whether the unit root belonged in the confidence interval. Though the asymptotic distribution of the $RLRT$ at the unit root is not exactly χ_1^2, the simulations in Chen and Deo (2006 b) indicate that using the χ_1^2 value at the unit root results in a test that is almost correctly sized albeit a little conservative up to the 10% level of significance. (i.e. the test will reject the null hypothesis of a unit root less often than the stated nominal significance level). We also used the critical values of the unit root $RLRT$ distribution under trend that were obtained by simulation in Francke and de Vos (2006) to decide whether a unit root should be rejected or not. (Though Francke and de Vos (2006) simulated this distribution for an $AR(1)$, we expect the distribution to remain the same for higher order processes). Rejection of the unit root using the critical values from this unit root $RLRT$ distribution is denoted by an asterisk in the Tables. As can be seen from the Tables, there are very few situations where the unit root is rejected by the critical values of the unit root $RLRT$ distribution but not by the χ_1^2 critical values. This is in keeping with the observation above that up to the 10% level of significance, the χ_1^2 critical values are very close to the unit root $RLRT$ distribution and will yield almost exactly the same inference for the unit root.

We start by inspecting the maximum RL point estimates of $\sum_{i=1}^{p} \alpha_i$ for the extended series in Table II and comparing them to the corresponding estimates in Andrews and Chen (1994) and Hansen (1999). Our estimates are consistently higher than the corresponding point estimates reported in Table 6 of Hansen (1994) (except in the situations where our estimates are on the unit root boundary and Hansen's estimates are greater than 1). On the other hand, our point estimates are very similar to the median unbiased point estimates reported in Table 4 of Andrews and Chen (1994).

The confidence intervals for $\sum_{i=1}^{p} \alpha_i$ generated from the $RLRT$ always lie in the parameter space by construction and hence have a natural upper limit of 1. By contrast, the intervals generated in both Hansen (1999) and Andrews and Chen (1994) can have not only the upper end point, but also the lower end point exceeding 1, with the net result being a confidence interval that lies entirely outside the permissible parameter space (Note that both Andrews and Chen (1994) and Hansen (1999) do not allow for explosive processes and hence their permissible parameter space for $\sum_{i=1}^{p} \alpha_i$ is bounded above by unity). Indeed, the 90% interval for the bond yield (original series) in both Hansen (1999) and Andrews and Chen (1994) lies entirely above 1, whereas the corresponding 90% $RLRT$ interval in Table I is (0.96,1]. Similarly, the 90% interval for the extended Consumer Prices series in Andrews and Chen (1994) lies entirely above 1, while the corresponding interval in Hansen (1999) has an upper end point of 1.018. On the other hand, the $RLRT$ based interval for this series, reported in Table II, is (0.983, 1]. The fact that the Hansen (1999) and Andrews and Chen (1994) intervals can contain regions greater than 1 also may be the reason why they have significantly lower power in rejecting the unit root when the alternative is stationary, compared to the power of the $RLRT$ interval.

Our 90% intervals reject the unit root for the four extended series Real GNP, per capita GNP, Employment and Unemployment. In contrast, Andrews and Chen (1994) reject the unit root at the 10% level for only three of these series, viz. Real GNP, per capita GNP and Unemployment, whereas Hansen (1999) rejects it at the 10% level for only two of these series, per capita GNP and Unemployment. In our analysis, the rejection of the unit root continues to hold for the 95% intervals for the extended versions of three out of these four series. It is interesting to note that our 95% confidence intervals do not have a width that is significantly greater than the width of the 90% intervals. One interesting consequence of this fact is that the unit root continues to be rejected at the 5% level for three of the four extended series for which it was rejected at the 10% level.

6 Algorithm for computing and optimising the restricted likelihood:

In principle, expression (6) can be computed quite easily since there are well known expressions for $\left|\mathbf{\Sigma}^{-1}(\boldsymbol{\alpha})\right|$ as well as simple algorithms for computing bilinear forms $\mathbf{a}'\mathbf{\Sigma}^{-1}(\boldsymbol{\alpha})\,\mathbf{b}$ (Galbraith and Galbraith 1973). However, the restricted log-likelihood may become unstable near the unit root since both the terms $\left|\mathbf{\Sigma}^{-1}(\boldsymbol{\alpha})\right|$ and $\left|\mathbf{W}_r'\mathbf{\Sigma}^{-1}(\boldsymbol{\alpha})\,\mathbf{W}_r\right|$ become zero when the $AR(p)$ process has a unit root. (This is seen most easily for the $AR(1)$ with intercept model, where $\left|\mathbf{\Sigma}^{-1}(\boldsymbol{\alpha})\right| = 1-\alpha^2$ and $\left|\mathbf{W}_1'\mathbf{\Sigma}^{-1}(\boldsymbol{\alpha})\,\mathbf{W}_1\right| = 1-\alpha^2+(n-1)(1-\alpha)^2$). Thus, it is useful to have an alternative expression for the restricted log-likelihood that does not possess this singularity at the unit root. A second difficulty with the concentrated restricted log-likelihood as given in (6) is that the optimisation involved is over the parameter space $\mathbf{A}$ stated just after Lemma

1. The parameter space $\mathbf{A}$ is awkward to impose when maximising the likelihood in practice, particularly for higher order models. However, Theorem 2 shows that the space $\mathbf{A}$ is in one-to-one correspondence with the space $\mathbf{\Phi}$ given by

$$\mathbf{\Phi} = \{\boldsymbol{\phi} : (\phi_{11}, ..., \phi_{pp}) \in (-1, 1] \times (-1, 1)^{p-1}\},$$

which is a simple space over which to optimise the restricted likelihood function. Using this re-parametrisation, we next present the algorithm for computing and optimising the restricted likelihood given in (6).

For a candidate $\boldsymbol{\phi} \in \mathbf{\Phi}$, define the corresponding vector of AR coefficients $\boldsymbol{\alpha}(\boldsymbol{\phi}) = (\alpha_1(\boldsymbol{\phi}), ..., \alpha_p(\boldsymbol{\phi}))$ through the recursions (8) and (9). Henceforth, we suppress the dependence of $\boldsymbol{\alpha}(\boldsymbol{\phi})$ on $\boldsymbol{\phi}$ and simply write it as $\boldsymbol{\alpha}$. Noting that (see Barndorff-Nielsen and Schou, 1973)

$$\left|\mathbf{\Sigma}^{-1}(\boldsymbol{\alpha})\right| = \prod_{i=1}^{p} \left(1 - \phi_{ii}^2\right)^i,$$

we see from (12) below that the ratio $\left|\mathbf{\Sigma}^{-1}(\boldsymbol{\alpha})\right| (\left|\mathbf{W}_2'\mathbf{\Sigma}^{-1}(\boldsymbol{\alpha})\mathbf{W}_2\right|)^{-1}$ can be computed as

$$\frac{\left|\mathbf{\Sigma}^{-1}(\boldsymbol{\alpha})\right|}{\left|\mathbf{W}_2'\mathbf{\Sigma}^{-1}(\boldsymbol{\alpha})\mathbf{W}_2\right|} = \frac{\prod_{i=1}^{p}(1+\phi_{ii})\prod_{i=2}^{p}\left(1-\phi_{ii}^2\right)^{i-1}}{C_1C_3 - (1-\sum_{i=1}^{p}\alpha_i)\,C_2^2}$$

where, defining $\alpha_0 = -1$, we have

$$C_1 = -\sum_{k=0}^{p}(n-2k)\,\alpha_k,$$

$$C_2 = \frac{n+1}{2}\left\{\sum_{k=0}^{p}(n-2k)\,\alpha_k\right\},$$

and

$$C_3 = \sum_{j=1}^{p}\left\{\left(\sum_{k=0}^{p-j}(k+j)\,\alpha_k\right)^2 - \left(\sum_{k=0}^{p-j}(p-k)\,\alpha_{k+j}\right)^2\right\} + \sum_{t=3}^{n}\left(t - \alpha_1(t-1) - \cdots - \alpha_p(t-p)\right)^2.$$

For the model with only intercept, the ratio $\left|\mathbf{\Sigma}^{-1}(\boldsymbol{\alpha})\right| (\left|\mathbf{W}_1'\mathbf{\Sigma}^{-1}(\boldsymbol{\alpha})\mathbf{W}_1\right|)^{-1}$ is computed as

$$\frac{\left|\mathbf{\Sigma}^{-1}(\boldsymbol{\alpha})\right|}{\left|\mathbf{W}_1'\mathbf{\Sigma}^{-1}(\boldsymbol{\alpha})\mathbf{W}_1\right|} = \frac{\prod_{i=1}^{p}(1+\phi_{ii})\prod_{i=2}^{p}\left(1-\phi_{ii}^2\right)^{i-1}}{C_1}. \quad (11)$$

The terms in the quadratic form in (6) are of the form $\mathbf{u}'\mathbf{\Sigma}^{-1}(\boldsymbol{\alpha})\,\mathbf{w}$, which from page 70 of Galbraith and Galbraith (1974) can be computed as

$$\mathbf{u}'\mathbf{\Sigma}^{-1}(\boldsymbol{\alpha})\,\mathbf{w} = \sum_{r=1}^{p}\sum_{s=1}^{p} u_r w_s m_{rs} + \sum_{t=p+1}^{n}\left(u_t - \sum_{i=i}^{p}\alpha_i u_{t-i}\right)\left(w_t - \sum_{i=i}^{p}\alpha_i w_{t-i}\right) \qquad (12)$$

where

$$m_{rs} = \sum_{j=0}^{r-1}\alpha_j\alpha_{j+s-r} - \sum_{j=p+1-s}^{p+r-s}\alpha_j\alpha_{j+s-r}$$

for $1 \le r \le s \le p$. For numerical stability, we recommend that any calculation involving the vector $\mathbf{X}$ be done using the sample mean corrected version $\mathbf{X} - \mathbf{1}\bar{X}$, where $\bar{X} = n^{-1}\sum_{t=1}^{n} X_t$ (Note that this does not change the restricted likelihood since it is already invariant to an origin shift in the data). The next sub-section describes how one can obtain confidence intervals for $\sum_{i=1}^{p}\alpha_i$ from the *RLRT*.

6.1 Confidence intervals for the sum of the AR coefficients:

When estimating AR processes, one quantity of interest to econometricians is the sum of the AR coefficients $\sum_{i=1}^{p}\alpha_i$ which can be thought of as a measure of persistence in the process. A $(1-\theta)\,100\%$ confidence region C_θ for $\sum_{i=1}^{p}\alpha_i$ can be constructed as

$$C_\theta = \left\{a : RLRT\,(a) \le \chi^2_{1,\theta}\right\}$$

where $RLRT\,(a)$ denotes the likelihood ratio test based on the restricted likelihood for testing the hypothesis $H_0 : \sum_{i=1}^{p}\alpha_i = a$ versus $H_1 : \sum_{i=1}^{p}\alpha_i \ne a$ and $\chi^2_{1,\theta}$ is the $(1-\theta)\,100\%$ percentile of the chi-square distribution with one degree of freedom. From Lemma 2 testing $H_0 : \sum_{i=1}^{p}\alpha_i = a$ is equivalent to testing $H_0 : \prod_{i=1}^{p}(1-\phi_{ii}) = 1-a$. In order to carry out this test, we need to maximise the restricted likelihood in (5) under the constraint $\prod_{i=1}^{p}(1-\phi_{ii}) = 1-a$. When $1-a > 0$ (the stationary case), this is equivalent to the constraint $\sum_{i=1}^{p} w_i = c$ where $w_i = \log(1-\phi_{ii})$, $c = 1-a$ and $w_i \in (-\infty, \log 2)$. The restricted likelihood optimised under this constraint should be approximately the same as that optimised over the set $A_\varepsilon = \{\mathbf{w} : w_i \in (-\infty, \log 2)\,,\ c-\varepsilon < \sum_{i=1}^{p} w_i < c+\varepsilon\}$ for a suitably small $\varepsilon > 0$. Standard statistical software such as R easily allow optimisation of functions over sets such as A_ε.

APPENDIX

Proof of Theorem 1: Since the $RLRT$ is invariant to one-to-one re-parametrisation we choose to work with the parametrisation $\lambda = (\beta_1, \beta_2, ..., \beta_p, \tau_v) \equiv (\beta_1, \theta, \tau_v)$, where $\beta_1 = \sum_{i=1}^{p} \alpha_i$, $\beta_i = -\sum_{j=i}^{p} \alpha_j$ for $i = 2, ..., p$ and $\tau_v = \sigma_v^{-2}$. For the restricted log-likelihood given in (5), we will denote expectations of the log-likelihood derivatives as

$$\kappa_{rs} = n^{-1} E\left(\partial^2 L / \partial \lambda_r \lambda_s\right), \qquad \kappa_{rst} = n^{-1} E\left(\partial^3 L / \partial \lambda_r \lambda_s \lambda_t\right), \qquad \kappa_{rs}^{(t)} = \partial \kappa_{rs} / \partial \lambda_t.$$

We will provide the proof only for the intercept case since the trend case follows along similar lines. We also note that we may assume $X_t = U_t$ since the restricted likelihood for the intercept case is invariant to the mean. Since $Var\left(\sum_{t=1}^{n} v_t\right) = O(n)$ and $Var\left(\sum_{t=1}^{n} X_t\right) = O(n)$, we see from (11), (12) and the fact that $(\phi_{11}, .., \phi_{pp})$ are continuously differentiable functions of $(\alpha_1, ..., \alpha_p)$ (Barndorff-Nielsen and Schou, 1973) that

$$n^{-1} E\left(\frac{\partial^2}{\partial \lambda_r \lambda_s} \log \frac{\left|\mathbf{\Sigma}^{-1}(\boldsymbol{\alpha})\right|}{\left|\mathbf{W}_r' \mathbf{\Sigma}^{-1}(\boldsymbol{\alpha}) \mathbf{W}_r\right|}\right) = O\left(n^{-1}\right)$$

and

$$\begin{aligned} & n^{-1} E\left(\frac{\partial^2}{\partial \lambda_r \lambda_s} \mathbf{X}' \mathbf{\Sigma}^{-1}(\boldsymbol{\alpha}) \mathbf{W}_1 (\mathbf{W}_1' \mathbf{\Sigma}^{-1}(\boldsymbol{\alpha}) W_1)^{-1} \mathbf{W}_1' \mathbf{\Sigma}^{-1}(\boldsymbol{\alpha}) \mathbf{X}\right) \\ &= n^{-1} E\left(\frac{\partial^2}{\partial \lambda_r \lambda_s} \frac{\left(\mathbf{X}' \mathbf{\Sigma}^{-1}(\boldsymbol{\alpha}) \mathbf{1}\right)^2}{\mathbf{1}' \mathbf{\Sigma}^{-1}(\boldsymbol{\alpha}) \mathbf{1}}\right) \\ &= O\left(n^{-1}\right). \end{aligned}$$

It follows that the information matrix $\mathbf{K} = ((-\kappa_{rs}))$ is given by

$$\mathbf{K} = \begin{bmatrix} \tau_v \gamma_0 & -\tau_v \mathbf{\Gamma}' & 0 \\ -\tau_v \mathbf{\Gamma} & \tau_v \mathbf{\Xi} & \mathbf{0} \\ 0 & \mathbf{0}' & K_{\tau_v} \end{bmatrix} + O\left(n^{-1}\right), \tag{13}$$

where $K_{\tau_v} = 1/\left(2\tau_v^2\right)$, $\gamma_0 = E\left(X_t^2\right)$, $\mathbf{\Gamma} = \left(E\left(X_{t-1}\Delta X_{t-1}\right), ..., E\left(X_{t-1}\Delta X_{t-p+1}\right)\right)'$ and $\mathbf{\Xi} = Var\left(\Delta X_{t-1}, ..., \Delta X_{t-p+1}\right)'$. Let κ^{rs} denote the entries of $-\mathbf{K}^{-1}$ and $\tilde{\kappa}^{rs}$ denote the entries of $-\mathbf{K}_{22}^{-1}$, where $\mathbf{K}_{22}$ is the lower right $p \times p$ sub-matrix of $\mathbf{K}$ corresponding to $(\beta_2, ..., \beta_p, \tau_v)$. From the results of Hayakawa (1977, 1987), Cordeiro (1987), Barndorff-Nielsen and Hall (1988) and Chesher and Smith (1995), we have

$$P\left(R_n(a) \leq x\right) = P\left(\chi_1^2 \leq x\right) + C(\boldsymbol{\alpha}) n^{-1}\left[P\left(\chi_3^2 \leq x\right) - P\left(\chi_1^2 \leq x\right)\right] + O\left(n^{-2}\right), \tag{14}$$

where

$$C\left(\boldsymbol{\alpha}\right)=\frac{1}{2}\left\{\sum_{\lambda}\left(l_{rstu}-l_{rstuvw}\right)-\sum_{\left(\theta,\tau_{v}\right)}\left(\tilde{l}_{rstu}-\tilde{l}_{rstuvw}\right)\right\}, \tag{15}$$

$$l_{rstu}=\kappa^{rs}\kappa^{tu}\left(\frac{1}{4}\kappa_{rstu}-\kappa_{rst}^{(u)}+\kappa_{rt}^{(su)}\right),\qquad \tilde{l}_{rstu}=\tilde{\kappa}^{rs}\tilde{\kappa}^{tu}\left(\frac{1}{4}\kappa_{rstu}-\kappa_{rst}^{(u)}+\kappa_{rt}^{(su)}\right), \tag{16}$$

$$l_{rstuvw}=\kappa^{rs}\kappa^{tu}\kappa^{vw}\left(\frac{1}{6}\kappa_{rtv}\kappa_{suw}+\frac{1}{4}\kappa_{rtu}\kappa_{svw}-\kappa_{rtv}\kappa_{sw}^{(u)}-\kappa_{rtu}\kappa_{sw}^{(v)}+\kappa_{rt}^{(v)}\kappa_{sw}^{(u)}+\kappa_{rt}^{(u)}\kappa_{sw}^{(v)}\right) \tag{17}$$

and

$$\tilde{l}_{rstuvw}=\tilde{\kappa}^{rs}\tilde{\kappa}^{tu}\tilde{\kappa}^{vw}\left(\frac{1}{6}\kappa_{rtv}\kappa_{suw}+\frac{1}{4}\kappa_{rtu}\kappa_{svw}-\kappa_{rtv}\kappa_{sw}^{(u)}-\kappa_{rtu}\kappa_{sw}^{(v)}+\kappa_{rt}^{(v)}\kappa_{sw}^{(u)}+\kappa_{rt}^{(u)}\kappa_{sw}^{(v)}\right).$$

Using standard results on the inverse of a partitioned matrix, we can express $-\mathbf{K}^{-1}$ up to $O\left(n^{-1}\right)$ as

$$-\mathbf{K}^{-1}=\begin{bmatrix}-Q^{-1} & -\mathbf{a}'\\ -\mathbf{a} & -\mathbf{K}_{22}^{-1}-\mathbf{B}\end{bmatrix}+O\left(n^{-1}\right), \tag{18}$$

where $Q=\tau_{v}\left(\gamma_{0}-\boldsymbol{\Gamma}'\boldsymbol{\Xi}^{-1}\boldsymbol{\Gamma}\right)$,

$$\mathbf{a}=\begin{bmatrix}Q^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\Gamma}\\ 0\end{bmatrix} \tag{19}$$

and

$$\mathbf{B}=Q^{-1}\begin{bmatrix}\boldsymbol{\Xi}^{-1}\boldsymbol{\Gamma}\boldsymbol{\Gamma}'\boldsymbol{\Xi}^{-1} & \mathbf{0}\\ \mathbf{0}' & 0\end{bmatrix}. \tag{20}$$

¿From the lower right $p\times p$ sub-matrix of (18) we see that for $(r,s)\in(2,...,p+1)$ we have $\kappa^{rs}=\tilde{\kappa}^{rs}+b_{r-1,s-1}$, where b_{rs} is the $(r,s)^{\text{th}}$ element from $\mathbf{B}$. It thus follows from (15) that

$$C\left(\boldsymbol{\alpha}\right)=\frac{1}{2}\left(l_{1111}-l_{111111}\right)+\frac{1}{2}\sum_{\left(\left(\beta_{1}\right),\left(\theta,\tau_{v}\right)\right)}\left(l_{rstu}-l_{rstuvw}\right)+\frac{1}{2}\sum_{\left(\theta,\tau_{v}\right)}\left(l_{rstu}^{b}-l_{rstuvw}^{b}\right), \tag{21}$$

where $\sum_{\left(\left(\beta_{1}\right),\left(\theta,\tau_{v}\right)\right)}$ denotes that at least one index in the summand must be β_{1} and at least one index from (θ,τ_{v}) and l_{rstu}^{b} and l_{rstuvw}^{b} are defined in a manner similar to l_{rstu} and l_{rstuvw} in (16) and (17) but with at least one κ^{rs} replaced by $b_{r-1,s-1}$. We now proceed to simplify each of the terms in (21). From Lemma 6 and part (iii) of Lemma 5, we have

$$\lim_{\boldsymbol{\alpha}\to\tilde{\boldsymbol{\alpha}}}\boldsymbol{\Gamma}'\boldsymbol{\Xi}^{-1}\boldsymbol{\Gamma}<\infty. \tag{22}$$

We also note that

$$\kappa_{rst} = O\left(n^{-1}\right) \text{ and } \kappa_{rstu} = O\left(n^{-1}\right) \qquad \text{for } (r,s,t,u) \in (1,...,p).$$

Hence,

$$\begin{aligned}
\frac{1}{2}\left(l_{1111} - l_{111111}\right) &= \frac{1}{2}\left(\kappa^{11}\right)^2 \left[\kappa_{11}^{(11)} - 2\kappa^{11}\left(\kappa_{11}^{(1)}\right)^2\right] + O\left(n^{-1}\right) \\
&= \frac{1}{2}\left(\kappa^{11}\right)^3 \left[\kappa_{11}^{(11)}\left(\kappa^{11}\right)^{-1} - 2\left(\kappa_{11}^{(1)}\right)^2\right] + O\left(n^{-1}\right) \\
&= \frac{1}{2}\left(\kappa^{11}\right)^3 \left[-\kappa_{11}^{(11)} Q - 2\left(\kappa_{11}^{(1)}\right)^2\right] + O\left(n^{-1}\right) \\
&= \frac{1}{2}\left(\kappa^{11}\right)^3 \left[-\kappa_{11}^{(11)}\left(-\kappa_{11} - \tau_v \mathbf{\Gamma}'\mathbf{\Xi}^{-1}\mathbf{\Gamma}\right) - 2\left(\kappa_{11}^{(1)}\right)^2\right] + O\left(n^{-1}\right) \\
&= \frac{1}{2}\left(\kappa^{11}\right)^3 \left[\kappa_{11}^{(11)}\kappa_{11} - 2\left(\kappa_{11}^{(1)}\right)^2\right] + \frac{\tau_v}{2}\left(\kappa^{11}\right)^3 \kappa_{11}^{(11)} \mathbf{\Gamma}'\mathbf{\Xi}^{-1}\mathbf{\Gamma} + O\left(n^{-1}\right) \\
&= \frac{1}{2}\left(\kappa_{11}\kappa^{11}\right)^3 \kappa_{11}^{-3} \left[\kappa_{11}^{(11)}\kappa_{11} - 2\left(\kappa_{11}^{(1)}\right)^2\right] + \frac{\tau_v}{2}\left(\kappa^{11}\right)^3 \kappa_{11}^{(11)} \mathbf{\Gamma}'\mathbf{\Xi}^{-1}\mathbf{\Gamma} + O\left(n^{-1}\right) \\
&= \frac{1}{2}\left(\kappa_{11}\kappa^{11}\right)^3 \tau_v^{-1} \frac{\partial^2}{\partial \beta_1^2}\left(\gamma_0^{-1}\right) + \frac{\tau_v}{2}\left(\kappa^{11}\right)^3 \kappa_{11}^{(11)} \mathbf{\Gamma}'\mathbf{\Xi}^{-1}\mathbf{\Gamma} + O\left(n^{-1}\right).
\end{aligned}$$

¿From (31), (22) and Lemma 5, we have $\lim_{\boldsymbol{\alpha} \to \widetilde{\boldsymbol{\alpha}}} \kappa_{11}\kappa^{11} = 1$, $\lim_{\boldsymbol{\alpha} \to \widetilde{\boldsymbol{\alpha}}} \left(\kappa^{11}\right)^3 \kappa_{11}^{(11)} < \infty$, and $\lim_{\boldsymbol{\alpha} \to \widetilde{\boldsymbol{\alpha}}} \frac{\partial^2}{\partial \beta_1^2}\left(\gamma_0^{-1}\right) < \infty$ and hence, up to $O\left(n^{-1}\right)$ we get

$$\lim_{\boldsymbol{\alpha} \to \widetilde{\boldsymbol{\alpha}}} \frac{1}{2}\left(l_{1111} - l_{111111}\right) = \frac{\tau_v^{-1}}{2} \lim_{\boldsymbol{\alpha} \to \widetilde{\boldsymbol{\alpha}}} \frac{\partial^2}{\partial \beta_1^2}\left(\gamma_0^{-1}\right) + \frac{\tau_v}{2}\left\{\lim_{\boldsymbol{\alpha} \to \widetilde{\boldsymbol{\alpha}}} \left(\kappa^{11}\right)^3 \kappa_{11}^{(11)}\right\}\left(\lim_{\boldsymbol{\alpha} \to \widetilde{\boldsymbol{\alpha}}} \mathbf{\Gamma}'\mathbf{\Xi}^{-1}\mathbf{\Gamma}\right), \quad (23)$$

where each limit is finite.

We next turn our attention to the last term on the right hand side of (21) and note that from (31) and (22), we get $\lim_{\boldsymbol{\alpha} \to \widetilde{\boldsymbol{\alpha}}} Q^{-1} = 0$, which in conjunction with Lemma 5 implies that $\lim_{\boldsymbol{\alpha} \to \widetilde{\boldsymbol{\alpha}}} \mathbf{B} = \mathbf{0}$. Furthermore, for the summation of the indices over (θ, τ_v), Lemma 5 guarantees that the limit as $\boldsymbol{\alpha} \to \widetilde{\boldsymbol{\alpha}}$ of all the terms of the kind κ_{rstu}, κ_{rst}, $\kappa_{rst}^{(u)}$ and $\kappa_{rt}^{(su)}$ is bounded and hence it follows that

$$\lim_{\boldsymbol{\alpha} \to \widetilde{\boldsymbol{\alpha}}} \frac{1}{2} \sum_{(\theta, \tau_v)} \left(l_{rstu}^b - l_{rstuvw}^b\right) = 0. \quad (24)$$

Finally we consider the middle term on the right hand side of (21) which can be expressed as

$$\frac{1}{2}\sum_{((\beta_1),(\theta,\tau_v))}\left(l_{rstu}-l_{rstuvw}\right)=\frac{1}{2}\sum_{((\beta_1),(\theta))}\left(l_{rstu}-l_{rstuvw}\right)+\frac{1}{2}\sum_{((\beta_1),(\tau_v))}\left(l_{rstu}-l_{rstuvw}\right) \qquad (25)$$
$$+\frac{1}{2}\sum_{((\beta_1),(\theta),(\tau_v))}\left(l_{rstu}-l_{rstuvw}\right),$$

where $\sum_{((\beta_1),(\theta))}$ denotes that at least one index in the summation must be β_1 and at least one index must be from θ. We now note that for any $(r,s,t,u)\in(1,...,p)$, we have $\kappa_{rstu}=O\left(n^{-1}\right)$ and $\kappa_{rst}=O\left(n^{-1}\right)$ when at least one index in the subscript is different from the other indices. We also note from (13) that $\kappa^{r,p+1}=O\left(n^{-1}\right)$ for $r\in(1,...,p)$. These facts, in conjunction with (19), (31), (22), Lemma 5 and Lemma 4 imply that the limit as $\boldsymbol{\alpha}\to\widetilde{\boldsymbol{\alpha}}$ of each of the terms on the right hand of (25) is finite.

Lemma 2 *For ϕ_{kj} and α_j, $j=1,\ldots,p$ as defined in (8), we have*

$$(i)\quad 1-\textstyle\sum_{j=1}^{p}\alpha_j=\prod_{k=1}^{p}\left(1-\phi_{kk}\right)$$
$$(ii)\quad \phi_{kj}=\mathcal{P}_{kj}\left(\phi_{11},\ldots,\phi_{kk}\right)$$

where $\mathcal{P}_{kj}(\cdot)$ is a multivariate polynomial.

Proof. From (8), we have

$$1-\sum_{j=1}^{k+1}\phi_{k+1,j}=1-\phi_{k+1,k+1}-\left(1-\phi_{k+1,k+1}\right)\sum_{j=1}^{k}\phi_{k,j}=\left(1-\phi_{k+1,k+1}\right)\left(1-\sum_{j=1}^{k}\phi_{kj}\right).$$

Part (i) of the lemma follows by applying the above equation repeatedly on $1-\sum_{j=1}^{p}\phi_{p,j}$. Equation (ii) is also obtained by induction, for example, $\phi_{21}=\phi_{11}-\phi_{22}\phi_{11}$, followed by

$$\phi_{31}=\phi_{21}-\phi_{33}\phi_{22}=\phi_{11}-\phi_{11}\phi_{22}-\phi_{22}\phi_{33},$$
$$\phi_{32}=\phi_{22}-\phi_{33}\phi_{21}=\phi_{22}-\phi_{11}\phi_{33}+\phi_{11}\phi_{22}\phi_{33},$$

etc. □

Lemma 3 *If $\phi_{kk}=1$ (or -1) for some $k\in\{1,\ldots,p\}$ and $|\phi_{hh}|<1$ for $h=k+1,\ldots,p$ in (8), then the polynomial $g(z)=z^p-\alpha_1 z^{p-1}-\cdots-\alpha_{p-1}z-\alpha_p$ has exact k roots on unit circle with* *at least one real unit root 1 (or* -1 *if* $\phi_{kk} = -1$*) and the rest of* $p-k$ *roots of* $g_p(z)$ *are inside the unit circle. That is,*

$$|z_j| = 1, \quad j = 1, \ldots, k$$
$$|z_j| < 1, \quad j = k+1, \ldots, p$$

with $z_j = 1$ *(or* -1*) for at least one* $j \in (1, ..., k)$. *Furthermore, the second inequality is true if and only if* $|\phi_{hh}| < 1$ *for* $h = k+1, \ldots, p$.

Proof. We only give the proof for $\phi_{kk} = 1$ since the proof for $\phi_{kk} = -1$ is similar. Let

$$g_k(z) = z^k - \phi_{k,1} z^{k-1} - \cdots - \phi_{k,k-1} z - \phi_{kk},$$

it is sufficient to show that (a) $g_k(z)$ has all the roots on the unit circle with one unit root. (b) $g_p(z) = \mathcal{P}_{p-k}(z)\, g_k(z)$ where $\mathcal{P}_{p-k}(z)$ is a $(p-k)$th degree Schur polynomial, which has all the roots inside the unit circle.

If $\phi_{kk} = 1$, then $\phi_{k,j} = \phi_{k-1,j} - \phi_{k-1,k-j}$, $j = 1, \ldots, k-1$ by (8), thus

$$\phi_{k,j} = -\phi_{k,k-j}, \tag{26}$$

and

$$g_k(1) = 0.$$

Thus one of the roots of $g_k(z)$ is 1. The rest of roots are also on the unit circle is based on a simple fact that with coefficients (26),

$$g_p(z) = -z^k g_p(1/z). \tag{27}$$

That is, the roots of $g_p(z)$ are the same as the roots of $g_p(1/z)$ but two sets of roots are reciprocal of each other. This can only happen if the roots are of the form $e^{i\theta}$ (i.e. on the unit circle).

The algorithm in (8) with few steps of algebra gives

$$g_{h+1}(z) = z g_h(z) - \phi_{h+1,h+1} z^h g_h(1/z). \tag{28}$$

Combining with (27),

$$g_{k+1}(z) = (z + \phi_{k+1,k+1})\, g_k(z),$$

$$g_{k+2}(z) = \left\{z^2 + (\phi_{k+1,k+1} + \phi_{k+2,k+2}\phi_{k+1,k+1})\, z + \phi_{k+2,k+2}\right\} g_k(z).$$

By induction,

$$g_{k+u}(z) = \mathcal{P}_u(z)\, g_k(z), \quad u = 1, \ldots, p-k,$$

where the uth degree polynomial

$$\mathcal{P}_u(z) = z^u - \pi_{u1} z^{u-1} - \cdots - \pi_{u,u-1} z - \pi_{uu}.$$

Moreover, equations (28) and (27) give

$$\begin{aligned} g_{k+u}(z) &= z g_{k+u-1}(z) - \phi_{k+u,k+u} z^{k+u-1} g_{k+u-1}(1/z) \\ &= z\left(\mathcal{P}_{u-1}(z)\, g_k(z)\right) - \phi_{k+u,k+u}\left(z^{u-1}\mathcal{P}_{u-1}(1/z)\, z^k g_k(1/z)\right) \\ &= g_k(z)\left\{z\mathcal{P}_{u-1}(z) + \phi_{k+u,k+u} z^{u-1}\mathcal{P}_{u-1}(1/z)\right\}, \end{aligned}$$

and

$$\begin{aligned} \mathcal{P}_u(z) &= z\mathcal{P}_{u-1}(z) + \phi_{k+u,k+u} z^{u-1}\mathcal{P}_{u-1}(1/z) \\ &= z^u - \pi_{u-1,1} z^{u-1} - \cdots - \pi_{u-1,u-1} z + \phi_{k+u,k+u}\left(-\pi_{u-1,u-1} z^{u-1} - \cdots - \pi_{u-1,1} z + 1\right). \end{aligned}$$

Hence

$$\pi_{u,u} = -\phi_{k+u,k+u}$$

and

$$\pi_{u,j} = \pi_{u-1,j} + \phi_{k+u,k+u}\pi_{u-1,u-j} = \pi_{u-1,j} - \pi_{uu}\pi_{u-1,u-j}.$$

That is, the mapping of $-(\phi_{k+1,k+1}, \ldots, \phi_{pp})'$ to the coefficients of $\mathcal{P}_{p-k}(z)$ is an Durbin-Levinson algorithm in (8). Since $|\phi_{hh}| < 1$ for $h = k+1, \ldots, p$ by assumption, $\mathcal{P}_{p-k}(z)$ is a Schur polynomial from the proof of Theorem 2 of Bardorff-Nielsen and Schou (1973). The last statement of the theorem also follows from the same theorem of Barndorff-Nielsen and Schou.□

Proof of Lemma 1. Firstly, if $\boldsymbol{\phi} \in \boldsymbol{\phi}_2$ then the transform in (8) gives an $\boldsymbol{\alpha} \in \mathbf{A}_2$ by Lemma 3. Secondly, if $\boldsymbol{\alpha} \in \mathbf{A}_2$, then $1 - \sum_{j=1}^{p} \alpha_j = 0$. By (i) of Lemma 2, at least one ϕ_{kk} of the corresponding $\boldsymbol{\phi}$ is 1. But by Lemma 3, only $\phi_{11} = 1$ and $|\phi_{kk}| < 1, k = 2, \ldots, p$. Thus, the corresponding $\boldsymbol{\phi} \in \boldsymbol{\Phi}_2$. We have shown the first part of the Lemma.

Note that the Jacobian for the mapping $\mathbf{\Phi}_2 \to \mathbf{A}_2$ remains the same as $\mathbf{\Phi}_1 \to \mathbf{A}_1$ in equation 13 of Barndorff-Nielsen and Schou (1973), that is

$$\left|\frac{\partial \boldsymbol{\phi}}{\partial \boldsymbol{\alpha}'}\right| = \prod_{k=2}^{p} (1-\phi_{kk})^{k/2} (1-\phi_{kk})^{[(k-1)/2]} \neq 0, \quad \boldsymbol{\phi} \in \mathbf{\Phi}_2.$$

The second part of the lemma follows by the inverse function theorem. □

Lemma 4 *For $j, k, \ell = 1, \ldots, p$, we have*

$$\lim_{\boldsymbol{\alpha} \to \tilde{\boldsymbol{\alpha}}} \frac{\partial \phi_{kk}}{\partial \beta_j} < \infty, \qquad \lim_{\boldsymbol{\alpha} \to \tilde{\boldsymbol{\alpha}}} \frac{\partial^2 \phi_{kk}}{\partial \beta_j \partial \beta_\ell} < \infty.$$

Proof. Denote

$$\boldsymbol{\phi}_k = (\phi_{k1}, \phi_{k2}, \ldots, \phi_{kk}, \phi_{k+1,k+1}, \ldots, \phi_{p,p}),$$

so that

$$\boldsymbol{\phi} = \boldsymbol{\phi}_1 = (\phi_{11}, \phi_{22}, \ldots, \phi_{pp})'$$

and

$$\boldsymbol{\phi}_p = \boldsymbol{\alpha} = (\alpha_1, \alpha_2, , \ldots, \alpha_p)'.$$

Writing the recursive equation (8) in the other direction

$$\begin{aligned} \phi_{k,j} &= \phi_{k+1,j} + \phi_{k+1,k+1}\phi_{k,k+1-j} \\ &= \phi_{k+1,j} + \phi_{k+1,k+1}\left(\phi_{k+1,k+1-j} + \phi_{k+1,k+1}\phi_{k,j}\right), \end{aligned}$$

we get

$$\phi_{k,j} = \frac{\phi_{k+1,j} + \phi_{k+1,k+1}\phi_{k+1,k+1-j}}{1-\phi_{k+1,k+1}^2}. \tag{29}$$

This equation gives a formula for $\boldsymbol{\phi}_k$ as a function of $\boldsymbol{\phi}_{k+1}$ that is $\boldsymbol{\phi}_k = \mathcal{T}_{k+1}\boldsymbol{\phi}_{k+1}$,where $\mathcal{T}_{k+1}$ is a $p \times p$ matrix

$$\mathcal{T}_{k+1} = \begin{bmatrix} \mathbf{\Psi}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-k} \end{bmatrix},$$

and $\mathbf{\Psi}_k$ is a $k \times k$ matrix

$$\mathbf{\Psi}_k = \frac{1}{1-\phi_{k+1,k+1}^2}\begin{bmatrix} 1 & 0 & \cdots & 0 & \phi_{k+1,k+1} \\ 0 & 1 & & \phi_{k+1,k+1} & 0 \\ \vdots & & \cdot & & \vdots \\ 0 & \cdot & & \ddots & 0 \\ \phi_{k+1,k+1} & \cdots & \cdots & 0 & 1 \end{bmatrix}.$$

The central entry is $(1-\phi_{k+1,k+1}^2)^{-1}(1+\phi_{k+1,k+1})$ if k is odd. Using the chain rule we have

$$\frac{\partial \boldsymbol{\phi}_1}{\partial \beta} = \frac{\partial \boldsymbol{\phi}_1}{\partial \boldsymbol{\phi}_2}\frac{\partial \boldsymbol{\phi}_2}{\partial \boldsymbol{\phi}_3}\cdots\frac{\partial \boldsymbol{\phi}_{p-1}}{\partial \boldsymbol{\alpha}}\frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\beta}},$$

where

$$\frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\beta}} = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & -1 & 1 \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}.$$

and

$$\frac{\partial \boldsymbol{\phi}_k}{\partial \boldsymbol{\phi}_{k+1}} = \mathcal{T}_{k+1}' = \begin{bmatrix} \mathbf{\Psi}_k' & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-k} \end{bmatrix}.$$

By (29),

$$\frac{\partial \phi_{k,j}}{\phi_{k+1,k+1}} = \frac{\left(1+\phi_{k+1,k+1}^2\right)\phi_{k+1,k+1-j} + 2\phi_{k+1,k+1}\phi_{k+1,j}}{\left(1-\phi_{k+1,k+1}^2\right)^2},$$

$$\frac{\partial \phi_{k,j}}{\partial \phi_{k+1,j}} = \frac{1}{1-\phi_{k+1,k+1}^2},$$

$$\frac{\partial \phi_{k,j}}{\phi_{k+1,k+1-j}} = \frac{\phi_{k+1,k+1}}{1-\phi_{k+1,k+1}^2}.$$

In the case that k is odd and $j = k+1-j = (k+1)/2$, then

$$\frac{\partial \phi_{k,j}}{\partial \phi_{k+1,j}} = \frac{1+\phi_{k+1,k+1}}{1-\phi_{k+1,k+1}^2}.$$

By (ii) of Lemma 2 $\phi_{k+1,j}$ is a polynomial in $\phi_{11},\ldots,\phi_{k+1,k+1}$, hence each entry of $\partial\boldsymbol{\phi}_1/\partial\beta$,

$$\frac{\partial \phi_{kk}}{\partial \beta_j} = \frac{\mathcal{P}_{kj}\left(\phi_{11},\cdots\phi_{pp}\right)}{\prod_{h=2}^p\left(1-\phi_{hh}^2\right)^2} = \frac{\mathcal{P}_{kj}\left(\boldsymbol{\phi}\right)}{\prod_{h=2}^p\left(1-\phi_{hh}^2\right)^2}, \tag{30}$$

where $\mathcal{P}_{kj}(\cdot)$ is a polynomial. The first derivatives in (30) have finite limit as $\boldsymbol{\alpha} \to \widetilde{\boldsymbol{\alpha}}$ by Theorem 2. The second derivatives,

$$\frac{\partial^2 \phi_{kk}}{\partial \beta_j \partial \beta_\ell} = \frac{1}{\prod_{h=2}^{p}\left(1-\phi_{hh}^2\right)^2}\left(\frac{\partial P_{jk}(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}}\right)' \frac{\partial \boldsymbol{\phi}}{\partial \beta_\ell}$$
$$+ \frac{4P_{jk}(\boldsymbol{\phi})}{\prod_{h=2}^{p}\left(1-\phi_{hh}^2\right)^4}\sum_{i=2}^{p}\left(1-\phi_{ii}\right)\phi_{ii}\frac{\partial \phi_{ii}}{\partial \beta_\ell}\prod_{\substack{h=2\\h\neq i}}^{p}\left(1-\phi_{hh}^2\right),$$

also have finite limit as $\boldsymbol{\alpha} \to \widetilde{\boldsymbol{\alpha}}$ because of Theorem 2 and since the limit as $\boldsymbol{\alpha} \to \widetilde{\boldsymbol{\alpha}}$ of (30) is finite as stated above. □

Lemma 5 *Letting $\gamma_j = E\left(X_t X_{t-j}\right)$ and $\rho_j = \gamma_0^{-1}\gamma_j$, we have for $i, j = 1, \ldots, p$,*

$$\begin{array}{lll}
(i) & \lim\limits_{\boldsymbol{\alpha}\to\widetilde{\alpha}} \dfrac{\partial}{\partial \tau_v}\left(\dfrac{1}{\gamma_0}\right) < 0, & \lim\limits_{\boldsymbol{\alpha}\to\widetilde{\alpha}} \dfrac{\partial^2}{\partial^2 \tau_v}\left(\dfrac{1}{\gamma_0}\right) < 0,\\
(ii) & \lim\limits_{\boldsymbol{\alpha}\to\widetilde{\alpha}} \dfrac{\partial}{\partial \beta_j}\left(\dfrac{1}{\gamma_0}\right) < \infty, & \lim\limits_{\boldsymbol{\alpha}\to\widetilde{\alpha}} \dfrac{\partial^2}{\partial \beta_i \partial \beta_j}\left(\dfrac{1}{\gamma_0}\right) < \infty,\\
(iii) & \lim\limits_{\boldsymbol{\alpha}\to\widetilde{\alpha}} \left(\rho_j - \rho_{j-1}\right) = 0, & \lim\limits_{\boldsymbol{\alpha}\to\widetilde{\alpha}} \left(\gamma_j - \gamma_{j-1}\right) < \infty\\
(iv) & \lim\limits_{\boldsymbol{\alpha}\to\widetilde{\alpha}} \dfrac{\partial}{\partial \beta_j}\left(\gamma_j - \gamma_{j-1}\right) < \infty, & \lim\limits_{\boldsymbol{\alpha}\to\widetilde{\alpha}} \dfrac{\partial^2}{\partial \beta_i \partial \beta_j}\left(\gamma_j - \gamma_{j-1}\right) < \infty
\end{array}$$

Proof. From equation (8) of Barndorff-Nielsen and Schou (1973) and (i) of Lemma 2,

$$\frac{1}{\gamma_0} = \tau_v \prod_{i=1}^{p}\left(1-\phi_{ii}^2\right) = \left(\tau_v\right)\left(1-\beta_1\right)\prod_{i=1}^{p}\left(1+\phi_{ii}\right). \tag{31}$$

Since $\boldsymbol{\alpha} \to \widetilde{\boldsymbol{\alpha}}$ implies $\beta_1 \to 1$, (i) follows immediately from above. Part (ii) follows from Lemma 4. We now show parts (iii) and (iv).

Equation (4) of Barndorff-Nielsen and Schou (1973),

$$\phi_{k+1,k+1} = \left(\rho_{k+1} - \sum_{j=1}^{k}\phi_{k,j}\rho_{k+1-j}\right)\left(1-\sum_{j=1}^{k}\phi_{k,j}\rho_j\right)^{-1},$$

gives

$$\begin{aligned}
\rho_{k+1} &= \sum_{j=1}^{k} \phi_{k,j}\rho_{k+1-j} + \phi_{k+1,k+1}\left(1 - \sum_{j=1}^{k} \phi_{k,j}\rho_j\right) \\
&= \sum_{j=1}^{k} \phi_{k,j}\rho_{k+1-j} + \phi_{k+1,k+1}\left(1 - \sum_{j=1}^{k} \phi_{k,k+1-j}\rho_{k+1-j}\right) \\
&= \phi_{k+1,k+1} + \sum_{j=1}^{k} \left(\phi_{k,k+1-j} - \phi_{k+1,k+1}\phi_{k,j}\right)\rho_j \\
&= \sum_{j=0}^{k} \phi_{k+1,k+1-j}\rho_j = \sum_{j=1}^{k+1} \phi_{k+1,j}\rho_{k+1-j}.
\end{aligned}$$

Applying summation by parts on $\sum_{j=1}^{k+1} \phi_{k+1,j}B^j$, we have

$$\sum_{j=1}^{k+1} \phi_{k+1,j}B^j = \left(\sum_{j=1}^{k+1} \phi_{k+1,j}\right) B - (1-B)\sum_{j=1}^{k}\left(\sum_{h=j+1}^{k+1} \phi_{k+1,h}\right) B^j.$$

Let $\nabla\rho_k = \rho_k - \rho_{k-1}$, then

$$\begin{aligned}
\nabla\rho_{k+1} &= \left(\sum_{j=1}^{k+1} \phi_{k+1,j}B^j\right)\rho_{k+1} - \rho_k \\
&= \left(\sum_{j=1}^{k+1} \phi_{k+1,j}\right)\rho_k - \sum_{j=1}^{k}\left(\sum_{h=j+1}^{k+1} \phi_{k+1,h}\right)\nabla\rho_{k+1} - \rho_k \\
&= \left(-1 + \sum_{j=1}^{k+1} \phi_{k+1,j}\right)\rho_k - \sum_{j=1}^{k}\left(\sum_{h=j+1}^{k+1} \phi_{k+1,h}\right)\nabla\rho_{k+1} \\
&= -\rho_k \prod_{i=1}^{k+1}\left(1-\phi_{ii}\right) - \sum_{j=1}^{k}\left(\sum_{h=j+1}^{k+1} \phi_{k+1,h}\right)\nabla\rho_{k+1-j}, \qquad (32)
\end{aligned}$$

Since $\nabla\rho_1 = \phi_{11} - 1,\ \nabla\rho_{k+1-j}$ has a factor of $(1-\phi_{11})$ by induction and $\phi_{k+1,h}$ is a polynomial of $\phi_{11},\ldots,\phi_{k+1,k+1}$ by (ii) of Lemma 2,

$$\nabla\rho_{k+1} = -\rho_k \prod_{i=1}^{k+1}\left(1-\phi_{ii}\right) + \mathcal{P}_{k+1}\left(\boldsymbol{\phi}\right)\left(1-\phi_{11}\right)$$

where $\mathcal{P}_{k+1}(\cdot)$ is a polynomial. Part (iii) follows from (ii) of Lemma 2. We have

$$\nabla\gamma_{k+1} = \nabla\rho_{k+1}\gamma_0 = \left(-\rho_k \prod_{i=2}^{k+1}\left(1-\phi_{ii}\right) + \mathcal{P}_{k+1}\left(\boldsymbol{\phi}\right)\right)\left\{\prod_{i=2}^{p}\left(1-\phi_{ii}\right)\prod_{i=1}^{p}\left(1+\phi_{ii}\right)\right\}^{-1}.$$

Part (iv) follows from Lemma 4. $\square$

Lemma 6 *For the matrix* $\mathbf{\Xi} = Var\left(\Delta X_{t-1}, ..., \Delta X_{t-p+1}\right)'$ *defined in the proof of Theorem 1 we have* $\lim_{\boldsymbol{\alpha}\to\widetilde{\boldsymbol{\alpha}}} \mathbf{\Xi}^{-1} < \infty$.

Proof: Note that $\mathbf{\Xi} = Var\left(\mathbf{T}\mathbf{X}_p\right)$, where $\mathbf{X}_p = (X_1, ..., X_p)'$ and $\mathbf{T}$ is a $p-1 \times p$ matrix of contrasts with a representative row being of the form $(0, ..., 0, -1, 1, 0, ...0)$. From page 384 of Harville (1974), we have

$$\begin{aligned} |\mathbf{\Xi}| &= |\mathbf{\Sigma}_p| \left|\mathbf{1}_p'\mathbf{\Sigma}_p^{-1}\mathbf{1}_p\right| \left|\mathbf{1}_p'\mathbf{1}_p\right|^{-1} \\ &= p^{-1} |\mathbf{\Sigma}_p| \left|\mathbf{1}_p'\mathbf{\Sigma}_p^{-1}\mathbf{1}_p\right|, \end{aligned} \tag{33}$$

where $\mathbf{\Sigma}_p = Var\left(\mathbf{X}_p\right)$. From equations (5) and (8) of Barndorff-Nielsen and Schou (1973), we have

$$|\mathbf{\Sigma}_p| = \prod_{i=1}^{p} \left(1 - \phi_{ii}^2\right)^{-i} \tag{34}$$

while using the expression for $\mathbf{\Sigma}_p^{-1}$ on page 70 of Galbraith and Galbraith (1974) yields

$$\mathbf{1}_p'\mathbf{\Sigma}_p^{-1}\mathbf{1}_p = -\left(1 - \sum_{i=1}^{p} \alpha_i\right) \left\{p\left(\sum_{i=1}^{p} \alpha_i - 1\right) - 2\sum_{k=1}^{p} k\alpha_k\right\}.$$

Using part (i) of Lemma 2 and summation by parts, we can simplify the right hand side of the above equation to get

$$\mathbf{1}_p'\mathbf{\Sigma}_p^{-1}\mathbf{1}_p = -\prod_{i=1}^{p} (1 - \phi_{ii}) \left\{-p\prod_{i=1}^{p} (1 - \phi_{ii}) - 2\left(\sum_{i=1}^{p} \alpha_i + \sum_{i=2}^{p}\sum_{s=i}^{p} \alpha_s\right)\right\}. \tag{35}$$

¿From (33), (34) and (35), we get

$$|\mathbf{\Xi}| = p^{-1}\prod_{i=2}^{p} \left(1 - \phi_{ii}^2\right)^{-i} (1 + \phi_{11})^{-1} \prod_{i=2}^{p} (1 - \phi_{ii}) \left\{p\prod_{i=1}^{p} (1 - \phi_{ii}) + 2\left(\sum_{i=1}^{p} \alpha_i + \sum_{i=2}^{p}\sum_{s=i}^{p} \alpha_s\right)\right\}$$

and hence from Theorem 2 we get

$$\begin{aligned} \lim_{\boldsymbol{\alpha}\to\widetilde{\alpha}} |\mathbf{\Xi}| &= p^{-1}\prod_{i=2}^{p} \left(1 - \widetilde{\phi}_{ii}^2\right)^{-i} \prod_{i=2}^{p} \left(1 - \widetilde{\phi}_{ii}\right) \left(1 + \sum_{i=2}^{p}\sum_{s=i}^{p} \widetilde{\alpha}_s\right) \\ &= p^{-1}\prod_{i=2}^{p} \left(1 - \widetilde{\phi}_{ii}^2\right)^{-i} \prod_{i=2}^{p} \left(1 - \widetilde{\phi}_{ii}\right) \left(1 - \sum_{i=2}^{p} \widetilde{\beta}_s\right), \end{aligned}$$

where $\widetilde{\beta}_s = -\sum_{s=i}^{p} \widetilde{\alpha}_s$ for $s = 2, ..., p$. Since the polynomial $z^p - \sum_{i=1}^{p} \widetilde{\alpha}_i z^{p-i}$ has roots $|\widetilde{z}_p| \leq ... \leq |\widetilde{z}_2| < 1$ and $\widetilde{z}_1 = 1$, we have $1 - \sum_{i=2}^{p} \widetilde{\beta}_s \neq 0$. Noting that $\max_{2 \leq i \leq p} \left| \widetilde{\phi}_{ii} \right| < 1$, we conclude that

$$\lim_{\boldsymbol{\alpha} \to \widetilde{\alpha}} |\boldsymbol{\Xi}| \neq 0.$$

This fact, in conjunction with part (iii) of Lemma 5 implies that $\lim_{\boldsymbol{\alpha} \to \widetilde{\boldsymbol{\alpha}}} \boldsymbol{\Xi}^{-1} < \infty$.

References

Andrews, D. and Chen, H-Y. (1994): "Approximately Median-Unbiased Estimation of Autoregressive Models," *Journal of Business and Economic Statistics,* 12, 187-204.

Barndorff-Nielsen, O and Schou, G. (1973): "On the Parametrization of Autoregressive Models by Partial Autocorrelations," *J. Multivariate Anal.* 3, 408–419

Cheang, W.K. and Reinsel, G. (2000): "Bias Reduction of Autoregressive Estimates in Time Series Regression Model through Restricted Maximum Likelihood," *Journal of the American Statistical Association*, 95**,** 1173-1184.

Chen, W. and Deo, R. (2006a): "Bias Reduction and Likelihood Based Almost-Exactly Sized Hypothesis Testing in Predictive Regressions using the Restricted Likelihood". Working paper

Chen, W. and Deo, R. (2006b): "A Smooth Transition to the Unit Root Distribution via the Chi-Square Distribution with Interval Estimation for Nearly Integrated Autoregressive Processes". Working paper

Efron, B. (1975): "Defining the curvature of a statistical problem (with applications to second order efficiency)," *Annals of Statistics*, 3, 1189–1242.

Elliott, G. and Stock, J. (2001): "Confidence Intervals for Autoregressive Coefficients Near One," *Journal of Econometrics,* 103, 155-181.

Francke, M, and de Vos, A. (2006): "Marginal Likelihood and Unit Roots," Forthcoming in *Journal of Econometrics.*

Galbraith, R. and Galbraith, J. (1974): "On the Inverses of some Patterned Matrices Arising in the Theory of Stationary Time Series," *J. Appl. Probability,* 11, 63–71.

van Giersbergen, N. (2006a): "Bartlett Correction in the Stable AR(1) Model with Intercept and Trend," Working paper, Universiteit van Amsterdam.

van Giersbergen, N. (2006b): "Bartlett Correction in the Stable AR(2) Model with Intercept and Trend," Working paper, Universiteit van Amsterdam.

Hansen, B. (1999): "The Grid Bootstrap and the Autoregressive Model," *The Review of Economics and Statistics,* 81, 594-607.

Harville, D. (1974): "Bayesian Inference of Variance Components Using Only Error Contrasts," *Biometrika,* 61, 383-385.

Harville, D. (1977): "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems," *Journal of the American Statistical Association*, 72, 320-338.

Kalbfleisch, J. and Sprott, D. (1970): "Application of Likelihood Methods to Models Involving Large Numbers of Parameters," *Journal of the Royal Statistical Society B,* 32, 175-194.

Kang, W., Shin, D. and Lee, Y. (2003): "Baises of the Restricted Maximum Likelihood Estimators for ARMA Processes with Polynomial Time Trend," *Journal of Statistical Planning and Inference,* 116, 163-176.

Larsson, R. (1999): "Approximation of the Asymptotic Distribution of the Log Likelihood Ratio Test for Cointegration," *Econometric Theory,* 15, 789-813.

Marriott, F. and Pope, J. (1954): "Bias in the Estimation of Autocorrelations," *Biometrika,* 41, 390-402.

Nelson, C. and Plosser, C. (1982): "Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics,* 10, 139-162.

Phillips, P. C. B. (1977): "Approximations to some finite sample distributions associated with a first-order stochastic difference equation," *Econometrica* 45, 463–485

Phillips, P., Moon, H. and Xiao, Z. (2001): "How to Estimate Autoregressive Roots near Unity," *Econometric Theory,* 17, 29-69.

Rahman, S. and King, M. (1997): "Marginal-Likelihood Score-Based Tests of Regression Disturbances in the Presence of Nuisance Parameters," *Journal of Econometrics,* 82, 81-106.

Sprott, D. (1973): "Normal Likelihoods and their Relation to Large Sample Theory of Estimation," *Biometrika,* 60, 457-465.

Sprott, D. (1975): "Application of Maximum Likelihood Methods to Finite Samples," *Sankhya, Ser. B.*, 37, 259-270.

Sprott, D. (1980): "Maximum Likelihood in Small Samples: Estimation in the Presence of Nuisance Parameters," *Biometrika,* 67, 515-523.

Sprott, D. and Viveros-Aguilera, R. (1984): "The Interpretation of Maximum-Likelihood Estimation," *Canadian Journal of Statistics,* 12, 27-38.

Sprott, D. (1990): "Inferential Estimation, Likelihood, and Linear Pivotals," *Canadian Journal of Statistics,* 18, 1-10.

Tunnicliffe Wilson, G. (1989): "On the Use of Marginal Likelihood in Time Series Model Estimation," *Journal of the Royal Statistical Society Ser. B*, 51, 15-27.

Xiao, Z. (2001): "Likelihood-based Inference in Trending Time Series with a Root near unity," *Econometric Theory,* 17, 1082-1112.

Table I. Coverage Rates of Confidence Intervals for AR(1)

	intercept			intercept & trend		
α	90%	95%	99%	90%	95%	99%
0.9	.8975	.9496	.9888	.8974	.9501	.9894
0.95	.8994	.9495	.9888	.9097	.9565	.9909
0.99	.9153	.9596	.9905	.9173	.9558	.9912
0.995	.9173	.9597	.9909	9173	.9562	9911
1	.9138	.9557	.9906	.9197	.9595	.9902

Table II. Rejection Rates of Unit Root in AR(1)

	intercept				intercept & trend			
	$\alpha = .9$		$\alpha = .95$		$\alpha = .9$		$\alpha = .95$	
Procedure	90%	95%	90%	95%	90%	95%	90%	95%
RLRT	.6869	.4827	.3057	.1755	.3670	.2220	.1594	.0865
Andrews	.4870	.3160	.1879	.1004	.2366	.1365	.0973	.0523
Hansen	.3511	.2103	.1328	.0690	.1967	.1134	.0883	.0460

Table III. Mean and Standard Deviation of 90% CI Lengths of AR(1) with Intercept

	all CIs				CIs with true α			
	$\alpha = .9$		$\alpha = .95$		$\alpha = .9$		$\alpha = .95$	
	mean	S.D.	mean	S.D.	mean	S.D.	mean	S.D.
RLRT	.1721	.0354	.1309	.0462	.1703	.0287	.1271	.0398
Andrews	.1774	.0417	.1316	.0529	.1787	.0339	.1316	.0455
Hansen	.1794	.0439	.1312	.0538	.1810	.0371	.1315	.0472

Table IV. Mean and Standard Deviation of 90% CI Lengths of AR(1) with Intercept and Trend

	all CIs				CIs with true α			
	$\alpha = .9$		$\alpha = .95$		$\alpha = .9$		$\alpha = .95$	
	mean	S.D.	mean	S.D.	mean	S.D.	mean	S.D.
RLRT	.1973	.0461	.1569	.0575	.1960	.0406	.1511	.0525
Andrews	.1990	.0552	.1543	.0661	.2012	.0483	.1550	.0577
Hansen	.1995	.0569	.1536	.0666	.2021	.0503	.1546	.0586

Table V. Coverage Rates for AR(2) with Intercept

α_1	α_2	$\alpha_1 + \alpha_2$	1%	5%	10%
1.3	$-.4$	.90	.9884	.9489	.8974
1.55	$-.6$	.95	.9901	.9484	.8964
1.75	$-.76$	.99	.9917	.9505	.9006
1.775	$-.78$	.995	.9924	.9557	.9073
1.8	$-.8$	1	.9908	.9571	.9128

Table VI. Confidence Intervals of Nelson-Plosser Data

Series	p	n	$\sum \hat{\alpha}_j$	90%	95%
Real GNP	2	62	.870	(.767, 1]*	(.747, 1]
Nominal GNP	2	62	.928	(.852, 1]	(.838, 1]
per capita GNP	2	62	.866	(.761, 1]*	(.741, 1]
Industrial Prod.	6	111	.921	(.802, 1]	(.780, 1]
Employment	3	81	.896	(.805, 1]	(.787, 1]
Unemployment	4	81	.721	(.574, .881)*	(.545, .916)*
GNP deflator	2	82	.958	(.893, 1]	(.881, 1]
Consumer Price	4	111	.997	(.958, 1]	(.952 , 1]
Wages	3	71	.942	(.870, 1]	(.857, 1]
Real Wages	2	71	.904	(.800, 1]	(.780, 1]
Velocity	1	102	1	(.935, 1]	(.922, 1]
Bond Yield	3	71	1	(.961, 1]	(.950, 1]
S&P 500	4	100	.962	(.876, 1]	(.861, 1]

Table VII. Confidence Intervals of Extended Nelson-Plosser Data

Series	p	n	$\sum \hat{\alpha}_j$	90%	95%
Real GNP	2	80	.863	(.774, .965)*	(.757, 1]*
Nominal GNP	2	100	.970	(.912, 1]	(.901, 1]
per capita GNP	2	80	.858	(.767, .964)*	(.749, 1]*
Industrial Prod.	6	129	.926	(.814, 1]	(.794, 1]
Employment	3	99	.893	(.811, 1]*	(.796, 1]
Unemployment	4	99	.724	(.594, .861)*	(.569, .890)*
GNP deflator	2	100	1	(.966, 1]	(.958, 1]
Consumer Price	4	129	1	(.983, 1]	(.979, 1]
Wages	3	89	.973	(.917, 1]	(.907, 1]
Real Wages	2	89	1	(.897 1]	(.882, 1]
Velocity	1	120	1	(.965, 1]	(.957, 1]
Bond Yield	3	89	1	(.924, 1]	(.912, 1]
S&P 500	4	188	1	(.920, 1]	(.907, 1]

*rejections using critical values for RLRT from Francke and de Vos (2006)